6

PROPOZYCJE i MATERIAŁY

Compatibility and Integration of Order Systems

WYDAWNICTWO SBP

6

Compatibility and Integration of Order Systems

MIĘDZYNARODOWE TOWARZYSTWO ORGANIZACJI WIEDZY
STOWARZYSZENIE BIBLIOTEKARZY POLSKICH
TOWARZYSTWO INFORMACJI PROFESJONALNEJ

PROPOZYCJE I MATERIAŁY

Kompatybilność i integralność systemów porządkowania wiedzy

Materiały z seminarium zorganizowanego
w ramach Spotkania TIP/ISKO
Warszawa, 13-15 września 1995 r.

WYDAWNICTWO SBP

WARSZAWA 1996

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION
POLISH LIBRARIANS ASSOCIATION
SOCIETY FOR PROFESSIONAL INFORMATION

CONTRIBUTIONS AND MATERIALS

Compatibility and Integration of Order Systems

Research Seminar Proceedings
of the TIP/ISKO Meeting
Warsaw, 13-15 September, 1995.

WYDAWNICTWO SBP

WARSAW 1996

Publikacja wydana przy pomocy finansowej Fundacji Współpracy Polsko-Niemieckiej ze środków Republiki Federalnej Niemiec

Mitfinanziert von der Stiftung für Deutsch-Poinische Zusammenarbeit aus Mitteln der Bundesrepublik Deutschland

Projekt graficzny okładki i strony tytułowej
Wydawnictwo SBP

Redaktorzy merytoryczni
Ingetraut DAHLBERG
Krystyna SIWEK

Bibliografia adnotowana
Kompatybilność i integralność systemów porządkowania wiedzy
1960-1995
Ingetraut DAHLBERG

ISBN 83-85778-62-4

CIP - Biblioteka Narodowa
Compatitibility and integration of order systems : research seminar proceedings of the TIP/ISKO Meeting, Warsaw 13-15 September 1995 / International Society for Knowledge Organization, Polish Librarians Association, Society for Proffessional Information. - Warszawa : Wydaw. SBP, 1996. - (Propozycje i Materiały / Stowarzyszenie Bibliotekarzy Polskich ; 6)

Wydawnictwo SBP. Warszawa 1996. Wydanie I.
Ark. wyd. 13,50 Ark. druk. 15,25
Skład i łamanie AK LAND s.c.
Druk i oprawa: Warszawska Drukarnia Naukowa, ul. Śniadeckich 8, 00-656 Warszawa

WSTĘP

Międzynarodowe Towarzystwo Organizacji Wiedzy (International Society for Knowledge Organization ISKO) powstałe w 1990 r. jest związkiem wybitnych przedstawicieli świata nauki i praktyki (informatyków, logików, matematyków, językoznawców, bibliotekoznawców) zajmujących się problematyką organizacji wiedzy.

Organizacja wiedzy jako jeden z najważniejszych i najpilniejszych problemów współczesnej nauki o informacji wymaga nadal wielu badań naukowych i rozstrzygnięć. Główne problemy, które skupiają prace wielu środowisk naukowych dotyczą m.in. przepływu informacji w dynamicznie rozwijających się ostatnio sieciach informacyjnych w świecie, eliminacji barier językowych, budowy odpowiednich systemów klasyfikacji wiedzy i informacji, uporządkowania występującego chaosu wśród sieci informacyjnych.

Forum dyskusyjnym w tym zakresie są międzynarodowe seminaria ISKO. Ostatnie, odbyte w Warszawie, (po raz pierwszy) w październiku 1995 roku poświęcone było kompatybilności i integralności systemów wiedzy. Współorganizatorem seminarium było, działające w Polsce, Towarzystwo Informacji Profesjonalnej.

Referaty wygłoszone na warszawskim seminarium można uporządkować w 3 grupach:

I. Problematyka barier językowych.

Zaliczamy tu wystąpienia Winfrieda Schmitz-Essera („Język ogólnej komunikacji i pojęcie kompatybilności"), Gerharda J.A. Riesthuisa („Teoria kompatybilności języków informacyjnych"), Ingetraut Dahleberg („Zalecenia kompatybilności — rewizja").

II. Szczegółowe problemy metodologii i metodyki budowy i funkcjonowania narzędzi organizacji wiedzy.

W tej grupie mieszczą się dwa wystąpienia:

1. Dagobert Soergel: Model bazy danych dla zintegrowanego tezaurusa
2. Mieczysław Muraszkiewicz, Henryk Rybiński, Wacław Struk: Problemy oprogramowania dla tworzenia wyników badań tezaurusa wielojęzycznego.

III. Największa grupa referatów była prezentacją wyników badań szczegółowych:

1. Ewa Chmielewska-Gorczyca: Kompatybilność narzędzi indeksowania w środowisku wielu baz danych.
2. Giliola Negrini: Ku strukturalnej kompatybilności systemów pojęciowych.
3. Eugeniusz Ścibor: Kilka uwag o tworzeniu konkordancji między systemem klasyfikacji uniwersalnej i tezaurusem interdyscyplinarnym na przykładzie Polskiej Klasyfikacji Tematycznej i Tezaurusa Zagadnień Wspólnych.
4. Stephan Hoppe: Model integracji wiedzy w zakresie medycyny.
5. Stella G. Dextre Clarke: Integracja tezaurusów z zakresu nauk agrarnych.
6. Corentin Roulin: Praktyczne aspekty integracji tezaurusów wielojęzycznych.
7. Harold H. Zimmerman: koncepcja i zastosowanie konkordancji klasyfikacji w środowisku publicznie dostępnych katalogów online.
8. Barbara Sosińska-Kalata: UKD jako międzynarodowy standard organizacji wiedzy w bibliograficznych bazach danych i katalogach bibliotecznych.
9. Jadwiga Woźniak, Teresa Głowacka: Kartoteka wzorcowa KABA jako przykład integracji polsko--francusko-angielskiej systemów haseł przedmiotowych.
10. Wiesław Babik: Terminologia jako poziom kompatybilności języków indeksowania.
11. Pavla Stanicková: Międzynarodowe zintegrowane systemy baz danych podłączone do wielojęzycznych tezaurusów z zakresu nauk rolnych i ochrony środowiska.
12. Tomáš Samek: Integracja języków indeksowania z tezaurusem „EUROVOC" w Czechach.
13. Krystyna Siwek: Uwarunkowania kompatybilności między polskimi i obcymi bazami danych.
14. Wiesław Gliński, Mieczysław Muraszkiewicz: Inteligentny interfejs dostępu do systemów informacji.

Sądzę, że seminarium można uznać za bardzo udane. Liczne wystąpienia przedstawicieli Polski są dowodem na to, że polskie środowisko aktywnie włączyło się w prace Międzynarodowego Towarzystwa Organizacji Wiedzy.

Na zakończenie pragnę podziękować Fundacji Współpracy Polsko-Niemieckiej, która sfinansowała ze środków Republiki Federalnej Niemiec seminarium oraz niniejszą publikację.

Marcin Drzewiecki

INTRODUCTION

International Society for Knowledge Organization — estabilished in 1990 — is a association of the outstanding scientists and practisioners (in the fields of computer science, logics, mathematics, linguistics and library science) who deal with problems of knowledge organization.

Knowledge organization as a one of the most urgent and important problems of modern information science still requires many researches and solutions. The most important problems many scientific societies focus on are:the flow of the information in the recently dynamically developing information networks, elimination of the language barriers, building of the adequate knowledge and information classifications systems, struggling against the chaos emerging in the computer networks environments.

The international forum where all those problems could be discussed are ISKO seminars. The last one held in Warsaw in Oktober 1995 was devoted to compatibility and integrality of knowledge systems. The Warsaw's seminar was co-organised by Polish „Towarzystwo Informacji Profesjonalnej" („The Society for Profesional Information").

The seminars participants' papers might be arranged in 3 groups:

I. The problems connected with language barriers:

Winifred Schmitz-Esser („Language of general communication and concept compatibility"), Gerhard J.A. Riesthuis („Theory of compatibility of information languages"), Ingentraut Dahleberg („The compatibility guidelines — a re-evaluation").

II. The detailed problems of methodology and methodics of the structure and functioning of the knowledge organization tools.

This group consists of two papers:

1. Dagobert Soergel: Data structure and software support for integrated thesauri
2. Mieczysław Muraszkiewicz, Henryk Rybiński, Wacław Struk: Software problems of merging multilingual thesauri

III. The most numerous group of papers has been devoted to the presentations of detailed research results:

1. Ewa Chmielewska-Gorczyca: Compatibility of indexing tools in multidatabase environment,
2. Giliola Negrini: Towards structural compatibility between concept systems,
3. Eugeniusz Ścibor: Some remarks on the estabilishment of concordances between a universal classification system and an interdisciplinary thesaurus,
4. Stephen Hoppe: The UMLS — a model for knowledge integration in a subject field,
5. Stella G.Dextre Clarke: Integrating thesauri in the agricultural sciences,
6. Corentin Roulin: Bringing multinlingual thesauri together: a feasibilty study,
7. Harold H. Zimmerman: Conception and application possibilities of classification concordances in an OPAC environment,
8. Barbara Sosińska Kalata: The universal decimal classification as an international standard for knowledge organization in bibliographic databases and library catalogues,
9. Jadwiga Woźniak,Teresa Głowacka: KABA subject authority file an example of an integrated Polish-French-English subject headings system,
10. Wiesław Babik: Terminology as a level for the compatibility of indexing languages. Some remarks,
11. Pavla Stanciková: International integrated database systems linked to multilingual thesauri covering the fields of environment and agriculture,
12. Tomáš Samek: Exing languages integration and eurovoc thesauri in the Czech Republic,
13. Krystyna Siwek: Compatibility discrepancies between Polish and foreign databases,
14. Wiesław Gliński, Mieczysław Muraszkiewicz: An intelligent front-end processor for accessing information systems.

I consider the seminar as a very succesful event. The numerous Polish participants' papers prove that Polish scientific society has actively joined the efforts of the International Society for Knowledge Organization.

Finally I would like to express the thankfullness to Foundation for Polish-German Co-operation whose financial donation made from the sources of German Federal Republic has enabled the Seminar organization and publishing of this book.

Marcin Drzewiecki

WYSTĄPIENIE PRZEWODNICZĄCEGO KOMITETU BADAŃ NAUKOWYCH PROF. ALEKSANDRA ŁUCZAKA

Panie Przewodniczący, Panie i Panowie,
Mam zaszczyt powitać Państwa na seminarium zorganizowanym przez Towarzystwo Informacji Profesjonalnej oraz Międzynarodowe Towarzystwo Organizacji Wiedzy.

Witam przedstawicieli nauki i przemysłu informacyjnego z wielu krajów Europy, a także ze Stanów Zjednoczonych.

Profesjonalne przekazywanie informacji jest w tej chwili jedną z najważniejszych dziedzin życia i niezwykle istotnym czynnikiem procesów integracyjnych w Europie. Cieszę się, że możemy gościć wielu tak wybitnych specjalistów. Polska co prawda nie jest potentatem, jeśli chodzi o przemysł informacyjny, ale robimy wszystko, aby znaleźć się na tym rynku jako odpowiedzialny i ważny partner. Wierzę, że spotkanie to stanowi istotny krok w tym kierunku.

Mam nadzieję, że atmosfera tego budynku, siedziby Komitetu Badań Naukowych, któremu mam zaszczyt przewodniczyć, będzie sprzyjała owocnej wymianie poglądów i dyskusjom. Życzę Państwu udanych obrad i miłego pobytu w Warszawie.

OPENING ADDRESS BY THE HEAD OF THE STATE COMMITTEE FOR SCIENTIFIC RESEARCH PROF. ALEKSANDER ŁUCZAK

Mr Chairman, Ladies and Gentlemen,
I have the honour to welcome you to the seminar organized by the Society for Professional Information and the International Society for Knowledge Organization.

I welcome representatives of the information science and industry from many countries in Europe as well as from the United States.

The professional information exchange constitutes at present one of the most important domains of our life and a significant factor of integration processes in Europe. I am glad that we can greet many so outstanding representatives here today. Poland is not a power as far as the information industry is concerned, but we are doing our best to enter this market as a responsible and important partner. I believe that this one as well as similar conferences are an important step in this direction.

I do hope that the atmosphere of this building, the building that houses the State Committee for Scientific Research which I have the honour to chair, will be conducive to the fruitful exchange of ideas and discussions. I wish you successful debates and a pleasant stay in Warsaw.

GREETING WORDS FROM THE ISKO PRESIDENT

On behalf of the ISKO Executive Board and especially its Vicepresidents, Dr. Robert FUGMANN and Professor Hanne ALBRECHTSEN with four further Board Members present — Dr. Hellmut LÖCKENHOFF, Drs. Gerhard RIESTHUIS, Dr. Pavla STANCIKOVA and Prof. Dr. Winfried SCHMITZ-ESSER, I am happy to cordially greet cordially our Polish colleagues and our dear guests at this opening hour of our first meeting in Poland.

I should like to express our deepest gratitude to our host in this beautiful conference hall of the State Committee for Science and Technology, Professor Dr. Aleksander ŁUCZAK and to the organizing Society for Professional Informationof this meeting presided over by Professor Dr. Włodzimierz GOGOŁEK and his able assistant, our dear Polish ISKO Coordinator, Ms. Krystyna SIWEK!

I would also like to greet all our colleagues who had come from such distances as the USA and a number of European countries and to extend our thanks for their readiness to feed their knowledge, experiences and expertise into the necessary discussions.

The topic outlined in our Call for Papers and reprinted in the conference brochure treats a problem that has been with us already since about 35 years — ever since thesauri have been created for the indexing of the scientific and technical literature. And it was exactly 25 years ago, that Rasmus MÖLGAARD-HANSEN, then the Danish Chairman of FID/CR delivered a lecture at the Meeting of the Polish Classification Research Group in Warsaw, Oct.9, 1970 on — what? — „Compatibility and Complementarity between Indexing Languages"! Thus we are celebrating now so-to-speak a silver-jubilee event to commemorate this former meeting of experts and to start afresh so solve the still unsolved problems and to reevaluate the work which has been done in the meantime, in order to come to solutions and recommendations for our future work in this very area of highly topical problems..

I should also like to mention that we had prepared an annotated bibliography on the topic of the conference beforehand and had sent it to some 30 colleagues for comments. The feedback was quite satisfying, also we found for ourselves more references so that by now the bibliography has grown from 230 to almost 500 items. Of course we hope to publish the results in the processings volume.

Last but not least I need to state that this Research Seminar was made possible by the generous support from the European Community and the Foundation for German-Polish Cooperation. We should like to thank these institutions very much indeed and express our hopes that the results of the Seminar will bear visible fruits for our profession at large and doe our future work. May GOD bless this meeting and all of you!

Ingetraut Dahlberg

Włodzimierz Gogołek

KEY NOTES
INFORMATION DEMOCRACY

The importance of the meeting is stressed by the fact, that in Poland there are only few events, which are devoted to professional information — a very important factor ofEuropean and global integration.

At the moment we are the creators of new democracy — information democracy.

fortunately, naerly all country borders for information traffic are now gone. There is an easy access to European information files in Asia and in America as well.

The digital form of information, that is a common factor of all forms of information — text, audio, graphics, video — is one of the most significant conditions for the development of the information industry.

It means that all of these forms of information can be collected and stored in computer's memory, processed and distributed in electronic form. Thanks to the modern technology all of these operations are very cheap and easy.

the main result of this progress in information managing is anew role of digital data in our firm, home and school. We see a new Gutenberg phenomenom for a much wider scale — because digital information storage covers not only text but picture, voice and video as well. It gives the possibility to process nesarly each form of information. the universality of digital information rapidly creates a huge market — multimedia market. Last year the value of the market was 30 times bigger than Poland's gross domestic product.

Multimedia are popular in our homes, e.g. compact discs, but multimedia are chiefly used in the computer technology.

today, computer systems use friendly interface, are independent from hardware and from operating systems. It means that we have good conditions to use the latest achievements in using the modern forms of professional information.

a fast development of the information industry is possible because of good digital telecom — mainly INTERNET.

Unfortunately, the development of multimedia, a very cheap access to telecom, caused jams on the domestic and international information highways. Designers and owners of the highways have to take very difficult decision — to introduce commercial telecom, I mean, commercial INTERNET.

The reason of the need of Internet's commercialization lies also in the growing number of electronic documents. Paradoxically, the digital form of documents creates the impression, that producing and distributing the documents is free of charge. So each day we are producing much more and longer documents. Many gigabytes of worthless documents wpuld be never produced if their authors had to pay full telecom and information storage costs.

today's computers make us much less concrete during our work with computers. the example of this trend is the popularity of menu, aspecially graphic menu. Short command has been changed into megabytes set of icons. It is a temporary substitute of the natural language. but the substitute is very expensive. this thesis is corroborated e.g. by the experiences of online users who have to pay for the usage of computer host time — although they have a possibility to use menu, the mostly use the language of commands. It is much more faster and cheaper.

paralelly with the information democracy development we can see the progress of a new region of „human computer behaviors" which together with soft — and hardware infrastructure makes „information culture". The culture means creation and effective usage of digital potential of telcom and information resources.

The culture consists of: legal regulations including copy rights, rules of creation, access and usage of information and universal education which covers these problems. The importance of this culture is shown by the contens of some information sources, ways of using them and many information crimes. Last year the estimated cost of one average crime in United Kingdom was 30.000 pounds.

The current step of the expansion of information democracy is sometimes called „The wild west". The task of the creators of democracy is to force thet the step should be a very short one.

closing my paper, I would like to say some words about our domestic information industry.

In Poland we have an outstanding chance to create a good market for the modern information industry. paradoxically, it is the result of Poland's delay in this fied. In our country it is possible to create users of modern technologies almost instantly because of the fact that we do not have the habbits of looking for information using relatively old technologies, e.g. „yellow pages".

Trainig is a very important factor in the Polish information industry. today in Poland there are few universities where the subject of thisindustry is taught.

we are still short of educational materials in these subjects.

These years are a turning point of the information democracy development. We are a unique chance to use it especially for the european integration process. so I'm sure that our meeting will be an important step towards information democracy in Europe and in the World.

THE COMPATIBILITY THEORETICAL BACKGROUND

Winfried Schmitz-Esser

LANGUAGE OF GENERAL COMMUNICATION AND CONCEPT COMPATIBILITY

At first glance, it might seem to be easier to achieve compatibility and integration of language-based concept order systems when the language used in such systems is as close as possible to that used in general communication.

However, experience tells us that order systems work better when based on more specialized terminologies such as those used in defined subject areas, or designed for small, specific worlds of application. There is a better chance in these cases of achieving concept compatibility, and of integrating existing order systems.

Problems arise rather more readily in the unconfined, global concept systems which are needed in order to deal with the masses of general knowledge of our time. It seems as if in these vast areas of fact, matter and thought non-compatibility of concepts, and thus restriction to local solutions only, is the price we have to pay for achieving at least some sort of order, and with it acceptable information retrieval performance.

This is basically the condition we are faced with when creating information systems which are needed by editors in the press and other media with regard to the published utterances of our time, in picture and film data banks, in information systems for parliamentary work, in museums of contemporary art or history, in non-specialized libraries, etc.

Computer assisted systems of this type have been my favourite playground for more than two decades, and it is with this professional and practical background, (and with an inclination towards linguistics), that I am going to explain here how I feel about the question of concept compatibility and term order integration.

One of the prime reasons for the difficulties encountered, especially in general, non-specialized systems, is the open nature and virtual unpredictability of natural language discourse. As in all non-numerical information systems, this phenomenon works at both ends, i.e. on the input side as the direction of discourse of the author, and on the users' side, as the terms used in any query. Concept systems and/or ordering systems are designed to help to create a consensus in understanding the meaning so that information may be retrieved.

SMALL, SPECIALIZED SYSTEMS NEARER TO COMPATIBILITY

Take, for example, an information system on lung diseases, forestry, or architecture. You will find that many of the fundamental coefficients relating to the meaning of search terms are readily defined by what Hjørland and Albrechtsen in a recent paper call „thought or discourse communities 1), 3) which they say are parts of society's division of labour", — peer groups of specialists, opinion leaders and their audience, readership, etc. Common understanding and use of a database sometimes are part of their group identity. In these communities, concept compatibility is rarely a major problem — at least among the happy insiders. The most amazing example of this that I ever saw was the European gene mapping project, where the scientists exchange their search results via a Wide Area Network.

As long as the meaning of a term is defined by the scientists and this definition is accepted by the scientific community, one is on safe ground. Words like ISOTOPES, COHERENT LIGHT, LINEAR PROGRAMMING, therefore pose no semantic problem. And it is within this well defined world of concepts that semantic relationships can readily be established. The term ISOTOPES used in concept system A will therefore be identical, or at least compatible with that same term used in system B. Difficulties arise when new terms crop up which are awaiting acceptance by the scientific community. The disadvantage of this is that it is those neologisms which usually constitute the more interesting part of our information systems, because they are in the forefront of scientific advance.

However, this is not limited to scientific definition. Market dominance, and a promising sales strategy will also suffice. Terms like CONNECTIVITY, WORK FLOW, WINDOWS, LEKTRIEVER, all have their commercial origin. These terms come and go with the life-cycle of the product, or merely with the respective marketing campaign, but also may have historic value.

Sometimes it is necessary to identify a hitherto unmentioned firm or product type in order to understand what is meant by HIGH RESOLUTION, FULL COLOR, etc. The problem then would be that the meaning itself shifts with changes in the product, and the term HIGH RESOLUTION to-day may mean double what it was five years ago.

However, the most important clarifying influence in such specialized systems is the domain itself. The domain decides what the meaning most probably will be, or can be; whether, e.g. INITIATION means the beginning of a process (in physics) or the introduction into manhood (in sociology); whether POSITIVE means something good (in ethics), something bad (in medicine), or a piece of film (in reprography).

Compatibility of terms will then largely depend on the definition and delineation of the domains covered by the different term systems to be considered or merged. If one of the systems covers more than the domain, then serious problems of open semantics occur. In a system of vehicle design, BODY may mean the physical frame of a car. If the larger system spans all aspects of art and design, then the word BODY requires discrimination between a multiplicity of meanings, including MAN, HUMAN BEING, CORPSE, DEAD, GLOBULE, ASTEROID, COLOR, MATTER and COUNCIL.

The same would occur with INFORMATION. Transferred from a safe place in signal theory, the word is being used with a dozen or so other meanings.

Concepts are addressed by terms, and since concepts only reside in the brain it is very difficult to find out whether a given term really reflects the same concept in different brains. So, when I talk of concept compatibility I know I simplify hugely when I take it as „using the same term".

DEFINITION OF SEMANTIC STRUCTURE A MUST

However, concept compatibility alone is not all that is needed in order to successfully integrate two or more language-based concept order systems. What is also needed is a defined, common semantic structure and a common repertoire of semantic relationships which have to be equally well defined.

If in thesaurus A the concept of BOXER MOTOR is defined as a narrower term (NT) of PISTON ENGINE, and PISTON ENGINE as a NT of COMBUSTION ENGINE, whereas in thesaurus B the PISTON ENGINE and the ROTATION ENGINE are dealt with as NTs of POWER UNIT, then the merging program will be at an impasse, since COMBUSTION ENGINE is set at a different level of logic to POWER UNIT.

Inconsistency will also occur if thesaurus A stipulates that BOOK is a NT of PRINT PRODUCT, whereas thesaurus B says that BOOK is part of (PO) TRAINING PROGRAM, and thesaurus C finds that BOOK is related to LIBRARY LENDING, since in thesaurus A no partitive relationship exists as required by B, and the partitive relationships of B must be expressed as hierarchical ones (NTs, BTs).

Thesauri designed for information retrieval usually are very coarse and highly abbreviated projections of concepts and concept structure, encountered in the depicted worlds. The DIN and ISO standards for retrieval thesauri are far from effective in guaranteeing common semantic structures.

In them, the common repertoire of semantic relationships is limited to three: equivalence, hierarchy, and proximity, which are all lacking clear, crisp definition. Equivalence may be anything between precisely the same meaning to quasi-synonyms; hierarchies may be of generic or constitutive type; and what is found as near or related to another term in thesauri of that type more often than not resembles the contents of a concept rubbish container!

So I make no secret of the fact that I am very sceptical about term concept integration in present retrieval systems, even in small, limited areas of application. This is due to the lack of common term system structures, not so much lack of concept compatibility. But that can be changed by constructing more articulate terminology system structures, and by defining them more clearly.

A group of German experts is currently working on the question of improved and graduated terminology system definitions. They expect to present a joint proposal on this issue at the ISKO/KTF meeting in Trier next month.

SEMANTICS IN GENERAL COMMUNICATION: NO CONCEPT COMPATIBILITY IN SIGHT

One of the main characteristics of general communication is that a participant in public dialogue, e.g. a journalist, a student, a politician, is in no way constrained to use a particular set of words. Not even words are always needed. What really matters is the successful accomplishment of the intended act of communication. Insofar as language communication is concerned of course, compliance is required with the basic rules of the language (e.g. spelling, grammar).

This means that any term and any combination of terms offered by the language may occur now and in the future. The system is then very large indeed, and open-ended. The problem here is much more severe. It begins with concept compatibility, but many different other factors are at work, within the total picture.

In the first instance, semantics is largely dictated by social and cultural factors, but the knowledge is dispersed among many different communities which indulge in a much more heterogeneous and chaotic discourse. Main boundaries are language, history and culture. What is TEA TIME, and what is MORNING (as compared to MORGEN and VORMITTAG), what is considered to be MACHO and what is considered to be SMART, what KITSCH or PATRIMOINE, are issues which are met and defined best in the socio-cultural context. Who knows why VEGETABLES is not quite the same as GEMÜSE, potatoes being an extra item you have to order in Germany!

The matter is indeed, highly complex.

What does NATIONAL mean to a German with respect to Germany to-day?

The common perception of historical experience, knowledge acquisition and social reflection work together to condition historical concepts.

In present day Germany, the term NATIONAL is still OK when used as a qualifier for the uprising against Napoleon. Looking at the students' movement following the two decades or so after the Vienna Congress, the term NATIONAL may be just acceptable, but PATRIOTIC is the preferred term. Today, after two world wars under the nationalist banner, the term NATIONAL is „out", the preferred adjective being FEDERAL or, after the unification of the two Germanies, just DEUTSCH, which is now fashionable. A decent member of German post-war society is a FELLOW CITIZEN (Willy Brandt) or an INHABITANT, or a CITIZEN, or a VOTER (unthinkable to address him as DEUTSCHER!), and what the state expects from him/her is to be a CONSTITUTIONAL PATRIOT, not plain PATRIOT — to be assured that he or she is clinically free from the slightest revisionist infection.

There are also breaks in concepts which are directly dictated by the course of history, other concepts remaining unchanged. But there is no universal rule. What in older papers and books is the ENEMY is now the WESTERN ALLIES.

What was experienced by the older generation as OCCUPATION and DEFEAT is now seen by a happier generation, and the majority of the population, as LIBERATION. However, when it necessary to address the LIBERATION of Mussolini on the Gran Sasso one cannot help but use this term in a literature search for the event up to the present day. The term FEDERAL also determines physical space. It just could not have meant the two Germanies before 1990.

This is by no means everything on this subject. Many, many terms commonly used in general communication are lacking common consensus on semantic definitions. Examples are words like LIBERALISM, CONSERVATIVE, CAPITAL, RESOURCE, POWER, DEMOCRATIC, DEVELOPMENT, HELP, HUMAN, etc.The power of these words resides in their shear „fuzziness", and their use is a normal part of the social game.

Quite frequently the meaning of a term depends on the author. What the word IMPERIALISM is intended to mean in the title: *„German literature in the age of IMPERIALISM"* only becomes clear when you know that the writer was Georg Lucacs, a Marxist philosopher, and it is only when you know this that you realise which literature of which period he is probably dealing with in his book. *Max Weber's* IMPERIALISM, in contrast, tackles a quite different historical phenomenon.

Sometimes, the meaning of a term cannot be fully understood because the phenomenon addressed is too recent, and so insufficient information is available; thus establishing semantic relationships in a thesaurus would be premature. This hinders up-to-date indexing of world events as is required in e.g. the press and other media information systems.

Or, there may be interminable discussions on how a past event in history should be addressed. Whether the students' riots in Germany in 1968 were an OUTBREAK OF DISCONTENT, an UPRISING, or a REVOLUTION (as the students would have liked to have it), — this has been a matter of public debate right up to the present day.

Certain terms are highly problematic because just using them in a thesaurus would be compromising. Terms like LIQUIDATION, CLEANSING, WORKING CLASS, VOLKSTUM, BLOOD, are examples which it is impossible to use in a thesaurus of a museum, a contemporary library or documentation centre without painstakingly expressing their inner meanig. Many of these words are euphemisms like RE-ESTABLISHING ORDER, ETHNIC CLEANSING, BROTHERLY HELP, but they also occur on a much less martial level, with pleasant annotations as, e.g., GEMEINWIRTSCHAFT or HARMONISATION.

Sometimes, an object escapes the discipline of being grouped in a given order system, although all the information on the object (the concept) may in fact be available and undisputed. It is clear that Chancellor Kohl's Christian Democratic Party (DCU) is a right wing party of the conservative type, and that Rudolf Scharping's Social Democrats are leftists on the rather progressive side. Now, what about the Greens? That may have been apparent to the public only by the nature of the parliamentary seating arrangements!

But what about a post war phenomenon such as the GESAMTDEUTSCHE VOLKSPARTEI (GVP) of the late Gustav Heinemann, the former Federal

President and father of Uta Ranke-Heinemann, the well known proponent of modernized Catholicism.

It would be difficult, or even impossible to find a suitable place for this party in any given political spectrum. Classifying the GVP as „national", or „conservative", or „liberal", or „anti-clerical", or „progressive" would not correspond to what it truly was. The nearest would perhaps be pacifist, but this by no means covers it all.

What does this mean in our consideration of our subject of concept compatibility? If the GVP cannot be properly positioned in a given order system, then the situation will be no better in any other order system, and compatibility cannot be achieved.

Was the GVP a conservative party? Was it national? Party founder *Gustav Heinemann* no doubt pursued a national concept, saw himself following in the footsteps of the so-called *Bekennende Kirche* (Professing Church), and of *Karl Barth*, one of the life cells of protest against Nazi tyranny. *Helene Wessel*, the co-founder, also pledged a national concept, and a balanced German policy, sovereignty and equal treatment, while opposing the neutralisation of Germany. Both fought against the manipulation of the church for reasons of politics, and they resisted re-armament and struggled for peace.

You may object that this is a rather far-fetched historical example from the middle of the Fifties. However, looking more thoroughly at concepts addressed in politics, social science, art and even in our daily life, one is surprised to see how poorly they are usually defined, and how open they are to a variety of interpretations.

Definitions Required

— MODERN WEAPONS	— including arms systems?
	— including arms systems software?
	— including sheltered command posts?
	— including telecom satellites?
	— if so, military telecom satellites only?
— RULING STRATUM	— the government, certainly.
	— MPs?
	— the Bosses?
	— if so, how far down?
	— the New Rich?
	— the Mafia?
— COMMUNICATIONS WORKERS	
	— telecom-people, certainly.
	— the postal services as well?
	— if so, private parcel services?

— EXHAUST GAS TESTS
— a test, certainly.
— a part of the Environment Program, OK.
— a measure of administrative protectionism?
— a source of public income?

— PRIMITIVE ART
— African tribal art, OK
— if so, also North American indians?
— also Mexican folk art?
— Maya relics?

This does not mean, of course, that the organization of knowledge is impossible in these broad and open domains. *It is possible, but only at the price of non-compatibility.* One has to think of a thesaurus as a model of the real world, and all that one has to do in creating it is to take an individual, defensible standpoint for one's view.

Therefore, it is recommended that the ordering system be designed such that it will be acceptable to the user group which in the larger systems will equate, or almost equate to the general public. One will only get general acceptance when the thesaurus respects the views of a society which is pluralistic. The major common axes, and prevalent values of our society will then be part of that system.

Such a model need not be — and probably cannot be — true in a purely philosophical sense. But this is not important. The main thing is that it is credible in the eyes of the common user because he/she feels familiar with the visions encountered in the model, and is also satisfied that the model is logically consistent.

Thus, if the system is articulate and predictable enough so as to represent the subject of the collection, then an acceptable ordering and retrieval performance might well be achieved. However, the properties required for an acceptable level of performance for such a system most certainly *cannot* be suitable to be prosperous for concept compatibility and integration with other term order systems.

These are the type of contemporary general, universal information systems as used in the media, in politics, in defence and education. Systems of this type are typically pragmatic. They are never hard science — which is logical because of the mass of volitional stipulations contained in them.

For example, look at this query statement in the library system of Haus der Geschichte (History of the Federal Republic) in Bonn! With a single statement 13.4DDR21.53J you can ask for a „brochure on the people's uprising in the German Democratic Republic in 1953" and, irrespective of whether you start from AUFSTAND, SIEBZEHNTER JUNI, KONTERREVOLUTION, VOLKSERHEBUNG or any other terms likely to be used for this kind of an event, — you will always be guided to the notation of 13.4 which stands for the preferential word REVOLUTION. When you know that the event happened in 1953 you may stipulate.53, but if you don't know you would have to look up a pre-established table of time-spans. There you find as Group 20: *The period of the two German states (1949 - 1990), and as Sub-Group 21: Federal Republic up to the end of the Great Coalition (1949 - 1969).*

Of course, the result you get is clear-cut. But as a user you are forced to use the structure of that table. Setting it up was no easy task for the historians in Bonn, as you may imagine. It was specifically designed for a Museum of German post war history. If it had been on modern Polish history, that table would look markedly different, and if it was to serve both countries' most significant eras, very little in common, and useful, would be achieved.

In the Bonn system, the preferential term REVOLUTION stands for a whole cluster of terms in its conceptual vicinity. Don't expect clear boundaries of thought. The system stipulates that in your searches, the „17th of June" has to be dealt with by use of the word REVOLUTION. If you don't agree, insisting it was just an UPRISING which failed, you can't use this system. The term 13.4 also stands for VENDEE. It then has to be combined with EBF for France and the respective time period code. So, what about the above-mentioned, a 68 students' uprising in Germany which also failed? You will find it in that Museum, but you will be required to address it as STUDENTS' DISORDERS. You see, it's deliberate, volitional. Constructing such systems, you are in Pascal's situation: You have to place your bet. Otherwise you can't live.

What one most certainly cannot do is establish order by concept definition while remaining „open" for any view of the world that may be desired to-day or in the future. This statement is as basic as much as it is obvious, at least as long as information retrieval systems are in principle drawn up as concept ordering systems.

LACK OF ARTICULATIVE POWER IN PRESENT CONCEPT ORDER SYSTEMS

Shortcomings in the definition of concepts is one thing; lack of articulative power in present concept order systems is another. These ordering systems are mostly designed as a tool for information retrieval. Even at the sophisticated end, — appearing as thesauri, classauri, faceted thesauri, etc., — their design remains very simple. They are based on single words, or short noun phrases, at most. They do not provide near the precision and flexibility of natural language, as anywhere used in general communication.

So, with current term-based concept order systems, there is a high probability that in system A, an ill-defined subject is represented by a rather inarticulate term or set of terms, and that in system B an equally insufficiently defined version of the concept is represented in no better way, — in alternative terms which may well be no less fuzzy.

Imagine the complexity of a term order system needed to cope with semantic structures such as are encountered in the examples of NATIONAL and GVP, and compare! It's worlds apart.

I must admit that this is not fertile ground for what we all feel should be achieved: *term and systems integration*.

A NEW APPROACH BASED ON LINGUISTIC ENGINEERING

However, in the future, another approach may arise which hopefully will bear fruit: the computer linguistic approach, or what is called „Linguistic Engineering".

It is possible to define the meaning of terms and phrases according to a common understanding (paradigm), and to represent this semantic knowledge by use of a computer, whereby a much more homogeneous base could be established on which to build more sophisticated ordering systems.

Such language-engineered bases which would be published as machine readable encyclopaedias rather than classical thesauri would by their very construction be much more homogeneous and, of course, interchangeable, and thus open to integration. They could be used for all subject fields, as well as for the whole range of computer linguistic activities like machine translation, machine-aided abstracting, editing, text analysis, speech recognition, to name only the most important. Establishing term-based concept order systems, thus enabling information retrieval, would then be one further possibility.

A proposal for such an approach was presented by the German Committee for Thesaurus and Classification Research (KTF) of the German Society for Documentation (DGD) in Frankfurt a few years ago and is being discussed at the present time at both national and international level by linguists and information scientists.

CONCLUSIONS AT A GLANCE:

Finally, here are the properties, and ensuing potential, with regard to concept compatibility and the integration of term ordering systems:

Systems Covering **Properties**	**Small Worlds, Dedicated Solutions**	**Large Worlds, Universal Solutions**
Individual, isolated solutions	Yes (Such systems exist)	Yes (Such systems exist)
Concept compatibility	Yes	No
Systems compatibility	Rare	No
Term order systems integration	Possible (if systems are better defined)	Not possible (Even if systems would be much more articulate)
Term order systems integration via linguistic engineering	Yes (hopefully)	Yes (hopefully)

OUTLOOK

When, sometime in the future, we will have such machine-readable, multi-purpose, and multilingual, semantic encyclopaedias, then the problems of compatibility and integration will be largely removed, and the language used for purposes of order and information retrieval will finally be that of general communication.

LITERATURE LIST

1. Albrechtsen, Hanne/Hjřrland, Berger
   "Understandings of Language and Cognition: Implications for Classification Research". In: Advances in Classification Research: Proceedings of the 5th ASIS SIG/CR Classification Research Workshop. Held 16 October 1994 at 57th ASIS Annual Meeting, October 16-20, 1994, Alexandria (VA) / edited by Raya Fidel, ... — New Jersey: Learned Information, 1995
2. Buchan, R. L.:
   "Intertwining thesauri and dictionaries" In: Information Services & Use, 9 (1989), pp. 171 - 175
3. Dahlberg, Ingetraut:
   "Concept and Definition Theory" In: Classification Theory in the Computer Age — Conversations across the disciplines, Proceedings, November 18-19, 1988, Albany, New York, 1989
4. Dahlberg, Ingetraut:
   "Towards establishment of compatibility between indexing languages". In: International Classification, 8 (1981), no. 2, pp. 86-91
5. Dahlberg, Ingetraut:
   "Grundlagen universaler Wissensordnung — Probleme und Möglichkeiten eines universalen Klassifikationssystems des Wissens", DGD-Schriftenreihe, Bd. 3, Ed. Deutsche Gesellschaft für Dokumentation, Verlag Dokumentation, Pullach bei München, 1974
6. Fischer, D. H.:
   "Consistency rules and triggers for thesauri". In: International Classification. 18 (1991), no. 4, pp. 212-225
7. Hjørland, Berger/Albrechtsen, Hanne
   "Toward a New Horizon in Information Science: Domain-analysis". In: Journal of the American Society for Information Science, 1994, Wiley (1994)
8. Körner, Horst G.:
   "Notationssysteme für Dokumentationssprachen und Klassifikationen". In: Forschungsbericht ID 80-013 Information und Dokumentation, Bundesministerium für Forschung und Technologie. Karlsruhe: Fachinformationszentrum, 9/1980
9. Lei, Zeng:
   "Compatibility of Indexing Languages in an Online Access Environment — A review of the Approaches". In: Advances in Classification Research, Vol. 3,

Proceedings 3rd ASIS SIG/CR Classification Research Workshop, Pittsburg, PA, 1992, pp. 161-181
10. Martinet, André
"Syntaxe Générale". Armand Colin, Paris, 1985
11. Mili, H., Rada, R.:
"Merging thesauri — principles and evaluation". In: IEEE transactions on pattern analysis and machine intelligence. 10 (1988), no. 2, pp. 204-220
12. Rada, R.:
"Maintaining thesauri and metathesauri". In: International Classification. 17 (1990), no. 3/4, pp. 158-164
13. Roulin, C.:
"Sub-thesauri as part of a metathesaurus". In: Classification research for knowledge representation and organization. Proceedings 5th Int. Study Conf. on Classification Research, Toronto, Canada, 24.-28.6.1991. Ed. by N.J. Williamson u. M. Hudon. Amsterdam: Elsevier 1992. pp. 329-336
14. Schmitz-Esser, Winfried
"Thesaurus — frischer Anlauf: lexikographisch, mehrsprachig, maschinengängig, universal, für Informationslinguistik und Information Retrieval. Vorstellung eines im KTF-Komitee für Klassifikations- und Thesaurusforschung erarbeiteten, neuen Thesauruskonzepts". Proceedings Deutscher Dokumentartag, Jena, 1993; Wolfram Neubauer ed., DGD, Frankfurt, 1993, pp. 261-274
15. Schmitz-Esser, Winfried
"Publikumsfragen an Literatur zur Zeitgeschichte — Haus der Geschichte in Bonn will Kluft zwischen Wissenschaft und Öffentlichkeit schließen". In: Bibliotheksdienst, Berlin 27. Jg (1993), no. 6, pp. 872-883
16. Schmitz-Esser, Winfried
"Der Konflikt zwischen Dokumentationssprache und natürlicher Sprache in Medien-Datenbanken — eine Topografie und ein Katalog von Anforderungen an die Informationslinguistik". Medienarchive in Gegenwart und Zukunft, Mediendokumentation 4, Saur-Verlag München, 1983, pp. 137-145
17. Schmitz-Esser, Winfried
"Anforderungen an publizistische Ad-hoc-Informationsdienste". In: Diebold-Bildschirmtextkongreß, Frankfurt, 1982, pp. 231-244
18. Schmitz-Esser, Winfried
"Die Pressedatenbank für Text und Bild des Verlagshauses Gruner+Jahr". In: Nachrichten für Dokumentation 27, 1977, no. 3, pp. 124-131
19. Soergel, Dagobert
"Indexing Languages and Thesauri: Construction and Maintenance". Melville Publishing, Los Angeles, 1974
20. Vickery, A.:
"Integrating thesauri into intelligent front-ends". In: ASIS '90: Proceedings 53rd ASIS Annual Meeting, Toronto, 4.-8.11.1990. Medford, NJ: Learned Information 1990

SUMMARY

At first glance it might seem to be easier to achieve compatibility and integration of language-based concept order systems when the language used in such systems is as near as possible to that used in general communication. Experience tell us, however, that order systems work better when based on more specialized terminologies such as those used in defined subject arteas or designed for small, specific worlds of application. In these cases there is a better chance in achieving concept compatibility and integrating existing order systems. Problems arise rather in the unconfined, global concept systems needed to deal with the masses of general knowledge of our time.

A major reason for this is that there are inadequate definitions underlying the structure and content of such broad and general order systems. Another prime reason is the open nature and virtual unpredictability of natural language discourse. A third one may arise from the restriction of the target which, as a rule, is information retrieval. But this is not all. Historic and cultural heritage, social factors language, different levels of knowledge and communication, and a number of other factors, seem greately to impede compatibility and the potential of integration.

In this paper will not only be analysed the difficulties encountered in systems of general knowledge. also will be presented solutions able to find in major order system projects carried out by the author for print and TV publishers and for the museum of the History of the Federal Republic of Germany, in Bonn. The new approach to the problem by the use of linguistic engineering will be reported as well.

Gerhard J.A. Riesthuis

THEORY OF COMPATIBILITY OF INFORMATION LANGUAGES

INTRODUCTION

At the ISKO Conference in Copenhagen (1994) Pauline Atherton Cochrane reminded us that the problem of compatibility of information languages was already a topic at the Second International Study Conference on Classification Research, held in Elsinore, Denmark on 14-18 September 1964. Since then some achievements in this field can be shown. Cochrane[1] mentioned the UMLS (Unified Medical Language System[2]) and switching vocabularies. At the same time there are many OPAC's and other on-line bibliographic databases, wherein only by using several different information languages one by one a satisfying subject search is possible.

This problem becomes more and more urgent. More and more bibliographic databases are merged. This gives relatively few problems as far as the formal entries are involved[3]. The choice and form of formal entries is standardized to a great degree. This means that when I search with the name of an author or a word from a title I am searching in the merged database as a whole. For subject entries the situation is quite different. In many cases the resulting databank needs to be searched using the different information languages used in the separate databases one by one. With any information language I am searching in a subset only of the total database. Using a UDC notation I will never find a record that has only subject entries taken from the Library of Congress Subject Headings. Lancaster and Smith remarked:

„Perhaps somewhat surprisingly, while controlled vocabularies tend to promote internal consistency within information systems, they also tend to reduce intersystem compatibility[4]".

---

[1] P.A. Cochrane. Elsinore revisited. — In: Knowledge organization and quality management. Proc. 3rd. Int. ISKO Conference, 20-24 June 1994, Copenhagen, Denmark. — Frankfurt : Indeks Verlag, 1994. — p.11-15.

[2] B.L. Humphreys. Unified medical language system : progress report. — In: International Classification, 15(1988)2, p.85-86, and A.R. Aronson, Th.C. Rindflesh, A.C. Browne. Exploiting a large thesaurus for information retrieval. — In: RIAO 94 Conference proceedings Rockefeller University New York N.Y. Oct 11 — 13, 1994. — Paris : C.I.D., 1994. — ISBN 20905450-05-3. — Vol 1, p.197-216

[3] But see: M. Reichart und M.W. Mönnich. Dublettenkontrolie in bibliographischen Datenbanken. — In: Bibliothek, Forschung und Praxis 18(1994)2, p. 193-216

[4] F.W. Lancaster and L.C. Smith. Compatibility issues affecting information systems and services. Paris : Unesco / GIP, 1983. — (PGI-83/WS/23). — p.52

Is there a way to solve this paradox? There are two simple solutions. The first one is not to use a controlled information language at all, the other one is: get everyone to use the same information language.

Not using a controlled information language means to rely on words from titles and abstracts, if so desired enriched with extra words. In a multi-language environment translations of titles and abstracts can be added too if necessary. This solution has the great advantage that it is cheap, but it means also poor subject access.

The second solution is a better one, but raises two interwoven questions: which one to use and, in many cases, what to do with the many records already indexed with another than the common information language? Both questions are important and not easy to answer. Even when participants agree which information language to use in the future the compatibility problem remains for records indexed in the past. Only for relatively small databases re-indexing is a realistic possibility.

The remaining part of this paper will deal with the paradox that controlling subject entries on the one hand and trying to achieve compatibility on the other hand seems not to be compatible.

WHAT IS COMPATIBILITY?

The concept „compatibility" is not very clear. A often used definition is:

> „Compatibility means that for each term A of an information language P there is a term A' in information language Q with the same meaning, so that we can convert A into A' without changes in meaning."

In the conversion only the form, the look changes, without any loss of information. To give an example: The subject heading „Warsaw" has the same information as the subject heading „Warschau". In this example the definition given is usable, but in many cases conversion of terms of one information language into terms of another information language is only possible with loss of information, but still the result can be acceptable. A book from the Library of Congress with the LCC notation PN 4400 gets automatically the notation 17.82 of the Dutch Basisclassification according to the conversion table used in the Dutch cataloguing system. PN 4400 means *Letters (Literary history)* and 17.82 means *Literary genres. Theory of genre.* PN 4400 gives more specific information about the content of the book than 17.82. We see a loss of information and still we can claim that the conversion is correct. Why can we say that the conversion is correct? The answer is simple: If we classify the book with the Basisclassification the correct notation is 17.82.

The LCC notation PN 1024 *Poetry by women* also converts correctly into BC 17.82, as do many other notations of LCC.

This means that we have a problem when we try to convert from the Basisclassification to the Library of Congress Classification. The notation 17.82 converts to many different notations from the Library of Congress Classification. How to choose between them? It means that compatibility is directional or a-symmetric. Conversion from language P to language Q is not the same as conversion from language Q to language P.

In my previous examples the conversion is made from separate terms to separate terms. If we try to convert the LCC notation HE 2701 *Railways. General works. United States* to a Dewey Decimal Classification notation we come across a new problem. Railroad transportation is DDC 385. Converting the term HE 2701 to 385 means a loss of information. This is not necessary. We can use a DDC expression consisting of 385 with an addition that indicates United States. In our example HE 2701 should be converted into 385.09(73). The expression 385.09(73) has the same meaning as the concept of the notation HE 2701.
We can now give another definition:

> An information language P is fully compatible with information language Q when an sentence that denotes correctly — using the vocabulary and syntax of P — the subject A of a document M can be translated, without re-indexing, to an sentence that denotes correctly — using the vocabulary and syntax of Q — as if subject A was indexed with language Q directly.

Note that in this definition „sentence" means „term" or „expression". According to this definition compatibility is directional or a-symmetric, and only those information languages can be fully compatible that cover the same or smaller *domain*.

TYPES OF COMPATIBILITY

In the literature the term compatibility is often used in the meaning: compatibility on the level of terms: for each term of a language S a 'correct' term of language T should be available. This type of compatibility we shall call *term compatibility*.

A second form of compatibility is the compatibility shown in the example with LCC and DDC. An expression (term or sentence) of language P converts into a sentence of language Q. We coin this type *sentence compatibility*.

There is a third type of compatibility. That is the compatibility defined in the last definition in chapter 2 of this paper. It is not necessary that each expression of language P converts to an expression of language Q. In many cases the subject of a document is not expressed by one expression but by a combination of expressions. The whole combination can convert to one expression in another information language. A book containing LCC-Dewey and Dewey-LCC tables[5] has the Library Congress Subject Headings *„Classification, Library of Congress"* and *„Classification, Dewey Decimal"*. The correct notation in the Basisclassification is 06.72 Subject indexing. This also is a type of compatibility. The term we shall use for this type is subject compatibility.

Much of the confusion in the debate on compatibility is because this distinction in types of compatibility is not made. It is quite possible that two information languages are compatible according one or two types only, and not according the other type(s). I shall give an example in the next chapter.

---

[5] M.L. Scott. Conversion tables : LC-Dewey, Dewey-LC. — viii, 365 p. — Englewood : Libraries unlimited, 1993

INFORMATION LANGUAGES

Information languages are artificial languages used for indexing and searching of documents according to subject content. There are several types: classifications, subject heading systems, thesauri. For our purpose this distinction is not very relevant. Far more important is the degree of precoordination. I shall show this by an example.

I have two precoordinated (enumerate) classificationsystems for literature.

Classification 1 for literature.

1. English literature
1.1. English prose
1.1.1. English prose through 16th century
1.1.2. English 17th century prose
1.1.3. English 18th century prose
1.1.4. English 19th century prose
1.1.5. English 20th century prose
1.2. English poetry
1.1.1. English poetry through 16th century
1.1.2. English 17th century poetry
Etc.
2. French literature
2.1. French prose
2.1.1. French prose through 16th century
Etc.
2.2. French poetry
2.2.1. French poetry through 16th century
Etc.
3. Russian literature
3.1. Russian prose
3.1.1. Russian prose through 16th century
Etc.
3.2. Russian poetry.
3.2.1. Russian poetry through 16th century
Etc.

Classification 2 for literature.

7 Prose
7a Prose through 16th century
7b 17th century prose
7ba English 17th century prose
7bb French 17th century prose
7bc Russian 17th century prose
7c 18th century prose
7ca English 18th century prose
Etc.
7d 19th century prose
7da English 19th century prose

Etc.
7e 20th century prose
7ea English 20th century prose
Etc.
8 Poetry
8b 17th century poetry
8ba English 17th century poetry
8bb French 17th century poetry
8bc Russian 17th century poetry
8c 18th century poetry
Etc.
8d 19th century poetry
Etc.
8e 20th century poetry
Etc.

Both classifications are built using the same characteristics of division and both are correct according to the classical rules for enumerative classifications. The first one is better if we take literary warrant into account, but that is not relevant here.

On the lowest level both classifications are compatible in both directions in all meanings of the word compatibility. On the higher levels there is no compatibility at all.

I can also use a facet classification:

A	Language	B	Form	C	Period
Aa	English	Ba	Prose	Ca	through 16th cent.
Ab	French	Baa	Novels	Cb	17th century
Ac	Russian	Bab	Short stories	Cc	18th century
Ad	Spanish	Bb	Poetry	Cd	19th century
		Bc	Drama	Ce	20 century

It is easy to see that each notation from both enumerative classifications can be translated into (composed) notations of the facet classification. The notation 2.1.1 of *Classification 1* convert to AbBbCa of the facet classification (and to 7ba of *Classification 2*). Notation 8c of Classification 2 becomes BbCc in the facet classification and cannot be converted into a notation of *Classification 1*.

Term and sentence compatibility exist between the enumerative classifications and the facet classification, but the subject compatibility is partly only. When we consider notation 7c ('18th century prose'), we see that there is no way to find out without re-indexing which one of the three possibilities Ba, Baa and Bab is the correct one for a document classified in class 7c of enumerative classification 2. The facet classification has a higher specificity in chain (more hierarchical levels) than the enumerative classification. If the notations Baa and Bab are deleted from the facet classification then we can say that these enumerative classifications are fully subject compatible with the facet classification in every meaning of the word. The conclusion is that an enumerative classification can be fully sentence compatible with a facet classification with the same domain and the same or less

specificity. Given a well-formed facet classification this is true for many enumerative classifications.

In the other direction not all possible notations from the facet classification can be translated in a notation from the enumerative classification even if the specificity is not greater than that of the enumerative classification. In the example some (complex) notations of the facet classification can be translated into a notation of both enumerative classifications, some in one of one of the two classifications only. The facet classification is partly compatible with the enumerative classifications[6].

The mutual compatibility of facet classifications depends on the domain, the facets used and the specificity of the languages. A less specific language is never fully compatible with a more specific one, even when both languages cover the same domain. In general we can say that the chance that an enumerative classification is compatible with a facet classification (within the same domain and with the same or less specificity in chain) is much greater than the chance that the enumerative classification is compatible with another enumerative classification with the same domain and the same specificity in chain.

For information languages that use words from a natural language as entries (*word systems*) we can set up the same discussion.

The subject heading[7] *Barock / Literatur / Deutsch / Europa / Rezeption / Geschichte 1600-1800* has no direct equivalent in the LCSH. We can use *German literature — Early modern 1500-1700 / history combined with German literature — -18th century — history and criticism*, but even combined they have not the same meaning as the RSWK subject heading.

With a coordinated thesaurus based on facet classification[8] the situation is different: each part of the complex RSWK heading can be translated into a separate descriptor, if the specificity of the thesaurus is not too great. Of course we have to pay a price: the context between the parts of the subject heading is lost.

Conclusion:

Enumerative information languages are in almost all cases only for a small part compatible with each other[9] but often fully compatible with a facet

---

[6] Some argue that it is not necessary to choose between the three possibiblities in the facet classification when a conversion is made from notation 7c of classification system 2 into the facet classification. That is true as long as only term or sentence compatibility is wanted. If one aims at subject compatibility then the choice is inevitable.

[7] This example is taken from the German RSWK handbook. In the RSWK system chains of descriptors are handled as precoordinated subject heading.

[8] The restriction 'based on a facet classification' is necessary. There are thesauri based on semantic characteristics of the natural language they derived from, with sometimes quite a lot of precoordinated descriptors.

[9] In theory it is possible that a very detailled enumerative information language is fully compatible with a very broad enumerative information language. The Library of Congress Classification with about 400,000 notations is almost fully compatible with the Dutch Basisclassification with 2200 notations. Partial compatibility is almost always available. Even different editions of the same enumerative information languages are not always fully compatible with each other. The Spanish 20th edition of Dewey Decimal Classification is a translation of the English 20th edition, except for the geographical table. That table is translated from the geographical table that will be part of the English 21th edition.

classification or a faceted thesaurus, given the same domain and the same or less specificity in chain.

Information languages that are enumerated in some parts and faceted in others are potentially more compatible the less precoordinated they are.

We can now reword the paradox of Lancaster and Smith:

> Perhaps somewhat surprisingly, while controlled vocabularies tend to promote internal consistency within information systems the more they are precoordinated, they also tend to reduce intersystem compatibility the less they are faceted.

SUBJECT COMPATIBILITY

In most documents the subject is reflected in one or more sentences of an information language. The situation that one single term can adequately express the subject is an exception. Translating term by term is — as said before — often not possible. Converting from sentence to sentence is possible in theory in many cases, but the number of possible sentences is often very great. The LCC has more than 200,000 notations, but even in this classification that is often considered as the proto-type of an enumerative classification, the number of notations increases tremendously when all possibilities for number building are used[10]. In the Dewey Decimal Classification the number of notations in the tables is not very great, but the DDC has more possibilities to build new notations than the LCC. The number of possible sentences in the UDC runs into millions, taking into account all possibilities of parallel subdivision and combinations. The number of sentences in word systems that form precoordinated chains of descriptors can be very great too. Examples are the Library of Congress Subject Headings and the German RSWK.

The sheer number of notations and sentences makes the building of conversion tables for these large enumerated systems in practice impossible[11].

Conversion tables for subject descriptions are not feasible at all. The sheer number of possible combinations of sentences forbids it.

When we look at existing conversion tables, as the one made by Scott for LCC — DDC and DDC — LCC then we can conclude that using these tables for direct conversion means a great loss of information. These tables serve as a help for indexers who have to index the same document in both information languages, they will find out in which part of the tables of the other language to look for the adequate notation.

There is no reason to assume that direct conversion from one more or less enumerative classification or subject heading system to another information language of the same kind can be done without a considerable loss of information,

---

10 See footnote 5).

11 The conversion table between the LCC and the Dutch Basisclassification is made by first making a conversion table between the Conspectus and the BC. In a second round all cases where one Conspectus class had more then BC notation were compared with the LCC itself.

even though there are exceptions, where the degree of specificity in chain is very different.

There is another point: why should we want to convert from one more or less enumerative classification or word system to another one? In a time where most bibliographic files are machine searchable databases enumerative information languages are not the best possible instruments for searching. Postcoordinate information languages, that enable the use of Boolean operators, seem to be far more appropriate. When we have to do something, because bibliographic databases are merged, it is much more adequate to convert subject descriptions expressed in an enumerate information language into terms from a coordinated information language. The best solution is to convert into descriptors from a thesaurus based on a facet classification.

It is my conviction that such a conversion in many cases can be done automatically, using the complete subject description given to a document. The more faceted the source information language, the more easy it can be done. The following instruments are necessary:

— A conversion table between the separate terms of the source language and the target language (the thesaurus). In this table some terms will convert to a combination of descriptors.

— An algorithm to break sentences of the source language into its component terms or parts with as less precoordination as possible.

Given these instruments, the sentence(s) of the subject expressions are broken down into the separate terms, the doubles are deleted and the remaining terms are converted into descriptors. The algorithms are quite simple for faceted classifications and languages like the LCC that are almost complete enumerative, more complicated for information languages like the DDC and the UDC, which are in some respects enumerative and in other respects are faceted. It should not be concealed that in this process information is lost. The less precoordination in the information language the more false drops in searching. In an enumerative language the context of the concepts expressed is known, in a sentence of a coordinate information language there is often no context[12].

The vocabulary of the source language can be used to build a thesaurus[13]. This has the great advantage that „automatically" the thesaurus has the same domain and the same specificity as the source language. In this way we ensure a very high if not full compatibility. For precoordinated terms of the source language a combination of descriptors has to be given in the conversion table.

In a situation where we have more than one information language in the same database we can try to convert all to the same faceted information language. This faceted language needs a domain that covers the domains of all source languages and a specificity that is the same as that of most specific source language. The

---

[12] Devices to express context in coordinate information languages exist, but are not often used. Examples are rigid facet formula's, links and roles.

[13] J. Aitchison. A classification as a source for a thesaurus : The Bibliographic classification of H.E. Bliss as a source of thesaurus terms and structure. — In: Journal of documentation 42(1986)3,p.160-181 and G.J.A. Riesthuis and S. Bliedung. Thesaurification of the UDC. — In: Tools for knowledge organization and the human interface. Vol. 2. Proc. 1st. Int. ISKO Conference, Darmstadt 14-17 Aug. 1990. — Frankfurt : Indeks Verlag, 1991. — p.109-117.

most specific source language is in this case fully compatible with the target language, with all other source languages information is lost with conversion. The loss of information is greater as the difference in specificity is greater.

In this paper I have given a definition of compatibility based on the idea that full compatibility means that after the conversion the document has an expression of the target language that is same as the one it would have been given when it was indexed directly in the target language. Given this definition term-to-term compatibility is not very sensible in most cases. I further discussed the conditions for fully compatible information languages. Only a small proportion of all possible pairs of information languages is fully compatible and even then only in one direction. In the last part I discussed the possibilities to converted from all types of information languages into a faceted thesaurus or facet classification.

SUMMARY

Since the 1960's the idea of converting codes or terms from one information language to another has filled a bookcase. In many proposals for automatic conversions and conversion tables a great optimism is showed. But in practice there are only few examples of well working conversions. If automated conversion is used the loss of information is not neglectable in most cases.

In this paper the reasons for this discrepancy will be treated, based on theoretical literature and reports published since 1960.

Ingetraut Dahlberg

THE COMPATIBILITY GUIDELINES — A RE-EVALUATION

COMPATIBILITY FOR COOPERATION

Aware of the new freedom which subject indexers gained by giving up the „old" classification systems in the early sixties and using — instead of a notational or coded approach — natural language as the access medium to knowledge or information stored in the new literature, many thesauri were born and started to be used. However, it was also soon realized that the capability attached to the common use of an existing classification systems, namely to have for every user of such a system one and the same reference basis, this capability was lost with the diversity of the new thesauri. Was there a bad conscience in those who abandoned the old systems and apparently with them the basis for cooperation? Anyhow, soon the insight seemed to have prevailed that something had to be done in order to regain this lost capability of a common reference tool.
Thus, the idea of trying to establish compatibility between thesauri of a related subject field arose already after the first thesauri were elaborated. This idea had not been considered necessary in some earlier times, when in general universal or specialized classification systems were used for the classification of the current literature. The negative experiences which C.E.Müller (1) made (1961) when trying to establish a concordance between a subject classification (e.g. in the building field) and the UDC[1] resulted in her statement about the impossibility of such an undertaking, and she found these her insights confirmed in earlier attempts by Donker Duyvis in 1935 and George Lorphčvre in 1949. Apparently therefore no other projects were attempted to reconcile for instance the DDC with the LCC or with the RCC or BBC in the years before the thesaurus movement started, which was after 1958. However, in the years thereafter, we do find a relatively great number of papers on this subject and even a conference (2) in 1965 and already a literature review on these questions (3) in 1966.

COMPATIBILITY FOR ECONOMIC REASONS

In the UNISIST Report of 1971 (4) the question of compatibility or convertibility of indexing languages was stressed and recommendations were made more or less for economic reasons, facing the fact that the abundance of created thesauri

[1] The full names of the universal classification systems abbreviated here are given in Section 12.

called for some means for switching between systems and sharing the results of indexing work. Indeed, most of the institutions had recognized the necessity for itself to develop such a thesaurus for the sake of complying with its own special collection. But somehow these UNESCO recommendations seem to have been neglected by those institutions, especially in countries in which there was no necessity of collaboration for economic reasons. A quite different situation existed in the socalled Comecon countries where the socialist idea of collaboration — although mostly enforced by governments — resulted in a number of conferences and papers with proposals on how to proceed in this matter and with actual work towards an establishment of compatible ordering systems in all subject fields, even between universal systems such as the socalled Rubricators in the USSR on a national and international level. But we heard it from friends in these countries: the more such a collaboration was enforced from above the more the colleagues in charge would resist and try to keep their activities and knowledge for themselves.

RECOMMENDATIONS FROM A THESAURUS TEACHER

When in 1974 D.Soergel published his famous „Indexing Languages and Thesauri: Construction and Maintenance" with the very valuable last chapter K (Thesauri as a basis for cooperation in information services) (5) providing a detailed description on how to deal with the problems of compatibility and convertibility he gave his readers a number of practical guidelines on how to proceed with the conceptual approach to compatibility. But alas, these seem to have been likewise neglected, they too did not result in the necessary activities on the part of the centers concerned. Also the final proposal in his Chapter K in which he showed a further method on how to overcome the unhealthy situation of the existence of too many special thesauri by creating one Universal Source Thesaurus from which everybody could serve himself for his special need was — for the years to come — not considered as yet a worthwhile necessity.

EXPERIENCES IN COMPATIBILITY RESEARCH IN DOCUMENTARY LANGUAGES

This was the title of a paper given by G. Wersig at the 3rd FID/CR Conference at Bombay, 1975 (6) in which he outlined the theoretical framework proposed for the comparison and correlation of the different thesauri and other ordering systems of the German Federal Government Departments with the result of elaborating on the basis of the concepts of all of these systems a „Federal Macro Thesaurus". Although the proposal was well elaborated and richly documented, it did not find the support of the reviewers and the German Government, subsequently its practical aims could not be achieved. Nevertheless it still remains for us to consider Wersig's well-explained theoretical considerations in this Bombay contribution for any further attempts to elaborate guidelines in this area when advanced computer technology for such a macrothesaurus will be available.

REVIVAL THROUGH THE SOCIAL SCIENCES?

In the early eighties a short revival of the compatibility ideas and aims mentioned came from the part of the social sciences. A few conferences, such as the one in Prague in 1980 (7), at the UNESCO in Paris in 1980 (8), in Bielefeld during the CONTA Conference of 1981 (9), and in Columbus before the ASIS Conference in 1982 (10) brought with them also the new idea of Thesaurus Integration, meaning to accumulate a number of computerized thesauri in a subject area for the sake of creating a common pool of their elements so that users could share the concepts of such an area. Such an integrated thesaurus would also reveal the different relationships having been established by their thesaurus authors.
A series of articles appeared on this topic by Aitchison (11), Whitelock (12), and Sager/Somers/McNaught (13) and much hope was fostered for the clarification of concepts in just these very crucial fields of knowledge. The investigations by Sager et al. had already involved some pilot studies on which they reported in their threepartite series in IC 1981-82. But, for what reasons whatsoever, the Integrated Thesaurus of the Social Sciences could not find a sponsor either, thus the idea is still waiting for its realization.

A LITTLE TOOL AND ITS APPLICATION

The interest in a solution of the compatibility problems of indexing languages prompted the UNESCO in 1980 to sponsor a project called „Guidelines for the Establishment of Compatibility between Information Languages in the Social Sciences" (14). In these guidelines which were published in 1981 (15) and somehow supplemented in 1982 by a kind of specification in extending on „conceptual compatibility" (16) a tool was introduced called Compatibility Matrix. This matrix was to serve several purposes, first of all for the establishment of verbal compatibility (e.g. of the coincidence of the terms of different indexing languages), thereafter for the establishment of conceptual compatibility on different levels of the hierarchy. This tool was applied in a number of investigations, e.g. in the social sciences (17), in culture (18) and (19), and quite recently in Poland even for the comparison of three universal ordering systems in four subject areas (20). However, each of the cases mentioned here served just only as an example, showing what the method involved and demonstrating that it worked. It was not as yet — to my knowledge — applied to correlate in this way entire systems.

WHAT HAMPERED PROGRESS?

Neither was the idea of necessary cooperation, freely or imposed, strong enough to serve as „movers" in the many attempts to proceed in comparing and correlating the concepts of thesauri or classification systems in certain fields of knowledge, nor was it the need for economic considerations. I assume that in most of the cases the projects proposed were not granted the necessary financial help simply for reasons of the rather huge amount of money deemed necessary for their

realization. Also, the insight might have lacked about the usefulness of the results to be achieved. And last not least, on the part of the systems involved, there may have been reservations against sharing with others what one has stored, most of the centers would rather hide their „treasures", sometimes even for reasons of competition.
One might, however, take the fact that when in the late sixties many institutions with their special thesauri got together in order to create one big thesaurus for many subject fields, such as the Thesaurus of Engineering and Scientific Terms (TEST), published 1968 in cooperation between the Engineering Joint Council and many US Departments, also the US Defence Department (21) this could be regarded as a wayout or another kind of solution to overcome a missing network of correlated thesauri. Here the experts agreed to compromise for the sake of having just one big tool which included their special input instead of a plethora of small but always growing tools for their special purposes. In a similar way the Root Thesaurus of the British Standardization Organization (22) could be looked at as an offer to use the structure, contents and software of this systematically arranged and faceted thesaurus as a sort of a Universal Source Thesaurus in the sense meant by D. Soergel although in this case not really „universal". In both cases mentioned enormous efforts and much money had been invested for the creation of such macrothesauri, but at least they were „put into being" and are still used in the professional world, serving also as examples for similar efforts.

PROGRESS EXAMPLES HERE AND THERE

Niehoff reported already in 1976 on the computerized integration of 11 analyzed vocabularies in the energy field (23) and the creation of a prototype conversion guide (a „synonym table"). The methodology consisted of 1) establishing term selection criteria, 2) analyzing individual system vocabularies for energy-related terms, 3) processing energy subsets, and 4) reviewing the integrated product and generating a final vocabulary. In this case one did not deal only with entire systems but included more comprehensive thesauri, such as TEST and the INIS Thesaurus and needed to make selections from them.
Further examples of progress are the Unified Medical Language System (UMLS) (24) started in 1987 in the United States, the Universal Agricultural Thesaurus as a cooperative effort of institutions in USA and Europe which began 1989 (25). In the education field the European Education Metathesaurus was proposed by EURYDICE (1990) (26). We will hear about these further efforts later during this Seminar.
This Seminar was challenged by the fact that the rather difficult situation to create a comprehensive ordering system for knowledge organization in the environmental sciences as shown during the Conference on Environmental Knowledge Organization and Information Management held at Bratislava, Sept.1994 called for solutions to handle the mapping and networking of different thesauri, also in different natural languages for the purpose of indexing, storage, and retrieval (27),(28).

NEW GUIDELINES TO ESTABLISH COMPATIBILITY FOR DIFFERENT PURPOSES?

Among the purposes of the Seminar as stated in the Call for Papers we felt that we should make an endeavor to

„reassess and evaluate general and possibly also special guidelines for the establishment of compatibility and integration of new and existing order systems".

For those who had not seen the Unesco Guidelines mentioned their outline is provided as follows:

Guidelines for the Establishment of Compatibility between Information Languages in the Social Sciences

Prepared for the Unesco Division for the International Development of Social Sciences, Paris by I.Dahlberg. 28.11.1980

1. Purpose
2. Previous work
3. Scope, limitations, advantages
4. Methods for the establishment of verbal comparisons between information languages (IL)
4.1 Recording of the elements of an IL
4.2 Establishment of an alphabetical comparison matrix M1
5. Method for the establishment of conceptual comparisons
5.1 Comparing concepts on the verbal level
5.2 Conceptual reorganization of matrix M1
5.3 Establishment of a compatibility matrix M2
5.4 Additional information
5.5 Establishment of system-related matrices M3
5.6 Establishment of hierarchical matrices M4
5.7 Index to the compatibility matrix
6. Treatment of compatibility problems
7. Establishment of compatibility between ILs
8. Use of a compatibility matrix
9. Organizational problems

As this task was commissioned by the Social Sciences Division of Unesco, the examples provided for the different matrices related to one of their fields, i.e., Social Welfare. Figures 2-5 show 1) the fields of a concept record, 2) the occurrence of concepts in the Dewey Decimal Classification (DDC), the Bliss Bibliographic Classification (BBC), and the UNESCO Thesaurus (UNT), 3) an example of a verbal compatibility matrix, and 4) an example of a part of a conceptual compatibility matrix.

Now, after 15 years have passed and economic reasons cry, so-to-speak, for cooperation all over the world, it might be timely to reconsider whether these guidelines could be recommended for general use, whether they would need an updating, extensions, revision, etc. or whether they should be replaced by other tools which would serve better the intended purpose.

(1) Name of concept of class
(2) Notation
(3) Next broader concept
(4) Highest concept in hierarchy/subject category
(5) Indication of hierarchical level of concept
    (A) highest level
    (B) next lowest level
    (C) third level, etc.
(6) Number of subconcepts, if comparison only on a certain level, in brackets for each level
(7) Form category of concept
    (O) Object, entity
    (P) Process, activity, state
    (Q) Quantity, quality
    (R) Relation
    (S) Space-related concept
    (T) Time-related concept
    (W) Subject-field or discipline
(8) Definition of concept (if necessary and posible)
(9) Other names of concept or class
(10) Source of concept abbreviation of IL accord. to (29)
(11) Remarks

Fig. 1: Fields of a concept record

DDC:	361	Social problems and social weffare
	361.1	Social problems
	361.2	Social action
	361.3	Social work
	361.4	Group work
	361.6	Public action
	361.7	Private action
	361.8	Community action
	361.9	Historical an geographical treatment
BBC:	Q	Social Welfare
	QAG	Social Welfare Administration
	QD	Social Work
	QE	Social Services
	QF	Social Security
	QG	Persons in need, causes of need
	QN	Deviants
	QO	Crime, criminology
UNT:	R85/99	Social Welfare
	R86	Social welfare philosophy
	R87	Welfare policy
	R88	Social welfare planning
	R89	Social welfare economics
	R90	Social welfare administration
	R91	Social workers
	R93	Social work
	R94/99	Social services
	R95	Social security
	R96	personal social services
	R97	Health services
	R98	Disaster relief work

Fig. 2. 'Social Weflare' in the UDC, BBC and UNT.

No.	Name	DCC	BBC	UNT	vc
1	Community action	361.8	QDP	R93.20.10*	3
2	Crime. Criminology		QO	R75/78	2
3	Deviants		QN		1
4	Disaster relief work		QGN	R98	2
5	Educat. welfare		QEN-P	J14.15	2
6	Group work	361.4	QDN	R93.10	3
7	Health services			R97	1
8	Personal soc. serv.			R96	1
9	Persons in need		QG		1
10	Private action	361.7			1
11	Public action	361.6			1
12	Social action	361.2		P71.20	2
13	Social problems	361.1		R70/84	2
14	Social security		QF	R95	2
15	Social services	361-365	QE	R94/99	3
16	Social welfare		Q	R85/99	2
17	Soc. Welf. admin.		QAG	R90	2
18	Soc. welf. economics		QAT M/Z	R89	2
19	Soc. welf. philosophy		QAE	R86	2
20	Soc. welf. planning			R88	1
21	Social work	361.3	QD	R93	3
22	Social workers		QB	R91	2
23	Welfare policy			R87	1
					42

* A thesaurus without notations will be entered only through its descriptors.

Fig. 3: Example of verbal compatibility matrix

No.	Name	DDC	BBC	UNT	cc
65	Social Welfare	361 Social problems and social welfare	Q Social Welfare	R85/99 Social Welfare	3
65.1	Soc. welf. philos.	361.01 Philos. & theory of	QAE Philos. of soc. welf.	R86 Soc. welfare philosophy	3
65.2	Soc. welf. admin.	—	QAG Soc. administration Soc. welf. administration	R90 Social welfare administration	2
65.3	Welfare policy	>361.25 Action within establ. soc. framework (policy)	QAG P Policy (in Social Welfare)	R87 Welfare policy	3
65.4	Soc. welf. planning	>361.25 Act. within establ. soc. framework (planning)	QAH Planning for welfare, social planning	R88 Social welfare planning	3
65.5	Soc. welf. econom.	—	QAT M/Z Management of soc. welf.	R89 Social welfare economics	2 2

Fig. 4. Examples of section of a conceptual compatibility matrix

REASONS FOR WORKING WITH COMPATIBILITY MATRICES

The visual display is favored today

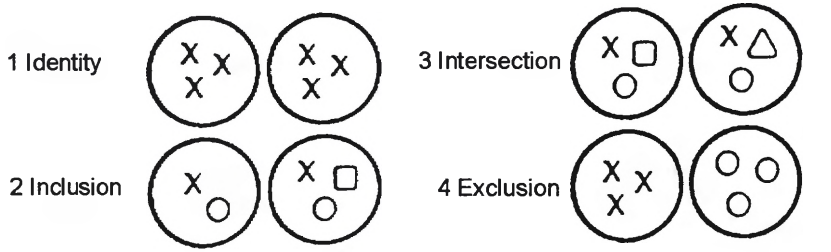
Mankind is living much more with pictures today than ever before and the computer screen is a picture too. If relationships are shown in a matrix form, they are much easier and faster understood than in a linear presentation. Therefore it seems that this holds also for the visualization possibility of verbal and conceptual compatibility.

The verbal and the conceptual display

Although the display of concepts on the verbal level seems to be useless as only the conceptual relationships matters, it is a necessary precondition for any work in one natural language in order to assess whether two or more different order systems are relating to the same subject field. If more comprehensive systems, as e.g. TEST is included in the comparison, selections of terms have to be made beforehand, possibly on the basis of the hierarchical groupings given with such a system.

With regard to the conceptual display it must be stated that the task of establishing conceptual compatibility for the purposes of indexing, storage and retrieval presupposes knowledge about the analysis and construction of concepts. Any mapping of terms is useless if their contents has not been established beforehand by a definition of their concepts. It does not really matter how a concept or a class in a classification system is called in one or in more than one natural language, what matters alone is its definition or its contexts of use. This was already stated by Soergel in 1974 and further explanations on how to understand concepts as knowledge units as well as their elements and their relationships were given in (29) and (30).

In looking at concepts as knowledge units comprising characteristics as their knowledge elements one can easily understand that a comparison of concepts according to their elements will show also the relationships between concepts and consequently also the networks or systems evolving from such relationships. In comparing concepts for reasons of compatibility one will, however, at first be interested in the formal relationships which can be derived from the situation of the concept elements to be stated. Figure 5 will show this clearly:



It was found in preliminary investigations as reported by Wersig in (6) that the relationship of inclusion was the one to have occurred most of the times in the comparisons of the elements of the 4 different systems studied. In our guidelines we had introduced the mathematical symbols < and > to indicate broader or narrower concepts in comparison, which equals the relationship of inclusion in two directions.

The faceted display

The Guidelines of 1980 favor the structuring of concepts according to categories or — with respect to their occurrence in subject fields — facets. A faceted arrangement of elements of subject fields can follow the representatives of concepts occurring in the propositions possible in one subject field. A generalized form of such a faceted structure has been introduced as the Systematifier Principle. Facets provide for mutual exclusiveness of concepts, i.e. they locate concepts only in one place and this takes care of a kind of reproducible order principle, highly welcome especially in cases were such arrangements have to be accomplished by different persons.

The complete display

If a number of order systems in one and the same subject field will be merged it goes without saying that the number of concepts will be more comprehensive and thus more complete with regard to the filling of facets with their necessary concepts. It would indeed systematically bring all concepts together which exist in a certain subject field. And not only this, it would also juxtapose to all the concepts existing in a field in faceted order in the left column of a matrix all those concepts which are used by the different systems included in the comparison.

ADVANTAGES OF AN ESTABLISHMENT OF COMPATIBILITY MATRICES

It can be imagined that to follow the guidelines and to set up the matrices in a faceted way as well as to reorganize all those concepts occurring in the order systems of a given subject field after careful examination of their definitions, will become a tremendous job. However, the result will be equally rewarding, as this will facilitate to survey the relevant existing concepts in a field and thus will be of great help in serving all the searchers in universal and special libraries, documentation centers and especially also terminology establishments, and it would not necessarily imply that the existing centers would have to give up their own ordering systems for the sake of the ordered arrangement supplied by the faceted structure of the matrices, nevertheless it will help them to realize

— how does one's own system compare with others?

— what are the concepts lacking in one's system?

— what are the concepts already existing in a field at all? and

— which concepts have not as yet been designated as such and may still be missing?.

The faceted arrangement of concepts demands provision of a notation. By a notation the system position of a concept is fixed and this means that the contents

of a given concept is also fixed by the hierarchy and sequence of the faceted structure. Thus, the conceptual contents is not only manifested in the verbal expression of a concept but also in the notation. This implies that one would not need to use terms to denote concepts but could use as well only the notation. Ultimately one would create by such matrices also the tools that are necessary for a comparison of concepts in different languages with all the benefits for terminologists and translators as well as of course for the indexing and retrieval purposes in information science.

PROPOSED STEPS TOWARD REALIZATION OF THE „BLACK BOX"

At present the published books and many other items are classed by at least six different universal classification systems at the same time, five of them in the English language, one in Russian. These are the

Dewey Decimal Classification (DDC)
Universal Decimal Classification (UDC)
Library of Congress Classification (LCC)
Bliss Bibliographic Classification (BBC)
Ranganathan's Colon Classification (RCC) and the Russian
Library Bibliographical Classification (LBC)

This is not such an exciting diversity as the diversity of natural languages which use these systems.
It would be foolish to assume that these systems could be merged into one as all of them have their own founders, owners and users and they have their history, their institutions, their way of life. And they are also competitive and struggle for survival.

However, it seems to be timely to be able to switch between these systems, at least on a higher level of abstraction. More than twenty years ago I published a dream — the realization of a Black Box to relate these six universal systems conceptually to one meta-system that comprises all of their concepts on a certain level of abstraction in order that a user can enter this Box with a term or a notation of one of these six systems and be advised under which notation of another universal system he/she would find further pertinent literature.

This dream could become a reality if the information profession wants it. It could be realized step by step and field by field, especially after the major systems are now being computerized.

If one would go about to relate one system to the next, one would get the following 15 comparisons:

DDC-UDC, DDC-LCC, DDC-BBC, DDC-RCC, DDC-LBC
UDC-LCC, UDC-BBC, UDC-RCC, UDC-LBC
LCC-BBC, LCC-RCC, LCC-LBC
BBC-RCC, BBC-LBC
RCC-LBC

However, if one would use a metasystem for switching, like e.g. the Information Coding Classification, ICC, one would need only the six relationships of each universal system to the ICC, as shown in Fig.6:

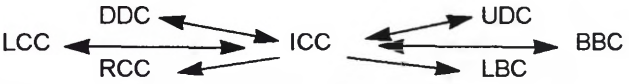


Fig.6: Switching between six universal systems via ICC.

As all the six systems are discipline-oriented the ICC must contain disciplines which it does, next to mission-oriented subject fields to correspond to the new needs not covered as yet in the main classes of the six systems. Indeed the ICC is based on the category of general objects which correspond to nine ontical levels of being and which are further differentiated by a faceted structure, a kind of application of the Systematifier mentioned.

Fig.7 shows this system in a matrix form with the first two levels of abstraction. The system has been elaborated in German with English equivalents for most of the ca. 6500 subject fields and would probably serve very well for such a switching purpose, especially since its decimal notation is very easy to handle[2].

Any other universal system such as the Japanese or Chinese universal classification systems could be introduced later on or attached via comparison with one of the six universal systems.

It is proposed to start eventually working for the Black Switching Box to relate between our main universal classification systems. May I ask, would this be a suitable project proposal resulting from this Research Seminar?

REFERENCES

1 Müller,C.E.: Konkordanzen zwischen Fachgebietsklassifikationen und der Internationalen Dezimalklassifikation. DK-Mitt. 6(1961)No.4, p.13-15

2 Newman, S.M. (Ed.): Information Systems Compatibility. New York: Spartan Books; London: MacMillan 1965. 150p.

3 Henderson, M.M., Moats, J.S., Stevens, M.E., Newman, S.M.; National Bureau of Standards: Cooperation, convertibility, and compatibility among information systems: a literature review. Washington,DC: U.S.Governm.Printing Office 1966. 140p. 660 refs.

4 UNISIST Study Report on the Feasibility of a World Science Information System. Paris: UNESCO 1971. 161p.

5 Soergel, D.: Indexing Languages and Thesauri: Construction and Maintenance. Los Angeles, CA: Melville Publ.Co.1974. 43+632p.

6 Wersig, G.: Experiences in Compatibility Research in Documentary Languages. In: Neelameghan, A.(Ed.): Ordering Systems for Global Information Networks. Proc. 3rd. Int. Study Conf. on Classif. Research, Bombay 1975. Bangalore: FID/CR and Sarada Ranganathan Endowment for Library Science 1979. p.423-430

[2] The ICC had been conceived in 1970-1974, was first spoken of at a conference in Ottawa, 1971 and its basis was published in 1974 (31). It was applied in several research projects and presented in its first two subdivisions 1980 in Salzburg. With its first three subdivisions it was published in 1982 in (32).

0 GENERAL FORM CONCEPTS	01 THEORIES, PRINCI-PLES	02 OBJECT, COMPO-NENT	03 ACTIVITY, PROCESS	04 PROPERTY ATTRIBUTE	05 PERSONS OR CONTD	06 INSTITU-TION OR CONTD	07 TECHNOLO-GY & PRO DUCTION	08 APPLICA-TION & DE-TERMINANT.	09 DISTRIBU-TION & SYNTHESIS
1 FORM & STRUC-TURE AREA	11 Logic	12 Mathematics	13 Statistics	14 Systemology	15 Organization Science	16 Metrology	17 Cybernetics, Control & Automation	18 Standardiza-tion	19 Testing and Monitoring
2 ENERGY & MATTER AREA	21 Mechanics	22 Phusics of Matter	23 General and Technical Physics	24 Electronics	25 Phisical Chemistry	26 Pure Chemistry	27 Chemical Technology & Engineering	28 Energy Science and Technology	29 Electrical Engineering
3 COSMO & GEO-AREA	31 Astronomy & Astrophysics	32 Astronautics & Space Research	33 Basic Geosciences	34 Atmospheric Sciences & Technology	35 Hydrospheric & Oceanol. Sc. & Technology	36 Geological Sciences	37 Mining	38 Materials Science & Technology	39 Geography
4 BIG-AREA	41 Basic biological Saiences	42 Microbiology and Cultivation	43 Plant Biology and Cultivation	44 Animal Biolog and Breeding	45 Veterinary Sciences	46 Agriculture & Horticulture	47 Forestry & Wood Sci. & Technology	48 Food Science and Technology	49 Ecology and Envitonment
5 HUMAN AREA	51 Human Biology	52 Health and Theoretical Medicine	53 Pathology and Practical Medicine	54 Clinical Medicine & Cure	55 Psychology	56 Education	57 Profession Sci., Labor, Leisure	58 Sport Science and Sports	59 Household and Home Life
6 SOCIO AREA	61 Sociology	62 State and politics	63 Public Administra-tion	64 Money and Finances	65 Social Aid, Social Politics	66 Law	67 Area Planning, Urbanism	68 Military Science and Technology	69 History Science and History
7 ECONOMICS & TECHNO-LOGY AREA	71 General and National Economics	72 Business Economics	73 Technology in general	74 Mechanical & Precision Engineering	75 Building	76 Commodity Science & Technology	77 Vehicle Science and Technology	78 Transportation Technology & Services	79 Utilities and Service Economics
8 SCIENCE & INFORMA-TION AREA	81 Science of Science	82 information Science	83 Informatics, computer science	84 Information in general	85 Communicat. Science	86 Mass Communica-tion	87 Printing and Publishing	88 Communication Engineering	89 Semiotics
9 CULTURE AREA	91 Language and Linguistics	92 Literature and Philology	93 Music and Musicology	94 Fine Arts	95 Performing Arts	96 Culture Sci-ences, nar-rower sense	97 Philosophy	98 Religion and Secret Teachings	99 Christian Religion

Figure 7: Information Coding Classification. Survey of Subject Groups.
©1982, rev. 1992 I.Dahlberg, Frankfurt/M, Germany

7 Palnicov, M.: ECSSID activities in the field of information languages compatibility. Paris: UNESCO 1980. 8p.
8 Litoukhin, J.: UNESCO's project for the establishment of an integrated thesaurus of the social sciences. In: (9), p.202-206
9 Riggs, F.W. (Ed): The CONTA Conference. Proc.Conf.on Conceptual and Terminological Analysis in the Social Sciences. Bielefeld, May 24-27, 1981. Frankfurt: INDEKS Verl.1982. 382p.
10 Svenonius, E.: Compatibility of retrieval languages. Introduction into a Forum. Int.Classif. 10(1983)No.1, p.2-4, 19 refs.
11 Aitchison, J.: Integration of thesauri in the social sciences. Int. Classif. 8(1991)No.2, p.75-85
12 Whitelock, P.J.: A descriptor bank of social science terms. Int.Classif. 9(1982)No.3, p.145-151
13 Sager, J.C., Somers, H.L., McNaught, J.: Thesaurus integration in the social sciences. Pt.1: Comparison of thesauri. Pt.2: Stages towards integration. Pt.3: Guidelines for the integration of thesauri. Int.Classif. 8(1981)No.2, p0.133-138; 9(1982)No.1, p.19-26; 9(1982)No.2, p.64-70
14 Dahlberg, I.: Guidelines for the Establishment of Compatibility between Information Languages in the Social Sciences. Prepared for the Unesco Div.for the Int.Development of Social Sciences, Paris Nov.1980. 25p.
15 Dahlberg, I.: Towards establishment of compatibitlity between indexing languages. Int.Classif. 8(1981)No.2, p.86-91
16 Dahlberg, I.: Conceptual compatibility of ordering systems. Int.Classif. 10(1983)No.1, p.2-4
17 Dahlberg, I.: The Broad System of Ordering (BSO) as a basis for an Integrated Social Science Thesaurus? Int.Classif. 7(1980)No.2, p.66-72 (based on contract work for Unesco-IDS)
18 Dienes, M.: Testing the compatibility matrix in the field of culture. In: (9), p.224-233
19 Dienes, M.: Structural differences in classification systems and the testing of the compatibility matrix in the field of culture. In: Perreault, J.M. et al: Universal Classification 2. Study Conf.on Classif.Research of FID/CR. Augsburg 1982. Frankfurt: INDEKS Verl.1983. p.112-125
20 Scibor, E., Tomasik-Beck, J.: On the establishment of concordances between indexing languages having a universal or interdisciplinary scope (Polish experience). Knowl.Org. 21(1994)No.4, p.203-212
21 Engineers Joint Council: Thesaurus of Engineering and Scientific Terms (TEST). New York: EJC 1968. 696p.
22 BSI ROOT Thesaurus. 2nd ed. 2 vols. Milton Keynes, Bucks.: Inform.Dept. Brit.Standards Institution, Linford Wood 1985 (1.1981). 580+620p.
23 Niehoff, R.T.: Development of an integrated energy vocabulary and the possibilities for online subject-switching. J.Amer.Soc.Inform.Sci. 27(1976)No.1, p.3-17
24 Humphreys, B.L.: Unified Medical Language System: Progress report. Int.Classif. 15(1988)No.2, p.85-86
25 Thomas, Sarah: Unified Agricultural Thesaurus discussed. Agricult. Libr. Inform. Notes 17(1991)No.1, p.12-14

26 Roulin, C.: Towards the European Education Metathesaurus. Brussels: Eurydice European Unit 1990.
27 Stancikova, P. et al (Ed.): Environmental Knowledge Organization and Information Management. Proc.1st European ISKO Conf. Bratislava, 14-16 Sept.1994. 2 vols. Frankfurt: INDEKS Verl. 1994. 232+96p.
28 International Society for Knowledge Organization (ISKO): Recommendations (from the Conference under (27)). In: (27), vol.2, p.70-74.
29 Dahlberg, I.: A Referent-oriented, analytical concept theory for Interconcept. Int.Classif. 5(1978)No.3, p.142-151
30 Dahlberg, I.: Conceptual Structures and Systematization. Int. Forum Inform. Doc. 20(1995)No.3, p.9-24
31 Dahlberg, I.: Grundlagen universaler Wissensordnung. München: Verlag Dokumentation 1974. XVIII,366p.
32 Dahlberg, I.: Information Coding Classification — principles, structure and application possibilities. Int.Classif. 9(1982)No.2, p.87-93. Also in: Classif.Systems and Thesauri, 1950-1982. Frankfurt: INDEKS Verlag 1982. p.107-132 = Int.Classif. & Indexing Bibliography, Vol.1.

SUMMARY

In 1980 „Guidelines for the Establishment of Compatibility between Information Languages in the Social Sciences" were submitted to UNESCO. A brief description of these guidelines in their different steps of developing matrices 1-4 and their index is given. The investigations made on the basis of these guidelines are considered. Although the task of establishing compatibility is a difficult and a time consuming concept-analytical work, it is a very rewarding one with respect to the venience and possibilities achieved for all appliers and users of any of the systems involved as well as for the general public. A re-evaluation of the necessary steps in the development with respect to the new technology available is presented.

THESAURI INTEGRITY — STRUCTURES AND SOFTWARE

Dagobert Soergel

DATA MODELS FOR AN INTEGRATED THESAURUS DATABASE

INTRODUCTION

This paper presents two data models for storing multiple thesauri in a single integrated database. Such a database can serve the following purposes:

— Aid searchers in finding the appropriate descriptors for a search, particularly a search in multiple data bases. The searcher could start from any term, or from general concepts, or from a known descriptor in another data base, to find the appropriate descriptors and/or free-text terms for the data base at hand.

— Produce indexing and searching conversion tables between index languages of different data bases and support semi-automatic indexing conversion and conversion of query formulations from one data base to another.

— Print subsets of the data base with subset-specific selection of preferred terms and descriptors. This can be used for producing any thesaurus that was used as a source in generating the data base, for improving and maintaining any such thesaurus, and for producing new specialized thesauri.

STRUCTURE OF THESAURUS DATA

A thesaurus deals with terms and concepts and the relationships among and between these entities. A term is a linguistic entity, a character string with meaning in a given language. If the same character string has two meanings, we have two terms (homonyms); most thesauri use parenthetical qualifiers to make each string unique. The same character string occurring in two different languages represents two different terms, even if the meaning is the same.

A thesaurus captures a great many relationships among terms, between terms and concepts, and among concepts. **Term-term relationships** include A *has morphological variant* B (such as *job* and *jobs*), A *has spelling variant* B

(such as *labor* and *labour*), and A *has synonymous term (ST)* B. More precisely, *has morphological variant* relates character strings that are derivative from the same stem. *Has spelling variant* relates stems and partitions the set of all stems into mutually exclusive groups. Each such group constitutes a normalized term; a preferred spelling variant can be selected to represent the term. (To keep matters simple, we mostly sidestep the spelling variant problem in this paper.) *Has synonymous term* relates normalized terms (preferred spelling variants) and partitions the set of normalized terms into mutually exclusive groups. A concept can be operationally defined as such a group of normalized terms. A preferred term can be selected from each group to uniquely designate the concept. All preferred terms may be used as descriptors, or descriptors may be further selected from the preferred terms. These considerations give rise to a status hierarchy among all terms (character strings).

The primary **term-concept relationship** is Term A *designates* Concept B; this relationship is implied by *has synonymous term* relationships, unless a thesaurus identifies concepts independently, for example through class numbers or notations. **Concept-concept relationships** include A *has broader term* B, A *has narrower term* B, A *has related term* B.

This simplified picture presents clear-cut distinctions, but reality is not that simple. Normalized terms often represent shades of meaning so that it is hard to tell whether two terms are synonyms or whether they represent closely related but different concepts. If the two concepts that are so closely related that to distinguish between them would not be useful for retrieval, some thesauri use the relationship *equivalent term* (ET) at least in their internal database (in the user version they may map ET to ST). The ET relation can be seen as partitioning the set of concepts into mutually exclusive groups. Each group corresponds to a newly formed ISAR (Information Storage And Retrieval) concept which is broader than any concept in the group. A preferred term can then be selected for each ISAR concept. The *equivalent term relationship* is at the borderline between term-term relationships and concept-concept relationships. For the term-based model (discussed below) this does not present a problem, but for the concept-based model one must decide whether to treat ET as a term-term relationship or a concept-concept relationship. This problem is particularly thorny for an integrated thesaurus database since terms that are equivalent from the point of view of one constituent thesaurus may need to be distinguished for retrieval purposes in another thesaurus.

Note. The term-term relationships and the equivalence relationship are equivalence relations in the mathematical sense (they are both reflexive and transitive) and thus partition the set in which they hold into mutually exclusive groups.

A complex thesaurus may further differentiate relationship types (for example, distinguish between genus-species and whole-part hierarchical relationships) and include other types of relationships. In addition, the selection of preferred terms and, from them, descriptors, gives rise to **instructions** which can be combined with the term-term and concept-concept relationships:

SEE refers from a non-preferred term to a preferred lead-in term
SF is the reciprocal

USE refers from a non-preferred term or from a preferred lead-in term to a descriptor
UF is the reciprocal.

Thus, an individual thesaurus at a given time is a complex, highly interrelated structure. If we add the time dependency of terms and concepts, their status and their relationships, complexity increases. If we integrate several thesauri into one data base while maintaining their individual identities, complexity increases still further. The data structure of an integrated thesaurus database must be able to handle this complexity efficiently.

DATA STRUCTURES FOR THESAURUS DATABASES

Some computer systems for thesaurus construction and maintenance use a record for every term with the information about the term, such as synonyms, broader, narrower, and related terms, stored in — usually repeating — data fields in the record (Figure 1a). Information is stored in large packages, and to access or change any piece of information we must get into the appropriate package. Even for an individual thesaurus such a structure is inflexible. For an integrated thesaurus data base it is unwieldy. For example, comparing two records for the same term from two different thesauri requires cumbersome processing of the two records.

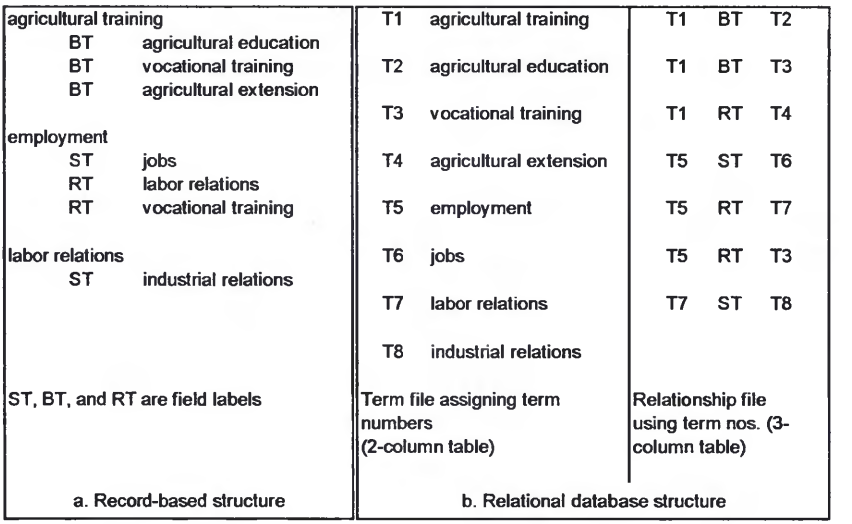


Figure 1. Data structures for a single thesaurus.

The relational approach to data base organization leads to a more elegant and efficient structure (Figure 1b). Information is stored in individual pieces that can be arranged in different ways. For example, *employment* RT *labor relations* is a piece of information that is stored by itself. Combining two thesauri stored in this format can be accomplished simply by putting all the pieces of information into one data base and eliminating duplicates. This structure has an additional advantage: Relationship types are not defined as fields in a record (and thus fixed in the database structure), but they are simply data values in a relationship record; thus new relationship types can be introduced with ease.

Furthermore, some thesaurus databases are fashioned after the structure of a printed thesaurus and use the full term string wherever a term is referred to. This introduces considerable redundancy: The same lengthy terms appear over and over again in cross references. In an integrated thesaurus database, the redundancy becomes even more severe; terms, concepts, and relationships are repeated. Efficiency of storage can be achieved by assigning each term a four-byte number and using these term numbers in all relationship pairs. (For clarity, term numbers will be represented in this paper as T1, T2, T3 etc., relationship numbers as R1, R2, R3, etc., and concept numbers as C1, C2, C3, etc. In a real system, these would be plain 4-byte numbers whose meaning is determined by the file and the data field in a record.) Figure 1 compares two data structures for a single thesaurus.

DATA MODELS FOR AN INTEGRATED THESAURUS DATABASE

A sample set of data to be stored in an integrated thesaurus database is given in Figure 2. It consists of terms and relationships structured according to the relational data structure, with terms spelled out (rather than represented by term numbers). The data in the data set have been collected from several thesauri, each identified by a three-letter abbreviation. The data from all thesauri were combined and the resulting pool sorted alphabetically by main term, relationship type, and cross term.

We will present two data models to handle these data, a **term-based data model** and a **concept-based data model**. **Both data models use the same method for organizing data about terms as illustrated in Figure 3**. The **term file** contains one record for each term stem. Thus if the term occurs in the singular in one thesaurus and in the plural in another, the integrated thesaurus database has only one record. The term record also gives the language of each term. (Remember that the same character string used in different languages represents different terms.) The **term source file** records for each term the thesauri in which it occurs; if a term occurs in three thesauri, it has three records in the term source file. Each record also gives the suffix that must be combined with the term stem to arrive at the exact form of the term used in the specific thesaurus. For example, UNE uses the form *jobs* while KAS uses the form *job*. In most cases, the stem is also the singular, but not always (for example, stem *hypothes*, singular suffix *is*, plural suffix *es*. This model of term data takes care of the morphological variation most important for thesauri, singular/plural variation. Other types of morphological variation could be accommodated through elaboration of the model.

Main term	Rel Type		
agricultural training	MT		UNE
agricultural training	MT		MAC
agricultural training	ST	farmer training	MAC
agricultural training	TR	formation agricole	MAC
agricultural training	BT	agricultural education	UNE
agricultural training	BT	vocational training	UNE
agricultural training	BT	vocational training	MAC
agricultural training	RT	agricultural education	MAC
agricultural training	RT	agricultural extension	UNE
agricultural training	RT	experimental farm	MAC
employment	MT		UNE
employment	ST	jobs	UNE
employment	MT		MAC
employment	MT		ERI
employment	RT	industrial relations	ERI
employment	RT	labor relations	UNE
employment	RT	labor relations	MAC
employment	RT	vocational training	UNE
job	MT		KAS
job	RT	employee relations	KAS
labot relations	MT		UNE
labot relations	ST	industrial relations	UNE
work	MT		DRI
work	MT		ZID
work	MT		DAS
work	ST	employment	DAS
work	RT	industrial relations	DRI
work	RT	labor relations	ZID
work	RT	labor relations	DAS

Figure 2. Data form several thesauri combined.

Term no.	Term stem	Language
T1	agricultural training	EN
T2	farmer training	EN
T3	formation agricole	fr
T4	agricultural education	EN
T5	vocational training	EN
T6	agricultural extension	EN
T7	experimental farm	EN
T8	employment	EN
T9	job	EN
T10	industrial relation	EN
T11	labor relation	EN
T12	employee relation	EN
T13	work	EN

One record for each term, each term identified by a term number.

a. Term file

Term no.	Thesaurus (source)	Suffix	Term Type
T1	UNE		DE
T1	MAC		DE
T2	MAC		NP
T3	MAC		DE
T4	UNE		DE
T4	MAC		DE
T5	UNE		DE
T5	MAC		DE
T6	UNE		DE
T7	MAC		DE
T8	UNE		DE
T8	MAC		DE
T8	ERI		DE
T9	UNE	s	NP
T9	KAS		DE
T10	UNE	s	NP
T10	ERI	s	DE
T10	DRI	s	DE
T11	UNE	s	DE
T11	MAC	s	DE
T11	ZID	s	DE
T11	DAS	s	DE
T12	KAS	s	DE
T13	DRI		DE
T13	ZID		DE
T13	DAS		DE

Multiple records for each term as needed, linked to term file trough term number.

b. Term source file

Figure 3. Both data models: term file and term source file.

The term file contains general information, the term source file contains information specific to each thesaurus. This distinction, which recurs in other files discussed below, is very important for compact storage of data. A source is any thesaurus in the integrated database, be it an independent thesaurus that was used in creating the integrated database or a new thesaurus created from the integrated database.

THE TWO DATA MODELS USE DIFFERENT METHODS FOR ORGANIZING DATA ABOUT RELATIONSHIPS.

Relationships in the term-based data model. All relationships are stored explicitly as they occur in the sources. Terms are referred to by their numbers. If

the same relationship (same terms, same relationship type) occurs in several thesauri, it is stored only once while preserving the information about the individual thesauri in which the relationship appears. This give rise to the structure shown in Figure 4. The **Term relationship file** stores each relationship once, identified by a relationship number. The **term relationship source file** stores for each relationship the thesauri in which it occurs.

Rel no	Cross term no			Rel no	Cross term no
R1	T1	ST	T2	R1	MAC
R2	T1	TR	T3	R2	MAC
R3	T1	BT	T4	R3	UNE
R4	T1	BT	T5	R4	UNE
				R4	MAC
R5	T1	RT	T4	R5	MAC
R6	T1	RT	T6	R6	UNE
R7	T1	RT	T7	R7	MAC
R8	T8	ST	T9	R8	UNE
R9	T8	RT	T10	R9	ERI
R10	T8	RT	T11	R10	UNE
				R10	MAC
R11	T8	RT	T5	R11	UNE
R12	T9	RT	T12	R12	KAS
R13	T11	ST	T10	R13	UNE
R14	T13	ST	T8	R14	DAS
R15	T13	RT	T10	R15	DRI
R16	T13	RT	T11	R16	ZID
				R16	DAS
One record for each relationship, each relationship identified by a relationship number.				Multiple records for each relationship as needed, linked to the relationship file trough the relationship number.	
a. Term relationship file				b. Term relationship source file	

Figure 4. Term-based model: Term relationship file and relationship source file.

In the term-based model, some of the stored relationships are term-term relationships (*has synonymous term* or ST for short), while others are concept-concept relationships. The concepts are expressed through the preferred term used in the individual thesaurus in which the relationship appears. Thus in the term-based model term-term relationships and concept-concept relationships take the same form. Concepts are not represented explicitly. Consequently, concept-term relationships are not made explicit but implied

A database structured according to the term-based model can be established very easily: Simply pool the data from various thesauri, each structured according to the relational structure described above, as follows: Keep a running list of

terms, sorted alphabetically, with term numbers (term file, Figure 3a); as a new thesaurus is added, check all terms against the list and add new terms. Record the thesaurus as a source for any term, existing or new (term source file, Figure 3b). In relationships, replace the terms by term numbers. Keep a running list of relationships, each identified by a relationship number, sorted by main term, relationship type, and cross term (term relationship file, Figure 4a). As a new thesaurus is added, check all relationships against the list and add new relationships. Record the thesaurus as a source for any relationship, existing or new (term relationship source file, Figure 4b)

The term-based model allows for great flexibility: It can capture relationships between various forms of the same term (spelling variants, abbreviations) and degrees of synonymity (*synonymous term*, *equivalent term*). It allows for one thesaurus having A ST B while another has A RT B; the administrator of the integrated thesaurus database need not make a decision between the two. The price for this flexibility is a lot of storage space — one and the same conceptual relationship is stored as often as it has linguistic expressions in the constituent thesauri — and less efficient processing — to collect all the occurrences of a conceptual relationship the program must trace all the terms for each of the concepts involved. To trace all terms for a concept, the program must start from one term, find all its synonyms, find all their synonyms, etc. This process is very laborious and error-prone: A single erroneous *synonymous term* relationship can lead to undesirable results. A program using this model should let the user specify that only relationships from a select list of thesauri should be used in the tracing process.

Relationships in the concept-based model. In the concept-based model, concepts are identified explicitly thorough concept numbers. Accordingly term-concept relationships (or concept-term relationships) are given explicitly in a **concept-term file** (Figure 5), which links each (disambiguated) term with exactly one concept. Thus the concept-term file has one record for each term and for each concept as many records as there are terms designating the concept. Term-term

Concept no		
C1	T1	agricultural training
C1	T2	farmer training
C1	T3	formation agricole
C2	T4	agricultural education
C3	T5	vocational training
C4	T6	agricultural extension
C5	T7	experimental farm
C6	T8	employment
C6	T9	job
C6	T13	work
C7	T10	industrial relation
C7	T11	labor relation
C7	T12	employee relation

Figure 5. Concept-based model: Concept-term file.

relationships (*has spelling variant*, *has synonymous term*) on the other hand are represented implicitly: All terms linked to the same concept are synonyms or spelling variants. Concept-concept relationships (BT, NT, RT) are established explicitly between concepts in the **concept relationship file** (Figure 6a), with sources for each relationship indicated in the **concept relationship source file** (Figure 6b).

Rel no	Main concept no	Rel type	Cross concept no
R1	C1	BT	C2
R2	C2	BT	C3
R3	C1	RT	C2
R4	C1	RT	C5
R5	C2	RT	C4
R6	C6	RT	C3
R7	C6	RT	C7

a. Concept relationship file

Rel no	Thesaurus
R1	UNE
R2	UNE
R3	MAC
R4	MAC
R5	UNE
R6	UNE
R7	UNE
R7	MAC
R7	KAS
R7	ERI
R7	DRI
R7	ZID
R7	DAS

b. Concept relationship source file

Figure 6. Concept-based model: Concept relationship file and relationship source file.

Establishing an integrated thesaurus database according to the concept-based model requires considerable effort: All terms for a concept must be brought together. While the ST relationships from the constituent thesauri are very helpful in this process, thesaurus 1 may contain term A and thesaurus 2 the synonymous term B without a relationship A ST B recorded anywhere; discovering this relationship, which is required for the proper construction of the concept-term file, takes intellectual effort

The concept-based data model is efficient for storage and processing, since each conceptual relationship, while it may expressed using various terms in the constituent thesauri, is stored only once in the integrated database. However, it is also inflexible: Decisions on term-concept relationships must be made once and for all and are then binding on all thesauri in the database; thus we cannot have A ST B in one constituent thesaurus and A RT B in another. There is also no distinction between spelling variants, and synonyms.

Some of the limitations of the concept-based model can be overcome at the cost of added complexity. A separate file connecting spelling variants to the normalized term would maintain the distinction between spelling variant and synonym. Equivalent terms, which are lumped together with synonyms in the simple model presented, can be kept distinct as follows: Include separate concept number for each of the equivalent concepts in a group and introduce a new concept which is above all the concepts in the group. Only the broad concept is selected as descriptor. As an example, consider the group of equivalent terms

(with their concept numbers) *Disease* (C35), *Illness* (C45), *Sickness* (C67), and *Ailment* (C73). Each would be treated as a distinct concepts. A new concept subsuming all of them would be introduced and called, perhaps, *Disease (broadly defined)* (C87). The concept relationship file would contain

C35 BT	C87
C45 BT	C87

etc. Most thesauri would select just C87 as descriptor, others might need the increased specificity of the narrower concepts, giving good scope notes that would help the indexers decide on the right descriptor in each case. On could introduce a more precise relationship BT-EQ; the rules for extracting an individual thesaurus from the integrated database might stipulate that BT-EQ should be converted to ST or ET.

CONCLUDING REMARKS

This paper described just the basic elements of each model. A real system needs to include a lot more data, for example the date when a descriptor was introduced and when it was discontinued.

A prototype software package for maintaining an integrated thesaurus database is described in Soergel 1994. This package uses the term-based model. It allows for a wide variety of outputs exploiting all the information in the database.

BIBLIOGRAPHY

Soergel, Dagobert 1995. **Software support for thesaurus construction and display**. Advances in Classification Research. v. 5 (Proceedings of the 5th ASIS SIG/CR Classification Research Workshop, Oct. 1994). Medford, NJ: Learned Information; 1995. p. 157-184.

SUMMARY

This paper presents two data structures for storing multiple thesauri in a single integrated database. Both data structures include an entity type term. For each term, they give the sources and the status (descriptor, lead-in term) in each source. Data structure 1 includes an entity type concept, identifiers each concept by a number, and links each (disambiguated) term with exactly one concept.

Synonymous Term relationships are represented implicitly: All terms linked to the same concept are synonyms. Conceptual relationships (BT, NT, RT) are established explicitly between concepts. This data structure is efficient for storage and processing, since each conceptual relationship, while it may expressed using various terms in the constituent thesauri, is stored only once in the integrated database. However, it is also inflexible. Data structure 2 does not include an entity type concept. All relationships are stored explicitly. This allows for great flexibility: It can capture relationships between various forms of the same term (spelling variants, abbreviations) and degrees of synonymity (Synonymous Term, Equivalent Term). It allows for one thesaurus having A ST B while another has A RT B. The price for this flexibility is more storage space - one and the same conceptual relationship is stored as often as it has linguistic expressions in constituent thesauri — and less efficient processing — to collect all the occurences of a conceptual relationships the program must trace all the terms for each of the concepts involved. The paper will also discuss various outputs from an integrated thesaurus database that would provide support for the construction of individual thesauri and of conversion tables between thesauri.

M. Muraszkiewicz, H. Rybiński, W. Struk

SOFTWARE PROBLEMS OF MERGING MULTILINGUAL THESAURI[1]

INTRODUCTION

Knowledge representation and knowledge engineering have been recognized by various information technology oriented communities as vital aspects of their activities. Practically, any computerized information system project includes a component related to the terminology establishment and its efficient use (Dik, S.C. [1987]). A thesaurus is especially a robust linguistic tool for a certain class of information systems and/or databases (e.g., Ciampi, C. et al. [1985], Fameli, E., Nannucci, R., Di Giorgi, R. [1983], Weihs, E. [1981]). This is even more valid in the context of international information systems and international communities. In particular, the problem becomes tremendously important when multilingual information systems are implemented. The needs for multilingual thesauri will be growing with the development of international hyper-text products like dictionaries, terminology vocabularies, etc.

Experience indicates that establishment of a multilingual thesaurus is not an easy task. Collecting the items in a multilingual environment is much more difficult than for one language. Moreover, the management and maintenance of multilingual thesauri require more sophisticated tools and skills. The process of multilingual thesaurus building is usually iterative and is a result of a consensus established among the subject specialists.

One of the methodologies in building multilingual thesauri consists in creating a new thesaurus on the basis of an experience gained with exploitation of various existing thesauri. In such a case existing thesauri are used as a platform for the new thesaurus. In addition one can observe recently a tendency of integrating existing information communities and system by means of standardizing linguistic tools that are used by the systems. For this approach there is a tremendous need for software tools simplifying comparisons of various thesauri and possibly transferring/merging some parts from one thesaurus to another.

Unfortunately, there are not many ready-to-use software tools to handle thesauri [Ritzler, 1991]. In Rybinski, et al [1993] a software for building multilingual thesauri has been presented. The software known as MULTHES/ISIS has been designed as a configurable system assisting a user in creating concepts, linking them by means of a set of predefined relations, and controlling the validity of the thesaurus structure. The software has shown valuable features in building

[1] The work reported in this paper has been jointly supported by grant no. 8 T11C 038 08 of State Committee for Scientific Research (KBN), Poland, and by the Institute for Theoretical and Applied Computer Science, Polish Academy of Sciences.

multilingual thesauri. It was successfully used for building the CEDEFOP thesaurus It is also used for maintenance of the OECD Macrothesaurus (known as MTM3.0) One of the main restrictions in using the software was a lack of tools supporting the methodology of merging essential parts of existing thesauri into one.

In this paper we present problems of merging multilingual thesauri. We describe a software which allows one to work with a number of thesauri, viewing them simultaneously, and creating a thesaurus as a result of merging essential material from existing thesauri. The previous version of software MULTHES/ISIS (Rybinski et al [1993] was taken as a starting point in developing the system which is called MTM 4.0.

The main features of the MTM 4.0 software are, inter alia, the following:

- thesaurus maintenance and support system;
- KWOC and full tree representation and navigation tools available on-line;
- KWIC, KWOC and full tree printouts;
- defining and customization of up to 100 conceptual relationship types;
- management of facets, codes (top classification), sources, regional variants, historical notes, etc.;
- support of the various types of authority files;
- computer assisted merging; thesauri comparison by means of windows
- support of the various alphabets;
- support of linguistic and orthographic variants;
- sorting facilities consistent with national standards;
- variable length data handling;
- flexibility in defining input and output forms.

From the terminal user standpoint MTM4 fulfills the following criteria:

- user-friendliness when entering, updating, deleting, merging, checking data;
- intelligent prompting of the end user whenever in doubt;
- powerful validation facilities covering proper structuring of a thesaurus (e.g. maintenance of relationship isomorphism between languages);
- features for documenting („keeping track") the history of the thesaurus evolution;
- availability of data protection facilities;
- availability of self-training and demonstration facilities;
- provision of a thesaurus publishing facilities at the professional level;
- modularity and openness to the further development.

The paper is composed as follows: the next section presents some basic notions concerned with the system MTM and its use to build multilingual thesauri. Then, some problems with thesauri merging are considered (Section 3). In Section 4 we briefly present the user interface for merging. The paper ends with remarks on possible further development.

BASIC NOTIONS

Besides standard thesaurus related terminology we are using in the paper some notions that are specific for multilingual thesauri applications. The basic information entities in the system are concepts, descriptors, and ascriptors (Lex [1987], Jackendoff [1989]). Any concept in a multilingual thesaurus consists of a

number of descriptors, each being a concept representation in a given language. The concepts may be linked to each other by means of semantic relationships (e.g. **part_of, is_a, broader, narrower etc.**). Thus, these relationships are multilingual and are reflected in all languages. On the other hand, synonyms of descriptors may be defined specifically in a given language as ascriptors, so monolingual links between ascriptors and descriptors may be set up to model synonymy relationships, specific for a given language.
Basic data types of MTM 4.0 are terms, facets and top classification elements.

Terms

The terms take the form of descriptors or non-descriptors (ascriptors). The terms may be categorized, for example one can distinguish in a thesaurus the following categories:

- generic; the terms which are specific and characterize the subject matter covered by the thesaurus;
- legal; the terms which have an official legal status and are defined in the various legal documents;
- standard; the terms which have an official status in standard documents;
- authority files; these are descriptors with special flags indicating their specific role in a thesaurus. It is recommended to store them in a special, substantive table. They should be also visible in the main data table. One can have for example *geographic terms, institutions, languages*.

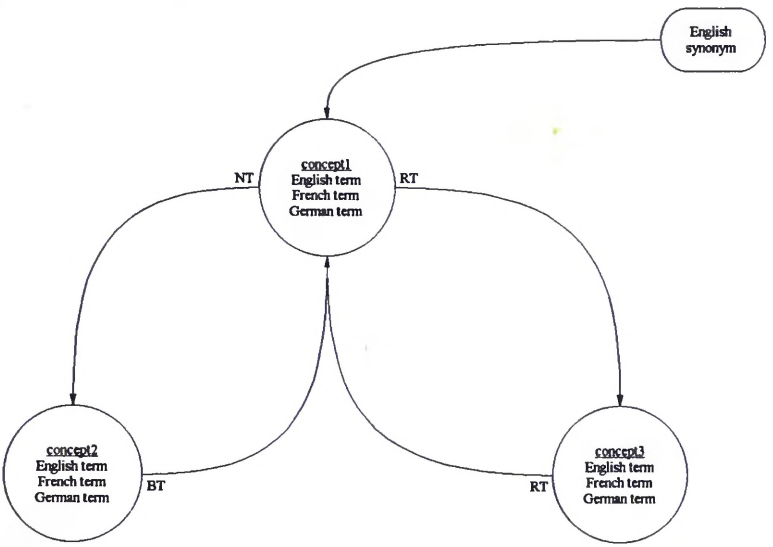


Fig. 1. Fragment of a thesaurus graph

In addition to the above categorization the notion of „Top Term" will be used. A top term is a descriptor that has no BT relationship. A special function providing a table containing only the top terms has been implemented.

Top Classification

The concept of Top Classification is proposed for MTM 4.0 in order to classify the top terms and to assist merging. The top classification records, whose number is accounted for 50-100, may be implemented as an authority file. The record consists of a code and the terms (one per language)

Concept structure — descriptor

Usually the descriptor record consists of the following data elements:

- concept number (record ID)
- terms (one for each language)
- top classification for top terms only
- various type scope notes, e.g. an explanation for indexers, another explanation for end-users,
- historical note;
- internal note used by a thesaurus builder only
- source
- remarks

In the approach presented the structure of Concept may be subject of adjusting to specific needs and aims of the given thesaurus.

A simpler structure is usually foreseen for non-descriptors (ascriptors); they play a role of synonyms which refer to the appropriate descriptors. The ascriptors may be monolingual or multilingual. Usually an ascriptor provides simple or more sophisticated indications which descriptor or list of descriptors should be used as semantically equivalent. For example

Physics of semiconductors: ***Use*** *Semiconductor + Physics*

Relations

Relations within a thesaurus link the terms. Graphically, the terms and relations constitute a graph. In some cases it is recommended that the graph has a tree structure. The MTM 4.0 allows one to maintain polyhierarchy which means that a„son" object may have more than one„father" object. The establishment of relationships between terms and assigning them a certain meaning is part of the thesaurus builder's responsibility.

It is worthwhile to mention that MTM 4.0 provides a set of classic pre-defined relations such as broader, narrower, etc. and gives to the user's disposal a facility to define his/her own relations which might reach the number of up to 100. The

nature of the latter entirely depends on the user. It is up to the user to define and name the needed relations. The MTM 4.0 has pre-defined relations as follows.

Internal
BT, NT, RT, US, UF, LV (for local variants), OV (for orthographic variants), US+, UF+
External
CR (concordance relation) which links equivalent terms from different thesauri.

One of the main difficulty of thesauri building and maintenance is a number of links that have to be established. A usual method to simplify the task is that only one-way links are to be set. The opposite direction is to be set up automatically. So, if broader type relationship is established from a concept A to B, the narrower type link is automatically set up from B to A (Fig. 1).

The consistency of a thesaurus is a condition sine qua non of any valuable thesaurus usage. Formally a thesaurus is a digraph with the concepts as nodes and relations as edges. The main task is to avoid loops in the thesaurus graph. The validation process checking the thesaurus graph is responsible for the so called global consistency. In case of microcomputers it is rather difficult to check global consistency in dialog mode. A simpler solution is to run a batch process. Another possibility is to control the so called local consistency, which refers only to a neighborhood of the modified node. The local validation has been adopted in MTM4.0. The global validation procedure is for the time being semi-automatic. As a basic tool to control the thesaurus consistency one can use display facilities to view the thesaurus in a tree form.

PROBLEMS WITH THESAURI MERGING

Before we start describing this subsystem, let us provide basic notions dealing with merging of two thesauri.
Source thesaurus — the thesaurus which serves as a source for terms and relations
Target thesaurus — the thesaurus to which the terms are copied from the source thesaurus.
Node — a Concept record in the thesaurus with all the languages and relations
Top term — a term which has no BT relations and is flagged as a special term for indexing other terms belonging to a given subject of the thesaurus domain. It plays the role of a descriptor;
Branch — a subtree of the thesaurus consisting of the descriptors connected by the NT relation.

First of all we assume that the source and target thesaurus should have the same leading language. If the leading languages are different in two thesauri, then even in the case that some languages appear in both thesauri, one can expect that a lot of compromises have been done with those languages, and there are semantic gaps in such languages. So, assigning to such languages the role of a leading language may be very risky for the target thesaurus. In such a case one could first recommend to revise the thesauri from the point of view of a common leading language.

Provided that there is a common leading (or at least base) language, there are still semantic problems with merging the thesauri. The basic problems may result from the following:

• the source and target thesauri play different roles in various information systems, so, for example they may differ in depth, level of specificity, etc.

• the scopes of the thesauri are different

If the thesauri differ in depth but are semantically compatible it seems that in many cases some subtrees may be just copied from one thesaurus to another. If however the scopes of two thesauri are different, this may cause real problems with relationships when transferring concepts from one thesaurus to another. Namely, it happens very often that the same terms in various knowledge domains may have different meanings, so they may constitute different graphs in the two thesauri.

It is therefore necessary to foresee such options in the thesauri merging software, so that on the ne side, one is able to copy a whole subtree (from an indicated level); on the other side, there should be facilities enabling one to copy nodes without links.

While copying a concept from one thesaurus to another, homonyms in one of the languages may be detected. Let us consider the following possible cases:

1. While copying a term the homonyms in the leading language are detected. Some terms in other languages on the source and target side are different. In this case one can assume that the two nodes play a similar role in both thesauri. The differences result from other translations in some languages. If we accept this, we should decide which language versions should constitute the descriptor forms. The other translations may be used as an ascriptor form for the given term. The system will give us the possibility to decide which language versions should become ascriptors;

2. As under 1 but we find that the two homonyms play different roles in the two thesauri, and we would like to preserve the term as it is already in the target language and add a new term playing the role as made in the source (in this case possibly the differences will appear in other languages). In this case we should change the form of source and target terms, so that we eliminate homonyms. As a result we will have two varying nodes in the target thesaurus; the question arises, where to link the new node;

3. Homonyms were found in some languages other than the leading one. We could expect that the two terms are close. If we want to have two nodes in the target we follow as in the case (2) above, eliminating the homonyms in the corresponding languages. If however we see that the two nodes should be merged, we proceed as in (1), i.e. we create one node, possibly generating some ascriptors.

In either case we should be able to consult our decision by examining the existing environment of the given concepts in the source and target thesauri. In the case when a new concept results from our procedure we should have the possibility to indicate this, if we want to preserve the relations from the target and the source thesauri.

Another problem with merging thesauri is concerned with relations coming with the„imported" concept. It seems to be too difficult to control all relation types. Especially the role of the RT relation may be different in various thesauri. We

therefore have accepted a solution that only BT and NT type relations are processed by the system.

If we assume that two nodes from the source and the target thesaurus should become one node in the target thesaurus we actually decide that the existing NT relation will be possibly expanded by the coming new concepts. If in addition the new term has some links to broader terms in the source thesaurus it may be valuable to preserve these links. At least the user should have possibilities to decide about it. In Fig. 2 an example of„merging relationships" is illustrated

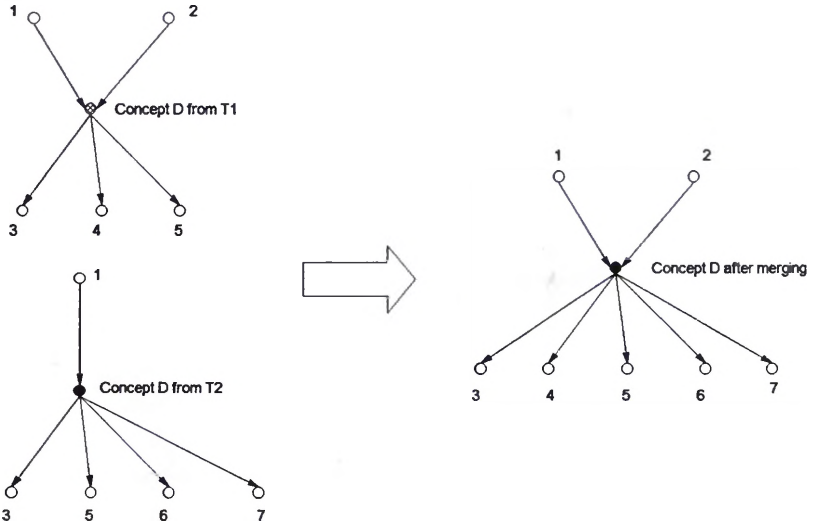


Fig. 2 Merging a concept from two thesauri

IMPLEMENTED FUNCTIONS OF THE MERGE SUBSYSTEM

The main menu of the merging subsystem contains the following options:

- Recall last merge operation. This option does not need any special comments. It will bring the user the state which was reached before leaving the last session with the Merge subsystem (Source and Target databases are reopened)
- Select source thesaurus. Once this option is selected the system displays a table as in Fig. 3.7, where the thesaurus can be selected as the source;
- Select target thesaurus. as above, but refer to the selection of the target database
- Create new database; this option allows one to create a new thesaurus database. With this option one can copy a structure of one of the thesauri defined by now and modify it, or one can create a new thesaurus definition from scratch (usually the first possibility is recommended);
- Unselect both; when needed, one can unselect source and target thesauri.

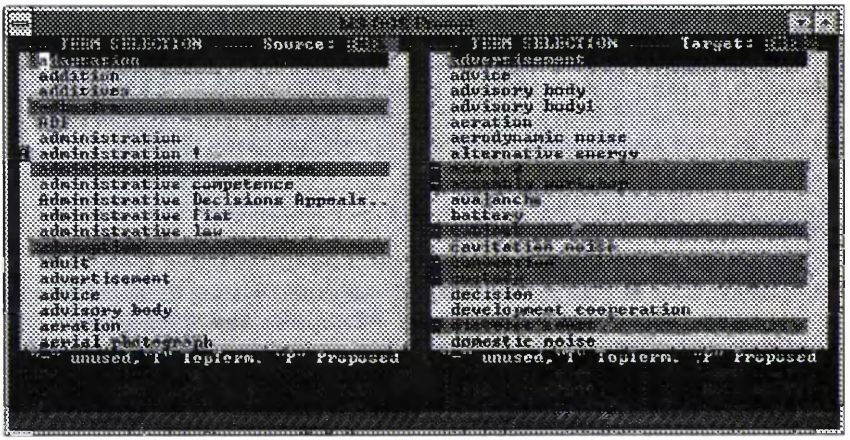


Fig. 3 Merging operation

Once the source and the target thesaurus is selected, the screen shows both thesauri in the boxes on the top. In addition, the system displays the two thesauri in the windows (Fig. 3). Below we will discuss the main merge operations. The following options are available:

- we can copy all records without asking for confirmation, until homonyms are detected — option A
- before copying every record we can edit it — option E;
- copy only that term, and ask for confirmation for all next ones (T);
- skip given term, go to the next (S);
- break the copying process (B)

If we are firm that the whole tree should be copied, wherever it is new in the target we select the fastest option A. Otherwise we can go one by one (E or T);

FINAL REMARKS

The MTM 4.0 software has been implemented as a tool for the construction of the Environment Thesaurus, being a result of merging two existing thesauri. The software presented is an advanced tool conceived to create a working background for a flexible thesaurus package. Presently the user can already adjust print formats, worksheets, alphabet's collating sequence and, obviously, one can configure the system for up to 9 languages. Especially, by defining the worksheets and record structure one can design term records tailored to thesaurus needs.

A list of further extensions and improvements of the software can be set up. One of the main tasks that should be defined is to provide the software with a mechanism to detect the loops within the thesaurus graph. For the time being, it is possible to verify the structure using the function of the tree display. Another

option that is planned for implementation is to link the thesaurus with the validation supported Input and Update functions.

Multilingual thesaurus building is a multivalent and intercommunity task. Therefore, it seems that the software presented can be helpful in supporting this process. It may significantly improve the quality of traditional information systems, as well as more sophisticated multilingual knowledge based systems.

REFERENCES

1. Ciampi, C. et al. [1985], THES/BUILD: an expert system for constructing a computer based thesaurus for legal informatics and computer law, [in:] Computing power of legal reasoning, ed. Charles Walter, New York, West Publ. Co., 1985, pp. 375-412.
2. Dik, S.,C. [1987]. Linguistically motivated knowledge representation, [in:] Language and Artificial Intelligence, Amsterdam, North Holland, pp. 154-170
3. Fameli, E., Nannucci, R., Di Giorgi, R. [1983]. Documentation in Legal Informatics and the International Bibliography on Computers and Law, Informatica e diritto, No. 3, pp. 183-239.
4. Jackendoff R. [1989]. What is a concept that a person may grasp it?, Mind and Language, 1989, vol. 4, No. 1/2 pp. 68-102.
5. Lex, W. [1987]. Representation of concepts for their computerization, Int. Classif., 1987, vol. 14, No. 3, pp. 127-132.
6. Ritzler, C. [1991]. Comparative Study of PC supported Thesaurus Software, Proceedings of International Conference on Knowledge Organization & Terminology, Bratislava.
7. Rybinski H., Muraszkiewicz M., Struk W., [1993]. MULTHES-ISIS: A Flexible Software for Multilingual Thesaurus Building, TKE '93, Koeln, Frankfurt: INDEKS Verlag 1993, pp. 152-159
8. UNESCO [1989]. Mini-Micro CDS/ISIS — Users Manual, PGI, UNESCO, Paris
9. Weihs, E. [1981]. Environmental Thesauri Construction, categories and function in Bavarian Land Information System, Proceedings of International Conference on Knowledge Organization & Terminology, Bratislava, pp. 104-115.

SUMMARY

Knowledge representation and knowledge engineering have been recognized by various information technology oriented communities as vital aspects of their activities. Practically, any computerized information system project includes a

component related to the terminology establishment and its efficient use (Dik, S.C. [1987]). A thesaurus is especially a robust linguistic tool for a certain class of information systems and/or databases (e.g., Ciampi, C. et al. [1985], Fameli, E., Nannucci, R., Di Giorgi, R. [1983], Weihs, E. [1981]). This is even more valid in the context of international information systems and international communities. In particular, the problem becomes tremendously important when multilingual information systems are implemented. The needs for multilingual thesauri will be growing with the development of international hyper-text products like dictionaries, terminology vocabularies, etc.

Experience indicates that establishment of a multilingual thesaurus is not an easy task. Collecting the items in a multilingual environment is much more difficult than for one language. Moreover, the management and maintenance of multilingual thesauri require more sophisticated tools and skills. The process of multilingual thesaurus building is usually iterative and is a result of a consensus established among the subject specialists.

One of the methodologies in building multilingual thesauri consists in creating a new thesaurus on the basis of an experience gained with exploitation of various existing thesauri. In such a case existing thesauri are used as a platform for the new thesaurus. In addition one can observe recently a tendency of integrating existing information communities and system by means of standardizing linguistic tools that are used by the systems. For this approach there is a tremendous need for software tools simplifying comparisons of various thesauri and possibly transferring/merging some parts from one thesaurus to another.

Unfortunately, there are no many ready-to-use software tools to handle thesauri [Ritzler, 1991]. In Rybinski, et al [1993] a software for building multilingual thesauri has been presented. The software known as MULTHES/ISIS has been designed as a configurable system assisting a user in creating concepts, linking them by means of a set of predefined relations, and controlling the validity of the thesaurus structure. The software has shown valuable features in building multilingual thesauri. It was successfully used for building CEDEFOP thesaurus. It is also used for maintenance of the OECD Macrothesaurus (known as MTM3.0). One of the main restrictions in using the software was a lack of tools supporting the methodology of merging essential parts of existing thesauri into one.

In this paper we present problems of merging multilingual thesauri. We describe a software which allows one to work with a number of thesauri, viewing them simultaneously, and creating a thesaurus as a result of merging essential material from existing thesauri. The previous version of software MULTHES/ISIS (Rybinski et al [1993] was taken as a starting point in developing the system.

Ewa Chmielewska-Gorczyca

COMPATIBILITY OF INDEXING TOOLS IN MULTIDATABASE ENVIRONMENT

The necessity of cooperation in the field of subject approach (indexing) and the profits of it have long been recognized and frequently discussed in the papers on information retrieval. However, usually they dealt with the problem of two or more separate institutions (agencies) having their own indexing tools and trying to make them compatible or integrated. What I am going to discuss is the problem of subject indexing within one institution having several databases. The discussion will be illustrated with the examples of the Sejm Library (the library of the Polish parliament).

BENEFITS OF COOPERATION

With the increase in cooperation between information services and the simultaneous use of several databases, cooperation in the field of subject indexing becomes more and more important. The benefits achieved through cooperation are numerous, e.g.

a) save effort in construction, maintenance and up-dating indexing tools;

b) share the results of subject indexing;

c) make possible switching subject searches from one database to another (facilitate the running of a search request in different databases);

d) increase the consistency of indexing between institutions.

Obviously, it is much easier to have one indexing language (all services use the same, or at least highly compatible, vocabulary) than to maintain separate languages and try to make them compatible. However, it is a common understanding now that an indexing language or thesaurus should be uniquely tailored to the document collection it is to cover (literary warrant), and to the interests (needs) of the user community it is to serve (users' warrant). A common thesaurus would in many cases make it impossible for the participating institutions to follow this principle, since each organization tends to have its own particular viewpoints. It is therefore unlikely that the adoption of a general scheme, in toto, will completely satisfy the individual needs of specialized agencies involved in the joint information processing venture.

The cooperation in the field of indexing is even more important in the case of various divisions (departments) of one institution. Often the different services within one institution develop their own vocabularies, the result of this being the heterogeneous systems and databases and a cumbersome process of searching through them for the user. Development of a single system out of such separate

indexing tools would promote consistency of indexing and effectiveness of retrieval (easy and convenient access to the user).

SITUATION IN THE SEJM LIBRARY

At the beginning of computerization four indexing languages were used in the Library for four separate files:

a) the modification of UDC (Universal Decimal Classification) for the card catalogue of books;

b) subject headings of the National Library used as the second parallel language in the OPAC;

c) a crude form of home-made classification scheme for audiovisual materials;

d) another home-made classification for the articles from Polish periodicals.

Apart from these the Computer Center of the Sejm Chancellory maintained a database of Polish legal acts, and for indexing them developed their own thesaurus with a very simple structure (only alphabetical list and limited network of cross-references).

The idea that appeared two years ago in the Chancellory of having one common indexing language covering the scope and fulfilling the needs of all those services seemed very exciting and promising. The users of parliamentary databases would be able to search different computer files existing in the network with the same set of indexing terms. The issue of updating (revision of the thesaurus) would be dealt with only in one center, and not in four different ones. Since there are some databases and other information sources used in the Library that are indexed with the European Parliament thesaurus EUROVOC, it was decided to develop an information retrieval language based on and compatible with this vocabulary. In this way, the automatic or semi-automatic switching from our own databases into those prepared by the European Union institutions and agencies or other parliamentary libraries using translations of EUROVOC would be possible. Initially, the idea was to translate EUROVOC into Polish but as the process of translation and indexing advanced it became obvious that simple translation had to be replaced by adaptation of EUROVOC which would meet our own specific needs.

SYSTEM OF THESAURI

The result of a project lasing almost for two years was the system of thesauri called STEBIS, consisting of ten separate vocabularies:

GEO — Thesaurus of Geografical Terms
TIO — Thesaurus of International Organizations
LAW — Thesaurus of Law
PAR — Thesaurus of Parliamentary Affairs
EBU — Thesaurus of Economy and Business
POL — Thesaurus of Politics

SEC — Thesaurus of Science, Education, Culture, Arts,Religion
TIC — Thesaurus of Information and Communications
SOC — Thesaurus of Social Policy and Environment
TIA — Thesaurus of Transport, Industry and Agriculture.

The first two thesauri serve as adjunct thesauri and can be added to (combined with) any other „topical" thesaurus of the system. Actually, the combination of those two thesauri exists as a separate file as well, i.e. geographical and international organizations names put together. The combination allows to link additionally some terms with RT-relationship that was not possible when they existed in separate thesauri, e.g.

NATO
RT countries of Nato

Thus, merging of those two thesauri enriched the vocabulary and now it serves as a list of identifiers (proper names) that can be added and used in any other thesaurus, not necessarily from the STEBIS system but from an external one as well, without requiring any changes in the structure and in the relationships between concepts. It cannot be used independently of a topical thesaurus.
The advantages of having an adjunct thesaurus isolated from a topical thesaurus are numerous, one of them being the possibility to take off the burden (overload) of many terms from the topical thesaurus. However, there are also some disadvantages, e.g. inability to link some terms with the RT-relationship that is worth indicating, e.g.

POL
reunification of Germany
status of Berlin
Germans

GEO
Berlin
Federal Republic of Germany
former GDR
German Democratic Republic
regions of Germany
West Berlin

or

POL
Yugoslavian question

GEO
territories of the former Yugoslavia
Yugoslavia

or

TIO
EFTA

EBU
free-trade agreement
free-trade area,

or

TIO
World Bank

EBU
development bank
economic relations
financial cooperation
international monetary system

The reason for maintaining eight topic thesauri instead of one common vocabulary is that they are tosupply the needs of different information services (politics-oriented, law-oriented, etc,). Actually, all of them use the same terminology (or rather highly overlapping) but are structured differently. In other words, the

same set of terms (based on EUROVOC) was restructured every time so as to represent the specific needs of various services and domains. For example in the „Law Thesaurus" we have the terms like „family law", "environmental law", "educational law" (thus representing the branches of law) at the highest level of the hierarchy (as top terms). However, in the politics thesaurus (POL) those terms are regarded as lower to the descriptors „family policy", "environmental policy", "educational policy", etc. In the social affairs and environment thesaurus they can be even lower in the hierarchy, this time the top terms being the descriptors like „family", "environment protection", education", etc. Thus in this thesaurus we can have the descriptor entries:

environmental protection
NT environmental law
environmental policy
family
NT family law
family policy,

whereas in the law thesaurus the hierarchy is reversed, e.g.

environmental law
NT environmental policy
environmental protection
family law
NT family
family policy.

The use of those terms (and their scope) is similar if not identical in each subsystem, in other words the use of them for indexing the same documents in the same manner is not hindered by a different place in the hierarchy. The only difference is that the users have a systematic display arranged according to the semantic structure of the field they are interested in, a feature that should enable them to navigate more easily in the network of vocabulary terms and relationships.

DIFFERENCES AND SIMILARITIES BETWEEN THESAURI OF THE SYSTEM

As was already mentioned above, for all thesauri the terminology is drawn (extracted) from the same set of descriptors and restructured according to their specific domain. The result is that they overlap considerably, there are terms that appear in more than one thesaurus, there are terms, however, that are typical only to one field (discipline) and are not likely to appear in the other thesauri. For instance, "ecological vehicle" is likely to be placed (put) both in transport (TIA) and in environment (SOC), "plant-health product" — both in agriculture (TIA) and in environment (SOC), while terms like „organization of transport" will probably be needed only in transport, "fodder crops" only in agriculture, and „atmospheric pollution" only in environment. However, "industrial pollution" would equally well suit environment and industrial thesaurus, and „motor vehicle pollution" — both environment and transport.

The terms appearing in more than one thesaurus (in fact the overwhelming part of them) have in each case the same form and the same set of nondescriptors

leading to them. The relationships to other descriptors can vary, however, especially broader and narrower terms, e.g. in TIC

data-processing law
BT computer science
NT computer crime
copyright
data protection
intellectual property

whereas the same term in LAW has an entry:

data-processing law
BT copyright
NT access to information
computer crime
computer equipment
data protection
dataprocessing
software

The differences in related terms are not so evident (conspicuous), sometimes they are similar or even identical, but sometimes there are discrepancies, e.g.

LAW	SOC
women's rights	**women's rights**
RT care of mothers and infants	RT equal remuneration
equal remuneration	female migrant
equal rights of men and women	maternity leave
equal treatment	parental leave
woman	participation of women
women's discrimination	position of women
women's services (military)	women's movement
women's work	women's unemployment

The differences in hierarchical relationships (broader and narrower terms) are more frequent at the higher level of the hierarchy, especially at the level of the top terms. The lower we move in the hierarchy, the fewer differences in the structure are likely to occur, e.g. "plant-health product" has got different broader terms in SOC (protection of plant life) and TIA (means of agricultural production), but narrower terms are the same since — regardless in what field they are used — the set of plant-health products is the same, i.e.

plant-health product
NT herbicide
insecticide
pesticide

THE SOURCE THESAURUS

From the description of the STEBIS system one can see that there was no source thesaurus from which terms could be derived and which integrated all the terms in one file. The idea was to have separate topic thesauri combined into one system and to create an umbrella index (guide) leading to all of them. The index was to consist of all descriptors and nondescriptors put together into one

alphabetical list, in this way cumulating the terms from all thesauri into one file, supplied with the indicators of the thesaurus in which a particular term appeared, e.g.

plant-health product (LAW, SOC, TIA).

The indicator (the source of a term) should be given in an abbreviation that can be immediately recognized by the user. When seeing those indicators the user can switch to the thesaurus in greatest degree reflecting his scope of interest (point of view), e.g. go to the SOC thesaurus, and check the dictionary context of the term. If not satisfied, he can switch directly (i.e. without going back to the index) into another thesaurus containing this term and check its environment (all broader, narrower, related terms, scope note, indexing instruction).

The practical considerations, like difficulties with implementing the system of thesauri into the database management system operating with only one thesaurus, or the convenience of users (both searchers and indexers) were the reasons for changing this idea. It was decided to make one of the system thesauri a source thesaurus cumulating in one file all the terms. The law thesaurus was chosen for this task, being the main one for the law library (which actually the Sejm Library is) and having the broadest scope (covering in fact all the domains of human activities). For most of the terms there was no trouble in finding the adequate (suitable) branch of law, nevertheless, there were some terms that were needed in cataloguing but seemed bizarre when added to any branch of law, e.g. types of documents (atlas, biography, conference proceedings), political ideologies (anarchism, communism, nationalism), sciences (humanities, earth sciences, social sciences), etc. To avoid artificial hierarchical assignment of terms, the separate part „FRINGE AREAS" was created for those topics. Thus, the full law thesaurus consists of three parallel parts (coincidentally equal in size):

— law,
— fringe areas,
— geographical and international organizations names.

This thesaurus is treated as the source for all the others, i.e. if a term is added, deleted or changed it is done firstly in LAW and than the change is transferred to the appropriate thesaurus of the system. This thesaurus contains all the terms, even if they were not used till now in cataloguing. The source thesaurus can be considered as a vocabulary comprising many specialized thesauri that can be extracted according to the needs, with the condition that their semantic structure (e.g. order of subject fields or facets) can be changed. This is more economical that storing many specialized thesauri separately. In fact, the law thesaurus can be considered as a cumulative thesaurus, according to the definition given by Soergel [1]: „A cumulative thesaurus is a thesaurus that cumulates the information contained in a number of thesauri or classification schemes serving as sources in its construction".

DIFFERENCES IN THESAURI BEING THE RESULT OF VARIOUS SCOPE AND TYPES OF INDEXED DOCUMENTS.

As was already mentioned, all topical thesauri of the STEBIS system — apart from the „Law Thesaurus" turned into the source thesaurus — contain only

the terminology typical for the field they are representing; hence the differences in their set of descriptors and their semantic structure. Besides the differences between thesauri resulting from their scope, there are differences in the use of one topic thesaurus in various services being the result of different materials (types of documents) indexed by various services. For example, within law there are three databases: one for books (library catalogue), one for articles (press cuttings), and one for legal acts (bills, laws, administrative orders, etc.). Preliminary research showed already that those services are not likely to use the same set of indexing terms, even if covering the same subject field, because the content and scope of the documents treated by them vary. For example, for the library catalogue there was a need for the term „Roman law" or „comparative law", as there are many documents on these subjects in the library collection. On the other hand, it is not likely that these terms will ever be required when indexing press cuttings (not many of them having historical perspective), and even less likely to be used in the legal acts database (it is hardly imaginable to have Polish legal acts concerning Roman or comparative law). For the press cuttings database very specific terms are needed quite often (so called „hot topics", names of persons and institutions), that are not likely to appear in the book catalogue. The greatest deviation (specific selection of indexing terms) is expected to take place in the database of audiovisual materials (TV & radio programmes, recordings of parliamentary proceedings, etc.). Thus, even if having the same topical thesaurus for several databases we have to accept some discrepancies in its use.

There are two ways of solving this problem:

— to adopt to the particular database only the terms actually used in indexing, while deleting or not copying all the others;

— to have a copy of the whole thesaurus in every database, assuming that some terms will never be used (which will be indicated by the lack of any number of documents indexed by it).

Both methods have some advantages and disadvantages, nevertheless, even if the end-user may be sometimes frustrated having found the term in an index with zero hits (leading to no documents since there are no bibliographic records attached to it in a particular database), it is much easier to maintain one thesaurus for all databases, and thus the second option seems worth recommending. In this situation we have to accept that a global thesaurus can produce blind references unless a special version of the software is used to skip over blind references for patrons and display them only for indexers.

Perhaps the reasonable solution would be to treat the thesaurus as an index above all the databases using, An indicator would be added to each term listed in it, showing in which database it was already used in indexing, similarly as the index-guide for several thesauri discussed earlier showing in which of them the given term appears. From the indicators given in each entry, one can see at which database (institution, agency) to look for material on a certain subject. Thus for the term „Roman law" in the law thesaurus there would be a reference „LIB" indicating that it was used in the library catalogue, and one can expect to get only books when doing a search on this subject. There is no reason to approach databases for legal acts, press cuttings or audiovisual materials as there are no documents indexed by this term. For other terms, e.g. "illegal abortion" there will be a reference „PLA" (Polish legal acts) indicating that only this database can be

searched for such specific topic (though nothing prevents this term from being used in the future in other databases as well). Still for the others, like „insurance" the two or three references (LIB, PLA, PRE) would be present, indicating that three of the databases (library catalogue, legal acts and press cuttings) would give positive results in answering to this query.

To this kind of index-guide (thesaurus in the function of an index leading to various databases, reflecting the contents of various databases) several other information sources can be added (external one as well, and even in different languages), provided that they use the same or highly compatible terminology, e.g. databases of other parliamentary libraries using a translation of EUROVOC for indexing. The other information tools can also be linked to this kind of thesaurus, as for example, the content of encyclopaedias or terminological dictionaries (in electronic form) giving the definitions or crude information on the thesaurus term (that in many cases will be the sufficient answer for a search).

The user, when doing a search in the system of that kind, will have to take two additional steps:

— selecting the thesaurus in the highest degree reflecting his needs (showing specific semantic structure characteristic to a given domain);

— selecting the database covering the specific topic from among those available for the given thesaurus.

It will be compensated, however, by ensuring that the chosen database will be the most suitable one (no zero-hits and the proper type of documents given in answering to his question), and that the searching aid (its semantic structure) will be tailored to his needs (reflecting his point of view, interests, knowledge, etc.).

REFERENCES

(1) Soergel, D.: Indexing languages and thesauri: Construction and maintenance. Los Angeles, CA: Melville 1974.

SUMMARY

The system of compatibile thesauri for multidatabase institution meeting demands and needs of different information services (literary and user warrants). The idea of an index-guide to many thesauri used in one system. Separating list of international organizations and geographical names as abiunct thesaurus applicable to any topic thesaurus . Creation of a source thesaurus from which various indexing tools can be extracted for different databases and information services. Changes in the hierarchy and associative relationships in the same set of terms reflecting the scope of the thesaurus and the kind of documents to be indexed. The idea of a thesaurus in the function of a guide to various databases. The discussion ilustrated with examples of the system of thesauri maintained in the Chancellary of the Sejm.

Giliola Negrini

TOWARDS STRUCTURAL COMPATIBILITY BETWEEN CONCEPT SYSTEMS

FIELD OF KNOWLEDGE AND CONCEPT SYSTEM

An analysis of particular domains, to which we sought to attribute a concept structure using fundamental common principles, generated important considerations concerning the characteristics of a specific domain. We thus discovered that, by using rigorous arrangement criteria, it is possible to organize conceptual systems whose structure highlights a domain „physiognomy", unknown even to scholars of the subject.

The structural principles which made these conclusions possible are form categories, manifested in the socalled „Systematifier", which establishes the system's external structure or „Gestalt" and provides it with stability for subsequent development. (2, p.106)

The term „system" defines „the *form of arrangement* that brings a set of elements into a structural relationship, a relationship that, considered in isolation is often itself defined as a system". (9, p.156) Hence the form of arrangement that puts concepts of a particular domain into a structural relationship is the concept system of the domain itself.

THE CREATION OF THE CONCEPTUAL SYSTEM

Any field of knowledge is related to an object of reality together with its predicates and related concepts. The form category model (7) which establishes the structure of the system allows us to define the object rigorously.

Definition of the field of knowledge

The identification of the *object* is not always immediate and transparent. Thus, for example, in the case of *science research* it was necessary to make a distinction between the concept of science and that of *research*. This is no easy matter in so far as the term „science", understood as „scientific activity", is to be regarded as having consolidated its position in the world of research, where it is normal to speak of science policy, national and international science, the forecasting and assessment of science and technology, big science and little science, science parks, science in the market and so on. Fig. 1 summarises

some of the often contradictory meanings of *science* which emerged from an etymological and a historical analysis of the term. (10, p.294)

Although „science" represents the concept of the *rigorous, methodical and systematic knowledge* that is open to critical, conceptual and experimental verification — ie, unified knowledge, true knowledge — the term is often also used to identify the process for knowledge. This meaning refers to scientific activity, irrespective of the results it has achieved.

Hence the general agreement that it is more proper to use the term „science research" to describe *the action carried out for the improvement of the process of knowledge and the process of application.* It is a construction supplemented by different elements, fuelled also by external impulses, largely representative of a system representing the confluence of constantly evolving synergies. Research is thus a concrete activity which can be oriented, organized and directed, and which utilizes quantifiable resources. (10, p.297)

Equally evident in this action is the ulterior goal of integrating scientific and technological innovations into the social and environmental fabric.
The diagram (Fig.2) shows the perspective image of the *research system*, located on a plane parallel to those of science and of society with which it interacts (as we can see in the second image looking from above).

In the Latin culture *Scientia*, the etymological origin of *Science*, has two meanings.

„actus sciendi „ or action of knowing
(subjective sense)

SCIENTIA
etymological origin of *Science*

„doctrine", i.e. theoretical knowledge
(objective sense)

These two meanings of scientia reappear in the various European languages with the terms *Wissenschaft*, *science*, *scienza* and *ciencia*.

A) *Science* as **synonym** of *Knowledge*

(Philosophical) Knowledge → '*Science*' the highest expression of knowledge as opposed to common or vulgar knowledge

(Scientific Knowledge) → '*Science*' inquiry aimed at achieving a progressive knowledge of the world and its laws with the special characterIstic of the experimental method

These two concepts, both synonyms of '*Science*' are opposing concepts

Philosophical knowledge *Science (sing.)* *the Science* theory and abstraction	AS OPPOSED TO	Scientific knowledge : *Sciences (pl.).* *a Science* (a particular science) practice and concreteness.

B) Science as **opposite** of *Knowledge*

From the Latin etymology

Science from *'scio (divide)'* identifies the act of separating in order to analyse.

Knowledge from *'cognosco (cum+gnosco)'* identifies the act of bringing together that which is separate in order to represent unified knowledge.

In conformity with the principle of the *evolution* of scientific knowledge, *SCIENCE* comes to be recognized as: *specific form of 'knowledge' which encloses, in whatever fashion and to whatever extent, a guarantee of its own validity.* It represents: *Unified Knowledge, True knowledge*, i.e.: the rigorous, methodical and systematic knowledge that is open to critical, conceptual and experimental verification.

Science of *Science is* the discipline concerned with the „progress" of knowledge characterized by the continuous search for truth.

Fig.1 Definitions of Science

The analysis of these concepts was indispensable to define the term „research", to establish the contents of the „domain" and the demarcation of its boundaries, a task preliminary to the study of the formal structure of the field of knowledge, and to find inside of it the proper classification of the terms which form the domain.

Another case in which it was not easy to identify the object was that of *Italian Literature.* (11) It was possible to note how in the *Literature* domain the fundamental characterising element is not the literary work produced by the author, nor literature seen as a set of literary works, but the «literary fact» contained in the literary work, the author's message to his readers. This is an event which manifests itself in a certain place, in a certain chronological moment, and to which the author gives his own creative image.

These reflections suggested the criteria for the application of the Systematifier.

Structure

Identifying the object of concern of a field of study by defining the concept that represents it properly and precisely, allows us to organize the structure of a concept system, say, of a subject field.

The form categories (mentioned above) determine the facets of a subject field, they are the *structural elements* of such a field and create their own concept clusters. These indicate:

a) theoretical aspects of the object;

b) the typology of the object — ie how it presents itself, how it is made up and its forms;

c) how it evolves and develops, which processes it undergoes and how, with what means and on what conditions;

d) further more specific aspects.

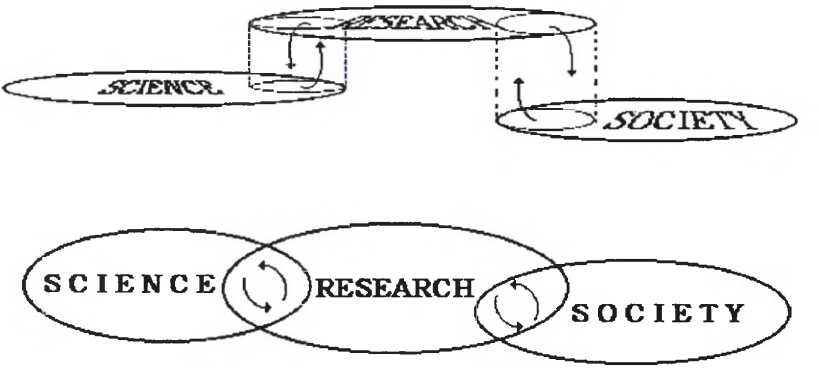


Fig.2 Interactions of „Research"with other systems

Using these criteria to analyse all the predicates of an object, concept clusters emerge which identify the basic aspects of its nature, its intrinsic features, its active or passive action and so on.

Each concept unit relates to a form category on the basis of its intrinsic *form characteristic*, a morphological and invariable characteristic of the concept. The relationship which exists between the concept and the correspondent category, the „categorial form relationship"(7, p.111), establishes the concept's belonging to the category certainly and stably.

Numerous other so-called *variable* or *content characteristics* may be attributed to a concept.

The concept which represents a variable characteristic, common to more than one concept, is a concept on a higher level. It creates a hierarchical structural relationship with the concepts which possess this characteristic and indicates a relationship of the genre-species or part-whole type.

In a particular concept system, where the form categories referring to the object are known as facets, the concept's role in the structure is to be seen in its intrinsic form characteristic and in the categorial form relationship which attributes it to a given facet. In this context, on the basis of its content characteristic, a hierarchical concept relationship is generated which creates a structure inside the facet. Following this procedure, it is possible to attribute a stable role to the concept in the concept system.

In the case of Italian Literature, the attribution of the terms *Handbook, Epitaph, Bestiary* and so on to the „object facet", according to the typology of the work, should be seen in terms of their main property: that is, they contain a literary fact. If these elements did not have this property, they could not be included among the works. In the term „Handbook", for example, the generic meaning of „document gathering fundamental information about a given subject" expresses different

content characteristics of this type of work, which are to be regarded as secondary to the fundamental characteristic of representation of a literary fact. In another concept system, referring, for example, to the subject addressed by the manual, the term might be part of a facet which envisages informational and cultural aspects of the object. It thus becomes clear how the same concept may, according to its revealed characteristics, cover different roles within single concept systems. In the Literature system itself, the term Handbook may acquire a further allocation, albeit secondary to the one above, since it consists of its characteristic as a tool of culture and of the teaching of literary matters. The different „aspects" of a single term in a system determine the different roles for the term, and need to be highlighted.

Scholars of the subject may be surprised to learn that literary criticism does not appear under literary works. It indicates the result of the study of the work carried out to investigate the literary fact and grasp the author's creative image. As such, it represents an *action performed on the literary work* and constitutes a special facet. The interpretative analysis which criticism performs and produces vis-à-vis the literary fact is the image of an unstoppable process. Criticism has a theory, a structure, a typology and a history all of its own.

Likewise, it may appear anomalous that the concept of „literary genre" and typologies of literary genre have not been introduced in the „object facet" relative to literary works and literature (set of works). We decided to keep these concepts distinct and to add to „Foundations of Literature" the term „literary genre" which, with its meaning of „variety of forms of literary expression", is a theoretical element in the analysis of the structural laws of literature and literary language, which are stable over and above contingent circumstances. To represent the various literary genres equivalent to the various typologies of literary genres (biographical, dramatic, narrative literature, etc.), we preferred to use the term „literature", which has been classified in detail according to content, form and mode of production and so on through particular characteristics of division.

In conclusion, in a concept system, which manifests itself through the organization of the terms of a domain, each term finds its own allocation according to the particular characteristics of the concept it represents. Thus different roles may emerge for the same term within the system (eg Handbook). It is possible to highlight the specific role of the term characterised by the proper relationship between its meaning and the system structure (Literary genre). When different terms have the same meaning, preference is given to the one which highlights most explicitly and incisively the characteristics of the object (Biographical literature — Biographical genre).

RELATIONSHIP WITH OTHER FIELDS OF KNOWLEDGE

Although the „*Theory of Science* demands that each scientific field be determined by its proper object area and proper methodology" (8, p. 61), this determination also takes place with the aid of concepts of an *exogenous* nature, imported from another domain.

The work carried out has allowed us to identify two different cases: a) concepts which, albeit belonging to other sciences, integrate directly in the

structure of the domain in question; b) concepts which intervene in the system as a cluster to represent knowledge of another domain.

In the Research system, an example of the first case is the estimate of science and technology research. The procedures to evaluate the *output* of research are performed using statistical, economic and bibliometric knowledge. In these cases the concepts acquired, studied, processed and applied to research are *instrumental* in so far as they represent methods of evaluation of research, and are *new* in so far as they reproduce new knowledge (technology balance of payments [TBP], patents, knowhow, TBP indicators, measures of research quality, innovation statistics, major product innovation, incremental product innovation, process innovation etc). These concept units must of course be integrated into the research system, where they find their natural allocation according to their distinctive characteristics. What basically is established is a sort of *transdisciplinarity,* a phenomenon characterised by the application of the methodologies of one field of knowledge to others. (8, p.63)

An example of case b) is to be found in the Research system where knowledge of a philosophical nature (epistemology of research) or economic, social, health, technological, environment and other policies is expressed by concept units relating to these domains. This knowledge is represented by a limited number of concepts which concisely express phenomena or problems of the domains themselves seen in relation to the object of the system.

The Systematifier model allows us to highlight, in a special facet, all knowledge flows of an exogenous nature fundamental to the concept system.

In the Literature system, to represent the literary fact and motivate the creative image of the author, we witness the intervention of concepts relating to the whole universe of knowledge, from psychology (emotional state, feeling) to politics (the historical moment experienced dramatically or with satisfaction), from religion to sociology, from geography to ethnology, from theatre to the domestic environment. The concepts relating to each of these sciences are allocated in the structure according to their characteristics recognized as literary fact.

We have seen, for example, how the term «flight», used by the author with the metaphorical meaning of „rising above reality", represented the aspiration of the human intellect and was to be attributed to the structure of the *psychology concept*. Likewise, the terms «lie» and «disguise» in their meaning of „dissimulation" have also been correlated to *psychology*.

In the same way, knowledge flows of an endogenous nature (methodologies, results of an activity etc.) are directed to other fields. The Systematifier helps us to identify such concepts highlighting how new knowledge introduced by the domain itself is useful in other fields.

INTEGRATION BETWEEN SYSTEMS

The research action, conducted rigorously, methodically and constantly in a specific field referring to a given object (eg cereals), generates the contribution of new connected knowledge. This field is pertinent to the discipline to which it is traditionally attributed (eg agricultural sciences), but for the development of new knowledge is open to other sciences and relates to them (eg the food industry,

dietetics etc). An interaction thus exists with the other domains for all fields of knowledge and this generates forms of *cross-disciplinarity*.

As we have seen, the Systematifier category model allows us to represent different forms of cross-disciplinarity within the context of a system. Nevertheless, our knowledge does not consist of independent concept systems but of an integrated whole of concepts.

Our mind receives vast, complex wholes of concepts which integrate in a disorderly way, one on top of another. They are interrelated but they appear with different features. This knowledge must necessarily be ordered into conceptual systems, if it is to be used. Systematised knowledge fields are limited knowledge parts which have been analysed and organized and must be integrated. The combination of these „primitive" systems provokes the emergence of complex formations.

Structural compatibilities between hierarchically related systems

The most natural type of combination which manifests itself among different knowledge fields is that of a hierarchical nature.

In defining the object of the system and seeking to define boundaries of the field, a relationship emerges between the specific field and the one above it. A relationship of hierarchical dependence has emerged between the Research system and the Science system, and between the Italian Literature system and the Literature system.

a) The case of „Science" / „Research"

The term „Science Research" is present in the *Science of Science* system and the same meaning as that acquired in the *Research* system should be attributed to it.

Integrating the knowledge of the two systems, correlated by the common term „Science Research", a complex formation appears. It consists on one level of a faceted polyhierarchical structure relating to the *Science* object, and on a second level by another polyhierarchical structure relating to the *Research* object. Two identical *forms of arrangement*, complete with their respective elements, are thus interlinked. Are these systems compatible? In what way can they agree and integrate with one another?

Since these are two hierarchically related *objects*, it is to be expected that they are compatible.

Analysis of the two systems throws up various common elements which may be compared one with another. First and foremost, the term „research": it is interesting to learn whether its meaning in the Science system coincides or is at least consistent with that in the other system, which facet it is attributed to and how it is subdivided. There is structural compatibility between the two systems only if meanings coincide or are coherent.

By definition, every concept unit present in both systems has an allocation of its own in each structure, according to the object of the system. It is thus necessary to assess whether the characteristics of the „research" concept and its subdivision in the Science system correspond to those in the Research system. If divergences

emerge, it is necessary to establish whether they are motivated and correct or whether contradictory concepts are present. In both cases, it will be necessary to establish the specifications or modifications necessary to make this knowledge compatible and combinable.

It is also important to consider that facets, form categories of each system, make it possible to compare and integrate concept clusters which respond to the same arrangement criteria. A common criterion is, for example, that of the theoretical principles of Science and Research, and, more precisely, „Philosophy of Science" and „Theory of Research".

Bearing in mind the definition of Science and Research, there can be no doubt that concepts relating to Theory of Research also relate to Philosophy of Science. This concept integration might be achieved through proper acquisition of concepts in a preconstituted structure, or by envisaging a *form of arrangement* whereby all the concept units belonging to the two clusters may be arranged in a single structure. Once again, the Systematifier may be a reliable guide in organising the concepts belonging to the two clusters.

Concepts such as logic, scientific method, theory, hypothesis etc, which may be present in both systems, must be identified univocally and have a precise allocation in the new structure.

A term might acquire different meanings in the two systems. It is necessary in this case to analyse divergences to distinguish whether: a) the term is being used improperly; b) both uses are proper; c) a particular meaning of the term has to be specified.

In the first case, it is necessary to establish the proper form of the term with respect to its concept content. In the second case, it is necessary to highlight the presence of a single term with different roles in the same structure, because the term represents different characteristics of the same concept. Finally, it is necessary to identify a form of the term which distinguishes its specific meaning and consequently highlights its particular role in the new structure. It would be advisable to make the same adjustments to the original system.

Using the same analytical process, it is necessary to record the use in the new structure of different terms which represent the same concept in the two systems. The arrangement of all the terms which make up the two facets of „primitive" systems generates a specialist, structured and univocal language.

The integration of two hierarchical systems through the introduction of the structure of the lower system must be performed gradually and rationally, analysing and comparing first the object and its typology, then other facets that are consistent and include common concepts.

The new system must reserve a proper allocation to particular facets (§ 2.2, d) according to the knowledge which they represent therein.

This laborious procedure makes it possible to gradually combine two different conceptual structures for a rigorous organization of a univocal specialized terminology.

b) The case of „Literature" / „Italian Literature"

In constructing the Thesaurus of Italian Literature it was necessary to take into account concepts regarding Literature in general, introduced in the particular system of national literature. The form of arrangement by facets makes it possible

to extrapolate general concepts from the particular system and organize the entire Literature system, founding it on the same principles. In this case, the particular system, devoid of all these elements and limited solely to literary concepts regarding national culture, would nonetheless maintain an identical „skeleton", valid for *all the other systems of national literature*.

It is thus possible to create a Literature system and its „satellite systems", regarding National Literature systems. The facet structure of the general Literature system will include principles, characteristics of works, problems, subjects, literary movements and so on common to the Literature of the various countries. While the facet structure of the National Literature systems will refer to concepts relating to the single national or linguistic culture. Relations between these national systems are the object of *Comparative literature*, an autonomous system which is also hierarchically dependent on Literature, which includes conceptual units which represent the elements for comparison between the various national Literatures.

Other structural compatibilities

The above reflections on the two hierarchically related systems may refer more generally to any concept system (eg system A) whose structure contains a *term* which is, in turn, the object of another system (system B). We can compare system A and system B containing the common term.

This case is very similar to that of the above-considered systems, but these systems are not hierarchically related. Once more a complex structure is generated combining two polyhierarchical structures. The integration of these systems seems more difficult to achieve. We distinguish two points of view:

1) *Point of view of the system B, having the common term as its object.*

Although this system has to include all the concepts which relate to its object, it might not encompass all the various aspects under which the object or problems regarding the object are seen and studied by other sciences. It will always be possible to acquire new knowledge of this object developed in any other domain. This knowledge, which might also clash with that already consolidated in the system, will increase the domain of the object adding new concepts to it if a part of the other structure is integrated.

2) *Point of view of the system A, containing the common term.*

The acquisition of system B in this system is more complex to achieve. Replacing the common term, which may have a very specific meaning and a special role, with the whole conceptual system or parts thereof does not appear to be an easy or rational procedure. It might be taken into consideration only if there were perfect consistency between the meaning and the use of the common term. It is, however, vital to be able to establish a conceptual relationship between the term and the entire system, which has the term as its object, with all knowledge relating to it duly organized.

Polyhierarchical structures established according to form categories are also helpful in organising cross-disciplinary domains. A vast, complex bibliography of Humanities Computing has revealed how the potential scientific interest, common to humanistic subjects and computer science transforms the disciplinary duality of

Humanities Computing into a new scientific domain. Its trends consist in defining itself autonomously with proper criteria and organization. (1, p.XIV) The humanities-structured index suggests a very complex classification of subjects in which each discipline is subdivided according to the main topics dealt with. In this case, common principles enable us to organize the various scientific domains homogeneously, and the support of structures already developed — such as those of Literature and National Literatures — may prove to be of fundamental importance in the integration process.

CONCLUSIONS

The combination of concept systems remains an extremely complex problem. The difficulty lies in the limited experience of constructing concept systems which can be compared one with another.

If systems are built with common principles and methodologies, it is possible to compare and consequently integrate different concept systems.

In this integration various semantic problems emerge, which in order to achieve a *univocal, specialist* and *structural language* can be solved. We have seen, for example, how the same term Handbook may have different roles in the new structure, because it represents different characteristics of the same concept.

In the study of the integration process of concept systems relating to the universe of knowledge, it is necessary to establish a general matrix presenting a framework of knowledge according to fundamental categories of form.

The ICC (International Coding Classification) (6, p.93) has a general structure enabling us to develop the organization of new knowledge fields. This basic framework and the construction of new knowledge systems, organized according to fundamental principles, will make it possible to compare systems relating to completely different domains to identify, record and systematize common concepts.

REFERENCES

1 Adamo G., Bibliografia di Informatica umanistica, Bulzoni, Roma, 1994, Informatica e Discipline Umanistiche, 5, 420 p.

2 Dahlberg I., Grundlagen universaler Wissensordnung. Probleme und Möglichkeiten eines universalen Klassifikationssystems des Wissens, K.G. Saur, München, Verl. Dokumentation,1974. XVIII, 366p

3 Dahlberg I., Ontical structures and universal classification, Sarada Ranganathan Endowment for Library Science, Bangalore, 1978, 64p.

4 Dahlberg I., A referent-oriented, analytical concept theory for INTERCONCEPT, in: International Classification, 5 (1978), n.3, p.142-151

5 Dahlberg I., Conceptual definitions for INTERCONCEPT, in: International Classification, 8 (1981), n.1, p.16-22

6 Dahlberg I., ICC — Information Coding Classification — Principles, Structure and Application Possibilities, in: International Classification, 9 (1982), n.2, p.87-93
7 Dahlberg I., Philosophical Foundations of Conceptual Ordering Systems, in: G. Negrini, T. Farnesi, D. Benediktsson (ed.), Proc. of CNR-Istituto di studi sulla ricerca e documentazione scientifica Conf., Rome, Dec. 3-4, 1990, INDEKS Verlag, Frankfurt, 1991, Advances in Knowledge Organization, 3, p.102-119
8 Dahlberg I., Domain Interaction: Theory and Practice, in: „Knowledge Organization and Quality" Albrechtsen H., Oernager S. (ed.) Third International ISKO Conference, Copenhagen 20-24 June 1994, Royal School of Librarianship-Copenhagen, INDEKS Verlag, Frankfurt, 1994, Advances in Knowledge Organization, 4, p.60-71
9 Diemer A., et al., System und Klassifikation. Versuch einer terminologischen und thematischen Klärung des Begriffsfeldes. In Diemer A. (Ed.): System und Klassifikation in Wissenschaft und Dokumentation, Meisenheim, Glan: Verlag A. Hain, 1968, p.150-156
10 Negrini G., Systematization of Scientific and Technological Research, in: „Cognitive Paradigms in Knowledge Organization", A. Neelameghan, M.A. Gopinath, K.S. Raghavan, P.Sankaralingam (ed.), Second International ISKO Conference, Madras 26-28 Aug. 1992, Sarada Ranganathan Endowment for Library Science, Madras, 1992, Ranganathan Birth Centenary Series, 5, p.293-307
11 Negrini G., Zozi P., et al., Thesaurus di Letteratura Italiana, CNR-Istituto di Studi sulla Ricerca e Documentazione Scientifica, Roma,1995, Note di bibliografia e di documentazione scientifica, 59, 279 p.

SUMMARY

„One of the most crucial problems in searching for the same concepts in different databases is the fact that concepts can be expressed in many different verbal versions unpredictably".

A reason for this could be that a different aim, a different analysis level, and a more advanced integration with another science can identify several fields, with their own identity and connected one another. In these fields different characteristics of the same concept are noted (remarked) and consequently diffferent roles are attributed to the same concept in the respective conceptual systems. all this justifies, in part, the many different verbal versions of the same concept. The term „System" defines the form of arrangement that brings a set of elements into a structural relationship, a relationship that, considered in isolation, is often itself defined as system". A structure capable of providing rational location

for its own constituent elements through fundamental principles, allows the organization of the system. These structural principles are categories, which establish a system's external structure or „Gestalt" and provide it with stability for subsequent development. This structure constitues the basis for a hierarchically subdivided conceptual system.

The model of categories called „Systematifier" has been studied and used in elaborating a concept system for Research Documentation and for Italian Literature. These categories play a fundamental role in organizing concepts of a knowledge field because they represent the criteria which identify the role of each conceptual unit in the system. Their use leads to important considerations with regard to the characteristics of the analyzed field and to its limits. In fact, the Systematifier principles require a profound analysis of the reference field in order to define it conceptually and to characterize it, bringing out all its nature and essence, its evolutive tendencies and its relationships with other knowledge fields.This structural organization, of which the seeker is usually only partly aware of, can be very precious help for him.Concept clusters, created by these categories and their subsequent structuringg define each filed within which each concept has his own allocation.

Knowledge flows of an exogenous nature, which however are essential to the development of the field, are acquired by the system. In the same way, knowledge flows of an endogenous nature (methodologies, results of an activity, etc.) are directed to other fields. Systematifier helps in identifying such concepts. Nethertheless our knowledge does not consists of independent concept systems but of an integrated whole of concepts. Systematized knowledge fields are limited knowledge parts which have been analyzed and organized and must be integrated.The combining of these „primitive" systems provokes the emergence of complex formations.If „the form of arrangement" is the same for each field, i.e. each field is structured according to the same categories, the integration will be represented by a complex multidimensional system where two knowledge fields are linked together by a concept. This concept is allocated according to its own characteristics in one system while constituting the object of the specific field in another. Different types of integration will be analyzed in the paper.

Eugeniusz Ścibor

SOME REMARKS ON THE ESTABLISHMENT OF CONCORDANCES BETWEEN A UNIVERSAL CLASIFICATION SYSTEM AND AN INTERDISCIPLINARY THESAURUS (ON THE EXAMPLE OF THE PTC AND THE TCT)

PRELIMINARY CONSIDERATIONS

The investigations that are presented in this paper have encompassed the establishment of concordances between the schedules of the Polish Thematic Classification (PTC) and the Thesaurus of Common Topics (TCT) as well as the description of findings resulting from the assignment of the TCT descriptors to the PTC headings (PTC heading = PTC number + wording). The said investigations were carried out as a part of a large research project called „Concordance Dictionary of Indexing Languages" (CDIL) which mainly consisted in the establishment of concordances between the PTC which was chosen as the master language and three other indexing languages (ILs) which were considered as target languages. These were: the descriptor language based on the TCT, the Universal Decimal Classification (UDC) and the Subject-Heading Language of the National Library in Warsaw. The results of the whole project were presented in the article published last year in „Knowledge Organization" (1). Thus the material contained in the present paper constitutes a part of the material already presented in the aforementioned article; however, this fragmentary article has been significantly enriched by adding some details which had not been presented in „Knowledge Organization".Both ILs under investigation, i.e. the PTC and the descriptor language based on the TCT, have been described with full particulars in the aforementioned article on the results of the CDIL as well as in some previous papers (2,3,4) so there is no need to describe them in this paper. However, it might be a good thing to remind the readers of the fact that the PTC is a universal hierarchical classification system while the TCT is an interdisciplinary thesaurus; the PTC — like most other classification systems — is rather a precoordinated indexing language whereas the descriptor language based on the TCT is characterized — just as other descriptor languages — by a rather high degree of post-coordination. Therefore, in the author's opinion, the results obtained when establishing concordances between the PTC schedules and the TCT may to a great extent coincide with the outcome which would be obtained when establishing concordances between the schedules of another universal classification system and another interdisciplinary thesaurus. The opinion of the author is reflected in the title of this paper („on the example of ...").

CONDUCT AND RESULTS OF INVESTIGATIONS

Thus in the framework of the CDIL project an attempt was made to assign TCT descriptors to all 1330 headings contained in the main PTC database which had been established at the Institute for Scientific, Technical and Economic Information (ISTEI) in Warsaw using the Micro CDS/ISIS 3.0 software; this database covered the PTC main table, i.e. it did not encompass the PTC auxiliary tables. When assigning the TCT descriptors (and the lexical units of the two other target languages) to the PTC headings three kinds (or degrees) of equivalence were distinguished, namely:

full equivalence — when the TCT descriptors were completely or almost completely equivalent to the given PTC headings;

partial equivalence — when the TCT descriptors which would be completely or almost completely equivalent to the given PTC headings were lacking so there were assigned the ones which were broader or narrower than the given PTC headings ; when neither broader nor narrower descriptors were to be found there was made an attempt to find otherwise semantically related descrptors, i.e. descriptors related to the given PTC headings in a non-hierarchical way[1].

Zero equivalence — when both fully and partly equivalent TCT descriptors were

Table 1 shows the results ot the assignment of the TCT descriptors to the PTC headings from the view-point of the appearance of one of the aforementioned kinds of equivalence[2].

The results presented in Table 1 show that the number of cases when the partial equivalence occured was over twice greater than the number of cases when the full equivalence appeared. The proportions of the occurence of the zero equivalence allow us to consider it as rather negligible phenomenon. The relatively high number of cases when a broader TCT descriptor (or set of descriptors) had to be assigned to a PTC heading is mainly due to the fact that in the TCT many subject-fields — such as Philosophy, Mathematics, Physics, Chemistry, Astronomy, Geology, Biology, Medicine, Agriculture, Chemical industry, Construction, etc. — which together occupy a considerable part of the PTC main table are represented in the TCT only to a very small extent, very often only by the name of the given subject-field and by the names of a few more important subfields.

---

[1] The PTC and the descriptor language based on the TCT are two different languages so the statement that a TCT descriptor is broader or narrower than a PTC heading is a simplification done for stylistic purposes. Such statement means in fact that a natural language expression which corresponds to a TCT descriptor is broader or narrower than another natural language expression that corresponds to a PTC heading to which this TCT descriptor should be assigned. Analogically, the statement that a TCT descriptor is related to a PTC heading in a non-hierarchical way means that a natural language expression which corresponds to a TCT descriptor is related in a non-hierarchical way to another natural language expresion which cooresponds to a PTC heading to which this TCT descriptor should be assigned.

[2] The results shown in Table 1 slightly differ from those presented in the article (1), Table 8 but in meantime a few errors were found and corrected.

Kind of equivalence		Occurences	
		number	%
Full equvalence		404	30,38
Partial equiva-lence	TCT descr. (set of descr.) broader than head.	380	28,57
	TCT descr. (set of descr.) narrower than head.	74	5,56
	TCT descr. (set of descr.) otherwise related to PTC head.	391	29,40
	Partial equiv. in general	845	63,53
Zero equivalence		81	6,09
Total		1330	100,00

Table 1. Kinds of equivalence of TCT descriptors in relation to PTC headings

It has been already stated that the PTC is rather a pre-coordinated language whereas the descriptor language based on the TCT is rather a post-coordinated one. Therefore we should not be surprised by the fact that in most cases the assignment of two or more descriptors to a single PTC heading was necessary. Table 2 shows the number and the percentage of PTC headings to which different numbers of descriptors have been assigned.

Number of descriptors assigned to particular PTC headings	Occurences	
	number	%
0	81	6,09
1	612	46,02
2	373	28,04
3	90	6,77
>3	174	13,08
Total	1330	100,00

Table 2. Number and percentage of PTC headings to which different numbers of descriptors have been assigned

Since in many cases two or more combinations of TCT descriptors were assigned to one PTC heading it might happen that the same descriptor could appear at the same time in a few combinations assigned to a given PTC heading. Each occurence of the same descriptor in the whole set of descriptors which was assigned to one PTC heading was counted as the occurence of a separate descriptor. Obviously, such solution resulted in a certain increase of the number of cases which were considered (and counted) as the assignment of three or more than three descriptors to one PTC heading. However, the number of cases when three or more than three descriptors were assigned to a single PTC heading was relatively small in comparison with the number of cases when the assignment of one or two descriptors took place.

The problem of the combination of descriptors will be more fully explained in a further part of this paper.

By combining both parameters (i.e. the kind of equivalence and the number of TCT descriptors assigned to one PTC heading) the results were obtained which are shown in Table 3.

The results presented in Table 3 show that there is no clear correlation between the occurence of different kinds of equivalence and the number of descriptors that have been assigned to one PTC heading, except that: 1) there is a significant positive correlation between the situation when the semantic equivalent of a PTC heading in the TCT is broader than this heading and the assignment of only one descriptor; 2) the percentage of cases when the semantic equivalent expressed in the TCT language is partly equivalent to a PTC heading and related to it in a non-hierarchical way slightly increases as the number of descriptors assigned to one PTC heading becomes higher.When establishing concordances between the schedules of the PTC and the TCT a grammar was adopted which consisted in forming the products ("combinations") of two or more descriptors. Altogether 619 combinations were formed consisting of 2 descriptors, 83 combinations composed of 3 descriptors and 6 combinations consisting of 4 descriptors. Sometimes a few combinations were assigned to one PTC heading, e.g. to the heading 67.08.00 Programming and designing of apartments, buildings and housing estates five combinatiopns were assigned consisting of 2 descriptors, five combinations composed of 3 descriptors and one combination formed of 4 descriptors. However, sometimes such combinations could not be set up because in the TCT some very general descriptors were lacking, such as „construction", „methodology", „system", „theory", etc., which would be easily combined with many other descriptors. E.g. the meaning of the PTC heading 11.09.00 Pedagogical systems could be expressed only by assigning the TCT descriptor PEDAGOGICS; the formation of the combination PEDAGOGIGS^SYSTEM which would exactly express the meaning of the aforementioned PTC heading was not possible because the descriptor SYSTEM (or SYSTEMS) did not exist.

SOME PROBLEMS ENCOUNTERED AND SOME SOLUTIONS ADOPTED WHEN ESTABLISHING CONCORDANCES

One of the problems encountered when assigning the TCT descriptors to the PTC headings was the boundary between the partial equivalence and the zero

Kinds of equivalence		Number of TCT descriptors assigned to one PTC heading					
		0	1	2	3	>3	Total
Full equvalence		—	168 12,63%	133 10,00%	34 2,56%	69 5,19%	404 28,57%
Partial equiva-lence	TCT descr. (set of descr.) broader than PTC head.	—	302 22,71%	64 4,81%	6 0,45%	8 0,60%	74 5,56%
	TCT descr. (set of descr.) narrower than PTC head.	—	23 1,73%	30 2,25%	7 0,53%	14 1,05	74 5,56%
	TCT descr. (set of descr.) otherwise related to a PTC head.	—	119 8,95%	146 10,98%	43 3,23%	83 6,24	391 29,40%
	Partial equiv. in general	—	444 33,39%	240 18,04%	56 4,21%	105 7,89%	845 63,53%
Zero equivalence		81 6,09%	—	—	—	—	81 6,09%
Total		81 6,09%	612 46,02%	373 28,04%	90 6,77%	174 13,08%	1330 100,00%

Table 3, Confrontation of the kinds of equivalence with the numbers of descriptors assigned to particular PTC headings

one. This boundary was not very clear so the decision on the occurrence in a given case of the partial equivalence or of the zero one had to be very often taken more or less arbitrarily. Sometimes the decision on the existence of zero equivalence was taken when there was a great semantic distance between a given PTC heading and any TCT descriptors which could be considered as being somewhat semantically related to it. E.g. in the PTC schedules there is the heading 16.00.00 Theory of literature. Literature which has some narrower (subordinate) subheadings, i.a. the heading 16.02.19 Monographs of writers, whereas in the TCT there appears only the descriptor THEORY OF LITERATURE and any other descriptors which would concern literature do not exist. Therefore, the descriptor

THEORY OF LITERATURE was assigned to the heading 16.00.00 Theory of literature. Literature is narrower in relation to it because is covered only a part of the scope of the said PTC heading. Instead, the heading 16.02.19 Monographs of writers was recognized as equivalent to the descriptor language based on the TCT in a zero degree because the semantic distance between it and the nearest semantically related TCT descriptor THEORY OF LITERATURE was considered as too great.

Some difficulties have arisen from the differences between the hierarchical trees which exist in the two ILs under investigation. E.g. to the PTC heading 10.51.00 Applied psychology corresponds the TCT descriptor which has the same shape as the wording of this heading, i.e. the descriptor APPLIED PSYCHOLOGY. However, this descriptor has been recognized as being narrower in relation to the said PTC heading because in the PTC schedules appears also the subheading 10.51.35 Psychology of labour, which is subordinated to the heading 10.51.00 Applied psychology whereas in the TCT there exists the descriptor PSYCHOLOGY OF LABOUR which is situated on the same hierarchical level as the said descriptor APPLIED PSYCHOLOGY. Thus this last descriptor — contrary to the PTC heading 10.51.00 Applied psychology — does not encompass the concept of psychology of labour and therefore has a scope which is narrower in relation to the scope of this heading.

The executor of the investigations under discussion has tried to take into account the context in which the given PTC headings appear. E.g. to the PTC heading 73.17.27 Trademarks corresponds the descriptor TRADEMARK in the TCT, which could be considered as fully equivalent to this heading. However, the said heading is located in the PTC section (main class) 73.00.00 Foreign trade and therefore its scope should be considered as limited to the trademarks used in foreign trade procedures. For that reason the assignment of the sole descriptor TRADEMARK to the aforementioned PTC heading had been found insufficient and finally a combination of descriptors TRADEMARK^FOREIGN TRADE was assigned to the PTC heading 73.17.27 Trademarks. As it was stated in the paper (1) some PTC headings are provided with so-called contents descriptions containing some expressions which aim is to describe the scopes of the given PTC numbers more fully than this is done by the wordings of these numbers; we may say that a contents description contains some information which supplements the information contained in the wording. When the wording of a PTC heading had no equivalents in the TCT then an attempt was made to assign the equivalents of the expressions contained in the contents description. E.g. in the TCT there could not be found the equivalents of the wording of the PTC heading 73.17.25 Kinds of international goods traffic. However, this heading was provided with the following contents description:

Contents of the heading: — export
— import
— transit
— reexport

Each of these expressions had its exact equivalent in the TCT so the TCT descriptors EXPORT, IMPORT, TRANSIT and REEXPORT were assigned to the PTC heading 73.17.25 Kinds of international goods traffic.Sometimes — in the case of partial equivalence — a choice had to be made of one of the existing

possibilities of expressing the meaning of a given PTC heading. E.g. the meaning of the PTC heading 71.21.00 Telephony could be expressed either by the pair of broader descriptors TELECOMMUNICATIONS (science) and TELECOMMUNICATIONS (process) or by the single descriptor TELEPHONE (between the terms „telephone" and „telephony" there exists an associative, i.e. non-hierarchical relationship). The second possibility had been chosen so finally the descriptor TELEPHONE was assigned to the PTC heading 71.21.00 Telephony.

Some minor differences occuring between the formulation of a given PTC heading were neglected and that of a TCT descriptor which would be considered as equivalent to this heading. E.g. the TCT descriptor TRADE UNIONS was recognized as fully equivalent to the PTC heading 09.33.00 Organizations of employees (trade unions).

4. CONCLUSIONS

The results of the establishment of concordances between the PTC schedules and the TCT have shown that the degree of compatibility between the TCT and the descriptor language based on the TCT is rather low. However, this degree would considerably increase if to the vocabulary of the TCT some descriptors were added denoting: 1) a few very general concepts such as construction, methodology, system, theory, etc.; 2) the fundamental notions of some very important subject-fields, hitherto represented in the TCT to a very limited degree, such as Philosophy, Mathematics, etc.

The results of the investigations that have been described above show that the table of concordances between a hierarchical classification, such as the PTC, and a descriptor language based on an interdisciplinary thesaurus, such as the TCT, cannot serve as tool for the automatic translation of entries formulated in of the ILs in question into the other language. However, the establishment of concordances may contribute to the better understanding of the nature of the ILs between which the concordances have been established and the already established table of concordances can serve as a very valuable material for the revision of these ILs.

REFERENCES

1 Ścibor, E., Tomasik-Beck, J.: On the establishment of concordances between indexing languages of universal or interdisciplinary scope (Polish experience). Knowl. Org. 21 (1994) No 4, p. 203-212

2 Ścibor, E.: Polish subject-field classification — broad ordering system for use on a national scale. In: Perreault, J.M., Dahlberg, I. (Eds): Universal Classification Subject Analysis and Ordering Systems. Proc. 4th Int. Study Conf. on Classification Research, Augsburg, 28 June — 1 July 1982. Vol. 2. Frankfurt/M : INDEKS Verl. 1983 p. 133-138

3 Ścibor, E., Jabrzemska, E.: Thesaurus of Common Topics (TCT): origin, functions, structure. Inf. Process. Manag. 21 (1985) p. 5-10

4 Ścibor, E., Jabrzemska E.: Thesaurus of Common Topics. In: Kent, A. (Exec. Ed.): Encyclopedia of Libr. & Inf. Sci. Vol. 44, Suppl. 9. New York, NY : Marcel Dekker 1989, p.388-395.

SUMMARY

There are formulated some remarks on the establishment of concordances between the tables of the Polish Thematic Classification (PTC) and the Thesaurus of Common Topics (TCT), which was carried out in the framework of the research project called "Concordance Dictionary of Indexing Languages" (CDIL). Several kinds and degrees of equivalence occuring between the PTC headings and the TCT descriptors, which were assigned to them, are analysed. There are noted some technical problems connected with the establishment of above-mentioned concordances (choice of one from two or more TCT descriptors, which are more or less equivalent to the given PTC heading, neglect of minor differences occuring between the formulation of the given PTC heading and that of the TCT descriptor, which can be considered as equivalent to this heading, etc.). There is stressed the difficulty, which arises from the fact, that some lexical units which could be considered as semantically equivalent, belong in each of the indexing languages under investigation to a different hierarchical tree. Taking as starting point the fact, that the PTC is a shallow universal classification system and the TCP is an interdisciplinary thesaurus, an attempt is made to formulate some generalizations and conclusions relating to the more general problem of establishing concordances between universal classification systems and interdisciplinary thesauri.

DISCIPLINE ORIENTED THESAURI

Stephan Hoppe

THE UMLS — A MODEL FOR KNOWLEDGE INTEGRATION IN A SUBJECT FIELD

1. INTRODUCTION

Medicine or health care delivery have to be seen as a multidisciplinary area, where information and knowledge of a wide range of heterogenous sciences has to be integrated.

If, for example, a patient has to be observed, aethilogical knowledge, anatomical knowledge, and knowledge about diseases and symptoms has to be considered. For his treatment additionally e.g. pharmaceutical, physical, and knowledge about surgery technics is required. In most cases financial and legal problems have to been taken into consideration as well. All this information and knowledge is required to be absolutely up-to-date.

For this, especially in the medical subject field tools are necessary that help to overcome the boundaries which make the differences between these sciences in the matter of language, i.e. the vocabularies that are used to name concepts in different information sources in different subject fields. To overcome this conceptual barrier between biomedical information sources the National Library of Medicine started the UMLS project.

The Unified Medical Language System is a long-term research and development project, conducted by the NLM (National Library of Medicine, Bethesda, MD, USA) since 1986.

The purpose of the UMLS is to improve the availability of machine-readable information sources for both retrieval and integration of biomedical information. For that are to overcome two significant barriers to effective retrieval. The first is the variety of ways the same concepts are termed in different information sources and by different people. In this respect it can be spoken of the „Tower of Babel" of medical terminologies. The second is the distribution of useful information among disparate databases and systems.

Modern technologies concerning workstations and network-structures are necessary but not sufficient to connect different users to distributed available machine-readable information sources. A not less important and most difficult requirement is the conceptual connection between user queries and available information which allows a user to receive correctly that answer he expected.

Timely access to accurate, i.e. relevant and up-to-date, information is an important point to improve decision-making and the quality of patient care and research. The number of useful biomedical information sources is increasing rapidly and their distribution among different databases and systems as well. This is underlined by the fact of the increasing number of sources accessible via Internet.

The UMLS strategy recognizes that many of the differences in terminologies used in different databases and by different users reflect significant distinctions in purpose and perspective. In order to fulfill all such different goals the UMLS task will become one of great complexity, even if current efforts in standardization of record structures, transmission formats, and terminology of specific types of biomedical information may help reduce it.

The NLM understands its mission as to support biomedical research and to improve health care delivery by providing ready access to published biomedical information, even though disparities in the way concepts are expressed will continue to exist in different sources. Current projects as Index Medicus®, MeSH®, MEDLARS® and others shall contribute to this mission [2, 5a].

In the subsequent chapters the UMLS project is explained in detail. At first the need for and the initial steps of the project are stated, and thereafter detail accounts of the contents, the structure, and the purpose of UMLS are given. The current state of the ongoing project is described and the plans for the future are described as well.

2. HISTORY AND STRATEGY

2.1 The UMLS Partners

The development of the UMLS is a distributed U.S.-national experiment with a great element of international collaboration. The NLM has composed an own multi-disciplinary research group and additionally contracted with a range of medical informatics research groups distributed throughout the United States. The current UMLS contractors are Lexical Technology Inc., Massachusetts General Hospital, Brigham and Women´s Hospital, the University of Pittsburgh, Columbia University, Stanford University, the Mayo Clinic, and Kaiser-Permanente. The UMLS project has also sought input from a wide range of intended users of UMLS products, including many from outside the United States. Up to date, more than 600 institutions and individuals throughout the whole world are experimenting with the UMLS Knowledge Sources.

The general strategy is to develop the UMLS components through a series of successive approximations of the capabilities ultimately desired. Rapid development and broad distribution of early UMLS products will allow subsequent expansions in scope and complexity to be based on feedback from real applications in a variety of biomedical environments. The NLM sees the success of the UMLS also as dependent on collaborators who are willing to apply their experimental products [2].

During the first phase of the UMLS project (1986 to 1988), the research team concentrated on the investigation of user needs, the identification of required UMLS capabilities, the examination of alternative methods for delivering these

capabilities, and the development of tools for the research effort. Two things were considered as necessary: new machine-readable *Knowledge Sources* and sophisticated user interface programs. The latter should use the highly structured information about biomedical terminology and databases contained in the Knowledge Sources for understanding and interpreting user inquiries, to identify and locate relevant information sources, and to execute successful searches.

2.2 The UMLS Knowledge Sources

The NLM and her partners concentrate mainly on the development of the Knowledge Sources while the interface programs to them should be implemented by users experimenting with these Knowledge Sources [2].

From the start, it was assumed that a „Metathesaurus" would be needed to link terminology and concepts from multiple vocabularies and classifications. As the UMLS team had no preconceptions neither about the specific form, nor of the method for building such a Knowledge Source, several different proposals appeared during progress of research, e.g. to develop a new canonical classification of biomedical concepts. Finally, it was discovered that the desired benefits would not be guaranteed by most of the approaches proposed, or their results would not justify the efforts needed, respectively. Direct linking of alternative names or concepts taken from existing machine-readable vocabularies emerged as a potentially viable way to build the *Metathesaurus*. This approach exploited both automated lexical matching techniques and the structured knowledge embedded in existing biomedical vocabularies, classifications, and databases such as MEDLINE® or SNOMED.

Once the structure and the scope of the *Metathesaurus* was defined, the UMLS project team deemed that a separate, associated UMLS *Semantic Network* was needed; not one of the individual concepts in the *Metathesaurus,* but of Semantic Types or Categories to which concepts from the *Metathesaurus* would be linked. The assignment of Semantic Types to concepts in the *Metathesaurus* provides a high level of categorization of these concepts and also links them to the biomedical „common sense" represented by the relationships among Semantic Types in the Network.

The third UMLS Knowledge Source, the *Information Sources Map* (ISM), should contain both human readable and machine-processable information about the scope and content of publicly available machine-readable biomedical information sources. This information is needed to support automated or semi-automated source selection. Another important component of the ISM records will be procedural information useful in performing successful automated searches of the selected sources.

With the 1994 edition of the UMLS for the first time an additional Knowledge Sources was released which had appeared to be helpful: the *SPECIALIST Lexicon.* This is an English language lexicon with many biomedical terms containing syntactic and other information about terms that is not covered by the *Metathesaurus.*

2.3 Building and Distribution of the Knowledge Sources

The second UMLS development phase (1989 to 1991) included the production of initial versions of the Knowledge Sources. The first experimental edition of the

Metathesaurus and the *Semantic Network* was issued on CD-ROM in 1990. One year later, the first version of the *Information Sources Map* was released, along with updated versions of the *Metathesaurus* and the *Semantic Network*. New experimental editions of the these three Knowledge Sources appear on CD-ROM annually, since 1994 the SPECIALIST Lexicon is released as well.

Broad dissemination of the early versions of the Knowledge Sources promotes the development of prototypes of the interface programs required to deliver the UMLS functionality to the end users. The combination of centralized development of the core of the Knowledge Sources and decentralized development of the applications programs that make use of them was considerd likely to foster progress towards the complex goals of the UMLS project.

The current version of the UMLS, the 6th edition issued in April 1995, is delivered on five ISO 9600 CD-ROMs including an ASCII relational format and the Abstract Syntax Notation One format (ASN.1; a formal language for describing structured and potentially complex data that allows unambigious data exchange across a communication medium such as Internet) for all Knowledge Sources. For the Metathesaurus and the Semantic Network, browsable versions are delivered, also on an additional Apple® Macintosh® Disc. Lexical programs useful in matching local vocabularies, free-text, or user queries to the Metathesaurus are also distributed on the CDs. 1995 for the first time UMLS users may also access the complete UMLS files via an Internet Knowledge Source Server.

The priorities of the third phase (1992 up to now) of the UMLS project are, besides the further development of the Knowledge Sources, to develop an array of useful applications that rely on the UMLS Knowledge Sources to expand and refine the Kowledge Sources based on feedback from early applications. [1, 5a]

3. UMLS KNOWLEDGE SOURCES

The UMLS Knowledge Sources contain the information needed for developing and realizing useful interfaces to biomedical information systems. The knowledge stored in the *Metathesaurus, Semantic Network*, and the *SPECIALIST Lexicon* is to help such interfaces to map user queries to information found in multiple biomedical information systems. The knowledge stored in the *Information Sources Map* should help to identify the most appropriate information source or sources for a query posed.

3.1 The Metathesaurus

Up to now, there have been developed a great number of different vocabularies and classifications according to diverse intentions and requirements. In order to map these information sources to allow universal access, the Metathesaurus, as the central vocabulary component of the UMLS, may be seen as a thesaurus transcending and covering all included individual vocabularies by means of the lexical and semantic links that it provides.

The Metathesaurus contains information about biomedical concepts and terms from a number of controlled vocabularies and classifications. The names, meanings, hierarchical contexts, and inter-term relationships from the source vocabularies are preserved, while new information is added and new relationships

between terms from different sources are added as well. It also includes use information, including the names of selected databases in which the concept appears, and, for MeSH terms, information about the qualifiers that have been applied to the terms in MEDLINE. Co-occurence-information of concepts in MEDLINE and AI RHEUM™ is also included.

3.1.1 The Contents and Scope of the Metathesaurus

The *Metathesaurus* is produced by automated processing of machine-readable versions of the source vocabularies, followed by human expert review. Ist scope is determined by the combined scope of its source vocabularies [3, 5a, 5b].

The 1995 edition of the UMLS Metathesaurus contains 222,927 biomedical concepts named by 478,562 strings from more than 30 source vocabularies. It contains all terms from the 1995 MeSH, NLM's Medical Subject Headings; DSM-IIIR and DSM-IV, the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Third edition (revised) and Fourth editions; the Classification of Nursing Diagnoses; the Home Health Care Classification of Nursing Diagnoses and Interventions; the Nursing Interventions Classification, and the Omaha System: Applications for Community Health Nursing; UMDNS, ECRI's Universal Medical Device Nomenclature System; AI Rheum, the NLM Rheumatology expert system; and WHOART, the WHO's adverse drug reaction terminology. It contains all preferred terms from COSTART, the FDA's Thesaurus of adverse reaction terms; all preferred names of diseases and procedures from ICD-9-CM, the International Classification of Diseases, 9th edition, Clinical Modification; and all procedures and topography terms from the 1993 SNOMED International. It contains selected terms from other vocabularies, including LCSH, the Library of Congress Subject Headings; CRISP, the USPHS Thesaurus for indexing scientific projects; DxPLAIN, Massachusetts General Hospital's expert diagnostic system; PsycInfo, the APA's Thesaurus of Psychological Index Terms; OMIM, Online Mendelian Inheritance in Man; and a set of clinical terms frequently used at three COSTAR sites. [1, 5b]

The MeSH Main Heading terms from the Metathesaurus are in the meantime translated into French, Spanish, and Portuguese language.

3.1.2 The General Structure

The *Metathesaurus* is organized by concept or meaning. Its purpose is to link alternative names and views of the same concept together. This is done by use of a three-level-hierarchy: Each individual name or string in the *Metathesaurus* has a unique (string-) identifier and, for English language only, is linked to all its lexical variants by means of a common term identifier. The same string in different languages (e.g., English and Spanish) has a different string identifier for each language. Different terms with the same meaning are linked to the same concept which has assigned a unique concept identifier. For all strings linked to one term and all terms linked to one concept, respectively, the „preferred form" is stated. The designation of the preferred names is done by an algorithm based on an order of precedence among the source vocabularies.

Additionally, useful relationships between different concepts are to be found and stated. Many of these relationships are derived directly from the source

vocabularies. For example the fact that there is a relationship between „Atrial Fibrillation" and „Arrhythmia" is derived from the hierarchical tree structures in the Medical Subject Headings (MeSH). Other, or more specific relationships are added. In this case, for example, the exact name of the relationship has to be „is_a". Relationships between concepts from different source vocabularies were created and added as well.

All other pieces of information in the *Metathesaurus* are either

- attributes of a concept, e.g. Semantic Type, example: „Disease or Syndrom";
- attribute of a term, lexical tag, example: „Eponym"; or
- attributes of a string, e.g. source of string, example: „MSH95|MH|D000735".

The UMLS project created a word index for all particular words out of any string additionally.

3.1.3 Strings with Multiple Entries

In some (not many) cases the situation may occur, that the same string names different concepts. For example, in MeSH, the string „Cold" is a name of a temperature. In SNOMED II, „Cold" is an alternative name for the „Common cold". As a result the string „Cold" has two different string identifiers, each linked to a different concept in the *Metathesaurus*, and each with an „ambigious string" indicator. To overcome this situation two distinct „Cold" strings, „Cold <1>" and „Cold <2>", have been created for the *Metathesaurus* to allow a clear machine-readable representation of the multiple meanings of „Cold". The plain string „Cold" is linked to both of the qualified strings. In future editions of the *Metathesaurus*, a different approach to distinguish the strings is likely to be adopted, e.g. „Cold (temperature)" and „Cold (disease)" [1].

3.1.4 The Metathesaurus Physical Formats

The six CD-ROMs coming with the 1995 edition of the UMLS Knowledge Sources contain four formats of the *Metathesaurus*: an ASCII Relational format, the ASN.1 format, a HyperCard® application named MetaCard™ to browse the Metathesaurus on an Apple Macintosh computer, and the DOS Coach Metathesaurus Browser [5b, 5c].

3.2 The Semantic Network

The *UMLS Semantic Network* consists of a set of Semantic Types, or Categories, and Relations between these Categories in order to provide a consistent categorization of all UMLS concepts. Each concept forming part of the *Metathesaurus* is assigned to one or more of these Categories. In this way, all information about the concepts themselves is stored in the *Metathesaurus*.

The 1995 edition of the UMLS Semantic Network includes 133 Semantic Types which are related by the hierarchical „is_a" relationship and 49 non-hierarchical relationships [1, 5d].

3.2.1 Structure and Contents of the Network

The Semantic Types are the nodes of the network, and the relationships between them are the links. Major groupings of Semantic Types for „organisms", „anatomical structures", „biologic function", „chemicals", „events", „physical objects", and „concepts or ideas" are done. The level of granularity varies across

the Network — a point not to be seen as a shortcoming, but as a compromise on the way in establishing a set of Semantic Types that will be useful for a variety of tasks, without introducing undue complexity.

Via the „is_a" link the hierarchy of Types within the *Network* is established, and it is used for defining and for finding the most specific Semantic Type to assign to a Metathesaurus concept. The non-hierarchical Relations are grouped into five major Categories which are themselves relations: „physically related to", „spatially related to", „temporally related to", „functionally related to", and „conceptually related to". So, the non-hierarchical links build part of a hierarchy themselves. The Relations provided in the Network do not represent all possible relationships between the defined Semantic Types. Rather, they should be understood as an initial step in the development of a set of relationships which will be useful in fulfilling the purposes of the UMLS project.

3.2.2 Inheritence

The relationships within the *Semantic Network* are stated between high level nodes whenever possible, and generally inherited along the „is_a"-link. The Relation „process of", for example, is defined between the two Semantic Categories „Biologic Function" and „Organism". Because of inheritence, it also holds between the Types „Organ or Tissue Function" which is a „Biologic Function" (via „Physiologic Function") and „Animal" which is a „Organism".

In some cases there may occur a conflict between the placement of a Type in the *Network* and the Relation to be inherited. For example, following the just mentioned relationship, this would lead to the expression „Mental Process" is „process of" „Plant" which is, in fact, nonsense, since plants are not sentient beings. Therefore the inheritence of this link is canceled — it is said to be blocked.

In other cases a link may be defined between two specific Semantic Types, but generally be blocked for all children of these *Types*. An example for this is the Relation „conceptual part of" which links „Body System" and „Fully Formed Anatomical Structure", but should not link „Body System" to all children of „Fully Formed Anatomical Structure" for e.g. a body system is not a conceptual part of a cell.

3.2.3 Assignment of Concepts to Categories

Each indivual concept of the *Metathesaurus* is assigned to one or more Semantic Types of the *Semantic Network*. The assignment of concepts to Types involves both algorithmic procedures as well as extensive review by subject matter experts. Default Semantic Types are assigned to concepts automatically wherever possible. But these default assignments are always additionally reviewed by experts to improve their correctness.

The Relations which are stated between the Semantic Types of the *Semantic Network* do not necessarily apply to all concepts assigned to this Type, but may or may not hold between any two particular concepts. The Relation „evalution_of" holds between the Semantic Categories „Sign" and „Organism Attribute", for example, but not every particular sign may be an evaluation of every particular organism attribute. Thus, signs such as „overweight" and „fever" are evaluations of the organism attributes „body weight" and „body temperature", respectively.

However, „overweight" is not an evaluation of „body temperature" and „fever" not of „body weight".

3.2.4 The Network as a Means of Abstraction

The *Semantic Network* provides not only an indication of the meaning of individual concepts, but an overall semantic structure for *Metathesaurus* concepts. Since *Metathesaurus* concepts are derived from quite different thesauri with different structures, the Network serves as a unifying force. Concepts which share a particular Semantic Type are grouped together forming a generalization for these objects. Thus, in this way concepts derived from multiple source vocabularies are linked together and categorized.

3.2.5 The Semantic Network Physical Formats

The UMLS *Semantic Network* is provided in four formats: an ASCII relational table format, an ASCII unit record format, the ASN.1-format, and a Macintosh application for viewing and browsing the network.

3.3 The Information Sources Map

The *Information Sources Map* (ISM) is the third UMLS Knowledge Source and makes use of the UMLS *Metathesaurus* and *Semantic Network* to support the following facilities:

- Determine which information sources are likely to be relevant to a particular inquiry
- Supply human-readable information to users about the scope, probable utility, and access conditions of particular sources
- Automatically connect to relevant information sources
- Automatically conduct a successful retrieval session on one or more sources

To achieve these aims the *ISM* is under development as a two-components-Knowledge Source. The so-called descriptive component is to describe electronically available biomedical information sources in such a way as to allow a user to find out appropriate sources to a particular query. The current version of the *ISM* consists of a prototype *ISM* file which contains data on the major publicly available databases of the NLM and such from outside the NLM as well.

The second component is called the procedural one of the *ISM*. The purpose of this component is to allow automatic connection to the information source and the execution of a successful information retrieval session [1, 5e].

3.3.1 The Descriptive Component

Four elements are used to indicate the conceptual scope of the information sources: Relevant MeSH terms, MeSH subheadings which denote the contexts in which the main MeSH headings are applicable, Semantic Types from the UMLS *Semantic Network*, and the so-called „Semantic Type Relations" which link two Semantic Types with a Relation from the *Semantic Network* and can be thought of as two nouns (Semantic Types) connected by a verb (Relation). For example, the type relation „Hazardous or Poisonous Substance | causes | Pathologic Function" is used to describe the content of a toxicology database. This approach should expedite the development of software tools that allow the use of natural language queries to discover relevant sources.

This way of applying indexing terms to information sources is similar to the usual way of indexing biomedical literature, except that in the case of literature the most specific applicable term is chosen, while in the case of the ISM, the most generally applicable term is chosen.

Other items describing the contents of information sources by such indexes are, for example,

- a narrative description of each database;
- assumptions on the intended audience;
- the type of information that is contained, e.g. bibliographic database, knowledge base, etc.;
- the probable uses for the database, e.g. for clinical practice, health services research;
- the organization that provides the database;
- the names and addresses of contact individuals;
- the name of the host system;
- and sample records from the database itself.

3.3.2 The Procedural Component

The procedural tools for the *ISM* are under active development, employing network-based open system tools such as WWW (World-Wide Web) or Gopher. Currently a prototype system for automated source identification has been demonstrated, the so-called Sourcerer. Within the next year, the NLM plans to continue work on Sourcerer, as well as developing a network-based tool for remote registry of new *ISM* records, known as *Apprentice*. Because of these developments, the NLM anticipates that the definite version of the ISM will be available via Internet with periodic snapshots appearing on CD-ROM. The format of the ISM will change as required to bring it into harmony with evolving network information retrieval software standards.

3.3.3 ISM Distribution Format

The *ISM* ASCII representation consists of files, included the Main Information Sources Map record, a file describing ISM field definitions and formats, and a file containing valid values for *ISM* fields which require controlled vocabulary. The current release for the first time comprises an ASN.1 description of the files.

3.4 The SPECIALIST Lexicon

The *SPECIALIST Lexicon*, developed to provide the lexical information needed for the SPECIALIST Natural Language Processing System (NLP), is intended to be an English language lexicon with many biomedical terms. The 1995 version includes approximately 65,000 lexical records and over 130,000 forms [1, 5f].

3.4.1 The Scope and Contents of the SPECIALIST Lexicon

For each word or term the lexicon contains syntactic, morphological, and orthographic information. Lexical entries may be single or multi-word terms. Entries are linked to their base form, i.e. the uninflected form of the lexical item; that is, the singular form in the case of a noun, the infinitive form in the case of a verb, and the positive form in the case of an adverb or adjective. Entries which

share their base form and spelling variants are recorded in the same record, that means the lexicon is organized by strictly syntactic features. In this way its purpose is to provide syntactic information about terms which intentionally is not part of the Metathesaurus.

The information delivered with each lexical entry includes an unique identifier, the base form, a syntactic category code, certain agreement information, complementation information if relevant, and various other properties relevant to that particular entry. Eleven syntactic categories are distingiushed, e.g. verbs, nouns, adjectives, and others.

The lexical items coded and recorded in the *SPECIALIST* Lexicon are derived from various sources. Approximately 20,000 lexical items from the UMLS Test Collection of MEDLINE citation records together with lexical items that appear both in the UMLS *Metathesaurus* and the Dorland's Illustrated Medical Dictionary. Additionally a selection of words from the general English language has been included. Therefore are chosen the 10,000 most frequent words listed in the American Heritage Word Frequency Book, and the 2,000 words used in the controlled definitions in Longman's Dictionary of Contempory English.

Included on the CD-ROM containing the SPECIALIST Lexicon are some lexical programs, indexes and databases that may help users to work with the UMLS Knowledge Sources. A range of variations of English lexical items can be generated that may be useful for recognizing lexical variations in biomedical terminologies and texts.

3.4.2 The SPECIALIST Lexicon Physical Formats

The *SPECIALIST Lexicon* is provided in three formats: a relational table format, a unit record format, and in ASN.1. The lexicon relational format is not fully normalized. Duplication of data among different relations and within certain relations exist by design. Developers will need to decide about the extent to which this reduncy should be retained, reduced, or increased for their specific applications.

The unit record format is a frame structure. In this way the slots are the basic lexical attributes, and the fillers express the possible values for that particular lexical item. The ASN.1 format of the lexical data contains information similar to that in the unit record format.

4. APPLICATIONS

To check and to improve the capability of the UMLS Knowledge Sources the NLM with its UMLS project is dependent on a great number of users who develop intelligent user interface programs for building a range of diverse applications on these Knowledge Sources.

The UMLS model of biomedical information retrieval includes the Knowledge Sources, many target machine-readable information sources, smart interface programs, and an involved user. The user must be willing to interact with the smart programs to clarify situations of ambigious inquiries, to select among alternatives presented by the system, and to evaluate the relevance of the information found.

This view of the role of the user is similar to that usually thought about during development of information systems [2].

4.1 The Importance of Applying for Future Development

The UMLS project seeks for applications as many as possible based on the UMLS Knowledge Sources, for they are the best way to test and evaluate their usefulness. In this sense evaluation or application are seen as an important part of the development itself.

Two main questions can be discovered and answered by such doing: Who are the intended users of the UMLS Knowledge Sources, and what kind of information do they wish to be covered? In fact, it has been seen that the primary intention for the UMLS Knowledge Sources which meant them to be used for biomedical information retrieval has been changed to also wanting them to use for describing clinical actions, e.g. in medical records.

Only through evaluation by usage, can the extend to which the UMLS Knowledge Sources cover the topics of interest, and what kind of approaches will be most effective for specific tasks, environments, and users be evaluated effectively.

4.2 Evaluation of the Knowledge Sources

The NLM itself, its UMLS research contractors, and other institutions or individuals are among applying or evaluating the UMLS Knowledge Sources in order to improve the progress and the outcome of the UMLS project.

The NLM coordinates the further devolepment of the Knowledge Sources under respect of the feedback and results of any applications. In this way, institutions which run such applications may alternate the Knowledge Sources of their own to make them cover the features desired, but the NLM itself will review such alterations. If they are thought to be necessary or really useful they will be considered for the next release.

4.3 UMLS as a Means of Sharing across Systems

M. Tuttle and S. Nelson are reflecting (in [3]) about the importance of *sharing, re-use* and *re-purposing* of health care information for improving the quality in health care delivery. They argue that a medical information, once stored for one reason, should be made available for use by other users in other contexts and for other purposes as well. The UMLS Knowledge Sources are, up to date, the best means of facilitating such doing, i.e. making information about patients available for a wide range of varied purposes, because the UMLS is the 'most important single repository of sharable, e.g. re-usable, biomedical experience'.

4.4 The Role of UMLS in Exploiting High-Capacity Information Technics

The power of workstations and world-wide networks is rapidly increasing and the number of Internet users as well. A number of important tools available via Internet, e.g. Gopher, World Wide Web [4], are emerging to assist users in navigating the Internet and in identifying and locating potentially useful information sources. But the technical access to hundreds or thousands of important databases is of little use if there is no means for determing which ones contain information useful in the current circumstances. As technical barriers to information access disappear, better semantic connections become even more important.

The UMLS Knowledge Sources along with intelligent interface programs that make use of them can supply the conceptual link that is indispensable to find out relevant information in the masses of data available on the Internet, at least on the field of biomedical information. On the other hand, it is the Internet that provides the possibility for distributed (timely) access to the UMLS Knowledge Sources. The combination of network advances and the semantic connections provided by the UMLS will speed progress towards the goal of seamless retrieval and integration of biomedical information.

„The UMLS project has much to gain and much to offer in the new era of high-performance computing and communications. The UMLS development team is only beginning to identify and test the ways in which these technical developments can help to make the UMLS goals a reality for today's health professionals and biomedical researchers." [2]

5. CONCLUSIONS

The UMLS project aims to deliver facilities for integrating the many different information sources (data bases, knowledge bases, indexing vocabularies) that are nowadays available and necessary for retrieving sufficient information. That is, to allow uniform access to all these distributed information sources, so a user will find effective answers to a query posed. Modern technologies concerning workstations and network structures, and the increasing number of tools for navigating e.g. the Internet make connections to a great number of information sources possible. But to make retrieval sessions *effective* an additional conceptual (semantic) link is required to allow the user to not only find all information sources, but to find the relevant answers from relevant information sources according to his query. This conceptual link is provided by the UMLS Knowledge Sources.

The UMLS Knowledge Sources are still under development. They have been distributed as early versions, because the general developing strategy is to evolve the UMLS components through a series of successive approximation of the capabilities ultimately desired. Therefore the feedback from users is needed who made experiences with early versions. In this way the developers can receive knowledge about the intended users, and what kind of information they want to be covered by the UMLS Knowledge Sources.

So, the UMLS as a long term project is still ongoing and seeks for a broad distribution of the Knowledge Sources and applications as many as possible based on the Knowledge Sources, for they are the best way to test and evaluate their usefulness. Therefore the Knowledge Sources are available from the NLM free of charge. An experimental agreement form has to be signed that commits the recipients of the UMLS products only to redistribute them as integrated components of computer applications they developed themselves, and to provide feedback to the NLM on how the products are being used and on any changes or enhancements that would make them more useful [5a].

REFERENCES

1. UMLS Knowledge Sources, 6th Experimental Edition — 1995, Documentation, Bethesda MD: National Library of Medicine
2. Humphreys BL, Lindberg DAB, McGray AT. The Unified Medical Language System. Meth. Inform. Med., Vol. 32, No. 4, Schattauer 1993; 281-291
3. Tuttle MS, Nelson SJ,.The role of the UMLS in 'storing' and 'sharing' across systems. Int. J. Biomed. Comput., Vol. 34, Elsevier Science 1994; 207-237
4. Schwartz MF, Emtage A, Kahle B, Neuman BC. A comparision of Internet resource discovery approaches. Comput Systems, 1992; 5; 461-493(ed). Proceedings of the 16th Annual Symposium on Computer Applications in Medical Care. New York: MacGraw-Hill,1992; 306-313
5. UMLS Fact-Sheets, NLM; URL:ftp://nlmpubs.nlm.nih.gov

(UMLS files and documents are located in the /umls subdirectory)

5 a. Unified Medical Language System Fact Sheet, June 1995
5 b. UMLS Metathesaurus Fact Sheet, June 1995
5 c. Coach Metathesaurus Fact Sheet, April 1994
5 d. UMLS Semantic Network Fact Sheet, June 1995
5 e. UMLS Information Source Fact Sheet, April 1994
5 f. UMLS SPECIALIST Lexicon Fact Sheet, June 1995

SUMMARY

The Unified Medical Language System (UMLS) is a long-term research and development project conducted by the National Library of Medicine (NML), Bethesda MD, USA, since 1986. Its purpose is to improve the availability of machine-readable information sources for both retrieval and integration of biomedical information of multiple sources, irrespective of the variations in the way similar concepts are expressed in different sources and despite the scattering of useful information among disparate computer systems.The UMLS approach involves the development of so-called „Knowledge Sources" that can be used by a wide variety of application programs. Currently there are four UMLS knowledge sources under development: a Metathesaurus, the SPECIALIST Lexicon, a Semantic Network and an Information Sources Map (IMS).

The Metathesaurus, as the central vocabulary component of the UMLS, may be seen as a thesaurus transcending and covering all included individual vocabularies by means of the lexical and semantic links that it provides. It is organized by concept or meaning. Concepts of more than 30 biomedical vocabularies are included, additional information to these concepts is added while the original meaning is always preserved. The Metathesaurus is organized in a three-level hierarchy (concept — term — string). Relationships among terms

are included to support the mapping from the user's terms to appropriate controlled vocabularies.

The Lexicon contains syntactic information for terms: many Metathesaurus terms, component words, and additionally other English words that do not appear in the Metathesaurus for they are no concepts itselves, e.g. verbs. The UMLS Semantic Network consists of a set of semantic types or categories in order to provide a complete categorization of all UMLS concepts. It contains information about those types and the permissible relationship among these types. Each specific concept forming part of the Metathesaurus is assigned to one or more of these categories. The Information Sources Map contains both human-readable and machine-processable information about the scope, location, used vocabulary, syntax rules, and access conditions of biomedical databases in order to help a user to find out appropriate sources to a particular query. One goal is to allow automatic connection to the information source and the execution of a successful information retrieval session. The UMLS knowledge sources are released annually in multiple formats on CD-ROMs. The breadth and complexity of the UMLS requires the knowledge and skills of experts in many firelds. For this reason the development is a distributed national experiment with a great element of international collaboration. The general strategy is to develop the UMLS components through a series of successive approximations of the capatibilities ultimately desired. Therefore the UMLS project has sought input from a wide range of intended users of UMLS products, including many from outside the United States.

Stella G. Dextre Clarke

INTEGRATING THESAURI IN THE AGRICULTURAL SCIENCES

1. INTRODUCTION

Rather than give a report of research completed, this paper presents an account of progress on a long term project which still has some years to run. It then invites comment on the future direction of the project.

2. BACKGROUND DESCRIPTION OF THREE DATABASES

In the beginning there were three bibliographic databases in the field of agriculture: AGRICOLA, AGRIS and CAB ABSTRACTS. Fig 1 shows some of their vital statistics. Notice that although they all cover agriculture in a very wide sense, with many related fields, there are significant differences in the selection of "related fields" covered by each. For example, AGRICOLA has extensive coverage of home economics, not matched in the other two. CAB ABSTRACTS is strong in sciences such as parasitology, entomology and mycology with application in medical as well as veterinary fields. AGRIS has special advantages in its coverage of grey literature. Given the variations also in geographical coverage, in record structure and in policy for including an abstract in the bibliographic record, a search conducted in each of the databases often yields three very different results.

3. DIFFERENCES IN INDEX LANGUAGES

Furthermore, the databases use different index languages. AGRIS uses AGROVOC, a multilingual thesaurus containing approximately 15,000 descriptors. The French, Spanish and English versions of AGROVOC are maintained by a Working Group drawn from several countries, coordinated by FAO. Additional versions in German, Portuguese, Italian and Arabic are maintained by agencies with an interest in these languages.

Both AGRICOLA and CAB ABSTRACTS use the CAB Thesaurus, which with approximately 50,000 descriptors and 10,000 lead-in terms is three times the size of AGROVOC but exists in English only. Despite the tremendous advantage of sharing one thesaurus, the differences in scope between the two databases cause some difficulties for indexers and searchers. Consequently, a number of

descriptors in the thesaurus are tagged "(AGRICOLA)" to indicate that they are available for use in that database but not in CAB ABSTRACTS. A further complication is the differences in spelling noted on either side of the Atlantic. CABI's editorial house style is British, whereas NAL's is American. The thesaurus therefore provides for two different allowed variants of terms like Defence/Defense, Anaemia/ Anemia, etc. They are marked BF (British Form) and AF (American Form) respectively. So, although the same thesaurus is used for both databases, there are some differences in the vocabulary which is actually used for each, not to mention differences in indexing practice.

The first editions of both thesauri, published in 1982 and 1983 respectively, were compiled by variations on what I shall simplistically call the A-to-Z method. That is to say, one collects from a variety of resources all the vocabulary that might be useful, sorts the terms into alphabetical order and then sets up relationships between the terms. For both thesauri, the procedure also involved extensive consultation with experts, and reference to authoritative sources of terminology. In the case of AGROVOC, the terms were first grouped into approximately 18 classes. But the basic A-to-Z approach contrasts with what I call the classification method. In the latter approach, candidate terms are first sorted into systematic order based on a classification scheme. This assists the identification of both synonyms and omissions, and helps the compiler set up relationships systematically. Only after this are all the terms sorted into alphabetical order. The lack of such a systematic approach, especially in a very large thesaurus, inevitably leads to anomalies, for the connections between alphabetically remote terms of similar meaning are difficult to spot. The editors of both thesauri will testify to this, as they have spent the years since the first editions of the thesauri came out correcting a steady stream of anomalies.

It should be noted that the above remarks do not apply to one very important group of terms in the thesauri, namely the scientific names of organisms. Agriculture, horticulture, forestry and veterinary sciences are fundamentally preoccupied with plants and animals of all sorts: not just the (relatively few) animals and plants we eat and/or have domesticated, but all the weeds, fungi, bacteria, insects, protozoans and viruses that affect them. Much of the efficacy of cultural methods and pest control depends crucially on correct identification of the organism concerned. But the naming of organisms is a notoriously confused business, starting with the proliferation of common names in different languages, dialects and provinces and extending to the arguments which even now prevail among taxonomists. Any thesaurus purporting to assist the flow and beneficial use of agricultural information must offer its users a great deal of guidance to choosing the correct, accurate and scientifically recognised name for each relevant organism. And so, over 27,000 terms in the CAB Thesaurus are taxonomic names. From the start, the editors of the CAB Thesaurus placed great emphasis on incorporating the hierarchies of organism names according to the strict principles of taxonomy. So here we already have a very systematically organized sector in the thesaurus, amounting to approximately half of its content and giving it enormous strength as a tool for information retrieval.

A lot of the size difference between the two thesauri is accounted for by the organism names. AGROVOC has about 6,000 of these, amounting to much less specificity than is available in the CAB Thesaurus. However, those maintaining

	AGRICOLA	AGRIS	CAB ABSTRACTS
Producer	NAL (National Agricultural Library of US Department of Agriculture)	FAO (UN Food and Agriculture Organization) coordinates, from input provided by national centres, regional centres and some international centres	CAB INTERNATIONAL (an intergovernmental, not-for-profit organization)
Size (end of 1994)	over 3 million records	over 2.2 million records	over 3 million records
Date of earliest input	1970	1978	1972
Subject coverage	Agriculture and related fields	Agriculture and related fields	Agriculture and related fields
Geographical coverage	International, with the emphasis increasingly on US publications	International, with publications from all countries which cooperate	International
Type of coverage	Research and extension literature with some grey literature published in USA	Research literature, conventional and grey literature	Research literature; grey literature limited to conference proceedings
Availability of abstracts	present in English in a small but increasing proportion of records	present in 22% of records (proportion increasing); language optional	informative abstracts in English on over 90% of records
Distribution media	online, CD-ROM	online, CD-ROM, print	online, CD-ROM, print

Fig 1: Vital Statistics of Three Agricultural Databases

AGROVOC have used the CAB Thesaurus as a reference source to the taxonomic nomenclature and so there is substantial alignment between the two in this sector. It may be worth noting that both thesauri are highly respected and much used in libraries, information centres and systems worldwide.

4. DIFFERENCE: A STRENGTH OR A WEAKNESS?

I have tried to bring out some of the differences between the three large secondary sources and their indexing languages, but have perhaps omitted to emphasize their strengths. It is a remarkable thing that three such excellent international databases thrive in this one field. Although agriculture is one of the most basic activities of man, fundamental to the well-being of all but exceptional societies, it very often does not command the high profile it deserves. In many countries the competition for research funding sees Agriculture the poor cousin behind Defence, Manufacturing Industry, and many others. But those researchers who are attracted into agriculture can in fact avail themselves of excellent information resources. The practical differences between the databases in terms

of coverage, quality and pricing mean that the user can choose the source which best suits his research interests and financial means.

Does this happy situation breed total satisfaction? Of course not! While it is true that each of the three producers receives numerous compliments on the service provided, at the same time some users do find difficulty in coping with three different databases (1). Understandably, they would prefer not to have to learn three different access procedures, three different coverage policies, and three different indexing languages. While few would seriously suggest the databases should merge and form a monopoly, still there are ways in which things could be simplified. In particular users have urged the compilation of one "Unified Agricultural Thesaurus". In response to such feedback from the international research community over the last decade or more, the three producers CABI, FAO and NAL have tried increasingly to work together.

5. AGREEMENT TO COLLABORATE

In June 1991 representatives of the three database producers met in Bonn to develop a course of action "towards more consistent access to agricultural information". The meeting was organized by ZADI (Germany's Zentralstelle für Agrardokumentation und -information), with the participation also of PUDOC (the Dutch Centre for Agricultural Publishing and Documentation) and DG VI of the European Commission. Agreement was reached on the following mission and goals:

Mission Statement for the Unified Agricultural Thesaurus Project

The Unified Agricultural Thesaurus Project aims to provide users with a comprehensive automated multilingual thesaurus system to facilitate access to agricultural databases.

Goals

Thesaurus Unification

Create a comprehensive, multilingual agricultural thesaurus system that harmonises AGROVOC and the CAB Thesaurus in a way that improves access to current and retrospective agricultural information for searchers and indexers.

Thesaurus Management System

Develop a computer based system for creating, maintaining, distributing and otherwise providing access to the unified thesaurus.

Governance Structure

Develop a governance structure to support the unified thesaurus system which includes shared responsibility for its development and management.

Funding

Procure funding to support the initial thesaurus and systems development as well as for ongoing operational activities.

Participants noted the desired harmonization process would be difficult if based on the existing thesauri as structured in alphabetical order. It was decided to develop a classified structure for both the CAB Thesaurus and AGROVOC as the first step towards unification. Each of the thesauri would then be reorganized to slot them into this structure. This should facilitate comparison and analysis on a systematic basis.

The EC subsequently agreed to provide some funding towards the first stage of the activity, in the context of another project coordinated by ZADI, aiming to develop an expert system for use with agricultural research databases (2).

6. PROGRESS TO DATE

Work began. And it has been a huge amount of work, because the thesauri are so large. The CAB Thesaurus alone has over 50,000 descriptors, and thus is so large that it is difficult for an individual to retain a grasp of the overall scope and the extent of detail in each subject comprised. An essential preliminary task was to prepare a computer printout of all the "Top Terms" present, showing the full underlying hierarchies. This printout was analysed and a structure was developed which would be capable of accommodating all the hierarchies. AGROVOC was similarly analysed by NAL and FAO and the two structures were compared. In September 1991 agreement was reached on a very high-level Classification Outline which would be applicable to both thesauri. (See Fig 2.)

Plans were also laid for development work to proceed at CABI's headquarters in England and at NAL in the USA, with the cooperation of FAO.

The characteristics of the required thesaurus or thesauri were agreed, to make sure that the work done at separate locations would be broadly comparable. Broadly speaking, the main output of this stage of the work at each site would be a classified, polyhierarchical thesaurus in the English language. Readers familiar with MeSH (3), or Thesaurofacet (4), or the BSI Root Thesaurus (5) may draw an analogy between this output and the classified sections of any of these landmark thesauri.

It was resolved that work on the two thesauri should be phased so that, for any one section, first one party would work on it and then the others would attempt to analyse the other thesaurus following the same structure. Thus the development work would pass from one team to the other and back, with comments and feedback. A workplan was agreed for the sections that would be developed by CABI and NAL/FAO, respectively.

Thesaurus management software was needed in order to facilitate the work and store the results. Regrettably, no suitable package was found, despite exhaustive research. As an interim measure CABI decided to use the outliner package Grandview, while NAL adopted TCSP (Liu-Palmer Thesaurus Construction System, Professional Edition).

All three UAT partners recognized that, while the software packages mentioned above would serve the project initially, in the longer term more adequate software was required. A specification of the requirements was prepared and agreed. Interest was expressed by the Canadian organization IDRC (International Development Research Centre), who were hoping to develop a general purpose thesaurus management module for their software MINISIS. Agreement was reached with IDRC to develop the required software in support of the continuing UAT project.

As it has transpired, development of the software (to a very demanding specification) has proved more difficult than anticipated. We now expect to take

1. GENERAL
   - 1.1 Common terms
   - 1.2 Organizations
   - 1.3 Regional geography
   - 1.4 Research
   - 1.5 Methodology
   - 1.6 Communication and information
   - 1.7 Computer science
   - 1.8 Mathematics and statistics

2. PHYSICAL SCIENCES
   - 2.1 Physics
     - 2.1.1 by discipline
     - 2.1.2 by phenomena
     - 2.1.3 by properties
   - 2.2 Chemistry
     - 2.2.1 by discipline
     - 2.2.2 by phenomena and properties
     - 2.2.3 by entities

3. EARTH SCIENCES
   - 3.1 Geology
   - 3.2 Soil science
   - 3.3 Hydrology
   - 3.4 Oceanography
   - 3.5 Meteorology and Climatology
   - 3.6 Geography

4. LIFE SCIENCES
   - 4.1 Biology
     - 4.1.1 microbiology
     - 4.1.2 botany
     - 4.1.3 zoology
     - 4.1.4 biophysics
     - 4.1.5 biochemistry
     - 4.1.6 physiology
     - 4.1.7 genetics
     - 4.1.8 evolution
     - 4.1.9 biological structure and form
     - 4.1.10 taxonomy
   - 4.2 Ecology
   - 4.3 Behaviour and psychology
   - 4.4 Natural history
   - 4.5 Organisms
     - 4.5.1 taxonomic groupings
     - 4.5.2 non-taxonomic groupings

5. APPLIED SCIENCE AND TECHNOLOGY
   - 5.1 Health and pathology
   - 5.2 Applied human and animal nutrition
   - 5.3 Agriculture
     - 5.3.1 by discipline
     - 5.3.2 soil management
     - 5.3.3 plant production
     - 5.3.4 animal production
     - 5.3.5 breeding/applied genetics
   - 5.4 Forestry
   - 5.5 Aquaculture
   - 5.6 Fisheries
   - 5.7 Environmental science
     - 5.7.1 environmental control
     - 5.7.2 environmental management
     - 5.7.3 safety
   - 5.8 Technology and engineering

6. SOCIAL SCIENCES AND HUMANITIES
   - 6.1 Education and extension
   - 6.2 Human ecology
   - 6.3 Sociology
   - 6.4 Economics
   - 6.5 Administrative science
   - 6.6 Socioeconomic and political systems
   - 6.7 Law and legislation
   - 6.8 Culture and humanities

Fig 2: Classification Outline Scheme (1991 version)

delivery towards the end of 1995. Without it the classification work has proceeded at a slower rate than we might have wished.

But let us now return to 1991, when the workplan and Classification Outline were agreed. The next stage was to develop and agree a Classification Structure going to greater depth than the Outline, though not to the full depth of either thesaurus. Fig 3 shows the Life Sciences section of the Structure that was completed in 1993. Note how it relates to section 4 of the Outline in Fig 2. If we think of the classified CAB Thesaurus as an iceberg (made up of 60,000 terms and their interrelationships) then the Classification Structure (with around 500 terms)

- biology
  - microbiology
  - botany
  - zoology
  - biophysics
  - biochemistry
  - physiology
    - physiology by organism
      - microbial physiology
      - plant physiology
      - animal physiology
    - physiological requirements
    - metabolism
    - growth and development
    - reproduction
    - digestive and nutritive physiology
    - circulation, transport mechanisms and gas exchange
    - secretion and excretion
    - physiological regulation
    - movement, neurophysiology and electrophysiology
  - genetics
    - genetic theory
    - genetic makeup
    - molecular genetics
    - cytogenetics
    - population genetics
    - microbial genetics
    - immunogenetics
  - evolution
  - biological structure and form
    - biological structure and form by attribute
      - anatomy
      - morphology
      - plant habit
      - ultrastructure
    - biological structure and form by entity
      - cells
      - tissues, organs, parts and systems
    - taxonomy
      - biological nomenclature
      - classification methods
      - taxa
      - taxonomic status
      - taxonomic products
- ecology
  - ecology by discipline
  - biogeography
  - ecological processes and phenomena
  - levels of ecological organization
    - ecosystems
    - ecological communities
    - populations
  - environments and habitats
    - environments
      - aquatic environments
      - terrestrial environments
    - habitats
  - environmental factors
- behaviour and psychology
  - behaviour
    - behaviour by organism
    - basic and maintenance behaviour
    - social behaviour
    - abnormal behaviour
  - psychology
    - affective psychology
    - developmental psychology
    - higher mental processes
    - individual psychology
    - social psychology
- organisms
  - organisms, by non-taxonomic groups
  - organisms, taxonomically arranged
    - viruses
    - prokaryotes
    - fungi
    - plants
    - animals

Fig 3: Life Sciences section of Classification Structure

is the tip of the iceberg, while the Classification Outline is a thin layer of snow covering the tip. The same may be said of the classified AGROVOC except that it is not quite so large below the water-line. The very positive achievement is that the two tips visible above sea-level are identical, representing total harmonization at this high level.

After agreement was reached on the Classification Structure in mid-1993, it was used to classify both thesauri completely, by the end of 1994. The extracts in Figures 4 and 5 illustrate the results for the two thesauri respectively, and they do show how the thesauri can now more easily be compared. Inevitably the process revealed some difficulties which have led all parties to revisit the Structure. Some modifications to it have recently been agreed and are now being reflected back into

(Applied H uman and Animal Nutrition)

(characteristics of foods and feeds)

- food composition
- antinutritional factors
- chop length
- nutritive value
  - index of nutritional quality (agricola)
  - digestibility
    - energy digestibility
    - protein digestibility
    - in vitro digestibility
  - energy value
    - calorific value use energy value
    - metabolizable energy
    - digestible energy
    - starch equivalent
  - nutrient availability
  - nutritive ratio
  - protein value
  - protein efficiency ratio
  - vitamin content
  - total digestible nutrients
- food acceptability
- organoleptic traits
  - flavour
    - egg flavour
    - milk flavour
    - oxidized flavour
  - odours
    - aroma
  - palatability
  - taint
    - boar taint
    - milk taint
  - rancidity
  - tastes
    - bitterness
    - sweetness
  - texture
  - tenderness
- proximate analysis (agricola)
  - crude fibre
  - crude protein

(diet and feeding)

Fig 4: Extract from Nutrition schedule of classified CAB Thesaurus

8. COULD THERE BE ANOTHER SOLUTION?

Five years have passed since we formed the basic plan of action. Meantime, technology has moved on, people's expectations may have changed and many novel ways of accessing information have emerged. It is appropriate to step back and re-evaluate the priorities. Here are some of the options for circumventing the difficulties:

(APPLIED HUMAN AND ANIMAL NUTRITION)
. (CHARACTERISTICS OF FOODS AND FEEDS)
. . carbohydrate content
. . dietary fibres
. . feed quality
. . crude fibre
. . keeping quality
. . nutritive value
. . . digestibility
. . . digestible cellulose
. . . digestible fibre
. . . digestible nitrogen
. . . digestible starch
. . . energy value
. . . . starch equivalent
. . . feed units
. . . nutritional losses
. . . protein quality
. . . total digestible nutrients
. . organoleptic properties
. . . flavour
. . . . off flavours
. . . rancidity
. . . tenderness
. . . texture
. . proximate composition
. . . crude fibre
. . . crude protein
. . . dry matter content
. . . nitrogen content
. . . protein content
. (DIET AND FEEDING)

Fig 5: Extract from classified AGROVOC

AGROVOC	CAB Thesaurus
NUTRITIONAL REQUIREMENTS uf nutrient requirements	NUTRIENT REQUIREMENTS uf nutritional requirements
NUTRITIVE VALUE uf nutritive ratio	NUTRITIVE VALUE NUTRITIVE RATIO
nutrition, animal USE animal nutrition nutrition, clinical USE therapeutic diets nutrition, human USE human nutrition nutrition, plane of USE feeding level AND food composition nutrition, plant USE plant nutrition [There is no descriptor "NUTRITION"]	NUTRITION NT1 animal nutrition NT1 applied nutrition (agricola) NT1 basic nutrition (agricola) NT1 child nutrition (agricola) NT1 clinical nutrition (agricola) NT1 community nutrition (agricola) NT1 elderly nutrition (agricola) NT1 infant nutrition (agricola) NT1 maternal nutrition (agricola) NT1 mineral nutrition NT1 plant nutrition NT1 preventative nutrition (agricola) NT1 total parenteral nutrition NT1 undernutrition
food composition USE proximate composition	FOOD COMPOSITION

Fig 6: A few examples of simple differences between the two thesauri

1 Do not worry too much about achieving excellent retrieval in the backfiles. Go right ahead with production of a compromise thesaurus, which will be better than either of the constituent thesauri because of its systematic basis. All three databases change over to indexing with the new UAT. After a few years no one will care about the (25 years of) backfile.

2 Move over to the compromise thesaurus as in Option 1, but also re-index the backfiles. By the time this project is undertaken, over the three databases this could mean re-indexing 10 million records! Much of the re-indexing could be done automatically by algorithm. But for millions of records it would be necessary either to accept low-quality indexing or to inspect the records individually.

3 Forget the compromise thesaurus. Maintain the two thesauri for their respective databases. Develop algorithms which will convert search statements from the one to the other. The parallel classification structures should facilitate this task. Of course, every time either thesaurus is updated, the algorithms will have to be updated.

4 As in option 3, leave the two thesauri to be used side by side for their respective databases. Forget about converting search statements from one to the other and concentrate on a problem of much greater interest to end-users: that is, converting search requests from users' natural language to one or more of the databases/thesauri as appropriate to each given query. The hierarchical structure at the backbone of both thesauri, combined with the semantic network of associative (RT) relationships should be a marvellous tool for analysing a searcher's query on its merits, without pre-established algorithms.

All of the above options are to some degree painful or expensive. In trying to decide between them I find myself torn two ways. As a trained information scientist, I take it for granted that searches should aim to retrieve All the relevant records and Only the relevant records. To achieve good recall and precision, consulting a thesaurus seems obvious, natural and straightforward. But I notice, as I am sure you do too, that end-users are not interested in thesauri. They do not always seem to care greatly about retrieving All the relevant records! The people who have urged us to prepare a UAT are mostly not end-users but librarians, documentalists and information scientists. Astonishingly, even they have not raised any alarm bells concerning the implications of the proposed UAT for backfile retrieval. I am driven to the conclusion that the Thesaurus as a species is little understood. So why do we spend the huge resources involved in making a bigger and better thesaurus if the users do not really care? Are we like lost Crusaders pursuing the Holy Grail? Or are we hot on the scent of something which can be built transparently into user interfaces so that end-users can search comfortably and effectively in their own native languages? On that question I would like to open the floor to comments on the options, and perhaps some suggestions as to how we should proceed.

REFERENCES

(1) Deselaers, N.; The Necessity for Closer Cooperation among Secondary Agricultural Information Services: An Analysis of AGRICOLA, AGRIS and CAB. IAALD Quart. Bull. 31 (1986) No.1, p.19-26

(2) Kőnig, Dr E.; Application of an Expert System to Assist Retrieval from Agricultural Databases (AGRIRES). IAALD Quart. Bull. 37 (1992) No.1-2, p.31-34

(3) National Library of Medicine; Medical Subject Headings (MeSH). Updated annually. Bethesda, USA. National Library of Medicine

(4) Aitchison, J.; Gomersall, A.; Ireland, R.; Thesaurofacet: a thesaurus and faceted classification for engineering and related subjects. Whetstone, Leicester, England: English Electric Company Ltd. 1969

(5) British Standards Institution; ROOT Thesaurus, 2nd ed. Milton Keynes, England: BSI 1985. 2 vols.

SUMMARY

The databases AGRICOLA, AGRIS and CAB ABSTRACTS show significant overlap, although each has distinctive features. Two of them are indexed using the CAB Thesaurus; the other uses AGROVOC. To make it easier for users to search all three databases, the producers are working together on a „Unified Agricultural Thesaurus". The first step was to agree a classification structure appropriate to the two thesauri. Then all the terms in each thesaurus were organized in accordance with this structure. Systematic comparison and analysis are now possible. The options for future work are discussed.

ACKNOWLEDGEMENTS

Sincere thanks are offered to Monique Bonnichon (Senior Information Systems Officer (Training and Methodology)) of FAO and Lori Starr 9Head, Thesaurus Management Section) of NAL, for their helpful comments and corrections of an earlier draft.

Corentin Roulin

BRINGING MULTILINGUAL THESAURI TOGETHER : A FEASIBILITY STUDY

I. INTRODUCTION

The study of which the methodology and results are summarised in this paper was commissioned in 1992[1] by organisations responsible for the management of

• the *European Education Thesaurus* (known as the EUDISED Thesaurus before the 19912 edition) which is managed jointly by the Council of Europe and the European Commission (this thesaurus is referred to as „TEE" below);

• the *Multilingual Thesaurus of Vocational Training* of CEDEFOP, the European Centre for the Development of Vocational Tra.ining (this thesaurus is referred to as „VTT" below).

Examining the feasibility of „bringing together" two thesauri means attempting to determine the best balance that can be achieved between different categories of needs and different possible levels of"rapprochement", or integration, given the characteristics of the thesauri concerned.

This study thus included (the numbering corresponds to the sections in this paper):

2. a *comparative analysis of TEE and VTT* considered separately;

3. the definition and choice of the needs to be taken into account;

4.the definition of thesauri „rapprochement" and the description, and evaluation in terms of conditions and advantages, of its different possible levels;

5.the creation of a method and of software tools for an automated contents overlap analysis of TEE and ItTT;

6.a further elaboration of a *scenario for a rapprochement of TEE and VTT* including an evaluation of the organisation and effort required, the formulation of hypotheses on the relative importance of the different needs, and the analysis of the consequences of these hypotheses on the respective interests of the different possible levels of integration;

7. *conclusions* for TEE and VTT and beyond this particular case.

---

[1] The title of that study was „Comparative analysis of the European Education Thesaurus and the Multilingual Thesaurus of Vocational Training and feasibility study on their future development and compatibility" (unpublished). It has been carried out jointly by BJL Consult (Brussels), contracted by the European EURYDICE Unit for the European Commission, and Infoterm (Vienna), contracted by the CEDEFOP.

2. COMPARATIVE ANALYSIS OF TEE AND VTT

2.1 Purposes, contexts of use

Both thesauri are used to index bibliographical references. TEE is also used to index other types of documentary material: textual data (descriptions of, and questions and answers about, national educational systems), directories, etc.

Where the thematic coverage of VTT is clearly delimited to vocational training, the field covered by TEE is less strictly defined: it is education at large, with emphasis placed on research in education, educational policy and administration of education.

The thesauri differ in their respective contexts of use: where VTT is mainly used in a single documentary system and network (that of CEDEFOP), TEE is used in a wide variety of contexts, either international (EUDISED and EURYDICE networks) or national.

2.2 General characteristics

Both thesauri have comparable general characteristics:

- same types of *elements* (descriptors, non-descriptors, scope notes, domains), with identical characteristics;
- same types of *relationships* (hierarchy, association, inclusion, semantic equivalence, application, language equivalence), and same use of them (although there are polyhierarchies in TEE);
- partly different types of *presentations:* both thesauri have alphabetical and rotated presentations (with some differences) but there is a graphical presentation in TEE (in the form of terminographs) where VTT includes a systematic presentation.

2.3 Compared profiles

TEE is almost twice the size of VTT (more than 2.700 concepts against little more than 1.500) and includes 9 language versions to the 7 in VTT (no versions in Danish and Dutch).

VTT includes proportionally more non-descriptors, both thesauri being as rich (or as poor) in the number of scope notes.

The network of associative relationships is more dense in VTT; this characteristic is to be considered in the light of the specificities of the hierarchical structuring of each thesaurus:

- the vocabulary of TEE is distributed over a larger number of „domains" (42 „microthesauri" for 17 „subject-oriented sections" in VTT) — which is partly due to its size — each domain including more concepts on average (68 for 89 in VTT[2]);
- TEE includes more numerous hierarchical chains (254 for 36 in VTT) and these therefore contain less concepts (11 on average for 42 in VTT) and include less levels of specificity.

The „styles" of structuring are thus contrasted:

---

[2] However, for VTT this figure hides great differences: the numbers vary between 22 and 196 concepts per section where there are between 36 and 120 concepts per microthesaurus in TEE.

• larger distribution and more balanced repartition of the vocabulary but less associative relationships in TEE (because of the possibilities and constraints associated with the existence of a graphical presentation),

• great differences in the distribution of the vocabulary (a few very large hierarchical chains) but more associative relationships in VTT (as is, it would not be possible to have a graphical presentation),

which increases the difficulty of „reconciling" the structures for the levels of integration requiring it.

2.4 Management

The contexts and methods of management of the two thesauri are also contrasted: TEE is managed jointly by two international organisations and its main users belong to two European information networks whereas CEDEFOP is solely responsible for VTT which is mainly used by the members of its network only.

The management procedures are therefore more complex in the case of TEE: its managers receive requests from more users for more language versions. They have therefore intensified the formalisation of their methods and have set up numerous tools and procedures to control what is a complex task.

The managers of both thesauri have already been cooperating with each other for several years. The thesauri are maintained with different software but, at the time of the study, they both had their latest edition published and distributed by the Office for Official Publications of the EC.

3. NEEDS

3.1 Users „Needs and Managers" needs

Users making searches in documentary systems using different thesauri ("cross-searches") have to re-formulate the same query several times. Inconsistencies (e.g. different descriptors for the same concept) generate disorentation and inefficiency. The workload involved can be lowered if means or help to „translate" queries are made available.

Information systems managers who need a documentary language with a thematic coverage at the intersection of those offered by existing thesauri may use different tools jointly if the „links" they have with each other are strong enough to allow it. If they are already using a thesaurus and wish to exchange or merge data with partners using another one, they have to address the problem of „translating" indexing statements.

Thesauri managers may envisage some level of integration

• to be in a position to continue serving their present users, whose needs evolve, by ensuring possibilities for extending the thematic coverage offered to that of „linked" thesaus;

• to attract new users (such as the information systems managers referred to above);

• to improve the productivity of their management by sharing both means (methodological and informatic tools) and content, which are the result of costly and long term investments.

3.2 Categories of needs in the case of TEE and VTT

In the feasibility study, four categories of needs have been taken into account, they are listed below with the mention of possible indicators of their respective acuteness:

1.Needs of the users making searches in information systems where either TEE or VTT are used i.e. „cross-searches" needs (indicator: frequency of such searches).

2.Needs of the managers of information systems (either existing or envisaged) not presently using TEE or VTT but who might turn to them if their needs can be accommodated i.e. „new users" needs (indicator: difficulties expected by potential users, and met by present ones, concerning the thematic coverages of the thesaus, especially in their fringe areas).

3.Needs of the managers of information systems using TEE or VTT who want to exchange data within information networks (e.g. for cooperative indexing, merging of databases, common access to distributed databases, etc.) i.e. „data exchange" needs (indicator: frequency and advantages of present or envisaged data exchanges).

4.Needs of the managers of TEE and VTT who might consider ways of improving the cost/benefit ratio of their activity (especially important for the European Commission who is supporting, at least partly, both thesauri) i.e. „thesauri management" needs (indicator: cost of not sharing tools and content).

4. THESAURI „RAPPROCHEMENT"

4.1 Defimtion, levels

„Rapprochement" has been proposed (1) as a generic term designating the process involved when „coordinated development", „harmonisation", „linking", „integration" or „merging" between thesauri is undertaken, i.e. a set of actions aiming at removing inconsistencies between the thesauri concerned and organising their development so as to make possible their joint use and their mutually fertilised management.

It was proposed to distinguish five successive levels of rapprochement, the characteristics of which are presented according to an increasing level of ambition and benefit:

1. „Terminological harmonisation" aims at removing the most obvious inconsistencies in the wording of descriptors (such as abbreviation vs full-form, singular vs plural), notably in the „functional" vocabulary (such as geographic terms, languages, institutions).

2. Establishing „cross-links" aims at helping users familiar with a given thesaurus (source) to use another one (target) by providing them with indications, for concepts of the source, on the „semantically closest" concept(s) in the target. „Semantic proximity" relationships are not necessarily given for each source concept, and do not necessarily lead to only one target concept, so that cross-links only give help for re-indexing[3].

3 Note that the term „re-indexing" is used for „translating either indexing statements or queries".

3. „Weak compatibility" is the situation where re-indexing can be automated as „interthesaurus semantic eqvalence" relationships have been established from each concept of each thesaurus to one concept of the other(s) (or, when „multi-equivalence" is necessary, to two concepts to be used together). These relationships may link concepts at different levels of detail (specificity) and expressed by different descriptors.

4. „Strong compatibility" is achieved when, in the common languages, the common concepts are expressed by the same descriptors and identical descriptors designate the same concepts. For those concepts, re-indexing no longer applies; the hierarchical relationships between them may still be different in each thesaurus.

5. „Association" demands reconciling the hierarchical structures where the thematic coverages overlap. It creates a common nucleus on which are hooked various specialised extensions of the thematic coverage offered, corresponding not only to the original specialisations of the associated thesauri but also to the satisfaction of new needs.

As far as tools are concerned, the process of rapprochement requires loading the two thesauri in a single database where they can be manipulated together but only the last level, association, requires the use of a single management and maintenance system.

Regarding the content of the two thesauri, if the perspective of a rapprochement has to be taken into account from the first levels, it only progressively becomes binding: this relative constraint never prevents the development of each thesaurus to meet exactly the specific needs of its users. Likewise, the presentations of each thesaurus can remain different: a joint edition is not mandatory.

4.2 Conditions and advantages

General conditions to be met for close rapprochement of two thesauri are:

- thematic coverages overlapping significantly (see 5. below);
- analogous structures (same types of elements and relationships, common languages);
- managers willing to cooperate and the possibility of organising joint management.

More specifically, the conditions that can justify the effort to reach the different levels — which lead to a table where the advantages of each level for each category of needs is conventionally weighted (see Table 1) — are summarised in the list below:

- condition for terminological harmonisation: the will of the managers (in the perspective of a rapprochement and for their common public image);
- condition for cross-links: existence of users making cross-searches;
- conditions for weak compatibility: users making searches with both thesauri and/or exchanges of indexing statements and/or queries between systems using them;
- conditions for strong compatibility: an amount of re-indexing (i.e. of data exchanges) that justifies it and/or the perspective of a later association;
- conditions for association: numerous exchanges between databases and/or perspective of merging databases and/or need for new, specialised documentary

languages with a thematic coverage at the intersection, or in the fringe areas, of those of the thesauri and/or necessity of cuts in the costs of the maintenance of the thesauri. (This ultimate level of rapprochement gives the most significant benefit to thesauri managers as it implies fixll joint management and thus maximum sharing of tools and content.)

Levels / Needs	terminolog. harmonis.	cross-links	weak compatib.	strong compatib.	association
cross-searches	1	2	3	3	3
new users	0	0	0	1	4
data exchanges	0	1	2	3	3
thesauri managers	1	1	1	3	4

Table 1: Conventional weighting of the advantages[4] of the different levels for the different needs

5. CONTENTS OVERLAP ANALYSIS OF TEE AND VTT

5.1 Method

Estimating the difficulty of reaching the different levels of rapprochement in a given case requires a comparative analysis of the respective contents of the thesauri concerned.

In addition to a terminological, qualitative, analysis not presented here, the feasibility study included the delineation of similarities and overlaps for „measuring the proximity" between TEE and VTT using a computational approach providing indications with a quantitative dimension.

Tailor-made software tools have been used for manipulating the thesauri together in the same database. The tested approach assumes that having identical

[4] Verbal notation for conventional weightings:

Weights / Categories	0	1	2	3	4
advantages	no or insignificant advantage	advantage but not significant or indirect or expected	significant advantage	important, decisive advantage	conjunction of important advantages
costs	no cost	(comparatively) low cost	significant cost	high cost	very high cost
needs	needs it is chosen to ignore	needs it is decided to take into account	needs considered important	needs considered very important, priority	needs it is considered absolutely necessary to meet

or similar descriptors or nondescriptors in common languages gives an indication of the likeliness that the concepts concerned are common concepts.

A „similarity" relationship is established between pairs of concepts and is automatically given a „weight" which is maximum when the chosen descriptors are identical in all the common language versions, decreases as differences are more numerous, and is minimum when there is only a „resemblance" (same (non-empty) words in a different order or slight differences in the wordings) between non-descriptors in only one of the common languages.

"Multiple similarities" are also taken into account, that is to say situations where a concept of one thesaurus is expressed by descriptors which have been chosen to express two or more different concepts in the other thesaurus, depending on the language version.

These computations allow the precise delineation of the areas of the thematic coverages, defined at the domain or the hierarchical chain level, where there are strong overlaps (numerous high weight similarities), problematic overlaps (numerous low weight similarities combined with multiple similarities) or no overlap (rare or no similarities).

The tools developed could be used in the actual rapprochement process to control complex manipulations, investigate possibilities, and ensure the availability of the final product.

5.2 Quantification of overlaps

The thematic overlap has been quantified by computing the size of different „zones" (see Fig.1):

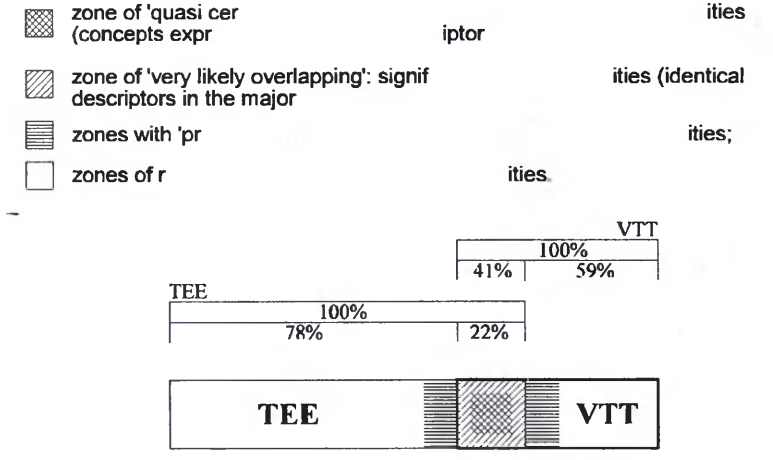


Fig.1: Quantification of overlaps

5.3 Localisation of overlaps

The thematic overlaps have been localised by determining the pairs of domains where the overlap is the most important (see Table 2 and Fig. 2 for an overall graphical display).

microthesauri TEE	number of concepts probably common	**'subject-oriented sections' VTT**
(41) Europe	59	<14> countries amd regions
(13) language sciences	27	<05> information, communication
(36) public authority	27	<11> administration, legislation, politics
(11) sciences and technology	25	<04> research, science, technology
(33) labour environment	25	<06> labour, employment
(04) system of education	23	<01> education
(34) profession and personnel	23	<08> occupations

Table 2: Localisation of overlaps

6. SCENARIO FOR A RAPPROCHEMENT OF TEE AND VTT

To formulate recommendations on the envisaged rapprochement of TEE and VTT it was necessary to evaluate the organisation and effort required, to make hypotheses on the relative importance of the different needs in this particular case, and to examine their impact on the respective cost/benefit balance of the different possible levels of rapprochement.

6.1 Actors, effort estimates

Three groups of actors would contribute to the actual rapprochement process work:

- the thesauri managers;
- the representatives of the „management groups" of each thesaurus (typically: one representative of the users for each language version of each thesaurus);
- a „project team" with terminology and multilingual thesauri specialists, with appropriate experience and tools;

the first two groups, assisted by the third one, would form a"joint management group".

Estimates of effort and duration are given — in terms of cumulated number of meetings of the „joint management group", cumulated effort of the project team (in person/months), and total duration (in months) — for each possible

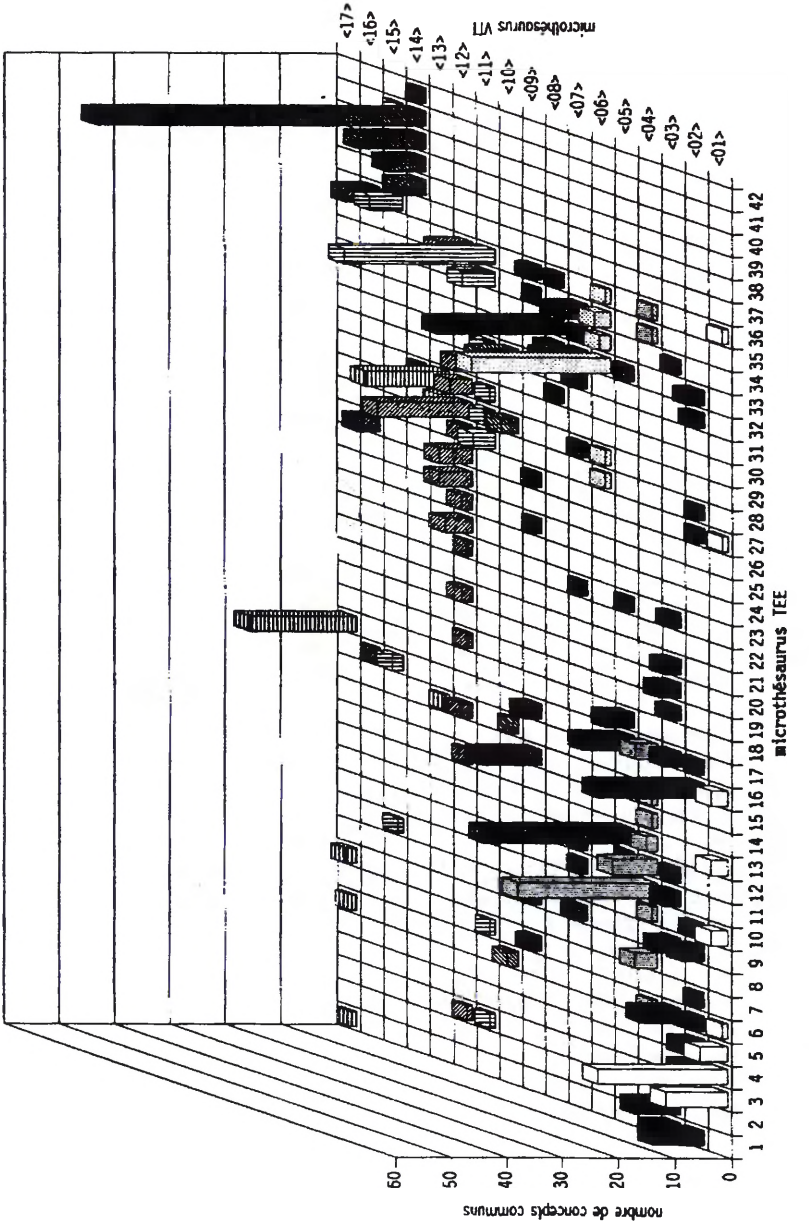


Fig. 2: Localisation of overlaps — graphical presentation

level of rapprochement, leading to a conventional weighting of associated costs (see Table 3).

Scenarios / Items	terminolog. harmonics	cross-links	weak compatib	strong compatig.	association
meetings joint group	1	1	3	6	12 to 15
effort project team	4	8	14	22	35 to 40
total duration	9	12	24	36	60 to 72
costs	1	1	2	3	4

Table 3: Estimates, and conventional weighting of the costs[4], for the different levels

6.2 Hypotheses on needs

In order to determine which levels of rapprochement would offer comparatively higher benefits

- taking into account costs and advantages as compared to the needs (see Table 1) — different hypotheses have been made on the relative acuteness or „urgency" of the four categories of needs considered (i.e. „cross-searches", „new users", „data exchanges", „thesauri managers"):

I. „Reference" hypothesis: all categories of needs are considered of equal importance.

II. Hypothesis: cross-searches and data exchanges are rare; no particular problems in the management of the thesauri; the priority is on gaining new users.

III. Hypothesis: there is an urgency to gain new users; the cost and quality of the thesauri management are put into question; the other two needs are put off temporarily.

IV. Hypothesis: numerous new databases in preparation: it is absolutely necessary to recruit their managers as TEENTT users; priority is also given to data exchanges; it is important to facilitate cross-searches; the needs of the thesauri managers are not a priority.

V. Hypothesis: merging and integration of databases are planned: data exchanges are of an absolute priority and cross-searches a priority; others two needs put off temporarily.

VI. Hypothesis: as the previous one but the management of the thesauri is seriously put into question; cross-searches become relatively less of a priority.

These hypotheses have also been represented using a conventionally weighting to allow integrated processing of the different dimensions considered (see Table 4).

[5] I.e. the cost establishing cross-links (incl. terminological harmonisation) is not significantly higher than the cost of the terminological harmonisation alone.

Needs / Hypotheses	cross-searches	new users	data exchange	thesauri managers
Hypothesis I.	1	1	1	1
Hypothesis II.	0	3	0	1
Hypothesis III.	0	4	0	3
Hypothesis IV.	2	4	3	1
Hypothesis V.	3	0	4	1
Hypothesis VI.	2	0	4	4

Table 4: Conventional weighting of the needs4 for the different hypotheses

6.3 Assessment of the relative interest of each level for each hypothesis

The results of the processing of all dimensions taken into account and conventionally weighted in a comparable and inter-related way (i.e. advantages, costs and needs) lead to a relative weighting of the interest of each level for each hypothesis (see Table 5).

The interpretation that can be made of the results is that, given the elements of the hypothesis, the levels with a positive value (shaded) are more accurate than those with a negative value.

Levels / Hypotheses	terminolog. harmonis	cross-links	weak compatib.	strong compatib.	association
Hypothesis I.	-0,3	0,2	-0,1	0	0,2
Hypothesis II.	-0,3	-0,3	-0,9	-0,2	1,7
Hypothesis III.	-0,2	-0,2	-0,9	-0,1	1,4
Hypothesis IV.	-0,4	0,1	-0,2	0	0,6
Hypothesis V.	-0,4	0,4	0,4	0,2	-0,6
Hypothesis VI.	-0,3	0,3	0	0,3	-0,2

Table 5: Results of the assessment of each level for each hypothesis on needs

7. CONCLUSIONS

7.1 Recommendations formulated for TEE and VTT

The relative importance of the different categories of needs as perceived by the persons responsible for the management of TEE and VTT at the time of the study lead to the following recommendations:

• steps should be taken towards a „coordinated development", i.e. the rapprochement, of TEE and VTT;

• in the short term (up to two years), the managers of these thesauri should aim to

— reach the level of cross-links (including terminological harmonisation),

— reach the most ambitious level of rapprochement (association) in selected domains, as a pilot to collect data for a later management decision on a possible further integration,

— budget for two meetings of the „joint management group" and 14 person/months effort of the project team.

7.2 Beyond this particular case

The feasibility study performed had two types of re-usable outputs:

• the software tools used, which allow:

— comparison of profiles of thesauri (in general),

— the precise delineation of the contents overlap between (multilingual) thesauri,

— the manipulation of thesauri as a support to rapprochement process work;

• the method for evaluating the relative interest of the different possible levels of integration,

which is very flexible as many inter-related parameters can be manipulated:

— the different levels of rapprochement and the different categories of needs,

— the weighting of advantages per level and per need, the weighting of costs per level,

— the weighting of needs according to hypotheses on their relative importance.

BIBLIOGRAPHY

1 Roulin, C., Cooper, C.: Bringing thesauri together: terminological problems and interests. In: Schmitz, K.-D. (Ed): TKE'93: Terminology and Knowledge Engineering. Proc. Third Int. Congress on Terminology and Knowledge Engineering, 25-27 Aug. 1993, Cologne, FRG. FrankfurtIM: INDEKS Verl.1993. p.251-259

SUMMARY

Presentation of the methodology and results of a comparative analysis of the European Education Thesaurus and the Multilingual Thesaurus of Vocational Training and feasibility study on their future development and compatibility. This

study was commissioned by the European EURYDICE Unit and the CEDEFOP to BJL Consult and Infoterm (the executive summary of the final report , not published, is available in French).

1. Comparative analysis of the two thesauri: purposes, contexts of use; types of elements and relationships, presentations; compared profiles (size, number of elements and relationships of the different types, ratios); contexts and methods of management

2. Assessment of the needs: users needs and managers needs ;typology of the needs (cross-searches, new users, data exchanges, thesauri management)

3. Objectives and means of a „rapprochement": definition of levels: 'terminological harmonisation", „cross-links", „weak compatibility", „strong compatibility", „association"; associated conditions and advantages

4. Contents analysis, analysis of overlaps: description of the method (tailor-made database and software tool); quantification of overlaps; localisation of overlaps; terminological analysis

5. Scenarios for a „rapprochement": actors, objectives and steps for each level, effort estimates; assessment of each scenario according to different hypotheses (methology and results)

6.Conclusions; recommendations formulated in the case presented; applicability of the methodology to other cases

LIBRARY CLASSIFICATION: COMPATIBILITY ISSUES

Harald H. Zimmermann

CONCEPTION AND APPLICATION POSSIBILITIES OF CLASSIFICATION CONCORDANCES IN AN OPAC ENVIRONMENT

1. INTRODUCTION

An online public access catalog (OPAC) is defined as an electronic data bank based on machine readable catalog sheets of a library.

Libraries normally used to provide such sheets (or cards) for internal administrational purposes as well as for users. The main function — similar to the „paper" or sheet version of the catalog — still is to identify the lending number of a book to support the lending process in a library.

Compared to data of special information services, the content of an OPAC is very poor (e.g. no abstracts). In addition, the data itself is based on monographs (normally, an OPAC doesn't refer to scientific articles), i.e. it doesn't represent the „world of knowledge" in an area. Instead, it represents the „books" available in a specific library collected by some librarians and institutes which are using this library as service institute.

Besides, the full text retrieval with an OPAC is limited to the words of the titles of a book. In addition, there might exist a kind of catch word or key word category so that the content might be represented by such words.

To some degree, also notations of classification systems (home-made or external ones) are used to describe the content of OPAC entries. But there is, at least in the area of „scientific librarianship" in Germany, no agreement about a standard classification system to be used.

In addition, within (smaller) libraries, domain specialists or experts are not available to be able to provide high-sophisticated classification. As a result, OPACs exist using different classification systems and with a different depth of classification.

2. OPACS AND THE FUTURE ROLE OF LIBRARIES IN CLASSIFICATION

The more a user is equipped with techniques like online access or CD-ROM facilities, the more he is looking for appropriate information systems.

Most of users (especially students) have in mind that „their" OPAC (for example, the OPAC of a university library), would provide an interesting tool for information. To some degree, i.e. for some „first" access to information, this might be true. But in general, such an „idiosyncratic" OPAC is not solving sufficiently the completeness problem of a data base, especially compared to other services, e.g. patent information or domain specific information services.

On the other hand, the „cumulation" of OPACs (i.E. the building of so-called shared catalogs) could reduce the problem of incompleteness, at least on behalf of monographs.

There are many reasons for cooperation between libraries, especially to reduce administration costs. The different classification systems used for several years are an obstruction for such a cooperation.

There are the following ways to solve the problem:

1. One is the „ideal" one, but very unrealistic: All libraries decide on an international standard of classification to be applied worldwide in their OPAC developments.

They agree on a distributed coding system where high-sophisticated specialists are coding „books" in there specific domains on the deepest level possible within the classification.

2. Libraries abandon the creation and coding of own classifications an rely on existing classifications in use at important information services and accepted by the experts of such an domain (like in medicine).

This would lead to different classifications in such a case where an object belongs to more than one of these „basic" domains, but it would allow the libraries to profit from the expert knowledge when building their (selected) OPACs.

In this case, the creation and use of concordances between the different classification systems is needed if a user accesses different OPACs or OPACs are cumulated.

3. Libraries continue in creating classifications of their own, but on a distributed way in the sense that they agree on the most-possible depth (within a specific classification) and on doing it by experts of the relevant domain. This leads to relatively „balanced" coding and high compatibility between the different systems and reduces the cost of coding.

Also in this case, if OPACs of different types are accessed or cumulated, relatively „simple" concordances will allow a content based query by the user (accessing the system by one of these classifications).

4. Libraries continue in creating classifications of their own on a very shallow level (in general) or in an unbalanced way, i.e. dependent on the experts or even specialisations existing (or not) within the relevant library (as today).

If OPACs of such types are accessed or even cumulated, highly sophisticated concordances have to be developed to reduce the content based query problems of the user.

3. THE USE OF A „CLASSIFICATION" THESAURUS AS A CONCORDANCE TOOL

In the following, a concordance (or correspondence) between different classification systems is considered as a specialised thesaurus in the sense that the classification elements (= notations) are handled as „artificial words" or entries of this thesaurus and relations are used to indicate the (type of) correspondence between the notations of different classification systems.

To explain this concept in general, the structure of a classification system is taken:

Let us assume that A is a class element (notation) of the classification C; A.A and A.B are also elements of this classification, also A.A.A, A.A.B, A.B.A and A.B.B are elements of this classification. By using the relations BT (broader term) resp. NT (narrower term), one can relate these elements in a hierarchical form by building the pairs

(I)	A		(BT)	A.A
	A	(BT)	A.B	
	A.A		(BT)	A.A.A
	A.A		(BT)	A.A.B
	A.B		(BT)	A.B.A
	A.B		(BT)	A.B.B

Let us assume that N is a class element of the classification K; N.A and N.B are also elements of this classification, also N.A.A, N.A.B, N.B.A and N.B.B are elements of this classification. By using the relations BT (broader term), one can relate these elements in a hierarchical form by building the pairs

(II)	N	(BT)	N.A	
	N	(BT)	N.B	
	N.A		(BT)	N.A.A
	N.A		(BT)	N.A.B
	N.B		(BT)	N.B.A
	N.B		(BT)	N.B.B

When creating a concordance between both classification systems C and K, the simplest case would be if only the names of the notations differ, but the content referred is the same. In this case, the synonym relation SY can be used to generate the concordance, where the reference to the relevant classification is indicated by an attribute:

(III)	A (C)	(SY)	N (K)	
	A.A (C)	(SY)	N.A (K)	
	etc.			

If the classification systems or parts of it differ in the sense that one notation has more than one representation within the other classification, but the

„cumulative content" is the same, the relation „partly synonym" (PSY) can be used to describe the relation:

(IV) A.B (C) (PSY) N.A (K)
A.B (C) (PSY) N.B (K)

If the classification systems or parts of it differ in the sense that one class name has one or more than one representation within the other classification and the „content" is overlapping with other classification elements of the related classification, the relation „quasi synonym" (QSY) could be used:

(V) A.B.A (C) (QSY) N.B.A (K)
A.B.A (C) (QSY) N.B.B (K)

What happens, if one classification (within the same domain) is much more differentiated on behalf of the depth (within a subclass)?

To demonstrate the possible solution of such a case, let us assume that classification C consists only of the category A, whereas classification K consists of at least the elements N, N.A and N.B.
Because there exists the relation BT:

(VI) N (K) (BT) N.A (K) and
N (K) (BT) N.B (K),

it will be sufficient to relate A (C) by the relation (SY) to N (K):

(VII) A (C) (SY) N (K),

if there exists a general rule that all narrower terms (notations) of a system where no explicit concordance relation exists are handled as being included in the relation of the notation which is in broader term relation to them.

If such a software doesn't exist, the rule can be used to build explicit PSY relations between all referenced elements of such a classifications:

(VIII) A (C) (PSY) N (K)
A (C) (PSY) N.A (K)
A (C) (PSY) N.B (K)

The last general rule to be handled is the case where there doesn't exist a relevant content classification within one of these classifications. This normally indicates that there will be no reference (document) with such a content available in the OPAC. For quality assurance reasons, a „NULL" relation could be used.

5. APPLICATION POSSIBILITIES

5.1 Application within a user query

To be able to use such a „classificational thesaurus" for data base access (by the user), the relevant OPAC must indicate the type (= attribute) of „his" classification system to the retrieval (and thesaurus) system.
If the user is asking a question by using „his" classification system, the retrieval system is applying the concordance rule to transfer this part of the query into the query relevant for the system. This transfer could be indicated to the user.

5.2 Application in creating cumulative OPACs

If cumulative OPACs are developed based on data from different libraries with different classification systems (e.g. in cooperation between libraries), there could be a decision made on applying one (the more explicit) classification out of the different „input" classifications, so that during the retrieval, the notation of this system would be the only resulting notation.

The alternative could be to cumulate all the different notations with the source (or type) indicator. This leads to more complicate queries (because all possible alternatives have to be created with „OR"), but the data are available in their „original" form.

5.3 Application within distributed OPACs

If queries are used to access different (distributed) OPACs by one query, the replacement rule mentioned above (5.1) could be used — dependent on the indicator of the OPAC actually accessed — to provide the „right" query.

5.4 Enrichment procedures

The concordance technique described above can also be used to „enrich" an OPAC data base with „deeper" notations.

Let us assume that „classification system C" is „shallow" in general (i.e. with very „broad" content notations, because there was no people or money available to describe the data on a very specialised level), and that there exists a classification system K which is more or less compatible to C in the sense that the notations of „C" used correspond to a high degree with the „broader notations" which are part of the more complex system K.

By using the concordance (and by replacing the classification system), the „weaker" notation of the OPAC could be replaced by the subtle notation, if the same document is detected.

6. INTEGRATION WITHIN „CLASSICAL" THESAURUS APPLICATIONS

The thesaurus based classification concept has an additional advantage: It can be used together with „classical" terms to provide the user with access tools from his „wording" (in natural language) — „via" a (natural language based) descriptor to the relevant (or possible) notation of the classification(s). High-sophisticated systems would even be able to provide this access in the user's natural language.

Therefore, guided by the „word-based" thesaurus, he or she will be able to select the relevant notation (by browsing techniques) and prepare the „right" search by using this notation.

7. FINAL REMARKS

The concept looks very simple, but the problems are in the detail. Fore instance, it will not make any sense to build concordances between a very week system and a highly sophisticated system, if the „enrichment rule" is not applied (or not possible).

The relations described have to be seen as examples; for practical purposes, this list might be extended.

In my opinion, it makes no sense to struggle for „the only general and universal system" of classification, because the „concordance classification thesaurus" is able to overcome, to a high degree, the problem on the user's side. On the other hand, the rule of „garbage in, garbage out" also is valid in this subject.

Barbara Sosińska-Kalata

THE UNIVERSAL DECIMAL CLASSIFICATION AS AN INTERNATIONAL STANDARD FOR KNOWLEDGE ORGANIZATION IN BIBLIOGRAPHIC DATABASES AND LIBRARY CATALOGUES

1. STANDARDIZATION AND COMPATIBILITY IN INFORMATION HANDLING

The importance of standardization in information handling is critical to the development of modern information services. It is obvious that standardization on an international scale is the basic requirement to achieve international compatibility of information systems and services. Generally compatibility issues affecting information systems are related to resource sharing which is the basic method of modern information activity at international as well as national and local level. In particular, this means some form of common access to information sources generated and maintained world-wide by numerous information centres, systems and services. Therefore, the rigourous application by each member of an information system of the same or compatible forms and tools of data preparation and presentation is one of the most important aspect of this method of co-operation. Computer and telecommunication technologies greatly facilitate resource sharing among information systems' members. In an automated environment the possibilities of co-operation are much greater than in a manual setting, but at the same time standardization and compatibility become increasingly important. Absence of compatibility always results in a loss of efficiency.

Rapidly increasing physical possibilities of wide access to various electronic information sources by modern computer networks do not mean real accessibility of those sources for a still extending population of their users. Development of networks is accompanied by an increase of language, cultural, educational, social and other kinds of diversity of the network's users. Therefore, it can be stated that paradoxically developments of computer and telecommunication technologies strengthen the information barriers which are the most difficult to overcome as they affect intellectual needs and abilities of information retrieval.

Considerable advances of standardization in the field of information handling have already been obtained in many areas. Unfortunately, there is still the lack of satisfactory solutions in the area of information preparation and organization concentrated on processing of informational content of documents (1). If an information system is to achieve its aim of providing users with a more extensive

range of services, it must have compatibility in the intellectual aspects of preparing the informational collections of the members of the system. Generally that means necessity of standardization of document description, subject analysis, classification, subject indexing and related to them methods of information retrieval. Except some advances in the area of standardization of cataloging rules and codes, international standards in the real sense of the word in those fields exists only sporadically.

2. THE USER'S APPROACH TO STANDARDIZATION OF KNOWLEDGE ORGANIZATION

Certainly it is a truism that any information service exists to satisfy the needs of the user. Hospitality to the user's approach is commonly accepted as one of the basic features of any modern automated system. It is vital that any modern computerized information system must be user-friendly. This includes many problems such as for example unified and easy to use command languages, extended help information or standardized keyboards. However, it should be noted that in this context rarely placed problems of organization of knowledge recorded in databases of various information systems also belong to aspects of the systems standardization which should be affected by the user's approach. Obviously users of information are hindered by various enquiry languages and keyboards varying from one system to another as well as by variety of languages and order systems used to present the information in various databases. Without the standarization of tools of knowledge organization, valuable strategies as those of multiple database searching lose their potential power and become supplementary methods used only for rare or untypical topics. As controlled vocabulary usually differs from one database to another, strategies of this kind are carried out rather with the use of uncontrolled natural language utterances than with retrieval languages basicaly allowing more precise and complete searching (2). The importance of standardization and compatibility of retrieval languages employed in various databases, systems and countries had already been well recognized over twenty years ago[1] but despite of many researches in the field there still is no generally accepted method or pattern.

In response to the user's enquiry, an ideal system of information retrieval -as it has been defined by C.N. Mooers — should retrieve all the information which the user would have selected himself as relevant to his search problem if he knew the content of the entire information collection (4). In the context of the user's approach to the tools of knowledge organization this classic statement could be interpreted as a recommendation of ensuring the accordance between specificity of

[1] Problems of compatibility were made one of the most emphasized issues in the UNISIST Report of the Unesco of 1971 (2). Since the beginning of the '70s many studies had been carried to establish methodological requirements of compatibility of retrieval languages and to design switching tools between languages most widely used at the international level. Considerable part of those efforts concerned the UDC and its ability to play the role of a pattern of semantic organization of natural language based vocabulary in thesauri (3).

identification of information needs and the structures of knowledge organization generally accepted by the users of a particular database and specificity of indexing or classification and structures of knowledge organization used in order systems implemented in that database. Hence, the content of the document should be described in the way the user would describe desirable information. This is the basis of a so called user-oriented policy of abstracting, indexing or classification considered as more effective than traditional subject-oriented policy (5). However, information users are better served if they are provided with access to various topically related databases produced and maintained world-wide, so consequently the potential user group of a particular database may be significantly differentiated. Networking and resource sharing demand a more flexible approach to representation of informational content of documents as such a document content description should suit information needs of various user categories in various countries and often in various specialist subject areas. Therefore, it is vital that the representation of document content ensures at least multiple aspect access and hierarchical information retrieval.

Equaling the subject approach used within one system, one country, or one specialist subject area, with that of another, faces many well known difficulties. To name the most important of them one may recall problems related to establishing the necessary and sufficient level or levels of specificity of information representation, dependency of relationships of some subjects and external contexts of knowledge perception in various environments, differences and changes in terminology used as the basis of retrieval vocabulary. The needs of the general user differ from the needs of the specialist as the first usually is satisfied by a broad subject analysis, while the other requires a highly specific analysis. The structures of knowledge organization generally accepted and understood by a group of users are affected by social, educational, political, and cultural factors which differ from one group to another. Terminology used in the same subject area may differ according to conventions accepted by various schools. Terminology used in various — even related — subject areas is usually polysemic and full of synonymous or quasisynonymous expressions. Also in some disciplines terminology is subject to change and development. Structures of terminological systems developed in various ethnic languages may also vary. This makes it difficult to establish real compatibility between various natural language based retrieval languages and decreases efficiency of natural language searching supported by multi-lingual terminological glossaries. That problem could be solved by the use of artificial codes of notions but on the other hand, the possibility of the user-system communication in the user's ethnic language (or at least in a language based upon its vocabulary) is vital for any modern information system in every dimension of that communication.

3. MODERN APPROACHES TO TOOLS OF KNOWLEDGE ORGANIZATION

Increasing availability of internationally produced and maintained information resources indicates two basic approaches to the tools of subject access to information. User-friendliness is the main factor affecting both of them. The first

approach is to ensure user-friendly presentation of information mainly by the use of natural language utterances in information retrieval. The second approach is to ensure a user-friendly semantic organization of search vocabulary which enables identification of vocabulary related to a specific search problem, and supports automatic retrieval at the level of specificity determined by the user and indexing policy employed in a particular database. Although these two approaches closely correspond to the traditional two methods of subject bibliographic data organization, the new quality may be found in the tendency to treat them as two necessary functions that should be supported parallel by modern instruments of knowledge organization.

The first approach is demonstrated in development of natural language based retrieval languages and in two ways that the development of international information systems have been attempted in that context. One requires the ability to handle the records of publications produced in many languages and to search the records in a variety of languages. An example of such a multi-lingual information system is that of the European Communities, where multi-lingualism is mandated by law. This stimulates the publication of multi-lingual glossaries and thesauri. This also resulted in attempts to improve language teaching throughout Europe, research in terminology and computer translation procedures, and development of terminological data-banks. Alternatively, as a policy decision for economic reasons, in some information systems one language is selected as the carrier for the system.[2] In fact, in such a system again retrieval must be supported by multi-lingual terminological glossaries that help the international user to select appropriate terms in the language of the system. In this approach, the development of multi-lingual thesauri is the basic form of standardization of the tools of knowledge organization.

The second approach to the design of modern instruments of knowledge organization in information systems emphasizes the user-friendly presentation of concept organization and easiness of identification of search terms to represent desirable concepts or search topics. This involves various issues related to development of effective methods of the indication and display of semantic relationships between concepts and topics represented by utterances of retrieval languages.

4. DEVELOPMENT OF THE UDC IN THE CONTEXT OF THE MODERN APPROACHES TO KNOWLEDGE ORGANIZATION

One of the fundamental principles of standardization in information handling is to take into account existing solutions and practices. This principle is of special importance to the area of subject analysis and description. There are at least two reasons why consideration should be given to the methods and systems of

2. The multi-lingual approach is adopted also in many European databases accessible in various information services, for example in French DELPHES EUROPEAN BUSINESS. An example of one language solution adopted by an international system is AGRIS — the International Information System for Agricultural Science and Technology which uses English as its language.

knowledge organization which are already in use. The first reason is related to the intellectual as well as time and financial costs of data preparation or necessary conversion to a new order system. The second reason deals with the habits of users which are formed by systems in use.[3]

The Universal Decimal Classification (UDC) belongs to the small group of three general bibliographic classifications that have become a kind of international standard on the basis of their world-wide use in library and bibliographic practice. It is used in 21 national bibliographies and in many non-English speaking countries is introduced as the official system of organization of classified catalogues in some kind of library networks (usually public and technical).[4] Unlike the other two classifications[5] the UDC has got unique advantages of great value to the purposes of standardization at an international level. Firstly, the UDC is a general scheme but in the majority of subject sections it is developed to uniquely deep details. The great specificity of the UDC allows to preserve coextensive indexing with the subject of the document and can be tailored to suit individual circumstances. This makes the classification suitable for general information systems as well as for specialist services.[6] Secondly, the UDC -as many other retrieval languages that use an artificial notation — is absolutely language independent but its schedules and alphabetical indexes exist in over 20 different language editions. There is no other order system that can serve so many various language user groups. There is no other system that can be used as a specific switching language between so many natural language based retrieval languages or as a multi-lingual access order system covering over 20 languages. Thirdly, the UDC numbers use clear and easy notation of the high level of expressiveness. This means that the notation not only may support automatic hierarchical retrieval but also allows to display other semantic relationships between classes (topics and concepts) represented by the numbers. However, it should be noted that the UDC notation does not always express hierarchy. Fourthly, the UDC is designed as a multiple access order system. There are no limits to represent any kind of the document content components that may be of interest to information users.

---

[3] It is worth to note that although modern bibliographical and library classification are designed as topical classification of documents by some scientists they are treated as a kind of pragmatical classification of sciences (6, p.268-270). For many users, those order systems are the basic filter of the perception of relationships between disciplines and notions.

[4] In Poland, UDC is the obligatory system used in classified catalogues in public and school libraries as well as in some special libraries mainly covering the area of engineering technology and sciences. According to the findings of the study on retrieval languages used in Poland 31,4% of the analyzed institutions use classification systems as its main retrieval languages and in 67,7% of them the UDC is the chosen classification. The rest of the studied libraries and information centres use: 29,8% — thesauri, 19,6% — subject headings, 18,1% and 1,1% — other kinds of languages ( ).

[5] Dewey Decimal Classification and Library of Congress Classification are the two other mentioned systems which informally are also accepted as a kind of international standards.

[6] In its specific version the UDC is used for example in such international specialist bibliographical services as Key to Economic Sciences (formerly Economic Abstracts), Bibliographia cartographica or Photographic abstracts. In the function of general bibliographical classification it is employed in 21 national bibliographies and some other general reference sources like for example Guide to reference material by A.J. Walford (8).

All those features of the UDC have been well known for long and to some extend they are successfully used in international exchange of information. In the early 70s, when the Unesco introduced the UNISIST Programme great attention had been paid to the ability of the UDC to play the role of a switching language. Although results of the studies undertaken in that time were not satisfactory and another system had been developed to fulfil the function,[7] the recommendations of the UDC development determined upon the basis of those studies and the model of a modern international retrieval language implicated by them were very close to those discussed and accomplished nowadays.[8]

The period since the end of the 70s to the end of the 80s was especially difficult for the UDC. Those difficulties are widely known and there is no place to comment them. During that time the revision of the scheme had been accomplished in very limited scope and the UDC had to wait until the beginning of the 90s for more radical decisions and solutions. In January 1992 FID transferred all the property rights and the responsibility for the UDC to the UDC Consortium (UDCC) — a non-profit making foundation under Dutch law. The founder members of the UDCC are FID and the five major publishers of the UDC are from Belgium, Japan, Netherlands, Spain and the United Kingdom.[9] All of those organizations are directly interested in the continuation of the UDC, so they are supposed to promote and control its development in the best way (10). Within four years of its existence the UDC Consortium determined the basic assumptions of the system's revision and development policy and dynamically proceeded to their realization. Basically the realized revision and studies related to them revert to the general development recommendations established by FID twenty years ago. Two dimensions of them directly correspond to two basic recommendations for development of modern user-friendly tools of knowledge organization. The first dimension deals with attempts to derive a thesaurus from the structure of the UDC and its alphabetical index. The second dimension is expressed in efforts to convert the UDC into a fully faceted system which enables explicit identification of simple and complex notions, and consequently enables

---

[7] Eventually, the planned switching language for the UNISIST took the form of a new roof classification — Broad System of Ordering.

[8] The main drifts of the discussion on the UDC revision in that time were the conversion of the scheme to a faceted structure and the conversion of the subject index into the form of alphabetical thesaurus. In March 1974, the conversion of the UDC into a faceted classification was accepted by FID/CCC as one of the most important directions of the UDC future revision. In November 1972 studies on possibilities of the conversion of the index were recommended officially by FID. In 1974, L. D'Haeneus and G. Lorphévre published the report of the research on conversion of the index to the section 330 Economics. Economic science concluding that the conversion is both possible and manageable although general revision of the language descriptions of class numbers is necessary (8). It should be noted that in the early seventies there were already several successful attempts to prepare so called UDC-thesauri for relatively narrow subject areas, for example DK-Spezialauszung Energie published in GDR in 1971 (9).

[9] Publishers of the UDC materials which are founder members of the UDCC: Centre de Lecture Publique de la Communauté (Belgium), Information Science and Technology Association INFOSTA-NIPDOK (Japan), Bohn Stafleu Van Loghum (Netherlands), Associacion Espanola de Normalizacion y Certificacion (Spain), and British Standards Institution (United Kingdom).

automatic recognition of related ideas and subjects represented by class numbers. Restructuring the UDC to some extend will use the structure of the Bliss Bibliographic Classification (BC2) as a framework. Conversion of the entire system into the form of a faceted classification will be followed by the development of a thesaurus which will replace the UDC subject index (11). In practical applications the thesaurus is planned as a basic tool of information retrieval used by the searcher. Descriptors selected by the user will be automatically converted into equivalent UDC class numbers offering advantages of automatic hierarchical retrieval.

It is obvious that the revision of the UDC is tightly connected to its application in computerized information systems. The main aim is to exploit all the advantages offered by the classification and to add to them those benefits which natural language based retrieval languages traditionally support. A good example of this type of the application of the UDC in a computerized retrieval system supporting logically structured searching performed by the use of controlled vocabulary based upon multi-lingual access is that of the ETHICS system (ETH[10] Library Information Control System) at the Library of the Swiss Federal Institute of Technology in Zürich (12).

General revision of the entire structure of both the UDC schedules and their natural language access tools which usually are still yet in the form of a subject index are necessary for successful application of the UDC in multi-lingual international information systems. Researches in both those dimensions are already in progress and seem to be highly promising. The rate of those studies is impressive but for the full success a wide international co-operation is needed. Natural language retrieval supported by the logical structure of the UDC requires development of the UDC-thesauri covering all the subject areas of the universe of knowledge and based upon various ethnic languages. According to the UDCC assumptions the UDC-thesauri should be developed for each subject area separately but all of them have to follow the same pattern of systematic organization reflecting the UDC revised structure. If such thesauri are to function in various natural languages in a relatively short time they may be developed only by designated organizations in countries interested in the wide application of the UDC — that means at least in each of the country that uses the UDC in its national bibliography or in which the national edition of the UDC schedules has been published. Otherwise the entire enterprise will result in the development of a multi-lingual macrothesaurus covering the universe of knowledge and only in several major languages such as English, German, French, Spanish or Japanese.

REFERENCES

1. Compatibility in information systems. In: Teaching package on standardization in information handling. Module 5. Paris, Unesco 1991. p.(5)1-(5)34. (PGI-91/WS/2)

[10] Eidgenössischen Technischen Hochschule in Zürich.

2. Harter, S.P.: Online information retrieval. Concepts, principles, and techniques. San Diego, Ca.: Academic Press, Inc. 1986. 259p.
3. UNISIST. Etude sur la réalisation d'un systéme mondial d'information scientifique, effectuée par l'organisation des Nations Unies pour l'education, la science et la culture et le Conseil international des Unions scientifiques. Paris, Unesco 1971. 181p.
4. Mooers, C.N.: The intensive sample test for the objective evaluation of the performance of information retrieval systems. RADC-TN-59-160 (TZB-132). Cambridge, Mass.: Zator Co. 1959 p.5
5. Lancaster, F.W.: Indexing and abstracting in theory and practice. Champaign, Ill.: University of Illinois 1991.
6. Kamiński, S.: Nauka i metoda. Pojęcie nauki i klasyfikacji nauk. (Science and method. The concept of science and the classification of sciences). 4th ed. Lublin: Towarzystwo Naukowe Katolickiego Uniwersytetu Lubelskiego 1992. 384p.
7. Tomasik-Beck, J. Języki informacyjne opracowane, adaptowane i wykorzystywane w Polsce (stan na 31.12.1994 r.). (Retrieval languages developed, adopted and used in Poland. State on 31 December 1994). Warszawa: IINTE 1994. 104p.
8. D'Haeneus L., and Lorphévre G.: Rapport sur la rédaction d'index "thesaurifiés" de la classification Décimale Univerelle. Bruxelles: FID 1974. 5p.
9. Ścibor E. Metody wiązania różnych języków informacyjnych ze szczególnym uwzględnieniem UKD i języków deskryptorowych. (Methods of linking different indexing languages with particular attention to UDC and descriptor languages). Warszawa: IINTE 1974
10. McIlwaine, I.C. and Buxton A.B.: Guide to the use of UDC. An introductory guide to the use and application of the Universal Decimal Classification. The Hague: FID 1993. 124p.
11. McIlwaine, I.C. and Williamson, N.J.: Future revision of UDC. Progress report on a feasibility study for restructuring. E&C 15(1993), p.11-17
12. Buxton, A.B.: Online applications. In: McIlwaine, I.C. and Buxton, A.: Guide to the use of UDC. An introductory guide to the use and application of the Universal Decimal Classification. The Hague: FID 1993. p.109-119

SUMMARY

Dynamic development of information technologies, especially information networks and multi-base information-retrieval services, causes rapid enhancement of technological possibilities of access to information sources generated all over the world in various ethnic languages and presented in various indexing and classification languages. This availabiblity is limited in practice by many non-

technological barriers. Some of them are the consequence of lack of internationally accepted standards of knowledge organization. Size of population and linguistic diversity of potential users of databases accessible by wide area computer network or in multi-database information-retrieval services is still growing. The main method of elimination of the language barriers is developnment of multi-lingual thesauri and subject headings lists. That expensive method faces many methodological problems implicated by incomparability of paradigmatic structures of various indexing languages and diversity of terminological systems developed in various natural languages. Increasing availability of international information resources indicates two basic directions of shaping the tools of subject access to information. The first direction is to ensure user-friendly presentation of information mainly by use of natural language utterances in information retrieval. The second direction is to ensure a user-friendly semantic organization of information resources which enables identification of vocabulary related to a specific search problem, and supports automatic retrieval at the level of specificity determinated by the user and indexing policy employed in a particular database.

The Universal Decimal Classification as one of the most widely used general as well as higly specific classification existing in over twenty different language editions has got an unique potential to become an international standard for knowledge organization in bibliographical and library databases. However, this feature of the UDC was well recognized and analyzed many times in the past without satisfactory results. Nowadays, in the context of last revisions of the scheme and new policy of its development carried out by the UDC Consortium, the question about the use of the UDC as the international standard again becomes actual. Two dimensions of those revisions and development policy of the UDC directly correspond with two basic recommendations for development of modern user - friendly tools of knowledge organization. The first dimension deals with attempts to derive a thesaurus from the structure of the UDC and its alphabetical index. The second dimension demonstrates in efforts to convert the UDC into a fully faceted system which enables explicit identification of simple and complex notions, and consequently enables automatic recognition of related ideas and subjects represented by class numbers. Researches in both those dimentions are already in progress and seems to be highly promissing. In the same time many of the institutions that have used the UDC for long time and now are adopting automatic systems are discussing necessity to replace it by subject headings or descriptor systems. The urgent problem becomes how to help the UDC to survive until its new version will be completed.

Jadwiga Woźniak
Teresa Głowacka

KABA SUBJECT AUTHORITY FILE An Example of an Integrated Polish-French-English Subject Headings System

INTRODUCTION

In 1990 preparatory work began in the Warsaw University Library (WUL) aiming at its automatization which included the building of an OPAC for the Warsaw University net of libraries. One of the more serious tasks faced was the necessity to design an indexing language for the OPAC .

At first utilization of a subject headings system was taken into consideration, which had been applied in the traditional WUL catalogue for over 60 years. However, studies on the vocabulary of the system which were carried out revealed its numerous deficiencies and weaknesses from the point of view of our purpose which was a *universal language with a several-level hierarchy of the terms specificity adjusted for subject cataloguing in an automated system which would have an authority file module*. It was decided then, that we should design a new indexing language with a view, among other things, to the highest possible compatibility of the designed language with one of the known subject headings systems which is widely spread geographically and linguistically.

In the first instance, the possibility of using the methodology and vocabulary of the Library of Congress Subject Headings (LCSH) was considered. However, after having got further acquainted with the French national level authority file RAMEAU[1] created in the Biblioth(que Nationale de France (BNF) and with its manual „Guide d'indexation RAMEAU" and after direct contacts with the authors of RAMEAU and librarians using that language in the current subject cataloguing of books in the BN-OPALE system (December 1990) we decided to create a subject heading system which would be compatible with RAMEAU[2] as much as possible.

---

[1] RAMEAU - Répertoire d'Autorité-Matiere Encyclopédique et Alphabétique Unifié.

[2] Comparable studies on LCSH and languages compatible with it: Répertoire de vedettes-matiere de la Bibliotheque de l'Université Laval á Québec (RVM) and RAMEAU showed that RAMEAU is a French language version of LCSH in which many deficiencies existing in the vocabulary of the obsolete LCSH were successfully eliminated.

Almost at the same time we were working on the translation of „Guide d'indexation RAMEAU" and getting acquainted with the available publications and elaboration of the first lists of free-floating subdivisions.

The simultaneous purchase of the VTLS system by the Warsaw University Library, the Jagiellonian University Library (JUL) in Cracow, the University Library in Gdańsk (ULG) and the Library of the Academy of Mining and Metallurgy (LAMM) in Cracow is the reason why, since 1992, all the works have been carried out jointly by several libraries. At the beginning of 1994 the Wrocław University Library (WrUL) adhered to the co-operation, too. Thanks to that co-operation the pace of the works, their quality and efficiency considerably increased.

In January 1993 an intensive team work was initiated over the methodology manual of the subject headings system that was being prepared and which we called KABA language (from: Katalogi Automatyczne Bibliotek Akademickich — *Automatic Catalogues of Academic Libraries*). In March'93 the building of the KABA subject authority file was commenced. The first to enter the file were the terms which had already appeared before and were elaborated in the lists of the free-floating subdivisions. All the works were ahead of the practical cataloguing of the items purchased by the libraries. At the end of 1993 nine lists of free-floating subdivisions were ready after being prepared on the basis of „Guide d'indexation RAMEAU '92" as well as the methodology manual[3].

Pragmatic reasons decided that we also drew from the experience of the Biblioth(que Nationale de France in the field of work organization. The RAMEAU file is created in a decentralized way, in co-operation with the university libraries of France but it is the Service AMA (Service d'Authorité Matiere) of the BNF which is responsible for the intellectual value of RAMEAU and which gives the headings an authority status. The KABA language is being created in a decentralized way, too, and the function of the central body is fulfilled by functioning at the WUL the Inter-Academy Team for the KABA Authority Headings File which consists of representatives of all the co-operating libraries[4].

All the year 1994 and the first half of 1995 were periods of an intensive work on the KABA file. Out of the teams which were doing the subject cataloguing in the co-operating libraries a Validation Commission was established, which at its regularly convened sessions collectively validated the headings proposed by the librarians. In 1994, 12 sessions were held and 16 in the first half of 1995 .

As our work progresses the Polish-French co-operation becomes closer and closer. One of the elements of the co-operation are practices which our employees go through in the Bibliotheque Nationale de France and working visits of the representatives of the BNF in Poland. We do not deny that we are keen on widening our co-operation and the international exchange of experience and information with other countries and languages.

---

[3] Authority File of the KABA Language, Part 1. Proper Names. Complete Work edited by J.Woźniak. Warszawa 1994.

[4] At present nine libraries co-create the KABA File: WUL, JUL, LAMM, GUL. WrUL and since 1995 four libraries from Lublin: the Lublin Catholic University Library, the Maria Curie-Skłodowska University Library, the Technical University Library and the Agricultural Academy Library.

THE KABA LANGUAGE AND ITS AUTHORITY FILE

KABA has a creative background in the vocabulary of three subject headings systems: LCSH, RVM[5] and RAMEAU. The priority source, however, is RAMEAU. The KABA language is created using the deductive-inductive method. The first lists of free-floating subdivisions were worked out in a deductive way. (The next ones are in a process of elaboration by using appropriate lists of RAMEAU subdivisions). The lexical stock of KABA itself is being created in a deductive-inductive way, as part of the introduced terms comes from the list of subdivisions and the other results out of the current subject cataloguing.

It should be emphasized that the work on the KABA language vocabulary may not be taken for a linguistic translation of the RAMEAU vocabulary; we create the KABA vocabulary by way of naturalization and adaptation of authority headings existing in RAMEAU, with preservation of the RAMEAU methodology but taking into consideration the requirements of the Polish language.

The items of headings for the authority file of KABA are worked out from the angle of their possible and acceptable conformity with their counterparts in RAMEAU. Practically, it means that: the lexical units of KABA language usually enter similar relations as those of their counterparts in RAMEAU, most often have identical indexing principles and functions in the language (heading, subdivision, heading and/or subdivision). Lexical units of the KABA language have been additionally connected by equivalence relations with respective French-language headings of RAMEAU and English-language headings of LCSH, usually fixed on the basis of two-sided English and French indexes of the RVM authority file. The VTLS system makes it possible to access the proper Polish-language headings in the KABA file via the special index „Z SEARCH" which shows all of their equivalents in the vocabulary of RAMEAU and LCSH.

Hence, one can say that the KABA language, by using the above mentioned solutions, creates a possibility of a computer-aided translation of subject headings created for documents in RAMEAU and LCSH languages. Thanks to that it is easier, cheaper and more efficient to use data from bibliographic bases in which the access points are headings formulated in the above mentioned languages.
Of course, now we are still in the initial period of our work. The vocabulary of the authority file of KABA amounts to about 13 000 terms, including 3 000 unified headings, 6 000 French-language and English language headings and about 4 000 Polish-language rejected terms. It should be emphasized that the KABA file is a central one for all co-operating libraries. It is stored on a mainframe in the Warsaw University Library. And it is available via INTERNET. The WUL server address:

limba.buw.uw.edu.pl 148.81.207.1

hello ID, user.clas02

ID=user's ID, e.g. hello jw,user.clas02

Among the unified headings non-complex headings prevail. We introduce the complex headings in accordance with the principles defined in RAMEAU, namely when they contain non-floating subdivisions (also chronological non-floating subdivisions), when they are quoted in other records as examples, or when it is

5 RVM — Répertoire de vedettes-matiere de Bibliotheque de l'Université Laval á Québec.

necessary for the fulfilment of the semantic relations in the language. The seemingly simple vocabulary of KABA is richer in practice, because many terms can be applied in subject headings in a double function — that of headings and subdivisions.

Recently, „The Lists of Free-Floating Subdivisions of the KABA Language" was published[6]. The lists have preserved the connection of the designed vocabulary with that of RAMEAU and the Polish-French and French-Polish subdivisions have been attached. Each of the lists has been provided with a respective code and number which are in accordance with the markings of the lists in RAMEAU and LCSH. Thanks to that, their tripartite comparison is possible.

The popularity of the KABA language is growing at an unexpected pace in Poland. Recently, further 15 libraries have declared their willingness to use KABA language in their online catalogues as well as the authority files of personal and corporate names, series titles and unified titles run by us. (In the end of July'95, personal and corporate names amount to about 36 000 headings, 3100 series titles and 500 unified titles). We are glad that so many Polish libraries are interested in the work carried out by us. That interest, in turn, makes us more responsible in front of the circles in question and creates more and more duties. We hope, however, to manage it and while taking care for the quality of the designed language we shall widen the co-operation to include other libraries in Poland and abroad.

SUMMARY

KABA authority file is planned for subject cataloguing and searching in OPAC of several academic libraries in Poland. Methods of KABA construction and utilization are compatibile with those worked out at the Bibliotheque Nationale de France. almost each concept in KABA authority file has a connection with its counterpart in French and English. French counterparts are derived from RAMEAU (Repertoire d'Autorite de Matiere Encyclopedique Unifie) - the French national level authority file. English counterparts are derived from Library of Congress Subject Headings Authority File via RVM (Repertoire de vedettes-matiere de Bibliotheque de l'Universite Laval a quebec). In the paper selected problems of the construction method and the establishment of integrity as utilized up-to now are discussed. the presentation covers also issues of international co-operation in the subject matter.

---

6 The Lists of Free-floating Subdivisions of the KABA Language. Elaborated by: Teresa Głowacka, Grażyna Jaśkowiak. Gdańsk 1995, Inter-Academy Coordination Team for Implementation of VTLS.

COMPATIBILITY PROBLEMS

Wiesław Babik

TERMINOLOGY AS A LEVEL FOR THE COMPATIBILITY OF INDEXING LANGUAGES. SOME REMARKS

1. INTRODUCTION

The application of terminology is not something entirely new in the theory and practice of constructing information retrieval languages (IRLs). For a long time already terminological dictionaries are used as one of the sources of lexical units for lexical systems of IRLs. Using specific terms in the same meaning is a necessary condition for efficient information exchange and scientific communication.

One of the most crucial problems in searching for the same terms in different information retrieval systems (IRSs) and databases is the fact that concepts can be and will be unpredictedly expressed in many different verbal forms unpredictedly in different natural and artificial languages. All the elements of the retrieval process: queries, messages, designations, semantic structures, etc. are the products of people, and are determined by the knowledge structures of people (public knowledge and personal knowledge). These structures must be taken into account in the retrieval process. To make computer-based retrieval systems more usable and more effective one must provide for greater interaction between, on the one hand, the knowledge structures incorporated in the semantic structures of IRL's and, on the other hand, in the knowledge structures of their users (11).

The aim of this paper is to emphasize the importance of terminology as a level for compatibility and integration of order systems and indexing languages as based on lexical units of natural languages. In this paper I will try to specify only such problems which require detailed scientific study and research.

2. CONCEPTS, TERMS, LEXICAL UNITS — WHAT ARE THEY?

There are various types of definitions of „terminology" and „term". Many of them indicate the aims, problems and scopes in terminology as a whole, while

others restrict themselves to the scope of this or that scientific discipline, for which they provide definitions of concepts or principles and rules for the correct formation of terms and terminological systems. We quote two such definitions of that „terminology" or „term" from selected books.

Terminology — in the words of H. Felber and G. Budin, the authors of „Theory and Practice of Terminology" , a book recently translated into Polish and published under the title „Teoria i praktyka terminologii", Warszawa 1994 — is „an ordered set of concepts of a given scope with signs attributed to them" (p. 27). A term is „a sign of a concept, consisting of one or more sequences of letters" (p. 26). A concept is „the idea [unit of mind] attributed to an object and replacing it in one's thinking".

According to the „Encyclopedic Vocabulary of Terminology of Information Retrieval Languages and Systems" (Słownik encyklopedyczny języków i systemów informacyjno-wyszukiwawczych. Warszawa 1993), edited by B. Bojar, terminology is: 1) a set of terms; 2) a part of linguistics. The „term" is „a lexical unit of any level of the structure of a language whose meaning is strictly determined in a given scientific or technical discipline" (p. 171).

These definitions illustrate the important differences in the approach of both disciplines to terms. In literature they are very often confused.

2.1. Approaches to terminology

There are two approaches to terminology, found and applied both in theory and in practice: a philosophical (epistemological/ontological/logical) approach, based on concepts (E. Wüster), and a linguistic approach, based on linguistic units (signs) (J. C. Sager). Information retrieval theory uses terminology as a set of terms forming an element of lexical systems of IRLs. Its approach to terminology is likewise a linguistic approach.

Information science and information activity perceive the terminologies of various disciplines as a basis and as basic tools for the lexical systems of IRLs. Information retrieval uses only a part of terms to build the lexical systems (lexicon) of IRLs. Terminology as a set of terms fulfils in IRLs two specific functions: a meta-informative function and a retrieval function.

In various object-oriented fields of knowledge the concepts have an extralinguistic form of existence, namely in linguistics and in the theory of information retrieval languages, where concepts have a meaning and significance. Concepts are the logical categories, while meaning is the linguistic category. Both levels are similar, but not identical. Here is how the differences are presented in scientific studies: terminology comes from concepts, linguistics (including also information retrieval languages theory) — comes from linguistic units. The philosophical approach to terminology regards the term as an extralinguistic object with extralinguistic relations to other objects. The linguistic approach to terminology looks at a term as the lexical unit of the given scientific language, in which there are interrelations at all levels of the structure of the linguistic system.

The first approach presented makes many difficulties in information retrieval, in particular in the field of the building of definition sequences and semantic nests. The basic difference between the approaches presented is that terminology builds terminological systems to solve many problems in scientific communication and to order the knowledge; whereas information science, that is: information

retrieval theory, only tries to apply those systems to optimize information retrieval processes.

2.2. Dimensions of terminology

Concepts are elements of the structure of knowledge. Knowledge is designated by lexical means. The items which are characterized by special reference within a discipline are the 'terms' of that discipline, collectively they form its 'terminology', and their totality form is the 'vocabulary'. A theory of terminology is primarily concerned with a referential system which relates knowledge structures to reality (cognitive dimension), then to lexical structure and defines the constituent elements of each type of structure (linguistic dimension). Terms are used with the specified fixed reference. The relationship existing between the term and the associated concept and proposing the term as a standard is publicly fixed.

The scientific process provides a systematic approach to the selection of characteristics and concept formation. Terminology relates terms to concepts (and not vice versa) and is therefore not concerned with absolute conceptual systems, but only with systems created for the specific purpose of assisting communication. The relationships between concepts we select and declare relevant for a particular purpose of ordering are therefore a subset of possible relationships.

Definitions provide the link between concepts and terms by means of an equation in which the definiendum is the term. A terminological definition provides a unique identification of a concept only with reference to the conceptual system of which it forms part, and it classifies the concept within that system. The definition fixes the intension, i.e. the reference of a term. Definition explicates the meaning of lexical items of a language by relating them to other items evocative of entities of the real world, of which they are symbolic representations. Terms are elements of language and could in principle be described purely linguistically by means of various relations.

On the basis of this criterion, Hayes and McCarthy (7) divide „knowledge" into two distinctive sections: 1. epistemological knowledge, which consists of data in specific data structures, and 2. heuristic knowledge, which describes how the data has to be handled. Concepts would fall under category 1, terms and terminological principles under category 2.

Terms are the linguistic representation of concepts. Concepts represented in terminological dictionaries are predominantly expressed by the linguistic form of nouns. Terms from 'natural' terminologies (incl. nomenclatures, etc.) often will correspond to items in a documentation language. But the function of documentation languages, having 'controlled vocabularies' as they do, is quite different, and it is advisable not to call this correspondence 'equivalency', but rather to treat the data realm of terminology differently from that of the documentation languages.

3. COMPATIBILITY AND CONVERTIBILITY OF VOCABULARIES

Terminology is the study and the field of activity concerned with the collection, description, processing and presentation of terms, i.e. of lexical items belonging to specialized areas of usage of one or more languages for special purposes.

From the point of view of terminology the lexicon of a language consists of many separate subsystems representing the knowledge structure of each subject field or discipline. Each knowledge structure consists of variously interlinked concepts. Approaching the study of terminology from its cognitive dimension requires an understanding of the structure of knowledge in order to obtain as complete and coherent a picture of the nature, behavior and interaction of concepts and their associated terms as possible. For maximum effectiveness, a controlled vocabulary must be carefully tailored to the requirements of a particular user community.

In a sense, the very existence of a controlled vocabulary implies a desire for compatibility among the terms used by authors, indexers, and searchers. Investigators have identified three major levels of compatibility. Descriptor A from one vocabulary may be mapped to descriptor x in a second vocabulary if: (a) A and x are identical, (b) A and x are synonymous, or (c) A is a species of the genus x./(6).

In this connection F.W. Lancaster (6) says about „similarity" between vocabularies. The relationship between a terminological vocabulary and an indexing vocabulary can be defined by the ratio

$$\frac{\textit{Terms for which equivalency exists in the two vocabularies (A + B)}}{\textit{Total of distinct A terms used in terminological vocabulary + total of distinct B terms used in indexing vocabulary}}$$

The compatibility existing between IRLs that are already in operation may be called a posteriori compatibility. If compatibility is secured from the beginning of the construction of an IRL's on the basis of the terminology of a given scope of science, especially of a terminological dictionary, then this kind of compatibility may be called terminological compatibility or a priori compatibility. Of course terminology is being understtood in a very wide sense here.

According to De Saussure's thinking in the light of structural linguistics, a terminological dictionary contains both terminological systems, i. e. terms on the expression level, and conceptual systems, i. e. concepts on the content level.

This cannot be — as it would seem — a simple matter of translation of terminological dictionary lexical units-terms based on natural language. In that case one would have to build a new language that is coherent with the source language, i.e. with the lexical system (terminological dictionary) of a scientific language, based on natural language. It is then a matter of denaturalization of terms in a given field of knowledge in opposition to the naturalization of one IRL's lexical system to a lexical system of the second language, assuming that both IRL's are based on quasinatural lexical units. This is very important, because — as already mentioned before — users in information retrieval use the terminology of their given field. This greatly faciliates information retrieval in comparison to retrieval with artificial lexical systems of IRLs, i.e. systems with the projected artificial semantics.

4. THE POSTULATES EMERGING FROM INFORMATION RETRIEVAL THEORY AS APPLIED TO TERMINOLOGY

The terminology collected in terminological vocabularies is selectively used as an important source of lexical units for lexical systems of information retrieval languages. Terms are lexical units having a special and conventional meaning and significance; they are the most informative elements of the language of science and of information retrieval languages. Therefore, in information retrieval, terms are most often used as keywords or descriptors, e.g. lexical units claiming the univocal meaning that is a non-synonym. Information retrieval tries to use structured terminological systems prepared from the terminologies of the given scientific disciplines, although the semantics of the lexical systems of information retrieval languages may be the systems with designed artificial semantics.

Application of prepared terminological systems makes the given information retrieval languages more user-friendly for those already familiar with them. It is essential that the terminology consists of well-structured terminological systems so that we may 1) be able to build definitions of terms in terms of lexical units, not only in terms of definitions of concepts; 2) have a criterion at our disposal to distinguish in our terms from any lexical units that are non-terms (3).

The aims of such terminological work and the functions of terminology standard will consist in 1) supplying the meaning of terms used in the given discipline; 2) ordering the terminology; 3) designing the terminological systems needed.

5. CONCLUSION

In conclusion let me note that if one would have to answer the question of the paper whether terminology may be a level of compatibility of IRL's, we can answer with „yes" on condition that terminology is understood here in a very wide sense as „the ordered set of concepts of the given scope with signs attributed to them" (G. Budin, H. Felber (5)) and is useful in the IRL's lexical systems.

REFERENCES

1. Babik W.: Terminology from the Point of View of the Information Science. Paper presented on 4th Congress and Symposium of The International Organization for Unification of Terminological Neologisms — IOUTN and the International Federation of Terminology Banks — IFTP. 22-25 September 1995 in Warsaw on the topic: „Worldwide Flow of Knowledge through Terminology". 4p.
2. Babik W.: Słownik języka informacyjno-wyszukiwawczego a słownik terminologiczny — zarys problematyki. [W:] III Sympozjum nt. „Zastosowanie mikrokomputerów w informacji naukowej, technicznej i ekonomicznej", Gliwice 23-24 listopada 1993 [Materiały konferencyjne]. Gliwice: Instytut Metali Nieżelaznych 1993, 12s.

3. Bojar B.: O normalizacji terminologii. Zagadnienia Informacji Naukowej 1987 nr 1(50) s. 23-35.
4. Dahlberg I.: Towards Establishment of Compatibility between Indexing Languages. International Classfification 1981 No 2 p. 86-91.
5. Felber H., Budin G.: Teoria i praktyka terminologii. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego 1994.
6. Lancaster F.W.: Vocabulary Control for Information Retrieval. Arlington: Information Resources Press 1986.
7. Nedobity W.: Terminology and Artificial Intelligence. International Classification 1985 vol. 12 No 1 p. 17-19.
8. Sager J.C.: A Practical Course in Terminology Processing. Amsterdam: John Benjamins 1990.
9. Słownik encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych. Red. nauk. B. Bojar. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego 1993.
10. Ścibor E., Tomasik-Beck J: On the Establishment of Concordances between Indexing Languages of Universal or Interdisciplinary Scope. Knowledge Organization 1994 No 4 p. 203-212.
11. Vickery B.C., Vickery A.: Information Science in Theory and Practice. London: Butterworths 1987.
12. Zeng L.: Achieving Compatibility of Indexing Languages in Online Access Environment. [In:] Encyclopedia of Library and Information Science. Ed. A. Kent. Vol. 50 Supplement 13. New York-Basel-Hong Kong: Marcel Dekker 1992, p. 1-24.

SUMMARY

Terminology has always been one of the central intellectual problems in theory of indexing languages. The last research directions are: 1) to consider indexing languages as specialized languages designed to optimize retrieval process, and to seek insights into their structure from the field of linguistics and terminology, 2) to interprete all the elements of the retrieval process; queries, messages, designations, semantic structures as the human products that are determinated by the structures of the public and personal knowledge represented in terminology. The paper presents the importance of terminology as a level of compatibility and integration of order systems and indexing languages.

Pavla Stančíková

INTERNATIONAL INTEGRATED DATABASE SYSTEMS LINKED TO MULTILINGUAL THESAURI COVERING THE FIELD OF ENVIRONMENT AND AGRICULTURE

INTRODUCTION

During the last two years CEIT assisted international bodies in further developing Micro CDS/ISIS software package applications dealing with multilingual problems

•conversion of data to Micro CDS/ISIS

•interface with DeskTop publishing

•conversion of data to Internet

Projects on multilingual applications were usually finalized by creating a multilingual database system and publications such as

• CEDEFOP thesaurus in 7 volumes: English, German, French, Italian, Greek, Spanish, Portuguese (project for TermNet/EC)

• Waste Vocabulary in 9 languages: German, English, Russian, Czech, Croatian, Polish, Slovak, Slovenian, Hungarian (project for TermNet/ÖNORM)

CEIT's latest achievements in Micro CDS/ISIS applications are as follows:

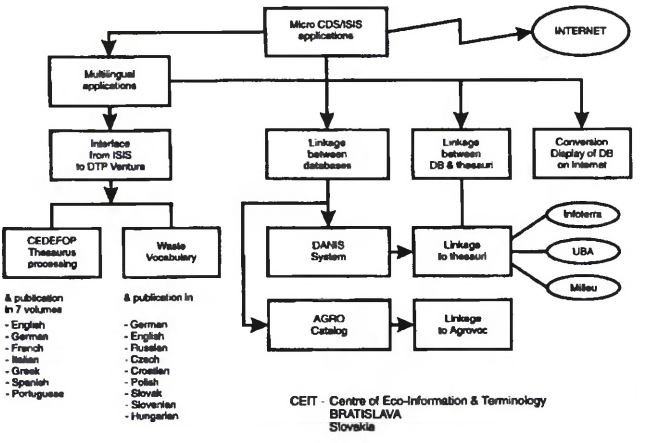


Fig. 1 Micro CDS/ISIS applications at CEIT Bratislava

• linkage between databases, supporting hypertext functions within ISIS
• conversion of terms collected in different thesauri in MTM4 software package
• linkage between databases and different thesauri.

All activities mentioned above are illustrated in the flow chart designated The latest results obtained by CEIT are described in more detail on two database systems:

1. DANIS — Information System on Human Resources, Standards and Legislation (Danube Programme Project supported by UNOPS New York)

2. AGRO Catalog related to AGRIS (Project launched at the Institute of Scientific and Technical Information for Agriculture, Nitra, Slovakia)

1. DANIS — THE INFORMATION SYSTEM ON HUMAN RESOURCES, STANDARDS AND LEGISLATION

DANIS, the Danube Information System on Human Resources, was launched in 1993. In September of that year, the United Nations Educational, Scientific and Cultural Organization (UNESCO) and in particular its International Hydrologic Programme (IHP) started to develop a Prototype Information System to support the Environmental Programme for the Danube River Basin. An implementation plan was adopted to establish a regional information network — the Expert Group — and to develop DANIS.

Cooperation among the Danube countries for the development of a regional information network and DANIS was further developed during the DMSG meeting held in Bratislava, 21-22 April 1995.

DANIS objectives

The objectives of DANIS are:

1. To provide access to existing primary information systems and databases developed for use in the Danube region.

2. To further develop DANIS as an information source and tool in order to assist the Programme Coordination Unit in its daily tasks.

3. To efficiently provide information to the public in all Danubian countries and other interested parties.

Data Management SubGroup (DMSG) agreed that all existing data collected and processed in DANIS had to be updated, edited, corrected and completed focusing on their availability on INTERNET. This task was included in Work Plan for 1995/96.

Priority of Data Collection

The responsibility of the DANIS System Manager is to manage and control the data entry process in such a way that priority will be given to:

• Working Groups defined within the Danube Programme (this relates to data entry on persons and organizations working within the Danube Programme)

• Task Force meetings and related events (this relates to a database on events, persons and organizations

• Danube PCU Library (all documents relating to the Danube Programme collected at the Danube Programme Coordination Unit in Vienna should be entered into DANIS)

• Standards relating to environmental management in the Danube River Basin

• Multilateral and bilateral legislation relating to Danube countries and the Danube Programme

• Related projects or programmes of interest to Danube countries and the Danube Programme (e.g. the Black Sea Programme)

The approach to data collection and information provided in DANIS was to give priority to information required by environmental professionals, institutions and organizations in the Danube countries. DANIS is an integrated database system available in its PC version or accessible on INTERNET. It is a powerful tool. DANIS gives access to relevant data, environmental activities and to the Danube PCU Library, as well as providing information about other programmes and relevant standards and legislation.

DANIS is an integrated tool based on linked databases :

• Danube PCU library documents (database DOCUM) — 381 entries

• Danube Programme Working Groups (database GROUP) — 10 entries

• persons involved and participating in the Danube Programme (database PERSO) comprising 312 records

• multilateral and bilateral agreements and conventions (database LEGIS) consisting of 204 records

• standards relating to the environment (database STAND) consisting of data collected at ON/INFOTERM Vienna, comprising 790 records

• events, mainly Danube Task Force meetings (database EVENT) comprising 13 records

• environmental programmes and projects (database PROGR) comprising 70 records

Data Sources

Database DOCUM (Danube PCU library) has been processed in the Vienna Danube PCU office. All activities relating to the database entry were supervised by the Centre of Eco-Information & Terminology (CEIT) Bratislava. The latest design of DANIS databases their linkage and display formats were further developed by CEIT Bratislava.

Data included in the database PERSO was converted from a text file created by Danube PCU Vienna. This database is not in its final form and requires further contact with persons entered in the DANIS database for more detailed information.

The database LEGIS, created by CEIT, is based on the following documents:

• „Bilateral and multilateral agreements and other arrangements in Europe and North America on the protection and use of transboundary waters". Senior Advisers to ECE Governments on Environmental and Water Problems. United Nations/Economic and Social Council/Economic Commission for Europe. ECE/ENVWA/32, June 1993, 40 p.

• „List of bilateral international agreements of the Slovak Republic in force until 31 December 1994". Ministry of Foreign Affairs of the Slovak Republic, Bratislava 1995, 226 p.

• „List of multilateral international treaties obligatory for the Slovak Republic by December 31", 1994, 103 p.

• „Directory of community legislation in force and other acts of the Community institutions". Official Journal of the European Communities, Volume 1. Analytical register, 24th ed., Dec. 1994

The records for the database STANDARD were selected in close cooperation with the International Information Centre for Terminology (Infoterm) at the Austrian Standards Institute (ON), Vienna. The data structure and layout was prepared by CEIT Bratislava.

Environmental Thesauri

Three main environmental multilingual thesauri were recognized as being relevant to the DANIS system:

1. INFOTERRA Thesaurus (UNEP, Nairobi)
2. Milieu Thesaurus (European Environment Agency, Copenhagen)
3. UBA Thesaurus (Umweltbundesamt, Berlin).

All environmental tools for indexing and searching documents within DANIS were converted into MTM4 software available for multilingual thesauri maintenance. This tool, consisting of three different environmental thesauri, can be used for

• further development, updating, maintenance of existing versions of multilingual thesauri

• entry of a new language version (for instance, national language)

The data entry module of the DANIS system supports a function which allows the user to select a term by „clicking" from any of those three thesauri. The data entry process for the descriptor field is supported by opening the window of the thesauri in alphabetical order or in a „tree" structure.

The DANIS information system consists of the following subsystems

1. Databases
   - persons
   - projects
   - organizations
   - events
   - standards
   - legislation
   - documents
2. Thesauri and classification schemes
   - INFOTERRA thesaurus
   - Milieu Thesaurus
   - UBA Thesaurus
   - CCF (Common Communication Format) Codes

2. THE AGRO CATALOG

The Institute of Scientific and Technical Information for Agriculture in Nitra launched a project for the creation of an integrated database system for agriculture in Slovakia. CEIT Bratislava was contracted to do the database design, conversion

of data to Micro CDS/ISIS application, conversion of data to Internet format and for the development of a tool for AGROVOC thesauri maintenance.

The AGRO database consists of three main information sources:

- library records
- research project reports
- mission reports

Selected records from AGRO database should be transferable to the AGRIS/FAO system.

The AGROVOC thesaurus was converted to MTM4 software in order to support the creation of a national (Slovak) version of the AGROVOC thesaurus. Simultaneously a linkage to the AGRO database was developed allowing the use of English equivalents in the thesaurus to create English descriptors compatible with the AGRIS system.

MTM4 software will be further used for

- maintenance/updating of the AGROVOC thesaurus
- creation of national (Slovak) version of the AGROVOC thesaurus.

The AGRO databases mentioned above are linked to the AGROVOC thesaurus. For more details see Fig.

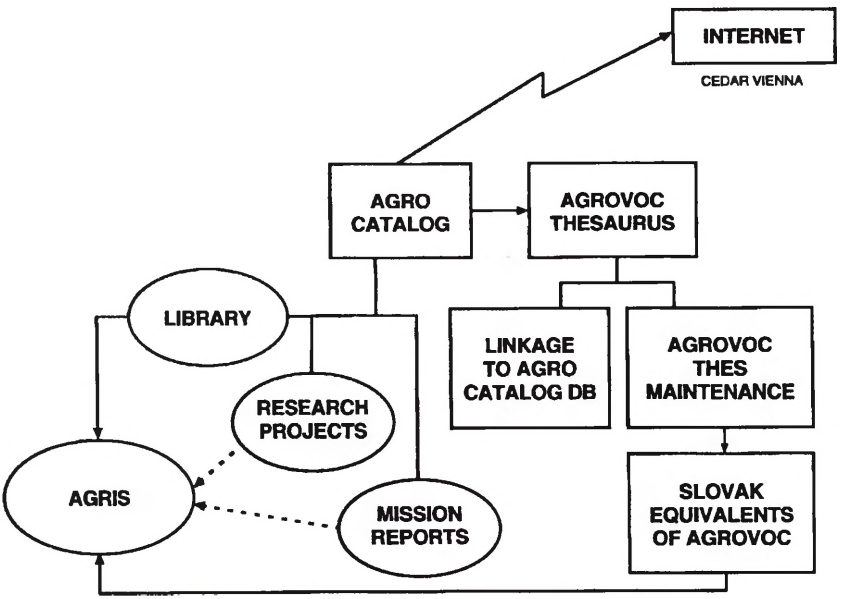


Fig. 2 AGRO catalog at UVTIP Nitra

3. DANIS ON INTERNET

- WWW format
- Gopher format
- BRS format to support efficient search

The linkages mentioned above are illustrated in the flow chart designated Fig.

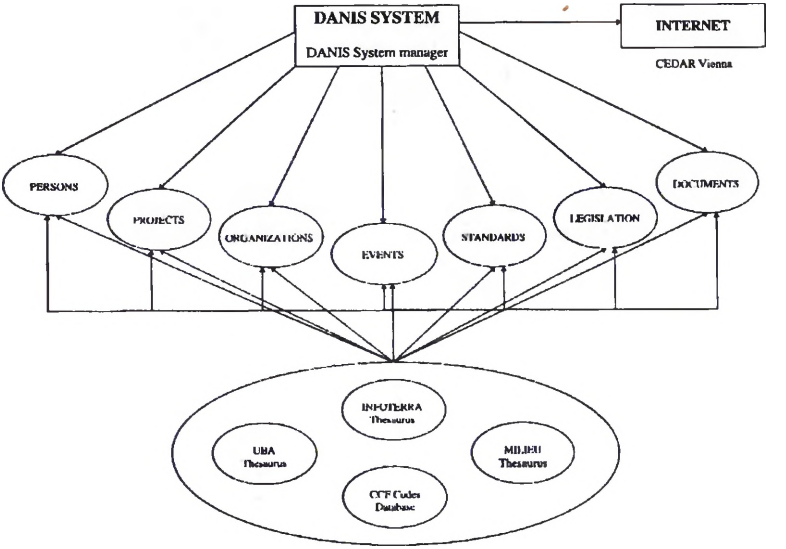


Fig. 3 DANIS — Information System on Human Resources of Danube Countries

CONCLUSION

The DANIS and AGRO catalogs could be taken as examples how problems of compatibility and integration of order system could be solved.
Such a solution requires the adoption of the following principles:

- To use existing thesauri as they are (from a content point of view)
- To create linkages from different fields to different thesauri
- To create/apply software for thesauri maintenance (including national equivalents)
- To use/create a user interface guide (user navigation guide)
- To use thesauri in combination with classification systems
- To create/use the potential of facet classification/facet thesauri for user navigation.

DANIS URLs:
www: http://www.cedar.univie.ac.at
gopher: gopher://gopher.cedar.univie.ac.at

SUMMARY

Reports on latest achievements obtained by CEIT in further developing of the DANIS system (the Danube Information System on Human Resources, Standards and Legislation) and the AGRO catalog related to AGRIS. Both database systems function as integrated systems with linkages to relevant multilingual thesauri. The DANIS system is linked to three available thesauri: The INFOTERRA Thesaurus (UNEP, Nairobi), the Milieu Thesaurus (European Environment Agency, Copenhagen), and the UBA Thesaurus (Umweltbundesamt, Berlin). The AGRO Catalog is linked to the AGROVOC thesaurus. This solution supports three main functions: control of data entry process, maintenance of multilingual thesauri and creation of a national (Slovak) version of an international multilingual thesaurus.

Tomáš Samek

INDEXING LANGUAGES INTEGRATION AND EUROVOC THESAURUS IN THE CZECH REPUBLIC

1. THE INFLUENCE OF AUTOMATION ON THE COMPATIBILITY OF INDEXING LANGUAGES

At a time when complete catalogues are being compiled in various libraries, at a time of an ever easier on-line access to various databases, the integration of order systems is a vital necessity. And yet, in the Czech Republic such integration is only in its very beginnings and its importance is still being underestimated: the individual libraries and information centers are mostly using mutually incompatible indexing languages.

In the course of automation, a great many libraries have come to acquire software with an inbuilt component for thesaurus administration (in the Czech Republic these were most often the *Tinlib* and *Aleph* systems), but mostly they were unable to feed this component with suitable data, because — except in a few isolated cases — no suitable thesaurus was available in Czech.

Automation thus confronted the libraries with the following basic questions:

1. whether to preserve continuity in the order system by transfering present headings and their references into electronic form, or

2. whether to give preference to some other order system.

2. ADAPTATION OF THE EUROVOC THESAURUS AT THE PARLIAMENTARY LIBRARY

2.1 Transition to the EUROVOC thesaurus

In 1992 the Czech Parliamentary Library (PL) was still using the subject headings created in the 1920s by its then director Zdenek Vaclav Tobolka, which system was being updated continually. With the transition to the Tinlib system, it was decided to abolish this catalogue and to adopt instead the polythematic multi-lingual thesaurus EUROVOC (1).

The main reasons speaking in favor of EUROVOC were:

1. the higher degree of control offered by the EUROVOC thesaurus in comparison with the Tobolka's system;

2. as a thesaurus compiled by the European Parliament for the indexing of all external and internal documentation, EUROVOC could be assumed to correspond to the needs of the Czech PL;

3. all European Union legislation (Celex database) was going to be EUROVOC-indexed in the near future.

Practice proved that the EUROVOC thesaurus does not fulfil all the original expectations to the degree hoped for expected. Especially the second assumption soon became a relative matter because the viewpoints of the European and the Czech Parliaments differed quite substancially. That was reflected of course in differences in vocabulary between the source thesaurus (the original EUROVOC) and the target thesaurus (the Czech version of EUROVOC).

2.2 Criteria for selecting a suitable source thesaurus

When translating and adapting a foreign-language thesaurus, or generally when adapting any source thesaurus, three sets of lexical units are created :

A. Descriptors of the source thesaurus that are necessary for the library for which the target thesaurus is being compiled (I shall abbreviate „library for which the target thesaurus is being compiled" to „target library").

• For the PL, for example, the following EUROVOC descriptors are necessary: parliament, executive power, neutrality, anarchism, etc.

B. Descriptors from the source thesaurus, now useless for the target library and thus not applicable in the target thesaurus.

• Examples from EUROVOC for the PL : sesame seed, rosé wine, etc.

C. Terms which the target library obviously needs for the indexing of documents, although they are not in the source thesaurus.

• The following expressions, for example, necessary at the PL are not included in the 2nd edition of EUROVOC: restitution, standing order, rehabilitation, descriptors marking the individual national languages, The Central-European Initiative, Czech regions, a more detailed period-sequence of history, etc.

Generally speaking, we can state: Let A, B and C be the sets of terms described above and $n(X)$ the number of elements of a set X as desired. Then the ratio

$$Pt = \frac{n(A)}{n(A) + n(B)}$$

determines the degree of applicability of the source thesaurus during t (period of time) for the target library.

After the target thesaurus is finished, other libraries with similar requirements for indexing who are likewise considering the suitability of the source thesaurus for their information system, will also find the following ratio important:

$$Qt = \frac{n(A)}{n(A) + n(C)},$$

which determines the degree of applicability of the source thesaurus to the target thesaurus.

In both cases, however, these are quantitative indices whose full value should not be overestimated and which should be regarded only as auxilliary parameters for the decision-making of the given information institute in selecting a suitable source thesaurus. Into the above equations one may insert gained by the processing of a merely statistically important sample of terms of the given thesauri,

because trying to ascertain suitability for indexation for all the terms in the given library would be extremely time-consuming.

There are many reasons why the given quantitative parameters have only a limited descriptive ability. One of these is the fact that they provide information only about the number of applicable descriptors and not about the semantic relations in the given thesaurus, likewise constituting an aspect of basic importance.

2.3 The compiler of the source thesaurus versus the compiler of the target thesaurus

The viewpoint of the compiler of the source thesaurus plays a very important role, particularly if it differs substantially from that of the target thesaurus compiler. Even though EUROVOC was compiled for the needs of the European institutions, which viewpoint at first sight is not so very different from that the Czech PL, it became obvious even during the very first sample translation that there are differences in the conception of even the most elementary descriptors.

Example: The descriptor paragraph for the term parliament contains only two narrower terms:

parliament
SN Use narrower terms.
NT national parliament
NT regional parliament

The documents which contain the standing order of the Czechoslovak Federal Assembly could be indexed in the following way in EUROVOC, considering that the 3rd edition of the thesaurus (2) already contains the descriptor standing order:

Czechoslovakia
national parliament
standing order

However, from the viewpoint of the bipartite Czechoslovak Federation the descriptor national parliament would not be acceptable, because by *national parliament* we used to understand either the Czech National Council or the Slovak National Council, i.e. regional parliaments, but not the Federal Assembly. Wishing as we did to adhere to EUROVOC to a maximum degree, the descriptor *parliament* was thrown out and a difficult search was started to find a good Czech equivalent for *national parliament*.

Quite generally, descriptors concerning ethnic aspects (nations, national languages, etc.) are so strongly marked in the original EUROVOC version by its Franco-centric viewpoint that they are useless for Central European conditions. For example, even in the 2nd EUROVOC edition, the term *nation* was included only as a nondescriptor:

nation
USE **state**

This is useless for any state with multiple nations. Terms for individual nations and national languages were absent totally, so that for example the German-Polish dictionary had to be indexed in this way:

Germany
Poland
national language
dictionary

That is why in the Czech version of EUROVOC we created new descriptors for all the languages we need for our indexing, and this in harmony with EUROVOC architecture. There were many changes of this type and supplements, but I trust that further examples will not be necessary.

2.4 The question of component languages in the source multilingual thesaurus

Even though the original EUROVOC thesaurus is in French, we primarily used the English and German versions for our translation. As all the EUROVOC languages have the same status, I considered it interesting to compare the English and German versions from the viewpoint of translation into Czech.

In selecting a foreign-language polythematic thesaurus as source thesaurus one should preferably orient oneself to language area whose semantic fields are closest to the national language of the target thesaurus, in our case the Czech language. Our experience has shown that of all non-Slavonic languages, German comes closest to Czech.

This is a fact often overlooked in the Czech Republic these days because of the international significance of English, which is why I consider it necessary to describe this fact in more detail here. In the late 18th century the Czech intelligentsia wrote scientific works in German, and only after some hesitation changed to Czech, thus creating the missing Czech equivalents. Although partly taken over from other Slavonic languages, these equivalents were always selected or coined under the strong influence of German terminological paradigmata.

As to its suitability for translation purposes, English can compete with German only in some of the new or swiftly developing areas (computers, genetics, microbiology, cybernetics, etc.). Even there we see the unfavorable influence of the gigantic degree of polysemia so characteristic of the English language as one of the semasiologically and onomasiologically richest languages in the world. If from a nondescriptor we provide a reference to only one descriptor (the case with EUROVOC), strongly polysemous expressions are practically useless, because we would have to refer the nondescriptor to several descriptors simultaneously. The richness of expression and meaning in English (many expressions for one meaning and many meanings to one expression) — paradoxical though it may seem — thus becomes a disadvantage for this language. This can be borne out quantitatively as well: in the 2nd edition of EUROVOC the German version had 5 200 nondescriptors, while the English version had 4 800.

REFERENCES

(1) EUROVOC. 2nd ed. Luxembourg, Office for Official Publications of the European Communities 1987. 5 vol.
(2) EUROVOC. 3rd ed. Luxembourg, Office for Official Publications of the European Communities 1995. 2 vol.

SUMMARY

Influence of automation on compatibility subject indexing languages. Adaptation EUROVOC thesaurus in the Czech Parliamentary library. Swich-over from Tobolka's subject headings to EUROVOC thesaurus. Quantitative criterion for choice of applicable source thesaurus. terminological implication of different viepoint of the authors of the source thesaurus (French EUROVOC) and the target thesaurus (Czech Pareliamentary Thesaurus). Importance of source thesaurus component languages for choice of the applicable source thesaurus. Connection between EUROVOC and harmonization of Czech law with EC law.

CLASSIFICATION COMPATIBILITY IN DATABASES

Krystyna Siwek

COMPATIBILITY DISCREPANCIES BETWEEN POLISH AND FOREIGN DATABASES

1. HISTORICAL BACKGROUND

Classification systems of scientific organizations, disciplines and specializations in databases on R&D in Poland represent institutional and disciplinary structure of Polish research and education. This structure is a result of various factors: historical, geographical, political, social, religious, cultural.
The milestones of the history of the Polish science are such events as foundation of the oldest Polish academic unit — Jagiellonian University in Cracow in XIV century, an important European centre in such fields as philosophy, astronomy, theology and law. In the end of XVI century were established academies in Wilno and in Zamość. In 1773 the Commission for National Education was created, fulfilling the function of the ministry of science and education. In 1800 come into being Societas Scientiarum Varsoviensis, while in 1873 Cracow established the Academy of Sciences and Letters. On the day of restitution of the independent statehood (1918) , in Poland there were four main universities: in Cracow, Lvov, Warsaw and in Wilno. Also several new specialized schools of higher education were called into being. After the II World War the science in Poland was rebuild, but according to Soviet model, specially in the field of social sciences. During that time research was conducted in 82 state schools of higher education.. International scientific contacts have been limited, although there were possible common, bilateral projects and publications. In some disciplines Polish research achievements were ranked highly. After 1989 organization of the Polish research infrastructure has been deeply changed, are being created conditions and possibilities for a further cooperation with research institutions world over.

2.THE ORGANIZATIONAL STRUCTURE OF THE POLISH SCIENCE

The institutional structure of the Polish science is divided into three main sectors:

— The Polish Academy of Science
— Universities and other institutions of higher education
— Industrial institutes and other R & D units

The Polish Academy of Sciences acts as a scientific body and as an national organization of research centres. According to historical tradition, the institutes of the Polish Academy of Sciences are mostly involved in basic research related respectively to the following seven academy divisions:

- social sciences and humanities
- biological sciences
- mathematics, physics and chemistry
- technical sciences
- agricultural sciences and forestry
- medical sciences
- Earth sciences and mining

The education sector is supervised by few different branch ministers. Besides there are about 60 private universities and colleges. The main sector is The Ministry of National Education, which coordinates education activities in Poland. All the universities, universities of technology and high schools of engineering , high schools of economy , agricultural universities and pedagogical colleges are supervised by the Ministry of Education. Medical academies are supervised by the Ministry of Health and Social Welfare. High schools of arts are in the gestion of the Ministry of Culture and Arts. Academies of Physical education are supervised by the president of Sports and Tourism Office. Two merchant navy colleges are supervised by the Ministry of Transport and Maritime Economy. There are schools of higher education subordinated to the Ministry of National Defence and to the Ministry of Internal Affairs It can distinguished classification of research and educational disciplines according to organizational structure. Universities and other schools of higher education are divided into faculties, which are subdivided into institutes, chairs and departments as lower levels of organizational classification.

Branch research is conducted in R & D institutes, centres and laborarories. Although the number of these units decreased lately they take important part in technology research because of close connections with industry. In Poland exist learned societies, sientific associations and foundations as well. There are regional associations, specialist associations, interdisciplinary associations, medical scientific associations, engineering societies and science foundations.

The division of sciences into basic and applied disciplines is not disjunctive. It depends mainly on concreteness of results and fulfilment of their practical functions. Thus the decision to ascribe particular scientific institutions to basic or applied sciences is arbitrary and rests on the assumptions, that all PAN institutions are classified as belonging to basic sciences and all sectoral units are assumed to belong to applied sciences. Among higher schools universities are thought to belong to basic sciences while higher technical schools and agricultural, economic and medical ones to applied sciences.

The science policy of government is realized by the State Committee for Scientfic Research (KBN) divided by commissions and sections according to disciplinary classifications.

The istinction ibetween basic and applied research is reflected in the structure of KBN, which consists of the Chairman of the Committee, the Committee for Basic Sciences and the Committee for Applied Sciences.
The Basic Research Commission works in the following six groups.

- humanities
- legal and economic sciences
- mathematics, physics and astronomy
- biological sciences, Earth sciences and environmental protection
- medical sciences
- agricultural sciences and forestry

The Applied Research Commission works in the following six groups as well.

- building. mechanics and architecture
- materials engineering and technology of materials
- chemical sciences, technical chemistry and process engineering
- electrical and power engineering
- automatics and electronics
- mining, geodesy and transport

3. SCIENTIFIC CARIEER

On bachelor's and master's of science levels a list of the disciplines of studies and titles is prepared by The Ministry of National Education. The list consists of 11 broad classes expressing organizational structure of universities and other schools of higher education:

— humanities
— nature, physics, chemistry, mathematics
— economy
— technology
— agriculture
— theology
— music
— fine arts
— theatrical arts
— medicine
— sport and tourism
— navigation

Above classes include 150 disciplines.

In Poland there are two levels of doctoral degree — doctor and doctor habilitatus in the range of:

— chemical sciences
— economic sciences
— pharmaceutic sciences
— physical sciences
— geographical sciences
— humanistic sciences
— forestry sciences
— mathematic sciences
— nature sciences

— agriculture sciences
— technological sciences
— theological sciences
— veterinary sciences
— political sciences
— legal sciences
— medical sciences
— military sciences
— physical education (sport) sciences

The same terminology is used for professors titles nominated by the President of Poland.

4. RESEARCH OUTPUT

Research activities conducted in mentioned above universities and RTD units include

- statutory R & D activities
- peer-reviewed research grants
- goal-oriented projects
- general science support activities

Statutory funding is the main source of R & D financing supports long-term research on the main research problems. Applications for statutory funding are submitted annually by each institution and are evaluated by KBN panels of experts. In the evaluation process institutional applicants are ranked by the mentioned above Commissions according to certain criteria based on categorization of the research institution. It allows achieve the greater compatibility of the evaluations of research bodies.

Research grants concern individual and collective research projects as an open competition among individuals and research teams, regardless of their institutional affiliation. Research proposals are evaluated through a two-stage review process. First, the proposals are send to qualified reviewers for their assessment and next examined by a panel within the appropriate group (according to discipline classification) of the KBN Commissions.

Research and development work important for social and economic reasons is stimulated by central or local administrative bodies to serve regional or sectoral scientific policy. Grants for goal-oriented projects aim to support innovativeness in economic entities by co-financing research designed to achieve results that will be used in production.

One of the most important factors in evaluation of RTD units and individual scientists are their publications — the research products and their citations — the research impact.

5. DATABASES AND DIRECTORIES ON RTD IN POLAND

The Information Processing Centre (OPI) as a unit aimed at gathering and supplying information on and for Polish science. It manages information systems

on R & D in Poland.. The information collections produced and maintained in the OPI cover data related to institutional and human resources as scientific institututes, research establishments and universities, scientists and researchers and data related to research products as the research works and publications. The essential task of the OPI is to provide services in the area of science and that of the applications of the scientific research results. OPI supplies also due information to the State Committee for Scientific Research. At present, the OPI operates nationwide databases in Polish and English version as SYNABA and DOKTORATY I HABILITACJE (Dissertations Database), databases on R & D institutions and researchers POLISH RESEARCH DIRECTORY. It manages NCR for Poland database on publications and citations of Polish authors (an extraction from ISI in Philadelphia resources).

SYNABA

The information system SYNABA is designed to compile information on current and completed research projects in Poland. The database comprises 100 000 records on over ten years on the broad scope of Polish science. From the beginning of 1996 some elements in English will also be included, such as: title, abstract and keywords. After political changes the structure of RTD is different, specially in social sciences and humanities.

So the same database, SYNABA, over 10 years, covers topics related to communist country, its politics and economy and, from 1990, specific to, so called, postcommunist country. Needless to say, that now SYNABA is bilingual — Polish and English, what involves new problems of translation of classification terms and keywords.

The Polish Thematic Classification (PKT), created by prof. E. Ścibor in The Institute for Scientific, Technical and Economic Information is used as a main linguistic tool together with keywords for research project abstracts. This classification has been modified for many years. Now the fourth version is in print. The differences between tha latest and earlier versions of the PKT are the most evident in social sciences and humanities.

Economic sciences and practice in communist countries were absoluely incompatibile to economic sciences in West Europe and USA, so even the same terms in classification schemes mean different concepts. There were many issues on centralized planning when in foreign databases there quite different concepts were illustrated by adequate terms.

In PKT Tables from 1985 there are such classes as
HUMANITIES AND SOCIAL SCINCES
01. MARXISM-LENINISM [Marksizm-leninizm]
02. PHILOSOPHY

in the newest verson of PKT Class 01 dissapeared
HUMANITIES AND SOCIAL SCIENCES
02. PHILOSOPHY

POLITICAL ECONOMY [EKONOMIA POLIYCZNA] (06) is replaced by ECONOMICS [EKONOMIKA]

Of course, there are many changes in other classes, caused by technological development, but they are not on the highest level of classification hierarchy.

Polish Research Directory

Polish Research Directory consists of four subbases. It includes SCIENTIFIC INSTITUTIONS file that stores information on scientific institututes of the Polish Academy of Sciences, schools of higher education, research and development units, libraries, musea, archives and other institutions conducting research activities.

Polish Research Directory subbase SCIENTIFIC INSTITUTIONS stores information on scientific institututes of the Polish Academy of Sciences, schools of higher education, research and development units, libraries, musea, archives and other institutions conducting research activities. The database is maintained in Polish and English versions. The information records of this database comprise the following items:

— name and address of an institution
— organizational structure of institution
— managerial staff
— main directions of research and didactic activity
— authorization to confer academic degrees
— availability and range of doctoral and postgraduate studies
— library and archival collections
— publication issued
— research workers holding the titles of professor and other academic degrees

Who is Who in Polish Science — SCIENTISTS AND RESEARCHERS subbase stores data on scientific workers. At present, the database contains information on 20 thousand professors and doctors habilitati and on 40 thousand doctors.

Additionaly, there is conducted subbase on learned societies and foundations, database on science policy bodies, database on Polish scientists abroad.
National Citation Report for Poland (from ISI — Philadelphia, USA) is the database on publications and citations of Polish authors.

All these mentioned above databases — SYNABA and POLISH RESEARCH DIRECTORY are similar in scope and structure to CORDIS (Community Research and Development Information Service) databases. Respectively SYNABA is similar to RTD-Projects and RTD-Results, although it is not fully compatibile.

Dissertations Database

Dissertations database stores information on academic degrees of doctor and doctor habilitatus conferred in Poland. It constitutes of dissertation titles register and is processed for statistical purposes as well.

The record fields include theme of dissertation, author, research unit which awarded the academic degree, place of employment of the author, name of the promotor, field of science in which the author has been conferred his academic degree. List of science fields is included in the chapter „Scientific carieer".

National Citation Report for Poland database

National Citation Report for Poland database is a selection of resources of ISI in Philadelphia for publications of the Polish authors. Information Processing Centre manages this database for individual and statistical analyses

Comparison between NCR for Poland database and Information Processing Centre data on research units and scientists is a question of different classification of scientific disciplines in Poland and in USA and a question of special method in ISI

First of all, paper (or other documents) in NCR database are classified through journal classification. i.e. there is not an individual classification per article. Sometimes it couses misunderstandings. All the „source journals" are classified by cathegory code (three letters) and product code (one letter)
Product code is used for edition of the „Current Contents" in 7 series:

— life science
— clinical medicine
— physical, chemical and earth science
— engineering, technology and applie sciences
— agriculture, biology and environmental science
— social and behavioral sciences
— art and humanities

There are 89 category codes in NCR for Poland database.

Tables of conversion between classification used in NCR database and the other service of Institute for Scientific Information in Philadelphia on worldwide ranking of publications and citations named National Science Indicators (NSI) include 18 NSI field codes and their relations to category and product codes.
e.g.

AGD = A/A, CMA, F
[Agricultural Sciences]=[Agriculture/Agronomy]+ [Agricultural Chemistry] + [Food/Nutrition]

CHD = ANL, CME, CML, CMP, ORG, PHC
[Chemistry] = [Analytical, Inorganic & Nuclear Chemistry] + [Chemical(T)] + [Chemistry(P)] + [Chemistry(S)] + [Organic Chemistry/Polymer Science] + [Physical Chemistry/Chemical Physics]

6. CONCLUSIONS

Experiences in managing NCR for Poland and other foreign databases in comparison with OPI databases prove, that discrepancies in linguistic tools are tentatively easy to solve. More difficult is to tackle with different concepts of science as a result of different concepts of world.

Neadless to say, for example, that general class LIFE SCIENCE is absent in all the classifications used in Poland.

More sound diffrences in concepts of science are generated by various ways of cultural and technological development then by classifcation methodologies. Historical, geopolitical, social and cultural environment is an important factor for various concepts of science.It may be the reason of incompatibile concepts under formally compatibile terms.

The Polish science renders diversity from one side and likeness from the other side to structures of science abroad.

In general, it is only a question of tables of conversion and switching language versions in using various Polish and foreign databases on RTD as integrated information and knowledge resources. Compatibility discrepancies should be avoided since there are sufficient software and linguistic tools to solve the problem of incompatibility on classification level.

The European Centre for Higher Education (CEPES) „Multilingual Lexicon on Higher Education" is one of examples of efforts on compatibility and integration in terminology according to RTD and education. Terms for each country are classified into nine sections: general, institutions and their structure, governing and administrative bodies, academic staff, students, access and admissions, teaching and learning, examining and evaluation, degrees and diplomas, where entries are given in the respective languages, followed by the English-language definition.

REFERENCES

T. Krauze: Struktura dyscyplinarna nauki polskiej w perspektywie porównawczej [Report], p. 69+121

Polska Klasyfikacja Tematyczna. II ed., Warszawa, 1985

Polska Klasyfikacja Tematyczna IV ed., Warszawa (in press)

Polish Research Directory, Warszawa, 1994

Multilingual Lexicon of Higher Education, München, 1993-95

J. Kozłowski: Miejsce Nauki Polskiej w Świecie. Na podstawie Science Citation Index, Warszawa, 1994

SUMMARY

Different national institutional structure of science is the result of social and political conditions. This structure is expressed in classifications used in research and development area, in databases on R & D. Classifications used in databases accessible in Poland on research and development, in bibliographic (abstracting databases) and in factual databases are described. The elements of identification and of characteristics (classification and indexing), in documents and facts description are compared. Descriptive elements of identifiers - corporate names, codes of organizations (eg. statistical code REGON) in some databases are analyzed.. Discrepancies in corporate names of research institutions in Poland - Polish and English versions, full names and abbreviations, incompatibility in corporate names of research institutions, classification of institutional activity and terminology of specialization of scientists affiliated to these institutions are presented. Some remedies are suggested concerning standarization of terminology and classification and indexing tools.

W. Gliński, M. Muraszkiewicz

AN INTELLIGENT FRONT-END PROCESSOR FOR ACCESSING INFORMATION SYSTEMS[1]

1. INTRODUCTION

Networking and integration of media are among the most rapidly growing chapters of computer research, applications and business [ENG95, LIU94]. The Internet is the buzzing term attracting attention of both researchers working in the domain of computer and information science and casual users looking for new opportunities when searching information [MAT95].

The Internet is a massive world wide network of computer networks comprising about 50,000 interconnecting networks, including about 3.2 million host computers directly connected to the net with more than 37 million users throughout the world. The rate of growth of the number of hosts is some 12 per cent per year. At present, the Internet spans more than 160 countries. As a virtual space of software, networking and computers, the Internet is infinitely renewable and infinitely adaptable, and it grows and changes every day.

The information resources available through the Internet are immense. The total volume of the files accessible via the Internet is counted in thousands of Gigabytes. Therefore, the main problem connected with the Internet is the flood of information. The most obvious is the difficulty of simply finding items in the vast seas of available material. Another issue is that not all of the information on the Internet is of equal quality or value. This is why, in spite of tremendous capacities of the Internet, the skills and knowledge on the part of its users will still remain a decisive factor.

The issue of identifying the relevant resources and accessing them is of crucial importance, in particular from a casual users' perspective. As a rule their specific knowledge about the network itself, and the distribution and contents of the information sources is limited. This paper addresses this issue by proposing a method helping the users to establish „good" queries and submitting them to the information resources (residing on the network) which are likely to contain the relevant documents.

The structure of this paper is as follows. In Section 2 we shall present the underlying concept of the front-end processor devised for facilitating the process of building and forwarding the queries. Preparation of a query for searching, which is the main operation executed by the front-end processor, will be described

[1] The work reported in this paper has been jointly supported by grant no. 8 T11C 038 08 of State Committee for Scientific Research (KBN), Poland, and by the Institute for Theoretical and Applied Computer Science, Polish Academy of Sciences.

in Section 3. The principle of *ADDRESSER's* functioning will be given in Section 4. The last Section will focus on the follow-up activities regarding the development of the front-end processor.

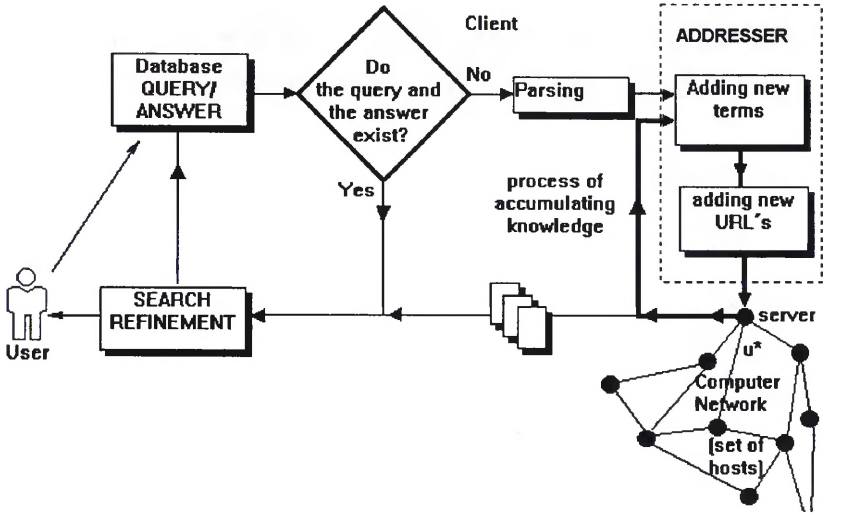


Fig. 1. Conceptual architecture of NetExp

2. THE CONCEPT

As mentioned in the previous Section the problem of formulating an appropriate query and launching it against the Internet resources is not an easy task even to an experienced information officer, without speaking about casual users. Besides technical obstacles connected to the use of the Internet client type tools such as Gopher or WWW, there are two significant reasons why the process of establishing good queries is difficult. The first reason is related to the fact that the user has to use some key-words for expressing his/her needs whatever the query language. At this point the user is never sure whether the terms used for setting up the query are actually those which are well „understandable" by the network. Secondly, the user in general does not know where are the information resources storing the documents which are likely to be relevant to the key-words. To summarise: when sitting in front of the Internet one can hardly formulate a „sharp" query and readily locate the pertinent databases.

The core idea of the proposed approach facilitating the process of establishing queries and efficiently accessing the information resources distributed across the Internet is to create a self-learning mechanism supporting the queries' establishment and forwarding them to the right places within the network. The mechanism, called *ADDRESSER,* has been conceived as a simple quasi

thesaurus which is composed of terms, URL addresses related to the terms and relations linking the terms. We assume that each Internet data server has its signature which is simply the sets of key-words characterising the information resources and/or documents placed on the server. The ADDRESSER is the main component of the front-end processor, called NetExp , whose structure is depicted in Fig.1.

Roughly speaking the process of preparing and launching the query looks as follows (see Fig.1):

1. A Boolean query is set up by the user. The user can either use key-words or pick up the terms from the quasi thesaurus.

2. The query is compared with the entries in the so called QUERY database where all the queries which have already been submitted have been registered along with the answers. Should the input query match one of the items kept in the Query database, the answer is picked up and immediately delivered to the user without accessing the Internet. Now, it is up to the user whether the query should be addressed to the Internet, hoping that additional (updated) information might be found, or not.

3. If there is nothing like the input question in the QUERY database, it is parsed and transferred to the *ADDRESSER* for semantic tuning and determining the set of URL addresses of the resources where the relevant information is expected to be available.

4. The servers indicated by the set of URL addresses are accessed from the user's server (marked as u*), and search is being done by the standard Internet tools. The documents found are moved back to NetExp , and after some refinement (e.g. formatting, editing) transferred to the user.

5.Having identified the pertinent servers on the Internet, any information in form of terms on the contents of the files residing on those servers is picked up and sent back to the ADDRESSER and added to the quasi thesaurus. This information is taken from the servers' signatures. This process can be considered as teaching the thesaurus.

3. PREPARATION FOR SEARCHING

The prototype of the front-end processor was implemented by means of the Toolbook ver.3.0 software. In this paper we shall not discuss the NetExp architecture in detail. We shall rather focus on how the preparation of the search process is performed by the NetExp and how the quasi thesaurus can learn from the network.

The user can formulate a question as a Boolean formula. Let us take an example. The query is:

POLAND ***AND*** *INFORMATION SCIENCE*

No doubt the query is made up of two terms:

term 1: POLAND
term 2: INFORMATION SCIENCE

Now, the process goes to the *ADDRESSER*. If the term which occurs in the query exists in the *ADDRESSER,* all the URLs related to this term are taken into account. If the NetExp cannot find any URL related to the term, it tries to find all synonyms of this term. Should the synonyms not exist or missing URLs are, the broader terms are looked for. Again, if these terms cannot be found or URLs are not specified the system seeks narrower terms. For instance, if the URLs for POLAND do not exist in the quasi thesaurus, the synonyms are considered

synonyms:
RZECZYPOSPOLITA POLSKA
POLSKA RZECZYPOSPOLITA LUDOWA
PRL

If there is not any URL for those synonyms, NetExp is looking for URLs of broader and narrower terms, which in our experimental quasi thesaurus were as follows:

broader terms:
EASTERN EUROPE
WARSAW PACT
narrower terms:
MAZOVIA
SILESIA
WARSAW

Thus, the following terms along with corresponding URL addresses were found:

:

TERM_TEMP	ADRES_TEMP
EASTERN EUROPE	http://148.81.213.3/info_4.htm
WARSAW PACT	http://148.81.213.3/info_2.htm
SILESIA	http://148.81.213.3/info_2.htm
WARSAW	http://148.81.213.3/info_2.htm
MAZOVIA	http://148.81.213.3/info_4.htm

For the term INFORMATION SCIENCE the system found 4 URL addresses:

http://148.81.213.3/info_2.htm
http://148.81.213.3/info_4.htm
http://148.81.213.3/info_5.htm
http://148.81.213.5/info_6.htm

Since our input query was formulated as a conjunction we get as a result the following table:

http://148.81.213.3/info_2.htm
http://148.81.213.3/info_4.htm

Now, the front-end processor can access the servers/files specified in the above table and get the relevant documents.

4. INTELLIGENT QUASI THESAURUS *(ADDRESSER)*

As it has been emphasised the intelligent quasi thesaurus is a core of the NetExp. It has the ability for accumulating knowledge about the network. Now, let us define the intelligent quasi thesaurus in a more formal manner. By SYG(u) we understand the signature of the server u, which is a set of all the terms indexing the files sitting on this server. Now, we can define the *ADDRESSER* as an ordered quadruple:

$$\mathbf{ADR=(T, A, \{B,N,S\}, t)}$$

where:

T - is a set of terms such that $T \neq \emptyset$;

A - is a set of URL addresses

{B,N,S} - sets of broader and narrower terms, and synonyms, respectively, in **T**

τ - is a relation $\tau \subseteq T \times 2^A$ such that for the server u with which URL addresses **A´** are assigned the following condition holds

$$\forall \tau \in T \; \forall A' \subseteq A \; \forall a \in A' \; (t \tau A' \rightarrow t \in SYG(u))$$

As a spin effect of the searching process, signatures of all the accessed servers are brought back to the front-end processor where the terms are added to the quasi thesaurus along with their URL addresses. The action of incorporating the terms into the quasi thesaurus and establishing the relations between them and the existing terms is assumed to be done by a person who is administering the NetExp. This is how the quasi *ADDRESSER* accumulates, or in other words learns, knowledge about the resources available on the network. Note that this process is by its nature monotonic since the knowledge can only increase or remains the same after any search (transaction) within the network. Fig.2 shows the learning curve of the quasi thesaurus; it also displays the fact that the knowledge about the network is finite.

It has to be stressed that before incorporating the quasi thesaurus to the front-end processor in question, a critical mass of knowledge has to be injected into the quasi thesaurus by the front-end processor administrator. One has to accept that it is impossible to start any kind of searching and learning without certain prior knowledge.

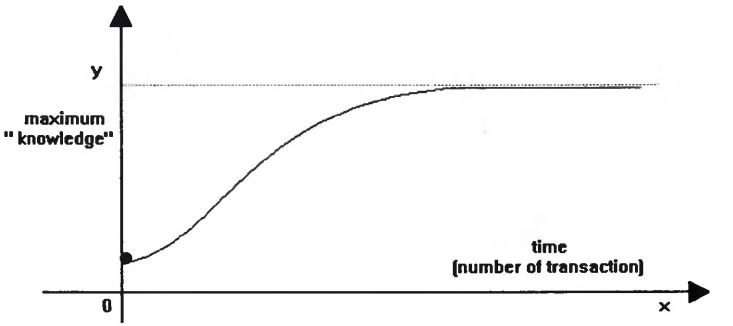


Fig. 2. Monotonic accumulation of knowledge by the quasi thesaurus

When designing the *ADDRESSER* we have encountered an interesting quantitative problem related to the approximation of original terms as given in the query submitted by the user. In other words, the issue is: how good is the approximation proposed by the quasi thesaurus. We have tackled this problem in the following manner. Let us denote the accuracy coefficient related to the term *t* as Θ_t. For Θ_t, the following assumptions were adopted:

1. $0 \le \Theta_t \le 1$
2. l is a number of all URL found of broader and narrower term (related to term t)

$$\lim_{l \to \infty} \Theta_l = 1$$

3. $\Theta_t \mid_{t=0} = 0$

At the beginning of our xperiments we decided to use the following function:

$$\Theta_l = \frac{\sqrt{(l-1)^2} - 1}{l-1}$$

However, during the experiment it turned out that the function was not a good one, so we devised another function (see Fig.3):

$$f(i) = 1 - \left[(1-\varepsilon)e^{\frac{1-i}{T}}\right]$$

where ***i*** is the number of URLs and the parameters take the following values:
ε **-0.1**, it means that for i=1, f(i)=0.123.
T - 100 , this factor is responsible for the function growth pace rate.

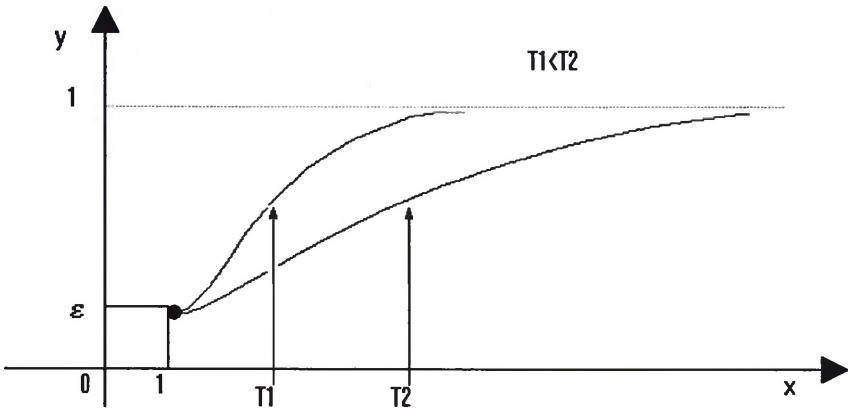


Fig. 3. Function $f(i) = 1 - \left[(1-\varepsilon)e^{\frac{1-i}{T}}\right]$

Thus, the function looks like

$$f(i) = 1 - \left[(1-0.1)e^{\frac{1-i}{100}}\right]$$

The coefficient of accuracy for the whole query composed of n terms is defined as follows:

$$\Theta_w = \frac{\sum_{i=1}^{n} \Theta_{t_i}}{n}$$

This measure proved to be a fair tool for estimating the approximation of original queries during our experiments with the front-end processor.

5. FINAL REMARKS

The paper constitutes an attempt to facilitate the process of interacting with large, geographically distributed, heterogeneous networks such as the Internet. In particular, in this context formulating queries is an error prone and difficult exercise. The work on the front-end processor, which has been implemented as an experimental project, proved usefulness of it for students and casual users. It is believed it may also be useful.

Two issues deserve further elaboration, namely:

- tuning of the accuracy function which is used when some terms provided in an input query are not covered by the quasi thesaurus;
- how large should be the prior knowledge to be given to the quasi thesaurus before starting its operation, and how the amount of „critical mass" of terms depends on the subject matter covered by the front-end processor.

Another important aspect is whether it is better to devise a general quasi thesaurus encompassing many domains or to maintain an array of narrow discipline oriented thesauri governed by a kind of a meta-thesaurus. These problems are addressed in the follow-up project that has just been started.

6. REFERENCES

[ENG95] Engst A.C, Low, C.S., Simon M. A., Internet Starter Kit, Hayden Books, 1995.

[LIU94] Liu C, Peek J., Jones R., Buus B., Nye A., Managing INTERNET Information Services, O'Reilly & Associates, Inc., 1994.

[MAT95] Mathiesen M., Marketing on the Internet, Maximum Press, 1995.

SUMMARY

Host computers in the Internet do not have indexes of their own information resources yet. This is one of the reason why the users can hardly find the information satisfying their requirements. NetExp is an intelligent expert system with automatic generation of inference rules which could assist users in retrieving information from large databases. The knowlwdge accumulated within the system is used for database selection, navigation, search formulation and search refinement. The prototype of this system is based on the object oriented methodology. Prospective application of the NetExp seems to be very useful in the process of integrating the various information resources aviable in the Internet. In particular it might help integrating thesauri and classification systems used in the United States and European countries.

LIST OF PARTICIPANTS

Dr Wiesław BABIK, Uniwersytet Jagielloński, Kraków, Polska
Dyr. Roman BENEDYKCIUK, Fundacja Naumanna, Niemcy
Mgr Lucyna BIELICKA, Biblioteka Narodowa, Warszawa, Polska
Mgr Elżbieta BORAJKIEWICZ, Akademia Rolniczo-Techniczna, Olsztyn, Polska
Dr Lech BOROWSKI, Instytut „Orgmasz", Warszawa, Polska
Mgr Piotr CELIŃSKI, Polska Agencja Prasowa, Warszawa, Polska
Mgr Barbara CHELT, Akademia Rolniczo-Techniczna, Olsztyn, Polska
Dr Ewa CHMIELEWSKA-GORCZYCA, Biblioteka Sejmowa, Warszawa, Polska
Dr Ingetraut DAHLBERG, ISKO President, Frankfurt, Niemcy
Ms Daniele DEGEZ-VATAIRE, Francja
Ms Stella G. DEXTRE CLARKE, CAB Int., Wielka Brytania
Prof. Marcin DRZEWIECKI, Uniwersytet Warszawski, Warszawa, Polska
Mgr Halina DUSIŃSKA, Główna Biblioteka Lekarska, Warszawa, Polska
Mgr Radosława DZIUBECKA, „T. Chlebowski", Warszawa, Polska
Prof. Jan GLIŃSKI, Francusko-Polska W. Szk. Nowych Technik Informatyczno-
-Komunikacyjnych, Poznań, Polska
Mgr Wiesław GLIŃSKI, Uniwersytet Warszawski, Warszawa, Polska
Dyr. Małgorzata GOGOŁEK, WIP, Warszawa, Polska
Prof. Włodzimierz GOGOŁEK, Prezes TIP, Warszawa, Polska
Red. Małgorzata GUTTMAN, PAP, Warszawa, Polska
Dr Stefan HOPPE, Inst. for Medical Informatics, Neuherberg, Niemcy
Mrs Wiltrud HOPPE, Neuherberg, Niemcy
Mgr inż. Barbara KOPER, OPI, Warszawa, Polska
Horst G. KÖRNER, Ms, ABD InfoConsult, München, Niemcy
Mgr Agnieszka KOZŁOWSKA, Uniwersytet Łódzki, Polska
Dr Jan KOZŁOWSKI, KBN, Warszawa, Polska
Mgr Andrzej LIBERA, Polska Agencja Prasowa, Warszawa, Polska
Dr Helmut LOCKENHOFF, ISKO Schatzmeister, Niemcy
Mgr Teresa ŁAGODZIŃSKA-MAŁYSZKO, Wyższa Szkoła Pedagogiczna, Olsztyn,
Polska
Mgr Elżbieta MAŃKOWSKA, PFUN, Warszawa, Polska
Mgr Marzena MARCINEK, Politechnika Krakowska, Kraków, Polska
Dr Katarzyna MATERSKA, Uniwersytet Warszawski, Warszawa, Polska
Dr Giliola NEGRINI, CNR, Roma, Włochy
Mgr Barbara NIEDŹWIEDZKA, Uniwesytet Jagielloński, Kraków, Polska
Mgr Irma PĘCIAK, CINP, Warszawa, Polska
Red. Piotr PŁÓCIENNIK, „INFORMACJA PROFESJONALNA" , Poznań, Polska
Dr Hanna POPOWSKA, Biblioteka Sejmowa, Warszawa, Polska

Mgr Zofia POWIDEŁ, Główna Biblioteka Lekarska, Warszawa, Polska
Drs. Gerhard RIESTHUIS, Univ. van Amsterdam Bibliotheek, Holandia
Dr Courentin ROULIN, BJL Consult S.A., Bruksela, Belgia
Mgr inż. Renata RYBARCZYK, OPI, Warszawa, Polska
Prof. Henryk RYBIŃSKI, ICIE, Warszawa, Polska
Mgr Bogumiła RYKACZEWSKA, Uniwersytet Warszawski, Warszawa, Polska
Dr Jadwiga SADOWSKA, Biblioteka Narodowa, Warszawa, Polska
Dr Tomáš SAMEK, Parliament of Czech Republic, Praha, Czechy
Prof. Winfried SCHMITZ-ESSER, DGD/KTF Hamburg, Niemcy
Mgr inż. Krystyna SIWEK, Koordynator ISKO-Poland, Warszawa, Polska
Prof. Dagobert SOERGEL, University of Maryland, MD, USA/Niemcy
Dr Barbara SOSIŃSKA, Uniwersytet Warszawski, Warszawa, Polska
Dr Pavla STANIKOVÁ, CEIT, Bratislava, Słowacja
Mgr Andrzej STYCZYŃSKI, Centralna Biblioteka Statystyczna, Warszawa, Polska
Mgr Beata SZAFRANEK, Uniwersytet Łódzki, Polska
Dr Wiesław SZCZEPANEK, Fundacja Biblioteki Narodowej, Warszawa, Polska
Dr Jan SZYDLAK, OPI, Warszawa, Polska
Dr Janusz ŚACH, Polska Akademia Nauk, Warszawa, Polska
Prof. E. ŚCIBOR, Instytut INTE, Warszawa, Polska
Mgr Barbara ŚWIĘCICKA, Centralna Biblioteka Statystyczna, Warszawa, Polska
Mgr Marek TALACHA, Polska Agencja Prasowa, Warszawa, Polska
Mgr Anna URYGA, Uniwersytet Jagielloński, Kraków, Polska
Dyr. Wacław WASIAK, OPI, Warszawa, Polska
Mgr Dariusz WICHNIEWICZ, ATM, Warszawa, Polska
Mgr Irena WILD-KAMIENIECKA, OPI, Warszawa, Polska
Dr inż. Marek WITCZAK, OPI, Warszawa, Polska
Dr Jadwiga WOŹNIAK, Uniwersytet Warszawski, Warszawa, Polska
Prof. Harold ZIMMERMANN, Universität des Saarlandes, Saarbrücken, Niemcy

COMPATIBILITY AND INTEGRATION OF ORDER SYSTEMS 1960-1995
An Annotated Bibliography
compiled by Ingetraut DAHLBERG

Introductory note:
A survey on what had already been published on the problems of compatibility and integration of order systems was deemed timely when planning for the Seminar in Warsaw, Sept.13-15, 1995 on this topic. Thus a preliminary bibliography had been compiled in alphabetical order before the event and could be supplemented by still very many titles after the Seminar. We feel, however, that there are still titles missing before the list could be regarded as final. Nevertheless we would like to publish this list in the proceedings volume, again in alphabetical order for easier identification of missing items before a separate publication in systematic arrangement with an author and a subject index will be issued.
Whoever will still notice a missing title is invited to contact the compiler and to help in making the list complete. Please write to I.Dahlberg at Woogstr.36a, D-60431 Frankfurt, Germany.
The notation, foreseen to sort the systematic display is given on the right side above each title. It is taken from Section 28 of the outline *of Knowledge Organization Literature* in the issues of the journal *Knowledge Organization*. The numbers denote the following subdivisions:

28 Compatibility and Concordance between Indexing Languages (IL)

28.016 Bibliographies on Compatibility
28.06 Conferences on Compatibility
28.07 Textbooks containing sections on Compatibility
28.08 Other monographs with such sections

281 Objectives and nature of systems compatibility
282 Intermediate languages
283 Compatibility in classing and indexing
284 Establishment of concordances
285 Correlative indexes. Mapping
286 Systems reconciliation, e.g. between classification systems and thesauri, linking of terms
287 Organized compiliation of compatible IL
288 Compatibility between systems in subject areas
289 Evaluation of compatibility.

001 **287**
Adam'janc,A.D.: Samojlova,L.M., Fastovich, N.L.: **Technological and software support for conversion of the All-Union Centre's automated information system to the use of All-Union classifiers.** (Orig. ru). In: Soversh. sistemy gos. peristracii NIR i OKR. Ispol'z.fondov VNTICentr. Moskva: 1982, p.86-88
The All-Union STI Centre had to rework its existing organizations and enterprises file in order that its classifiers of organizations and enterprises could be applied to new document forms for reasons of compatibility with other information systems. Description of the conversion process.

002 **287**
Agraev, V.A.: Kobrin, R.Ju., Shul'ts, M.M.: **Information retrieval system compatibility.** (Orig. ru). In: Autom.Doc.Math.Linguist. Vol.8 , No.2 1974, p.29-37, 39 refs.
Discussion of the problem of designing an intermediate language for combined information retrieval systems including thesaurus compatibility and combination of grammars.

003 **288-6**
Aitchison,J.: **Integration of thesauri in the social sciences.** (Orig. en). In: Int.Classif. Vol.8 , No.2 1981, p.75-85, 18 refs.

Objectives, characteristics, and compilation of the proposed integrated thesaurus of the social sciences are discussed. It is not intended to supplant well-established thesauri but to act as a master-reference tool, which could serve as an aid in searching across databases or as a switching mechanism. The first step in compilation should be the merging of terms to form a descriptor bank. This can reveal both matches and inconsistencies, but is limited by its verbal and alphabetical approach. To compare at the concept level needs a classification framework, preferably one with a faceted structure. The classification would serve as the master against which the terms from the merged IL, arranged in a compatibility matrix would be compared and modified.

004 **07.23**
Aitchison, J.: Gilchrist, A.: **Thesaurus construction. A practical manual. 2nd ed.** (Orig. en). London, GB: Aslib 1987, 173p. 109 figs. ISBN: 0-85142-197-0
This complete guide to the construction of thesauri for use in information retrieval covers also the special problems of multilingual thesauri. It also examines developments resulting from the use of thesauri in online retrieval, such as the merging and integration of thesauri and the design of searching thesauri.

005 **288-262**
Anan'eva, L.I.: Galina, M.B. et al.: **On automating nomenclature translation.** (Orig. ru). In: Vopr.inform.teorii i praktiki , No.32 1978, p.12-1, 17 refs.
A Name Structure automatic translation algorithm for the translation of the names of chemical compounds into the corresponding structural formulae and vice versa was developed embracing the items of several nomenclatures current in chemical publications.

006 **288-46**
Andre, P.Q.J.: **Universal agricultural thesaurus discussed.** (Orig. en). In: Quart.Bull.IAALD Vol.34 , No.3 1989, p.150--151

007 **286**
Andrian, A.C.: **Once more about the unificationof the UDC and the DDC.** (Orig. ro). In: Probl.de Inform.si Doc. Vol.26 , No.3 1992, p.155-167, refs.

008 **281**
Angell, R.S.: **Compatibility in subject access vocabularies: the role of relations between indexing terms.** (Orig. en). In: VINITI International Forum on Informatics, Vol.2 Moskva: VINITI 1969, p.243-261

009 **282**
Antopol'skij, A.V.: **ASTIS language apparatus. Structure and the problem of compatibility.** (Orig. (ru), en). In: Autom.Doc.& Math.Linguist. Vol.17 , No.4 1983, p.29-44, 64 refs.

010 **285**
Antopol'skij, A.V.: Salij, A.D., Fedosimov, V.I.: **A study of automatic methods for identification of lexical overlaps among information retrieval thesauri.** (Orig. ru). In: Nauchn.-tekhn.inform.Ser.2 , No.2 1979, p.14-17, 7 refs.

011 **288-6**
Artowicz, E.: **BSO - MISON Rubricator comparisons in Poland.** A report prepared for the WG3 Meeting in Prague,9-11 Sept.1980. (Orig. en). Prague: ECSSID WG3 1980, 19p. 2677

012 **281**
Artowicz, E.: **Possibilities for information languages integration with special reference to a descriptor language and a subject catalogue language.** (Orig. pl). In: Zag.inform.nauk. , No.1 1985, p.47-68, 9 refs.

013 **281**
Atanasiu, P.: Bazareson, G.: **Vergleichende Analyse von Thesauri und Möglichkeiten ihrer Vervollkommnung.** Comparative analysis of thesauriand possibilities for their improvement. (Orig. de). in: ZIID 1966.6 , p.87-109

014 **337**
Austin, D.: **The PRECIS system for computer-generated indexes and its use**

in the British National Bibliography. (Orig. en). In: Wellisch et al 1972 , p.99-115

015 **283**
Austin, D.: **Compatibilité, aux fins d'indexage, des informations sur la recherche et des ouvrages bibliographiques.** Compatibility for purposes of indexing of information on research and bibliographical works. (Orig. fr). In: Colloq. Int. Syst. Serv. Inform. concernant Rech. Sci. Cours. Paris: UNESCO 1975

016 **286**
Austin, D.: **Automation in subject cataloguing in the Bibliographic Services Division of the British Library.** (Orig. en). In: LIBER , No.21 1983, p.24-36 3 refs.
Among others the article contains a discussion of the compatibility problems between the Library of Congress Classification and Blaise.

017 **288-69**
Bandur, G.: **Nutzorientierte Erschliessung von Archivgut und die Anwendung des Geschichtswissenschaftlichen Thesaurus.** User-oriented analysis of archive materials and the utilization of the history thesaurus. (Orig. de). In: Archivmitteilungen Vol.27 , No.3 1977, p.106-107 2 refs.
Problems of compability between the history thesaurus and the one for the public state organs of the GDR are discussed.

018 **287**
Barth, C.: **Problems in the development of a set of compatible IR languages for the International Scientific and Technical Information System.** (Orig. de). In: DDR--Med.Rept. Vol.4 , No.2 1975, p.96-100

019 **287**
Bauer, G.: **Gedanken zur Theorie des Aufbaus und der Funktionsweise der in einemm Thesaurussystem zusammengefaßten Thesaurusarten.** Thoughts on the theory of the structure and functions of the different types of thesauri comprised in a thesaurus system. (Orig. de). In: ZIID 1966.6, p.1-52

020 **288-26**
Bauer, G.: Manecke, M., Schoenberg, D.: **Die Anwendung des Prinzips der Facettenklassifikation beim Aufbau des Thesaurussystems Chemie. 1. Teil.** Application of the principles of faceted classification in the construction of the thesaurus. (Orig. de). Berlin: Dt. Akad. d. Wiss., Forsch.Gr.TSC 1967

021 **288-26**
Bauer, G.: **Zur Methodik des Aufbaus koordinierfähiger Fachthesauri im Rahmen des Thesaurussystems Chemie.** On the methodology of the construction of compatible special thesauri in the framework of the thesaurus system chemistry. (Orig. de). In: ZIID 1967.4 , p.207-209

022 **281**
Bauer, G.: **Die Bedeutung der Kategorien als Ordnungsmittel in Thesauri und Speichern der elektronischen Datenverarbeitungsanlagen.** The importance of categories as ordering devices in thesauri and in computer storagek. (Orig. de). In: ZIID-Z. Vol.15 , No.2 1968, p.61-65

023 **288-26**
Bauer, G.: **Zur geeigneten Begriffsordnung im Thesaurussystem Chemie.** On the suitable conceptual structure in the thesaurus system chemistry. (Orig. de). In: Informatik Vol.16 , No.2 1969, p.35-41
The article is in two parts: Pt.1 treats Concepts not from structural chemistry; Pt.2 (which appeared in Informatik (1969)No.5, p.11-16 treats Concepts from structural chemistry.

024 **08.414**
Beck, H.: **Klassifikation und Informationswiedergewinnung. Zu aktuellen Problemen aus bibliothekarischer Sicht, Teil 5.** Classification and information retrieval. On pertinent problems from a librarian's point of view, Pt.5. (Orig. de). In: Zbl.Biblwesen Vol.92 , No.1 1978, p.6-17 30 refs.
Considers the problem of the uniform classification and the unification of subject cataloguing.

025 **286**
Beckett, R.: **The integration of nomenclature and classification: A collaboration of clinical experience, logic and computer science.** (Orig. en). In: Medical Informatics Vol.8 , No.1 1983, p.47--53

026 **282**
Belikova, A.Ya.: Zajtseva, O.V., Mironova, N.I.: **On an international information retrieval language.** (Orig. ru). In: Automatic text processing by methods of applied linguistics. Proc.Conf. 1977. Kisinev, RU: 1977, p.130-131

027 **282**
Beling, G.: Wersig, G.: **The new concept of an intermediary language system for information networks.** (Orig. en). In: In: Batten,W.E.(Ed.) EURIM 2. Proceedings, Amsterdam, 23-25 March 1976 London: Aslib 1977, p.117-121, 7 refs.
The Concept of an intermediary system is introduced which must fulfill the following functions: Orientation, model, standardization, derivation, selection, terminological control and mediation. Based on these considerations a concept for an intermediarysystem was designed, a Federal Thesaurus System in Germany with its subsystems.

028 **287**
Beloozerov, V.N.: **Integration of scientific-technical vocabulary collections.** (Orig. ru). In: Osnov.napravleniya razvitiya i soversh.rabot po standartiz.nauch.--tekhn.terminol. v XI pyatiletke Moskva: 1983, p.18-21

029 **286**
Beloozerov, V.N.: Nosikov, S.M., Fedosimov, V.I.: **Compatibility between the IR thesauri of GASNTI agencies.** (Orig. ru). In: Probl.sozdaniya retrosp.poisk. massivov v avtomatiz.centrakh NTI. Ch.2 Moskva: 1985, p.18-21

030 **288-26**
Berg Hansen, I.: **Subject compatibility between chemical abstracts subject sections and search profiles used for computerized information retrieval.** (Orig. en). In: J.Chem.Doc. Vol.12 , No.2 1972, p.110-113, 5 refs.

031 **284**
Bernhard, F.: Reul, H., Schulte-Tigges, F., Sunkel, H.: **Erstellung von Konkordanzen zu Sanskrit-Texten durch elektronische Rechenanlagen.** Establishing concordances to Sanskrit texts with computers. (Orig. de). In: Linguistics , No.22 1966, p.5--23

032 **285**
Bernier, C.L.: **Correlative indexes I: Alphabetical correlative indexes.** (Orig. en). In: Amer.Doc. Vol.7 , No.4 1956, p.283--288

033 **285**
Bernier, C.L.: Heumann, K.F.: **Correlative indexes. III. Semantic relations among semantemes and the technical thesaurus.** (Orig. en). In: Amer.Doc. Vol.8 , No.3 1957, p.211-220

034 **285**
Bernier, C.F.: **Correlative indexes: V. The blank sort.** (Orig. en). In: Amer.Doc. Vol.9 1958, p.32-41

035 **285**
Bernier, C.L.: **Correlative indexes VI: Serendipity, suggestiveness and display.** (Orig. en). In: Amer.Doc. Vol.11 , No.4 1960, p.277-287

036 **285**
Bernier, C.L.: Dyson, C.M., Friedman, H.J.: **Correlative indexes VII: Trope vocabularies and trope indexes for chemistry.** (Orig. en). In: J.Chem.Doc. Vol.2 1962, p.93-102, 7 refs.

037 **287**
Bielicka, L.A.: **Fundamentals of development of disciplinary and subdisciplinary thesauri for SINTO.** (Orig. pl). In: Aktual.probl.inform.i dok. , No.3 1979, p.47--48
Description of 16 thesauri developed or to be developed within a common national programme in Poland. They should be coordinated and brought together under a roof thesaurus together with a General Problem Thesaurus to include terms in

economics, labor organization, and hygiene, the science of science, scientific information, patents, and standardization.

038 **281**
Bielicka, L.A.: Paciejewski, J., Scibor, E.: **Problems of compatibility and methods for its attainment in information languages.** (Orig. pl). In: Or./IINTE , No.50 1984, p.1-47, 35 refs.

039 **288-51/4**
Blanken, R.R.: Stern, B.T.: **Structuring the Excerpta Medica Thesaurus via a polyhierarchic decimal classification.** (Orig. en). In: In: Batten,W.E.(Ed.) EURIM II. 23-25 March 1976, Amsterdam. London: Aslib 1977, p.148-150
Considers the problems of linking frequently used MALIMET terms with particular subcategories in the Excerpta Medical Classification (MALIMET = Excerpta Medica Master List of Medical Indexing Terms).

040 **289**
Bishop, C.W.: **Alternate approaches to a UMLS.** (Orig. en). In: Med.Decision Making , No.4 (S) 1991, p.99-102

041 **287**
Blankenstein, G.: **Erarbeitung von Thesauri unter Berücksichtigung der zukünftigen Entwicklung.** The construction of thesauri with a view to future developments. (Orig. de). In: ZIID-Z. Vol.14 , No.6 19967, p.177-180

042 **281**
Blumstengel, R.: Schüler, W.: **Paßfähigkeitsprobleme beim Übergang zu neuen Formen der Patentinformationsversorgung.** Compatibility problems in the transition to new forms of patent information. (Orig. de). In: Informatik Vol.22, No.6 1975, p.29-33, 6 refs.

043 **288-235**
Bojko, N.V.: Polozenceva, G.I.: **Structural and semantic compatibility of bibliographical data record formatsk.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.5 1980, p.17-25, 9 refs.
The record formats of INIS and INSPEC were analyzed in order that a domestic databank on thermophysical properties of matter may be supplemented by bibliographical data provided by these two systems.

044 **285**
Bonczek, R.H.: Whinston, A.B.: **A generalized mapping language for network data structures.** (Orig. en). In: Inform.Syst. Vol.2 , No.4 1977, p.171-185, 10 refs.

045 **286**
Borisova, N.D. et al.: **Some methodological principles behind the design of a table of correpondences between the GASNTII Subject Authority and the UDC.** (Orig. ru). Moskva: VINITI GKNT AN SSSR 1985, 363p.
Discussion of compatibility problems and presentationof methodologicalp principles underlying the design of a table of correspondences between the Subject Authority and the UDC as well as suggested UDC numbers for Subject Authority concepts. The appendix lists the tables of correspondences.

046 **284**
Buchan, R.L.: **Intertwining thesauri and dictionaries.** (Orig. en). In: Information Serv. & Use Vol.9 , No.3 1989, p.171-175 17refs.
Reports about the practice of combining a thesaurus with a dictionary by adding a definition part with definitions taken from different sources. The NASA Thesaurus contains over 3.300 definitions for its ca 17.000 terms.

047 **287**
Bukhtijchuk, N.A.: Mikhajlova, G.V., Yakovich, L.F.: **The problem of creating a unified coding scheme for STI agencies and libraries.** (Orig. ru). In: Probl.soversh.i razvitiya avtomatizir.inform.bibl.sistem GPNTB SSSR v 12 pyatiletke Moskva: 1987, p.98-107, 4 refs.

048 **289**
Burgun, A. et al.: **Designing a sub-set of the UMLS knowledge base applied to a clinical domain: methods and evaluation.** (Orig. en). In: Proc.Ann.Symp. Comput.

Appl.Med.Care 1994 Rennes, FR: Lab d'Inform.Medicale 1994, p.968

049 **284**
Busa, R.S.J.: **Rapidissima composizione di indice e concordanze di parole mediante schede perforate.** Quick compilation of indexes and word concordances with the help of punched cards. (Orig. it). In: La Documentazione in Italia 1952, p.95-97

050 **285**
Cahn, D.F.: Herr, J.J.: **Automatic database mapping and translation methods.** (Orig. en). Berkley, CA: Univ.of California, Lawrence Berkley Lab 1978, 13p.
Automatic methods for content-directed translation and information mapping among machine-readable databases are considered.

051 **288-51/4**
Campbell, J.R.: Payne, T.H.: **A comparison of four schemes for codification of problem lists.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1994. p.201-5 ISSN 0195-4210

052 **288-75**
Canadian thesaurus of construction science and technology TC/CS. (Orig. en). Ottawa, CA: Gov.of Canada. Dept.of Industry & Trade 1978, 220+350p.
The thesaurus is in two parts, alpha--hierarchical and alpha-permuted listing and consists of an English and a French version. Complete compatibility with other Canadian thesauri in this domain is ensured.

053 **288-51/4**
Carenini, G.: Moore, J.D.: **Using the UMLS Semantic Network as a basis for constructing a terminological knowledge base: a preliminary report.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1993. p.725-9 ISSN 0195-4210

054 **281**
Cavalcanti, C.R.: **Universal integrated media for information processing.** (Orig. en). In: In: Neelameghan, A. Ordering Systems for Global Information Networks. Proc.3rd Int.Study Conf. on Classif.Research, Bombay, 6-11 Jan.1975 Bangalore: Sarada Ranganathan Endowment f.Libr.Sci. 19799, p.

055 **288-6**
Centre for Scient. Inform. Czechoslov. Acad. Sci.: **Comparison of the 3rd version of the Rubricator MISON (MR) and UDC (frequency distribution of parallel classified sets.** (Orig. en). Paper presented at ECSSID WG3 Meeting, Prague, 9-11 Sept.1980 , 17p. 2676

056 **282**
Chamis, A.Y.: **The usefulness of switching vocabularies for online databases.** (Orig. en). In: ASIS'85. White Plains, NY. Vol.22 1985, p.311-314, 4 refs.

057 **282**
Chamis, A.Y.: **Selection of online databases using switching vocabularies.** (Orig. en). In: J.Amer.Soc.Inform.Sci. Vol.39 , No.3 1988, p.217-218, 3 refs.
A thesaural relationship model was developed to measure the degree of compatibility and switching capability of a thesaurus or of vocabularies, designated as compatibility and switching values (CSV). Groups of terms were matched with VSS (Vocabulary Swiching Systems). States that switching vocabularies have a great potential for facilitating the selection of appropriate search terms.

058 **07.752**
Chamis, A.Y.: **Vocabulary control and search strategies in online searching.** (Orig. en). New York: Greenwood Press 1991

059 **282**
Chanzhin, A.G.: **Some peculiarities of a descriptor IR language for large integral information systems.** (Orig. ru). Moskva: VINITI 19977, 20 p., 7 refs.

060 **286**
Chaplan, M.A.: **Mapping Laborline Thesaurus terms to Library of Congress Subject Headings: Implications for vocabulary switching.** (Orig. en). In: Libr.Quart. Vol.65 , No.1 1995, p.39-61, refs.
Terms from the Laborline Thesaurus were manually mapped to LCSH and an

INMAGIC database for the map was created in order to permit analysis of the patterns of matches. It was estimated that realistically 41.6 percent of the terms could be successfully switched automatically using currently proposed or available strategies

061 **281**
Chernyi, A.I.: **Information retrieval languages. Types, design principles, compatibility.** (Orig. en (trsl.)). In: Autom. Doc.& Math.Linguistics Vol.12 , No.1 1978, p.1-13 26 refs.

062 **286**
Chiapetti, F.S.: Serrai, A.: **Dewey, classification scheme and the semantic universe.** (Orig. en). In: Int.Classif. Vol.3 , No.1 1976, p.7-11, 11 refs.
Report of work toward reconciliation of the UDC with thesauri and exemplification of a thesaurus-like UDC in philosophy.

063 **288-51/4**
Cimino, J.J.: Johnson, S.B., Peng, P., Aguirre, A.: **From ICD9-CM to MeSH using the UMLS: a how-to guide.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1993. p.730-734 ISSN 0195-4210

064 **284**
Chikawa, S.: Mutsudo, Y.: **Changeover to standard classification schemes at Sophia University Library.** (Orig. ja). In: Annals of Japan Soc.Libr.Sci. Vol.28 , No.3 1982, p.110-116

065 **282**
Coates, E.J.: **Switching languages for indexing.** (Orig. en). In: J.Doc. Vol.26 1970, p.102-110

066 **283**
Colombo, D.S.: Niehoff, R.T.: **Design and development considerations during implementation of an online subject switching system.** (Orig. en). In: In: Martin, S.K. Information politics. Proc.ASIS 39th Ann.Meeting, San Francisco, CA, Oct.4-9, 1976 Washington, DC: Amer. Soc. Inform. Sci. 1976, p.116-

067 **286**
Comparison of UDC with descriptor systems. (Orig. en). In: FID News Bull. Vol.17 , No.10 1967, p.106

068 **281**
Compatibility in information systems. (Orig. en). In: Teaching package on standardization in information handling. Module 5 Paris: UNESCO 1991, p.(5)1-34

069 **284**
Department of Commerce et al: **Concordance between the Standard Industrial Classifications of the U.S. and Canada.** (Orig. en). Washington, DC: Dept.of Commerce, Bureau of the Census 1991?, 262p. PB91-155291ADR
Concordance between the 1987 U.S. SIC and the 1980 Canadian SIC (Standard Industrial Classification). From the Canadian side Statistics Canada collaborated.

070 **284**
Institute of Electrical Engineers: **Concordance to the INSPEC Classification 1969--1977.** (Orig. en). London: Inst.of Electrical Enggs. 1977, 88p.

071 **288-822**
Croghan, A.: **A thesaurus-classification for the physical forms of media.** (Orig. en). London,GB: Antony Croghan 1970, 47p.

071a **281**
Dahlberg,I.: **Major developments in classification.** (Orig.en). In: Advances in Librarianship Vol.7 1977, p.41-103 179refs.
Among others the problems of compatibility of thesauri are discussed.

072 **281**
Dahlberg, I.: **Einheitsklassifikation.** Uniform classification. (Orig. de). In: Alsheimer, R. (Ed.) Bestandserschließung und Bibliotheksstruktur. Rolf Kluth zum 10.2.1979 Wiesbaden: Harrassowitz 1979, p.31-49, 27 refs.
Treats the presuppositions for a uniform classification system which lead to a general compatibility between CS.

073 **282, 288-6**
Dahlberg, I.: **The Broad System of Ordering (BSO) as a basis for an integrated social sciences thesaurus?.** (Orig. en). In: Int.Classif. Vol.7 , No.2 1980, p.66-72
The SS coverage of the BSO was found to be relatively small compared with the corresponding classes of the DDC, the Bliss Classif. and the UNESCO Thesaurus. The BSO falls short also in its structural features. Not having been designed for the purpose ofintegrating other systems, it cannot be blamed for its non-suitability for such an application purpose.

074 **281**
Dahlberg, I.: **Guidelines for the establishment of compatibility between information languages.** (Orig. en). Frankfurt: 1980, 25p.
Project report prepared for the UNESCO Div.for the International Development of the Social Sciences. Published in Int.Classif. 8(19819No.2, p.86-91, 31 refs. under the title: Toward establishment of compatibility between indexing languages. Contents: Outline of previous work done in the field of compatibility between IL and description of the scope, limitations and advantages of establishing compatibility between IL. Methods for verbal and conceptual comarisons are outlined together with the generationof compatibility matrices. Description of their use and of organizational problems.

075 **281**
Dahlberg,I.: **Conceptual compatibility of ordering systems.** (Orig. en). In: Int.Classif. Vol.10 , No.1 1983, p.5-8, 16 refs.
On the basis of the guidelines of 1981 the section concerning conceptual work is treated in more depth taking into account the new developments of the past years toward thesaurus integration and integration computerization.

076 **28.016**
Dahlberg, I.: **Classification and indexing systems. Theory, structure and methodology (1950-1982).** (Orig. en). Frankfurt/Main: INDEKS Verlag 1985, 223p. ISBN: 3-88672-302-X
Among the ca 5600 references listed there are 212 on problems of compatibility, arranged under 28 and its subdivisions.

077 **286**
Davis, C.H.: **Integrating vocabularies with a classification scheme.** (Orig. en). In: Amer.Doc. Vol.19 , No.1 1968, p.101

078 **289**
DeHart, F.E.: **Topic relevance and BSO switching effectiveness.** (Orig. en). In: Int.Classif. Vol.9 , No.2 1982, p.71-76, 21 refs.
Report of a case study which attempted to obtain further insight into topic relevance, a concern of indexers, searchers, and those who construct thesauri and ordering schemes. 'Topic relevance' is distinguished from 'relevance to a user's need'. The experiment aimed at determining whether statistically significant differences existed in the proportion of matches against fifty articles on the topic and present in three information sources. The study also set out to examine the potential usefulness of the BSO as a switching mechanism on the Section Heading level relative to the topic. It was found that topic relevance may be a more serious problem in switching effectiveness than suspected.

079 **287**
Delgado, R.R.: **Unification of documentary languages.** Basis for the creation of common thesauri in Spanish. (Orig. es). In: Reuniber-78. Conf.iberoamer.sobre inform.y doc.cient. y tecnol. Madrid: Inst. Inform. Doc. Cienc. Tecnol. 1978, p.99-111, 22 refs.
The necessity of elaborating a documentary metalanguage which is universal and international is pointed out. Modalities for the generation of Spanish thesauri for many subject fields are proposed.

080 **287**
Delgado, R.: **Plan for development of coordinated thesauri in Portuguese and Spanish.** (Orig. es). In: Rev.Esp.Doc.Sci. Vol.2 , No.4 1979, p.337-355, 6 refs.

081 **287**
Delgado, R.R.: **The integration of documentation languages: the end of Babel.**

(Orig. es). In: Rev.Esp.Doc.Cient. Vol.3 , No.4 1980, p.333-340

082 **286**
Dewey, H.: **The relationship between the headings in the subject catalog and the classification numbers of the books.** (Orig. en). In: Perreault, J.M.(Ed.) Reclassification. Rationale and problems. Proc.Conf.on Reclassif., University of Maryland, April 4-6, 1968 College Park, MD: Univ.of Maryland 1968, p.57-75

083 **288-29**
Deweze, A.: **The realization of a thesaurus: compatibility with foreign systems in the frame of ELDOC.** (Orig. fr). In: 1er Congr. Natl. Franc. Inform. Doc. Commun. Paris 1974. Paris: CNRS 1974, p.335-343, 2 refs.
Short reminder of the objectives and principles of the THESEE (Thésaurus Electricité Electronique), in particular of the semantic constraints of a trilingual thesaurus and examination of its applications in national or multinational documentation systems

084 **286**
D'Haenens, L.: Lorphčvre, G.: **Rapport sur la rédaction d'index thesaurifiés de la Classification Décimale Universelle.** (Orig. fr). Brussels, BE: Fed.Int. de Doc. (?) 1974, 5p.

085 **289**
Diener, R.A.V.: Tsuffis, M.L.: **A method for evaluating the feasibility of the intermediate lexicon concept for existing information retrieval systems.** (Orig. en). In: Inform.Management 1980's. Amer. Soc. Inform.Sci. 40th Annual Meetg., Chicago 1977. White Plains, NY: Knowledge Ind. Publ. 1977, p.28
An intermediate lexicon is only valuable if there is a reasonable level of indexing consistency between centers. A method is proposed for testing the feasibility of the intermediate lexicon concept.

086 **288-9, 289**
Dienes, M.: **Testing the compatibility matrix in the field of culture.** (Orig. en). In: In: Riggs, F.W. The CONTA Conference. Proc.Conf.on Conceptual and Terminological Analysis in the Social Sciences, Bielefeld, May 24-27, 1981. Frankfurt: INDEKS Verlag 1982, p.224-233, 9 refs.

087 **288-9**
Dienes, M.: **Structural differences in classification systems and the testing of the compatibility matrix in the field of culture.** (Orig. en). In: In: Perreault,M. et Universal Classification 2. Subject Analysis and Ordering Systems. Proc.Int.Study Conf.on Classif.Res. 4, Augsburg,28 June - 2 July 1982 Frankfurt/Main: INDEKS Verlag 1982, p.112-125 17 refs.
The classification systems included in the research are the Unesco Thesaurus, the Soviet MISON Rubricator, the UDC and the Broad System of Ordering.

088 **283**
Dimitrova, I.: **Parallel indexing with the MISON Rubricator and BSO applied in Bulgaria.** (Orig. en). Paper presented at ECSSID WG3 Meeting, Prague, 9-11 Sept.1980. , 25p. 2679

089 **284**
Dixon, J.E.G.: **Concordances KWIC and complete. An appraisal.** (Orig. en). In: ALLC Bull. Vol.6 , No.1 1978, p.28-33, 5 refs.

090 **288-12**
Donovan, P.W.: Hunt, D.C., Mack, J.M.: **Mathematics in a major library using the Dewey Decimal Classification.** (Orig. en). In: Austr.Acad.Res.Libr. Vol.6 , No.2 1975, p,87-91, 2 refs.
A concordance was aimed at between the Mathematical Offprint Service (MOS) Scheme of the American Mathematical Society and the DC 18 class Mathematics 510-519, however, it was impossible to relate the MOS classes in a usable way.

091 **285**
Doszkocs, T.E.: **Automatic vocabulary mapping in online searching.** (Orig. en). In: Int.Classif. Vol.10 , No.2 1983, p.78-83, 7 refs.

092 **287**
Dvorakova, H.: **Descriptor-type IR languages and trend towards thesaurus standardization and compatibility.** (Orig.

cz). In: Techn.Knih. Vol.23 , No.10 1979, p.277-284, 19 refs.

093 **28.06.77-09-6/8**
Edinaja sistema informacionnopoickovykh yazykov. Unified system of information retrieval languages: Summaries of papers presented at an All-Union Conference, Yurmala, Latvian SSR, 6-8 Sept.1977. (Orig. ru). Riga: Latv.Universitet 1977, 170p.
Among the 54 papers presented the following are relevant in the context of this bibliography:

094
Vcherashnij, R.P.: Bunova, M.A.: **Compatibility among classificatory retrieval languages.**

095
Giljarevskij, R.S.: **A note on the compatibility between retrieval languages types.**

096
Glushkov, V.M.: Skorokhod'ko, E.F., Stognij, A.A.: **Measuring the compatibility level of document retrieval languages.**

097
Gruzitis, B.E.: Sover, N.B., Morozov, V.A.: **Some design concepts and the structure of a compatible language system of a Republic computer-based STI system.**

098
Danilov, M.P.: Lesochina, E.I., Smirenskij, V.B.: **The social sciences thesaurus system.**

099
Konshin, G.I.: **Contribution to the problem of compatibility between classificatory and descriptor languages.**

100
Konshin, G.I.: Moskalenko, Z.D.: **Information language compatibility problems in the context of an evolving information network in a field.** With special reference to the Geology automated STI system.

101
Kosmacheva, N.V.: Makurina, R.G.: **Vocabulary compatibility of different IR systems in a field.** With special reference to the Electrical Engineering and Source computer-based IR system.

102
Kochetova, S.M.: Murinson, E.A., Shenderov, V.Z.: **The philosophy of interfacing classificatory and descriptor languages.**

103
Shorokhov, I.M.: Shorokhova, L.I.: **Development of retrieval languages for the field of art as a component part of a unified retrieval language system.**

104
Kuznetsov, A.I.: **Toward the interfacing of primary techno-economic data with a basic classification.**

105
Negulyayev, G.A.: Tarasova, G.A.: **Concerning the effectiveness of joint use of retrieval languages in an automated patent search system.**

106
Orlov, A.N.: **Social sciences related problems of compatibility between a subject heading language and a descriptor language.**

107
Pshenichnaya, L.E.: **On the compatibility between thesaurus vocabularies in an IR system.**

108 **288-515**
Eichhorn, M.M.: Reinecke, R.D.: **Development and implementation of a thesaurus for the visual sciences.** (Orig. en). In: Bull.Med.Libr.Assoc. Vol.58 , No.1 1970, p.23-29

109 **286**
Ellmann, D.: **Der Rubrikator des staatlichen automatisierten Systems der wissenschaftlich-technischen Information der UdSSR.** The classification system of the national automated system of

scientific-technical information in the USSR. (Orig. de). In: Informatik Vol.30 , No.4 1983, p.16-19 14 refs.
The degree of compatibility between the national classification system (Rubricator) and the one of the international system (ISWTI) is investigated and assessed.

110 **286**
Ellmann, D.: **State and results of work with thesauri in the scientific and technical information system in the GDR.** (Orig. cs). In: Czech.Informatika Vol.18 , No.7/8 1976, p.193-195
Description of the thesaurus work in the GDR with special emphasis on the problem of compatibility with the mono- and multilingual thesauri in the international system of the Comecon countries.

111 **286**
Czechoslovak Academy of Sciences. Main Library: **Experiments in computerized conversion of UDC and MISON Rubricator in 1979-1980 in the Czechoslovak Academy of Sciences and major results.** Machine procedure and computerized conversion. (Orig. en). Prague: ECSSID WG3 Meeting, Sept.9-11, 1980 1980, 9p.

112 **284**
Farringdon, M.G.: **The preparation of a computer-generated concordance.** (Orig. en). In: Indexer Vol.12 , No.4 1981, p.185--188

113 **283**
Feitscher, W.: **Kompatibilität beim automatischen Indexieren und Recherchieren.** Compatibility in automatic indexing and retrieval. (Orig. de). In: VIII. Kolloquium über Inform.u.Dok. 1973. Themenkreis 3. Ilmenau, GDR: Techn.Hochschule. INER 1972, p.33-46

114 **285**
Fenske, R.E.: **Mechanization of library procedures in the medium-sized medical library: Correlations between National Library of Medicine classification numbers and.** MeSH Headings. (Orig. en). In: Bull.Med.Libr.Assoc. Vol.60 , No.2 1972, p.319-324, 1 ref.

115 **286**
Földi, T.: Sommer, K.: **Konkordanz zwischen Thesaurus documentaire d'économie und Universelle Dezimalklassifikation.** Concordance between the Thesaurus documentaire d'économie and the Universal Decimal Classification. (Orig. de). The Hague, NL: Fed.Int.Doc. 1970 Report to FID/C3

116 **288-51/4**
Fontalina, M.N.: **Convergence and compatibility analysis of the All-Union Book Chamber and medical libraries subject indexing.** (Orig. ru). In: Sostojanie predmet.katalogov v bibliotekach SSSR Leningrad: 1978, p.35-52

117 **288-51/4**
Foster, W.R.: **The coordinated development of health-related clearinghouse vocabularies: a proposed technique.** (Orig. en). In: In: Tally, R.D. Dueltgen, R.R. (Eds.) Information choices and policies. Proc.42nd Ann.ASIS Meeting, Minneapolis, MN, Oct.14-18, 1979. White Plains, NY: Knowledge Industry Publ. 1979

118 **288-51/4**
Friedman, C.: **The UMLS coverage of clinical radioloy.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1992. p.309-313 ISSN 0195-4210

119 **283**
Frost, C.O.: Dede, B.A.: **Subject heading compatibility between LCSH and catalog files of a large research library: a suggested model for analysis.** (Orig. en). In: Inform.Technol.Libr. Vol.7 1988, p.288-298

120 **283**
Fuellhart, P.O.: Weeks, D.C.: **Language of information scirence. Convertibility in information systems. Final report - Pt.2.** (Orig. en). Washington, DC: G.Washington Univ.Med.Center. 1971, 205p. Biol. Sci. Commun. Project

121 **283**
Furniss, P.: **A study of the compatibility of two subject catalogues.** (Orig. en).

Sheffield, GB: Sheffield Univ. Postgrad.SLIS 1980, 283p.

122 **281**
Gaaze-Rapoport, M.G.: **Integrated system of information languages: daydreams and real tasks.** (Orig. ru). In: Nauchno-tekhn. inform., Ser.2 , No.1 1978, p.32-34

123 **282, 288-82**
Gardin, Natacha: **The intermediate lexicon. A new step towards international cooperation in scientific and technical information.** (Orig. en). In: UNESCO Bull.Libr. Vol.23 , No.2 1969, p.58-63

124 **281**
Gendlina, N.I.: **Problems in optimizing semantic devices in computerized library systems.** (Orig. ru). In: Avtomatizir.bibl.-inform.sistemy Novosibirsk, RU: 1985, p.34--42, 23 refs.

125 **282**
Gilchrist, A.: **Intermediate languages for switching and control.** (Orig. en). In: Aslib Proc. Vol.24 , No.7 1972, p.387-399, 42 refs.

126 **282**
Giles-Peters, A.: **Experiments in the mechanical construction of cross-database thesauri.** (Orig. en). In: Online Information '93. Proc.17th Int.Online Inform.Meeting, London, 7-9 Dec.1993. Oxford, GB: Learned Inform.(Europe) 1993, p.137-147, refs.
Reports on progress in the development of AUSTRALIS PLUS, a front-end query translation system which aims to provide a uniform interface to the searching of the databases of the Australian database service AUSTRALIS.

127 **281**
Giljarevskij, R.S.: **On the compatibility of various types of information retrieval language (IRL).** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.1 1978, p.11-13, 9 refs.

128 **285**
Glasersfeld, E.v.: IDAMI Language Research Section: **Multistore - a procedure for correlational analysis.** (Orig. en). Milano, IT: Ist.Doc.Assoc.Meccanica Italiana 1965, 87p. Rpt.ILRS-T10-650120

129 **289**
Glushkov, V.M.: Skorokhod'ko, E.F., Stognij, A.A.: **Evaluation of the degree of compatibility of information retrieval languages of document retrieval systems.** (Orig. en from ru). In: Autom.Doc. & Math. Linguistics Vol.11 , No.1 1978, p.18-26
Originally published in Russian in Nauchno-tekhn.inform.Ser.2, (1978)No1, p.14-19, 7 refs.

130 **286**
Goltdammer, I.: Stein, G.: **Thesauri, Koordinierung, sowjetische Bibliothekarisch-Bibliographische Klassifikation, elektronische Datenverarbeitung.** Thesauri, coordination, Soviet Library--Bibliographical Classification, electronic data processing. (Orig. de). In: Natl.Symp. Thesauri f.mechanisierte Inform. Recherchesysteme. Berlin, GDR: ZIID 1967, p.252--266

131 **287**
Gopinath, M.A.: Prasad, K.N.: **Thesaurus and classification scheme: A study of compatibility of principles for construction of thesaurus and classification scheme.** (Orig. en). In: Documentation Research and Training Centre. Conference, Bangalore, 1-5 Dec.1975 Bangalore, IN: DRTC and INSDOC 1975

132 **281**
Gopinath, M.A.: Prasad, K.N.: **Compatibility of the principles for design of a thesaurus and a classification scheme.** (Orig. en). In: Libr.Sci. Slant Doc. Vol.13 , No.2 1976, p.56-66, 11 refs.

133 **286**
Gorol, A.: Mlodozeniec, W., Zolkowska, T.: **An attempt to unify the classification of documents according to the UDC system by adjusting its symbols to the descriptors of the thesaurus.** (Orig. ru). In: Aktual.Probl.Inform.i Dok. Vol.20 , No.3 1975, p.3-11, 10 refs.
The descriptors of a thesaurus of non-ferrous metals were reconciled with the given UDC-notations. In the majority of the

3543 cases descriptors were found to correspond to the UDC-numbers, in a certain number of cases, two or three UDC--numbers had to be composed. These were marked for non-use in classification.

134 **281**
Gornostaev, J.: Rappich, H.: **Methodische Grundfragen der Erarbeitung kompatibler Thesauren.** Methodical basic questions in establishing compatible thesauri. (Orig. de). In: Informatik Vol.19 , No.6 1972, p.23-29, 11 refs.

135 **284, 288-731, -718**
Greif, S.: Stifterverband f.d.Dt.Wissenschaft: **Patente und Wirtschaftszweige: Zusammenführung der Internationalen Patentklassifikation und der Systematik der Wirtschaftszweige.** Patents and industrial branches: Assemblage of the International Patent Classification and the Classification of Industrial Branches. (Orig. de). Köln, DE: Heymann 1990, 375p.
The book consists of two parts: Pt. 1 describes the indicator function of patents according to areas, efficiency of the indicator, and foundation and application of it. Pt.2 is concerned with the mapping of the IPC with the Classification for Industrial Branches (of the Federal Office of Statistics) with subchapters on the aims, the methodology, the results, and the possibilities for implementations and enlargements. Tables are given, appendices, and a bibliography.

136 **281**
Grygler, M.: **Aspekte der Information, die sich aus den unterschiedenen Festlegungen der Begriffe und Benennungen in Thesauri, Standards and Fachtexten ergeben.** Information aspects resulting from differing determinations of concepts in thesauri, standards, and special language texts. (Orig. de). In: Dokum./Inform. , No.51 1981, p.190-195

137 **283**
Gul'nitskii, L.L.: **Comparison of models for an information-retrieval language.** (Orig. en from ru). In: Automatic doc. math. linguistics Vol.16 , No.3 1982, p.61-71
Results from an experimental comparison of two systems, one existing and one proposed with different structures and different rules for indexing. No reworking of existing thesauri is required in order to transfer to this model and operate in a network. The independence of the paradigmatic relations is retained and the requirements for unifying the lexics is minimized. The query indexing rules are not altered, while he document indexing rules are supplemented with an algorithmized procedure.

138 **288-46**
Haendler, H.: **Dokumentationssprachen in heterogenen landwirtschaftlichen Dokumentationsverbundsystemen.** Information languages in heterogeneous agricultural information networks. (Orig. de). In: Mitt.Ges.Bibl.wes.Dok.Landbau , No.23 1975, p.7-14, 10 refs.

139 **284**
Hammond, W.: Rosenborg, S.: **Experimental study of convertibility between large technical indexing vocabularies with table of indexing equivalents.** (Orig. en). Silver Spring, MD: Datatrol Corp. 1962, 297p. Datatrol Techn.Rpt.IR-1

140 **284**
Hammond, W.: **Convertibility of indexing vocabularies.** (Orig. en). In: USAEC: The literature of nuclear science: its management and use 1962, p.223-234 TID-7647

141 **284**
Hammond,W.: **Common vocabulary approaches for government scientific and technical information systems.** (Orig. en). Silver Spring, MD: Datatrol Corp. 1963., 108p. Techn.Rpt. IR-10

142 **286**
Harrish, B.A.: **A comparison of Library of Congress and National Library of Medicine subject headings.** (Orig. en). Detroit, MI: Wayne State Univ. School of Medicine 1967, 21p. Report No.39

143 **283**
Heigl, F.J.: **Die Kompatibilität von Ordnungssystemen in der Praxis.** Compatibility of order systems in practice. (Orig. de).

In: Maschinenwelt-Elektrotechnik Vol.30 , No.6/7 1975, 2 p.

144 **281**
Henderson, M.M. et al: **Cooperation, convertibility and compatibility among information systems: a literature review.** (Orig. en). Washington, DC: US Dept.of Commerce. Natl.Bureau Stand. 1966, 140p.

145 **285, 288-254**
Hendrickson, J.B.: Miller, T.M.: **Reaction indexing for reaction databases.** (Orig. en). In: J.Chem.Inform.& Comput.Sci. Vol.30 , No.4 1990, p.403-408, 11 refs.
Description of a program to translate the reaction entries in the REACCS and SYNLIB databases into the same terms as those employed in the SYNGEN program for synthesis design.

146 **283**
Henshaw, M.: **Conversion sampler. Principles, examples and a design for developing conversion tables for book classification schemes.** (Orig. en). In: Library Journal Vol.92 , No.19 1067, p.3964-66

147 **285**
Herner, S.: Heatwole, M.H.: **An experiment in the design of a non-manipulative correlative index.** (Orig. en). Washington, DC: Herner and Co. 1959, 25p. Techn.Note No.2

148 **284**
Hines, T.C.: **Machine arrangement of alphanumeric concordance, thesaurus, and index entries - the need for compatible standard rules.** (Orig. en). In: In: Luhn, P. (Ed.) Automation and Scientific Communication. ADI-Proc. 1963, p.7-8

149 **284**
Hines, T.G.: Harris, J.L., Levy, C.L.: **An experimental concordance program.** (Orig. en). In: Comput. Hum. Vol.4 1970, p.161--171

150 **284**
Hochgesang, G.T.: **SOCCER - a concordance program in information storage and retrieval.** (Orig. en). Ithaca, NY: Cornell Univ. 1966, p.III-1 - III-23 Rpt.No.ISR-11

151 **288-51/4**
Hoepker, W.W.: **Struktur und Kompatibilität des Thesaurus der Medizin.** Structure and compatibility of the Thesaurus of Medicine. (Orig. de). In: Int.Classif. Vol.3 , No.3 1976, p.81-84, 6 refs.
The thesaurus described is an international thesaurus in the German language. It is compatible with the Clinical Key of Diagnosis (KDS), the International Classification of Diseases (ICD/E), and the Systematized Nomenclature of Pathology (SNOP). Its volume comprises 22,000 different terms.

152 **287**
Hoffmann, G.: **Erkenntnisse und Erfahrungen bei der Aufstellung von Teilthesauri.** Experiences in the development of constituent thesauri. (Orig. de). In: RGW Symposium. Entwickl.v. Deskriptorsystemen u.ihre Nutzung beim Wiederauffinden v.Inform., Teil 1 u.2. Berlin, GDR: ZIID 1966, p.174-184

153 **286**
Hoffmann, G.: **Zur Vereinbarkeit von Fachthesauri für das Patentschrifttum.** On compatibility of special thesauri for patent literature. (Orig. de). In: ZIID-Z. Vol.15 , No.5 1968, p.215-220, 1 ref.

154 **284, 2858-51/4**
Honeyman, J.P.: **Vocabulary mapping of subject headings of medical and health-related government documents at the University of Massachusetts Medical Center Library.** (Orig. en). In: Reference Librarian Vol.32 1991, p.229-2245, 27 refs.

155 **288-46**
Hood, M.W.: Ebermann, C.: **Reconciling the CAB Thesaurus and AGROVOC.** (Orig. en). In: IAALD Quart.Bull. Vol.35 , No.3 1990, p.181-185

156 **282**
Horsnell, V.: **An intermediate lexicon for information science.** (Orig. en). In: Informatics 1. Proc.Conf. Aslib Co-ordinate Indexing Group, 11-13 April 1973 at Durham

Univ. London, GB: Aslib 1974, p.17-31, 4 refs.

157 **282**
Horsnell, V.: **Intermediate lexicon for information science. A feasibility study. Final Report.** (Orig. en). London, GB: Polytechnic of N.L., School of Librnship 1974, 110p.
The results of the feasibility study showed that the greater the relative specificity and vocabulary size of the input language the better the switching performance.

159 **282**
Horsnell, V.: **The Intermediate Lexicon. The possibilities for information exchange networks.** (Orig. en). In: Proc.38th Ann.Meeting, Amer. Soc. Inform. Sci. 1975, p.79, 2 refs.

160 **282**
Horsnell, V.: **International cooperation via an intermediate lexicon.** (Orig. en). In: INFOTERM: Symposium on Int.Cooperation in Terminology. Vienna, 9-11 April 1975 Vienna, Austria: 1975, p.153-162

161 **282**
Horsnell, V.: **The Intermediate Lexicon: an aid to international cooperation.** (Orig. en). In: Aslib Proc. Vol.27 , No.2 1975, p.57--66, 4 refs.

162 **289**
Horsnell, V.: **Evaluation and implementation of an Intermediate Lexicon.** (Orig. en). In: EURIM II. Proc. Conf.23-25 March 1976, Amsterdam, NL London: Aslib 1977, p.110-112, 3 refs.

163 **289**
Horsnell, V.: Merrett, A.: **Intermediate Lexicon Research Project: Phase 2: Evaluation of the switching and retrieval performance of the Intermediate Lexicon for Information Sci..** (Orig. en). London, GB: Polytechnic of N.L. School of Librnship 1978, 100p. ISBN: 0-900639-11-3

164 **281**
Hoyle, W.G.: **A measure of overlap in classification systems.** (Orig. en). In: In: Neelameghan, A. Ordering systems for global information networks. Proc.3rd FID/CR Conf., Bombay, Jan.6-11, 1975 Bangalore, IN: Sarada Ranganathan Endowment Libr.Sci. 1979, p.261-265, 5 refs.

165 **288-51/4**
Hübner, W.: **Zur Methodik und Technologie eines Systems medizinischer Fachthesauri.** On the methodology and technology of a system of medical thesauri. (Orig. de). In: In: Hübner, W. Beiträge z.Inform.u.Dok. in der Medizin Berlin, GDR: Dt.Akad.d.Wiss., Abt.Med.Inform. 1968, p.33-48, 10 refs.

166 **288-51/4**
Humphreys, B.L.: **Unified Medical Language System: Progress report.** (Orig. en). In: Int.Classif. Vol.15 , No.2 1988, p.85--86
The National Library of Medicine (NLM) supports the establishment of a Unified Medical Language System (UMLS) in order to overcome the lack of precise link among related biomedical information in different automated systems. The goal is not to impose a single medical vocabulary on all users and systems, but to make the myriad of classifications of medical knowledge invisible to the user while providing a single logical path to a broad range of biomedical information sources. The report describes the UMLS knowledge sources, its features, the results of six contractor reports, and the NLM contributions.

168 **288-51/4**
Humphreys, B.L.: Lindberg, D.A.B.: **Building the Unified Medical Language System. Proc.13th Annual Symposium on Computer Applications in Medical Care.** (Orig. en). 1989, p.475-480

169 **289**
Humphreys, B.L.: Lindberg, D.A., Hole, W.T.: **Assessing and enhancing the value of the UMLS Knowledge Sources.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1991. 1991, p.78-82 ISBN: ISSN0195-4210

170 **288-51/4**
Humphreys, B.L.: **UMLS Semantic Network.** (Orig. en). In: NLM Fact Sheet ,

No.June Bethesda, MD: Natl.Inst.of Health 1995, p.1-2

171 **288-51/4**
Humphreys, B.L.: **UMLS Metathesaurus.** (Orig. en). In: NLM Fact Sheet , No.Aug. Bethesda, MD: Natl.Inst.of Health 1995, p.1--2

172 **288-51/4**
Humphreys, B.L.: **Unified Medical Language System.** (Orig. en). In: NLM Fact Sheet , No.Aug. Bethesda, MD: Natl.Inst.of Health 1995, p.1-2

173 **283**
Husain, S.S.: **An indexing system: Switching language for integrating automated information networks.** (Orig. en). Ann Arbor, MI: University Microfilms Int. 1981, 221 p.
PhD Thesis from the Case Western Reserve University. An account of the basic concepts, problem areas, and operational characteristics and techniques of switching languages, concentrating particularly on the design and development of Classified Concept Coordinate Indexing as providing the necessary switching mechanism for information transfer in a network. It is based on a synthesis of pre- and post-coordinate indexing and enables free coordination of diverse concepts, through keyword inverted indexing based on an apriori set of classes.

174 **286**
Hyslob, M.: **A compatibility study of two information systems.** (Orig. en). In: Amer. Doc. Vol.14 1963, p.292-298

175 **283**
Izmajlova, N.S.: Sakharnyj, L.V.: **Search convertibility of the All-Union Book Chamber and State Public Library subject headings in an automated IR environment.** (Orig. ru). In: Predmet.poisk v tradiciion.i netradicion.IPS Leningrad, Russia: 1981, p.171-200, 8 refs.

176 **288-51/4, 298**
Jachna, J.S.: Powsner, S.M., Miller, P.L.: **Augmenting GRATEFUL MED with the UMLS Metathesaurus: an initial evaluation.** (Orig. en). In: Bull.Med.Libr.Assoc. Vol.81 , No.1 1993, p.20-28

177 **281**
Jacob, W.: Rolland, M., Praetorius, C.: **Zur Erstellung von Hilfsmitteln für die Erhebung der formalen und inhaltlichen Elemente dokumentarischer Bezugseinheiten auf der Basis von ADEK.** On the development of tools for formal and contents-related bibliographic description according to ADEK, the General Data Input Catalogue. (Orig. de). Eggenstein-Leopoldsh: FIZ Energie, Phys.,Math. GmbH 1977, 92p. BMFT-FB ID 78-03
A method was developed for the description and definition of types of documents occurring in different types of information systems in order to facilitate data exchange between information networks.

178 **283**
Jaster, J.J.: **Subsumption scheme for dictionary of indexing equivalents.** (Orig. en). Silver Spring, MD: Datatrol Corp. 1963, 25p. Techn.Report IR-3 AD 299 236

179 **283**
Kalachkina, S.Ya.: **Algorithmic determination of descriptor equivalents in different natural languages.** (Orig. en from ru). In: Automatic Doc. Math.Ling. Vol.21 , No.4 1987, p.21-29, 6 refs.

180 **283**
Kautto, V.: **Linking equivalent terms in three languages with the aid of UDC--numbers.** (Orig. en). In: T.Dokum. Vol.37 , No.2 1981, p.41-43

181 **285**
Kalinichenko, L.A.: **Axiomatic extension of a conceptual data model which are equivalent to traditional data models.** (Orig. ru). Metody mat.logiki v probl. iskusstv. intellekta i sist. programmir. Ch.2. Vilnius, SU: 1980, p.196-198, 2 refs.

182 **283**
Kent, Allen: Perry, J.W.: **Encoding of information previously indexed.** (Orig. en). In: Perry et al 1958 , p.343-360

183 **286**
Kien, J.P.H.G.: **The relationship of the Library of Congress Subject Headings with the UDC in the frame of the PICA project.** (Orig. nl). In: Open Vol.3 1971, p.553-557

184 **281**
King, D.W.: Shumway, R.H.: **A model for translating from one classification system to another.** (Orig. en). In: ICIREPAT: 5th Ann. Meetg.of the Committee for Int.Cooperation in Inform.Retrieval among Examining Patent Offices, U.K.Patent Office,London 1965 Washington, DC: Thompson Books 1966

185 **283**
Kirtland, M.: Atherton, P., Harper, S.: **Integrating borrowed records into a database.** Impact on thesaurus development and retrieval. (Orig. en). In: Database Vol.3 , No.4 1980, p.26-33
If one merges parts of databases to socalled tertiary databases, 3 approaches to vocabulary development can be chosen: reindex some records, merge without reindexing, or initial standardization. The paper describes the impact of a merged file on indexing and thesaurus construction.

186 **284**
Knoll, H.: **Observations on concordances.** (Orig. en). In: World Patent Inform. Vol.5 , No.1 1983, p.19-34, 6 refs.
The limitations and pitfalls using concordances are illustrated by examples derived from concordances between the German Patent Classification, the U.S. Patent Classifiction, the International Patent Classification and the Universal Decimal Classification.

187 **287**
Ködderitzsch, W.: **Einheitliche Prinzipien der Erarbeitung von Informationssprachen und Terminologiestandards.** Uniform principles for the elaboration of information search languages and terminology standards. (Orig. de). In: Informatik Vol.27 , No.3 1980, p.32-35, 15 refs.

188 **287**
Kodola, V.E.: Skripkin, V.A.: **Compatibility problems in the CMEA countries' International Scientific and Technical Information System.** (Orig. de). In: Int. Forum Inform. Doc. Vol.3 , No.2 1978, p.18-23, 8 refs.
Description of the principal parameters affecting compatibility, i.e. those concerning information, information retrieval languages, information processing hardware and media. List of normative documents.

189 **281**
Kofnovec, L.: **Set of selection languages of the international system of scientific and technical information.** (Orig. cz). In: Cheskosl.Inform. Vol.17 , No.7/8 1975, p.200-205, 6 refs.
The article contains a schematic representation of the relations between the elements of sets of selection languages in the international system. Discussion of these relationships and functions.

190 **287**
Kolchinskij, M.L.: **A unified system of information languages for documentary retrieval in a system of automated information services.** (Orig. ru). In: Nauchnotekhn.inform.,Ser.2 , No.1 1978, p.24-31, 21 refs.

191 **28.06.73-11-28/30**
8. Kolloquium über Information und Dokumentation. Themenkreise 2 u. 3. Ilmenau, GDR, 28.11.-30.11.1973. 8th Colloquy on Information and Documentation. Session 2 and 3: Uniform Concepts for Communication in international Cooperation. Compatibility of methods and tools to facilitate cooperation. (Orig. de). Ilmenau, GDR: TH Ilmenau, INER 1973, 112 + 87p. Dok./Inform. Ser.24 and 25
The issues contain the following papers of relevance:

191a
Merta, A.: **Die Einheit der Begriffe als Grundlage des konzeptionellen Aufbaus von internationalen Systemen der wissenschaftlichen und technischen Information.** Unity of concepts as a basis for the conceptual construction of internatio-

nal scientific and technical information systems.

191b
Rosenbaum. H.-D.: **Methodische Grundsätze für das Erarbeiten mehrsprachiger Thesauren der Mitgliedsländer der RGW..** Methodical principles for the elaboration of multilingual thesauri in the COMECON member countries.

191c
Brendler, G.: **Probleme beim Aufbau eines zweisprachigen Thesaurus im Sprachenpaar deutsch-russisch.** Problems in constructing a bilingual Russian-German thesaurus.

191d
Pilz, J.: **Methodische Aspekte beim Aufsteilen und Benutzen mehrdimensionaler Strukturen im Thesaurus.** Methodological Aspects concerning the establishment and use of multidimensional thesaurus structures.

191e
Herrmann, P., Löschner, G., Rudorf, D.: **Zum Problem der Erarbeitung mehrsprachiger Thesauri**. Concerning the problem of elaborating multilingual thesauri.

191f
Feitscher, W.: **Kompatibilität beim automatischen Indexieren und Recherchieren** Compatibility in automatic indexing and searching.

191g
Mater, E.: **Elementarbegriffe als Grundlage maschineller Indexierung und Konvertierung** Elementary concepts as a basis for computerized indexing and correlating.

193 **288-124**
Korel'skaya, T.D.: **Translation algorithm of a natural language versus a predicate calculus language.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 Vol.6 1975, p.24-34, 9 refs.
The algorithm is founded on syntactical transformations in the field of geometry.

194 **281**
Kostyukova, L.S.: Lejchik, V.M., Petrova, Z.Yu.: **Compatibility problems of the terminology of an information retrieval thesaurus and of standards of terms and definitions.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.1 , No.3 1984, p.26-29

195 **286**
Kozlova, E.K.: Lesokhin, M.M.: **Integration of tree-like thesauri.** (Orig. ru). In: Inzh.lingv.i optimiz.prepodavanklja instr. yazykov Leningrad, Russia: 1980, p.85-89

196 **284**
Kralova, A.: **Experiments in the use of the Balex databases of developing concordances between indexing languages.** (Orig. cz). In: Kniz.a ved.inform. Vol.21 , No.3 1989, p.104-108, 6 refs.

197 **287**
Kristal'nyj, B.V.: Antopol'skij, A.B.: **Composition, structure, and functions of linguistic support for the network of automated STI Centers.** (Orig. ru). In: Tez. dokl.9-go nauch.seminara po system. issled.GASNTI. Ch.1 Moskva, Russia: 1979, p.43-46

198 **287**
Kristal'nyj, B.V.: Ilieva, B.M.: **Development of compatible subject authorities of the GASNTI in the USSR and the National STI System in Bulgaria.** (Orig. ru). In: Nauch i tekhn.inform.zu rubezhom. Moskva, Russia: 1984, p.28-38

199 **287**
Kristal'nyj, B.V.: **A uniform system of rubricators - linguistic basis for the network within GASNTI.** (Orig. ru). In: Nauchn.-tekhn.inform.,Ser.1 , No.7 1989, p.13-15

200 **281**
Krumholz, W.: **Kompatibilität, Normung und Terminologie.** Compatibility, standardization and terminology. (Orig. de). In: DK-Mitt. Vol.24 , No.4 1980, p.11-13
Typology of standards according to formal criteria and their special significance for establishment of compatibility between standards.

201 **285**
Kuhns, J.L.: Montgomery, C.A.: **Experiments in information correlation.** (Orig. en). In: AFIPS Conf.Proc. Vol.25, 1964., p.577-585

202 **286**
Kulik, A.N.: **On the feasibility of connecting the different descriptor information retrieval systems on the basis of a common language.** (Orig. ru). In: Nauch.-tekhn.inform.Ser.2 , No.11 1969, p.18-21

203 **282**
Kulik, A.N.: **Acquisition in a unified network of information retrieval systems and the thesaurus.** (Orig. en from ru). In: Autom.Doc. Math.Linguist. Vol.6 , No.4 1972, p.66-76, 30 refs.

204 **281**
Kulik, A.N.: **Problems of compatibility of information retrieval systems and requirements of the language of an information network.** (Orig. ru). In: Nauch.--Tekhn.Inform.Ser.1 , No.12 1973, p.9-16, 43 refs.

205 **281**
Kurbakov, K.I.: **The problem of compatibility of classification systems and one possible approach to its solution.** (Orig. ru). In: Nauch.-Tekhn.Inform.,Ser.2 , No.7 1968, p.24-26
The new approach uses the principle of complementarity.

206 **281**
Kurbakov, K.I.: Boldov,V.G.: **Problem of compatibility of information classification systems and some probable methods of its solution.** (Orig. en). In: In: Neelameghan, A. Ordering Systems for Global Information Networks. Proc.3rd Int.Study Conf. on Classif. Res., Bombay, 6-11 Jan.1975. Bangalore, IN: Sarada Ranganathan Endowment f.Libr.Sci. 1979, p.150-156, 16 refs.
Existing classification systems are described and compared. It is concluded that the problem to establish an optimum system for international exchange of STI must be solved in stages. A 3-sequence world system of classification is shortly introduced.

207 **284**
Lamb, S.M.: Gould, L.: **Concordances from computers.** (Orig. en). Berkley, CA: Univ.of California 1964, 90p.

208 **281**
Lancaster, F.W.: Smith, Linda C.: **Compatibility issues affecting information systems and services.** (Orig. en). Paris: UNESCO. Gen.Inform.Program 1983

209 **281**
Lancaster, F.W.: **Compatibility and convertibility. Chapter 14 of Vocabulary Control for Information Retrieval. 2nd ed. Arlington, VA: Information Resources Press 1986.**

211 **288-46**
Lehman, R.: **Setting up a system of compatible thesauri in the international system AGROINFORM: results and experiences.** (Orig. ru). In: Byul.NTI/MS AGROINFORM Vol.11 , No.1 1984, p.32-37

212 **288-76**
Lestang, Ph.: **Harmonizing international commodity classification: an urgent task.** (Orig. en). In: Forum Ware Vol.7 , No.3-4 1979, p.151-152
The paper shows a way out of the present diversity.

213 **288-384**
Lety Ringeissen, V.: **Compatibilité entre classification hierarchique et systčme terminologique pour un sous-ensemble d'un fonds documentaire de recherche routičre.** trčs spécialisé en métallurgie. Compatibility between a hierarchical classification and a terminological system for a sub-set of a documentation resource in road research. (Orig. fr). Paris: Mém.INTD 1976, 77p.
Presentation of the documentation service of the Laboratoire Central des Ponts et Chaussées.

214 **288-51/4**
Levesque, Y.: LeBlanc, A.R., Maksud, M.: **MD Concept: a model for integrating medical knowledge.** (Orig. en). In: Proc.

Ann. Symp. Comput. Appl. Med. Care 1994. p.252-256 ISSN 0195-4210

215 **289**
Levy, F.G.: **Compatibility between classification and thesauri: first evaluation of a study in the field of information storage and retrieval.** (Orig. en). In: Proc.33rd FID Conf., Tokyo 1967. Paper III g 2 , 15 p., 6 refs.

216 **281**
Light, U.E.K.: **The principle of interdependent incompatibility - a new postulate.** (Orig. en). In: In: Horsnell,V.(Ed.) Informatics II. Proc.Conf.Aslib Co-ordinate Indexing Group, Oxford 1974. London: Aslib 1975, p.81-88

217 **288-51/4**
Lindberg, D.A.: Humphreys, B.L., McCray, A.T.: **The Unified Medical Language System.** (Orig. en). In: Methods Inform. Med. Vol.32 , No.4 1993, p. 281-291

218 **288-6**
Litoukhin, J.: **Toward and integrated thesaurus of the social sciences.** (Orig. en). In: Int.Classif. Vol.7 , No.2 1980, p.56-59

219 **288-6**
Litoukhin, J.: **UNESCO's project for the establishment of an integrated thesaurus of the social sciences.** (Orig. en). In: In: Riggs, F.W. The CONTA Conference. Proc. Conf.on Conceptual and Terminological Analysis in the Social Sciences, Bielefeld, FRG, May 24-27, 1981 Frankfurt/Main: INDEKS Verlag 1982, p.202-206, 4 refs.

220 **286**
Liu-Lengyel, H.Y.: **The development and use of the Chinese classification system.** (Orig. en). In: Int.Library Review Vol.19 , No.1 1987, p.47-60, 29 refs.
The history of classification in China goes back to the years 26-6 B.C. It is outlined until the installation of the Chinese Library Book Classification in 1975. Its structure is briefly reported and its compatibility with LC is discussed.

221 **282**
Lloyd, G.A.: **The Universal Decimal Classification as an international switching language.** (Orig. en). In: In: Wellisch, H. (Ed.) Subject Retrieval in the Seventies. New Directions. Proc. Int.Symp., Univ.of Maryland, May 14-15, 1971. Westport, CT: Greenwood Publ. 1972, p.116-125

222 **286**
Löhr, H.: Güldner, R., Poehlmann, E., Haase, U.: **Anpassung eines Arbeitsthesaurus an übergeordnete Thesauren.** Adaptation of a working thesaurus to superordinated thesauri. (Orig. de). In: Informatik Vol19, No 6, p.30-34, 4 refs.

223 **286**
Lorenson, R.: **Adapting LC schedules to DC notation.** (Orig. en). In: Libr.Resources Techn.Serv. Vol.9 , No.2 1965, p.210-212

224 **288-5384**
Lunin, L.F.: Stovall, M.: **Thesaurus for literature analysis in a cancer research institution: Construction and correlation with three national thesauri.** (Orig. en). In: Automat. and Sci.Communic., ADI Ann.Meetg.1963. Short Papers, Pt.2 Washington, DC: Amer.Doc.Inst. 1963, p.189-190, 3 refs.

225 **286**
Malevich, E.N.: **Correlation between library catalogues and bibliographic editions.** (Orig. ru). In: Sov.Bibl. , No.6 1983, p.51-58

226 **281.233**
Mamedova,M.G.: Skorokhod'ko, E.F.: **A method of optimizing information retrieval thesauri.** (Orig. ru, en). In: Autom. Doc. & Math. ling. Vol.19 , No.4 1985, p.82--92, 6 refs.

227 **282**
Mandel, C.A.: **Multiple thesauri in online library bibliographic systems.** (Orig. en). Washington, DC: Library of Congress 1987, 108p. ERIC ED 291 390

228 **288-26/27**
Maneck, M.: **Thesaurussystem Chemie - ein Beitrag zur Schaffung eines ein-**

heitlichen Dokumentations- und Informationssystems auf dem Gebiet der Chemie. Thesaurus system chemistry - a contribution to the creation of a unified documentation and information system in chemistry. (Orig. de). In: Zentralinst. f.Inform.u.Dok. (ZIID) (Ed.) RGW-Symposium. Entwicklung v. Deskriptor-systemen u.ihre Nutzung beim Wiederauf-finden von Inform., Berlin, 28.6.-1.7.1966 Berlin, GDR: ZIID 1966, p.53-72 ZIID-Schriftenreihe 12

229 **285**
Masarie, F.E.Jr.: Miller,R.A., Bouhaddou,O., Giuse,N.B., Warner,H.R.: **An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies.** (Orig. en). In: Comput.Biomed.Res. Vol.24 , No.4 1991, p.379-400

230 **289**
Masys, D.R.: **An evaluation of the source selection elements of the prototype UMLS Information Sources Map.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1992. p.295-298 ISSN 0195-4210

231 **288-51/4**
McCray, A.T. et al: **UMLS knowledge for biomedical language processing.** (Orig. en). In: Bull. Med. Libr. Assoc. Vol.81 , No.2 1993, p.184-194

232 **288-51/4**
McCray, A.T.: Srinivasan, S., Browne, A.C.: **Lexical methods for managing variation in biomedical terminologies.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. 1994. p.235-239 ISSN 0195-4210

233 **288-35**
McCuaig, H.E.: Mercier, M.A.: **A UDC water thesaurus. Concordance, development, and use.** (Orig. en). In: 1st Open Conf.on Inform.Sci. in Canada, Montebello, Quebec 1973. Chalk River, Ont.: Atomic Energy of Canada Ltd. 1973, p.104-118

234 **286**
McCuaig, H.E.: Mercier, M.A.: **Use of a thesaurus/classification concordance for indexing and retrieval.** (Orig. en). In: Proc.ASIS, Vol.10 Westport, CT: Greenwood Press 1973, p.129-130, 5 refs.

235 **287**
Mechido, L.S.: Mishin, V.A.: **Linguistic support for the Lenin State Library automatic system in the context of its interaction with the network of automated STI Centres.** (Orig. ru). In: Tez.dokl.9-go Nauch.seminara po sistem issled. GASNTI, Ch.2 Moskva, Russia: 1979, p.106-109

236 **283**
Melton, J.: **Encoding of information previously classified.** (Orig. en). In: Perry, J.W., Kent, A.: Tools for machine literature searching. Semantic code dictionary, equipment, procedures. New York: Interscience 11958., p.361-378

237 **286**
Melton, J.: **A note on the compatibility of two information systems: Colon Classification and Western Reserve University (Encoded Telegraphic Abstracts) and the feasibility of interchanging their notations.** (Orig. en). Cleveland, OH: West.Res.Univ.Cent.Doc.Commun.Res. 1960, 7 p. Techn.Note 13
Also in: Libr.Science Today. Ranganathan Festschrift Vol.1. London, GB: Asia Publ.House 1965. p.49-62.

239 **287**
Menillet, D.: Grandjouan, J.M.: **Coordination des thesaurus sectoriels et spécialisés et comptabilité avec ie macrothésaurus dans le cadre national d'information scientifique et technique.** Coordination of thesauri in the frame of national scientific and technical information network. (Orig. fr). Paris: CNRS 1974, 41 p. DGRST-7371441
Coordination of existing thesauri may be done on the basis of the French standard Z47-100. In parallel a document for the elaboration of multilingual thesauri "Problems of linguistic correspondence" had been worked out.

240 **288-6**
Meyriat, J.: **Social science information languages: a comparative analysis.** (Orig. en). In: Int.Classif. Vol.7 , No.2 1980, p.60-65
Preliminary report of an analysis of some 60 existing information languages in the social sciences.

241 **288-46**
Mildner, H.: **Results of a cooperative work in establishing an agricultural thesaurus.** (Orig. ru). In: Bjull.nauchn.tekhn.inform. Vol.4 , No.1 1977, p.3-11, 12 refs.
Report on the elaboration of a bilingual Russian-German thesaurus in the field of agriculture and forestry as basis of a system of compatible thesauri for the members of the Comecon countries. In the first edition some 5000 terms were monohierarchically presented.

243 **286**
Mili, H.: Rada, R.: **Merging thesauri: principles and evaluation.** (Orig. en). In: IEEE Transact.Pattern Analysis Vol.10 , No.2 1988, p.204-220, 56 refs.

244 **288-51/4**
Miller, P.L.: Frawley,S.J., Wright,L., Roderer,N.K., Powsner,S.M: **Lessons learned from a pilot implementation of the UMLS information sources map.** (Orig. en). In: J.Amer.Med.Inform.Assoc. Vol.2 , No.2 1995, p.102-115

245 **287**
Mishin, V.A.: Sanches-Mechido, L.: **An integrated retrieval language of a general-purpose computer-based IR system.** (Orig. ru). In: Sov. Bibliotekovedenie , No.5, 1976, p.59-68, 10 refs.
Integration of LBC and UDC with a subject heading system. Developed for the Current Accessions Retrieval System and the Lenin Library, Moscow.

246 **285.371**
Mishin, V.A.: **Principles and methods of automatic reclassification.** (Orig. ru). In: Bibliotekoved. , No.3 1979, p.56-68, 14 refs.
The different possibilities of notation conversion of one system into another are considered.

247 **281**
Mölgaard-Hansen, R.: **On compatibility and complementarity between indexing languages.** Lecture delivered at the Meeting of the Polish Classification Research Group, Warszawa, Oct.9, 1970. (Orig. en). In: T.Dok. Vol.27 , No.5 1971, p.57-58, 15 refs.

249 **286**
Müller, C. E.: **Konkordanzen zwischen Fachge-bietsklassifikationen und der internationalen Dezimalklassifikation.** Concordances between subject classifications and the International decimal Classification. (Orig. de). In: DK-Mitt. Vol.6 , No.4 1961, p.13-15

250 **285, 288-51/4**
Muench, E.V.: **Biomedical Subject Headings. A reconciliation of National Library of Medicine (MeSH) and Library of Congress Subject Headings, 2nd ed..** (Orig. en). Hamden, CT: Shoestring Press 1971.79, 774p.

251 **285, 288-51/4**
Muench, E.V.: **A computerized English-Spanish correlation index to five biomedical library classification schemes based on MeSH.** (Orig. en). In: Bull.Med.Libr.Assoc. Vol.50 , No.3 1971, p.404-411

252 **285, 288-51/4**
Muench, E.V.: **Biomedical Classification Schemes: An English-Spanish correlation index to five library classification schemes based on MeSH.** (Orig. en). Ann Arbor, MI: Xerox University Microfilms 1974, 7 v.

253 **286**
Murinson, E.A.: Shenderov, V.Z., Kochetova, S.M.: **Reconciling a classificatory and a descriptor language.** (Orig. ru). In: Vopr.inform.teorii i prakt. , No.36 1978, p.31-50, 11 refs.
Intention to study experimentally the process of reconciling the IPC vocabulary and that of

the TEST (Thesaurus of Scientific and Technical Terms).

254 **286**
Nakamura, Y.: **Terms and classification systems: An introduction to the TOCS principle.** (Orig. ja). In: Libr.& Inform.Sci., Japan , p.95-106
Advocation of TOCS (Term-Oriented Classification System) which aims at making classification systems more easily accessible through terms and words by reforming the style and structure of the schedules. Example for TOCS applied to the UDC.

256 **283**
Negulyayev, G.A.: Tarasova, G.A.: **The integrated use of information retrieval languages in patent search.** (Orig. en). In: World Patent Inform. Vol.4 , No.2 1982, p.70-74

257 **288-51/4**
Nelson, S.J. et al.: **From meaning to term: semantic locality in the UMLS Metathesaurus.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1991. p.209-213 ISSN 0195-4210

258 **288-51/4**
Nelson, S.J. et al.: **The semantic structure of the UMLS Metathesaurus.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1992. p.649-653 ISSN 0195-4210

259 **286**
Neville, H.H.: **Feasibility of a scheme for reconciling thesauri covering a common subject.** (Orig. en). In: J.Doc. Vol.26 , No.4 1970, p.313-335

260 **283**
Neville, H.H.: **Aspects of indexing in compatibility.** Arbeitspapier 13. (Orig. en). In: ISO/TC 46: Int.Scientific Symp.on Multilingual Thesauri, Berlin, DE, 8-10 Oct.1973 , 10p., 2 refs.

261 **28.06.64**
Newman, S.M. (Ed.): **Information Systems Compatibility.** (Orig. en). Washington-New York: Spartan Books. London: MacMillan 1965, 150p.
Contains the 12 papers of the 6th Institute on Information Storage and Retrieval, Center for Technology and Administration, American University, Washington 1964. The following 5 papers are of relevance here:
261a Shilling, C.W.: **The need for compatibility in the biological sciences.**
261b Hammond, W.: **Dimensions in compatibility.**
261c Taube, M.: **A review of programs to achieve information center compatibility.**
261d Stegmaier, R.B.Jr.: **Programs to achieve compatibility.**
261e Atherton, P.: **Is compatibility of authority files practical?**

262 **288-28**
Niehoff, R.T.: **Integrated energy vocabulary 1976.** (Orig. en). Springfield, VA: NTIS 1976, 459p. PB 259 000
Compilation of some 30,000 scientific and technical terms in energy research and development from differnt information system vocabularies for use in indexing and as a reference tool for the development of microthesauri. It can also serve further studies of vocabulary compatibility and convertibility.

263 **283**
Niehoff, R.T.: Colombo, D.S.: **The meaning and significance of subject switching in relation to online data bases.** (Orig. en). In: In: Martin, S.K. Information Politics. Proc.ASIS, 39th Ann.Meetg., San Francisco, CA, Oct.4-9, 1976. Vol.13 Washington, DC: Amer.Soc.Inform.Sci. 1976

264 **288-28**
Niehoff, R.: Kwasny, S., Walker, A., Wessells, M.: **The design and evaluation of a vocabulary switching system for use in multi-base search environments.** (Orig. en). Columbus, OH: Battelle Columbus Labs. 1980, 84p.
The VSS (Vocabulary Switching System) is an experimental online automated switching system which produces semantic, syntactic, and generic relatives to search terms supplied by the user. It is based on existing thesauri or other controlled vocabularies.Discussion of results of 29 experimental stacks.

265 **281**
Niehoff, R.: Kwasny, S.: **The role of automatic subject switching in a distributed information network.** (Orig. en). In: Online Review Vol.3 , No.2 1979, p.181-192

266 **282**
Niehoff, R.: Mack, G.: **The Vocabulary Switching System. Description of evaluation studies.** (Orig. en). In: Int.Classif. Vol.12 , No.1 1985, P.2-6
The VSS is presented as an experimental system designed to enhance search strategies and ultimately retrieval performance for those who use online bibliographic data bases. It contains 15 indexing and retrieval vocabularies from 12 different suppliers. It is a stand-alone, online, database containing the subject descriptors and all the syndetic relationships found in the 15 vocabularies. Its major fields (modules) are: Physical Sciences, Life sciences, Social Sciences and Business Module. Evaluations were made and their results are summarized.

267 **285**
Niehoff, R.: **The optimization and use of automated subject switching for better retrieval.** (Orig. en). In: In: Benenfeld,A.R. Communicating Information. Proc. 43rd ASIS Ann.Meetg., Anaheim, CA, Oct.5-10, 1980. Vol.17. White Plains, NY: Knowledge Industry Publ. 1980, p.397-400

268 **282**
Niehoff, R.T.: **Development of an integrated energy vocabulary and the possibilities for online subject-switching.** (Orig. en). In: J.Amer.Soc.Inform.Sci. Vol.27 , No.1 1976, p.3-17, 8 refs.
11 vocabularies were analyzed and integrated: AIP, CA, API, GA, INIS, TEST, GEOR, EI, PA, DDC, and NASA. The methodology consisted of 1) establishing term selection criteria, 2) analyzing individual system vocabularies for energy-related terms, 3) processsing energy subsets and 4) reviewing the integrated product and generating a final vocabulary. A prototype conversion guide (synonym table) was constructed.

269 **283**
Norwood, J.W.: **Machine-aided retrieval.** (Orig. en). Alexandria, VA: Defense Documentation Center 1975, 45p. AD-A-006500
Comparison analysis of Machine-Aided Retrieval (MAR) and Machine-Aided Indexing (MAI) terminology for system compatibility with the Natural Language Database (NLDB).

270 **284**
Nöther, I.: **Modeli einer Konkordanz-Klassifikatioon für Systematische Kataloge, Teil 1.** Model of a concordance classification for classified catalogues. (Orig. de). In: Bibliotheksdienst Vol.28 , No.1 1994, p.15-33, 9 refs.
Part 1 (of 2) describes the typical form of a concordance classification.

271 **284**
Nöther, I.: **Modell einer Konkordanz-Klassifikation für Systematische Kataloge, Teil 2.** Model of a concordance classification for classified catalogues, Pt.2. (Orig. de). In: Bibliotheksdienst Vol.28 , No.2 1994, p.175-187, 7 refs.
Consideration of the optimal choice of a concordance classification system among the existing ones. Presentation of a project to test such a system in some subject fields.

272 **288-51/4**
O'Connor, J.: **Correlation of indexing headings and title words in three medical indexing systems.** (Orig. en). In: Amer.Doc. Vol.15 , No.2 1964, p.96-104, 13 refs.

273 **286**
Öhmann, E.: Olivecrona, Ch.: **Some notational, hierarchic and syntactic problems in connection with concordances between UDC and thesauri.** (Orig. en). In: UDC in Relation to other indexing languages. Proc.Int.Symp., Herceg Novi, Yu., 28 June - 1 July 1971. Beograd, Yu.: Yugosl.Center f.Techn.Scient.Doc. 1972, 48p.

274 **289**
Öhmann, E.: **Die DK-Kritik von H.Arntz im Lichte von Konkordanzuntersuchungen**

zwischen DK und Thesauren. The UDC criticism of H. Arntz in the light of concordance investigations between the UDC and thesauri. (Orig. de). In: Nachr.Dok. Vol.23 , No.2 1972, p.65-68

275 **283**
Osmolovskij, L.G.: Petuchov, O.A., Chodorovskij, L.A.: **introduction of data conversion operations into a data model algebra.** (Orig. ru). In: Vopr.Sistemotechn. , No.5 1980, p.41-44, 4 refs.
Defines unary operations in a set of information system schemes. Extends the data model algebra.

276 **284**
Painter, A.F.: **Convertibility potential among government information agency indexing systems.** (Orig. en). In: Library Res.Techn.Serv. Vol.7 , No.3 1963, p.274-281

277 **288-6**
Palnicov, M.: **ECSSID activities in the field of information languages compatibility.** (Orig. en). Paris: UNESCO 1980, 8 p.
A presentation to the UNESCO Consultative Meeting on the establishment of an integrated thesaurus of the social sciences, 9-11 June 1980.

278 **287**
Panov, A.A.: Suchmaneva, E.G.: **Efficiency of information systems interaction and the linguistic support of the BAN SSSR computer-based information retrieval system.** (Orig. ru). In: Sbor.nauch.tr.Biblioteki AN , No.5 1978, p.190-199, 9 refs.

279 **283**
Panov, A.A.: **Interaction of indexing languages.** (Orig. ru). In: Probl.vnedreniya BBK Moskva, SU: 1979, p.115-117, 1 ref.

280 **285**
Pauer, M.: **Deckung von aufstellung und systematischem Katalog.** Mapping of shelf-arrangement and arrangement in the systematic catalogue. (Orig. de). In: In: Alsheimer,R.(Ed) Bestandserschließung und Bibliotheksstruktur. Rolf Kluth zum 10.2.1979 (Subject cataloging and library structure. Festschrift) Wiesbaden, DE: O. Harrassowitz 1979, p.158-169

281 **284**
Perreault, J.M.: **On concordance between classifications.** (Orig. en). In: VINITI (Ed.) Int.Forum on Informatics. Vol.2 Moscow: VINITI 1969, p.230-241, 10 refs.

282 **284**
Perreault, J.M.: **Concordance between classification and union lists between collections.** (Orig. en). In: Int.Forum Inform.Doc. Vol.6 , No.1 1981, p.19-21, 5 refs.
The nature of relationships between various classification systems is examined focussing on extensional similarity. Some parallels are identified between the concordancing of classification systems and the compilation of union lists of various collections. Conclusions.

283 **283**
Petukhov, O.A.: Khodorovskij, L.A.: **Introduction of data conversion operations into a data model algebra.** (Orig. ru). In: Vopr.sistemotekhn. 1980, p.41-44, 4 refs.
Defines unary operations in a set of information system schemes. Extends the data model algebra.

284 **283**
Pflug, G.: **Classification in the libraries of the German Federal Republic.** (Orig. en). In: Int.Classif. Vol.9 , No.1 1982, p.27-29, 20 refs.
The history of initiatives toward a common system in German libraries is traced as well as the insight into the possibilities of common rules for subject cataloguing.

285 **282**
Plante, S.: Grandjouan, J.-M.: **Un macrothésaurus des techniques.** A macrothesaurus for technology. (Orig. fr). In: 1er Congr. Natl. Franc. Inform.Doc. Commun. Paris 1974. Paris: CNRS 1974, p.157-168
The general thesaurus of about 500 terms is meant for broad indexing and compatibility between different special thesauri

286 **285**
Pokras, Yu.L.: **One method for achieving linguistic compatibility for computerized STI systems.** (Orig. ru). In: Probl. avtomatizir. obrab. nauch.-tekhn. inform.: 4 Vses. nauchn.-tekhn. Konf. Moskva, Russia: 1983, p.60-61

287 **287**
Popowska, H.: **Construction of thesauri according to the method of A.V. Sokolov.** (Orig. pl). In: Aktual.Probl.Inform.Dok. Vol.24 , No.6 1979, p.21-26, 15 refs.
The construction method using facets in blocks is presented. The criteria for compatibility are shown and Sokolov's method is discussed as well as the possibilities of a realization of such a thesaurus.

288 **283**
Potapova, R.D.: **A study of the compatibility between the central subject indexing and the local library subject indexing.** (Orig. ru). In: Sov. Bibliografiya , No.5 1975, p.22-30

289 **281**
Potapova, R.D.: **A method to measure the degree of convertibility and compatibility between central and local subject cataloguing.** (Orig. ru). In: Vopr. premetn. katalog. Leningrad: 1976, p.38-56

290 **81.06.74-02-11/13**
Problems of Design and Development of International Scientific and Technical Information Systems. Vyp.1. (Orig. ru). Moskva: Mezdunar.Cent.Nauch.i Tekhn.Inform. 1974, 171p.
Publication of the following 23 papers presented at an international symposium, 1-13 Febr.1974 in East-Berlin on thesaurus construction with special regard to problems of compatibility and multilingualism.

291 Arsenov, C.T. et al.: **Language compatibility problems in the International Scientific and Technical Information System.**

292 Dimitrov, V.: **Some problems in constructing a German-Bulgarian computer thesaurus.**

293 Pulev, V.: **Developing a compatible thesaurus system for machine building technology as part of a national thesaurus for science and technology.**

294 Rosenbaum, H.-D.: **Methodological guidelines for compiling multilingual thesauri for CMEA countries.**

295 Bart, H.: **Compatible thesaurus development: experience and problems.**

296 Winkler, H.: **A note on the problem of compatibility of multilingual thesauri.**

297 Herrmann, P.: Löschner, G., Rudorf, D.: **Toward the problem of multilingual thesaurus construction.**

298 Kuklich, F.: **Building a thesaurus for the International Scientific and Technical Information System.**

299 Richter, G.: Weber, H.: **The significance of the subject-independent basic structure of a multilingual thesaurus.**

300 Gerisch, G.: **Experience in developing an international bilingual (Russian-German) thesaurus for construction.** First results of its testing as part of an international indexing experiment.

301 Heinemann, E.: Noack, H., Grunert, M.: **Practical experience in compiling a bilingual Russian-German (German-Russian) thesaurus for chemistry and chemical engineering.**

302 Mildner, G.: **On the concept of an international agricultural thesaurus.**

303 Pilz, J.: **Some problems and experience in the joint development of compatible thesauri for machine-tool and processing machine manufacture.**

304 Tomasik-Beck, I.: **Development of a general methodology for building national systems of compatible thesauri for the natural sciences and engineering.** A methodology for their application to text indexing.

305 Kazakov, E.N.: Kopylov, V.A.: **Use of computer indeveloping a compatible thesaurus system for the natural sciences and engineering.**

306 Lachuti, D.G.: Pevsner, B.: **On the use of a multilingual machine dictionary in a computer-based descriptor system for searching electrical engineering literature.**

307 Leonteva, T.M. et al: **Guidelines for construction and use of a bilingual chemistry thesaurus.**

308 Mechtiev, D.M.: Chrustaleva, R.A.: **The methodology of automating the compilation and editing of a four-language branch-oriented thesaurus for oil-production.**

309 Maitla, K.: **Building a Russian-Estonian thesaurus and term classification in the thesaurus.**

310 Shrejder, Yu.A.: **Information and meta-information in the thesaurus.**

311 Mojzhishek, I.: **Thesaurus construction methods.**

312 **08.28**
Problems in the design and development of the International Scientific and Technical Information System. vol.1. (Orig. ru). Moskva: Mezhdunarod.Centr. Nauch. i Tekhn. Inform. 1975, 288p.
The volume contains among its 35 contributions some 16 treating problems of compatibility, as e.g.

313 Sokolov, A.V.: **A system of functionally specialized languages for the integrated information system.**

314 Mol'ska, I.: **The SACO information retrieval languages.**

315 Tichonov, I.V.: **Development problems of compatible IR languages for the International Scientific and Technical Information System.**

316 Rosenbaum, H.-D.: **Development of compatible IR languages for the national and international scientific and technical information systems of the CMEA countries.**

317 Kofnovec, L.: **A global outlook for IR language applications and an intermediate language for the International and Technical Information System components.**

318 Keen, M.: **Prospects for IR language compatibility: an analysis of experimental findings. An abstract.**

319 Vilenskaya, S.K.: **Descriptor and hierarchical languages: The problem of compatibility within a common IR system.**

320 Wolff-Terroine, M.: **Multilingual systems and thesaurus compatibility.**

321 Sgall, P.: Benesova, E.: **Feasibility of information language compatibility.**

322 Finogenov, A.S.: **The development procedure of the international multilingual thesaurus for the construction field.**

323 Csendes, B.: **A motor-transport IR language and its compatibility with the IR language of the International Scientific and Technical Information System.**

324 Mildner, H.: **Some experience in the semantic coordination of a Russian-German vocabulary file with associated UDC and IPC numbers.**

325 Belopol'skaya, L.A.: Nefedova, T.A., Vashsh, E.D.: **Some experience in the development of a bilingual vocabulary for the trade and economic information communication network as part of Informchim.**

326 Leont'eva, T.M. et al.: **Main steps in the development of a bilingual (Russian-German, German-Russian) thesaurus for chemistry and chemical engineering.**

327 Gasilina, V.N.: Gavrilenko, B.I., Savel'eva, A.I.: **Salient features of a methodology for bilingual thesaurus construction (in the ferrous metallurgy field)).**

328 Mickov, C. et al.: **The methodology of use of the VNIIMI thesaurus descriptor fields in developing specialized medical thesauri.**

329 **285**
Probst, M.A.: **Thesauri in informatics.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.9 1979, p.14-20, 29 refs.
Questions of correlations between a thesaurus and different classifications, as well as between different branch thesauri are clarified.

330 **285**
Protz, A.: Loehr, H.: **Automatische Konvertierung von Deskriptoren und Notationen in Dokumentennachweisen des Typenprojektes R 300.** Automatic conversion of descriptors and notations in document retrieval systems of Project R 300. (Orig. de). In: Informatik Vol.25 , No.1 1978, p.39-43, 6 refs.

331 **281**
Rada, R.: **Connecting and evaluating thesauri. Issues and cases.** (Orig. en). In: Int.Classif. Vol.14 , No.2 1987, p.63-69, 36refs.

332 **282**
Rada, R.: **Maintaining thesauri and metathesauri.** (Orig. en). In: Int.Classif. Vol.17 , No.3-4 1990, p.158-164

333 **288-71/2**
Rakitin, Yu.M.: **Economic indicator table conversion language.** (Orig. ru). In: Nauchn.-tekhn.inform.,Ser.2 , No.4 1982, p.10-15
Presents a method for compiling and converting economic indicator tables based on the representation of indicator attributes as a composition of independent multitudes of basic trees. Describes the economic indicator table conversion and main implementation algorithms.

334 **284**
Ramus, J.E.: **CONCORD: a word index generator for arbitrary text strings.** (Orig. en). Livermore, CA: Univ.of California, Lawrence Liverm.Lab. 1975, 17 p. NTIS UCID-30126

335 **283**
Ranganathan, S.R.: **Conversion table.** (Orig. en). In: Library Manual. Ed.2. Bombay: Asia 1960, p.387

336 **286**
Ranganathan, S.R.: Neelameghan, A.: **Basis for study of compatibility, and of compatibility of the Colon Classification with the Universal Decimal Classification.** (Orig. en). In: Libr.Sci. Slant Doc. Vol.7 , No.2 1970, p.97-109, 8 refs.

337 **286**
Ranganathan, T.: Neelameghan, A.: **Compatibility of UDC with a freely faceted version of CC: a case study.** (Orig. en). In: DRTC Ann.Seminar 9: Abstracting, indexing, and reviewing periodicals. Comparative study of schemes for library classification Bangalore, IN: Doc.Res. and Training Centre 1975, p.405-433

338 **288-71/2**
Richard, P.: Virville, M.de: **L'automatisation de la construction de tables de correspondences entre nomenclatures économiques.** Automatic construction of correspondence tables between economic nomenclatures. (Orig. fr). In: R.A.I.R.O.,Informat. Vol.13 , No.4 1979, p.375-392

339 **281**
Richmond, P.: **A thesaurus within a thesaurus: a study in ambiguity.** (Orig. en). In: In: Rawski,C.H.(Ed.) Toward a theory of librarianship. Festschrift for Jesse H. Shera. Metuchen, NJ: Scarecrow Press 1973, p.268-301

340 **286**
Riesthuls, G.J.A.: Bliedung, S.: **Thesaurification of the UDC.** (Orig. en). In: In: Fugmann,R.(Ed.) Tools for Knowledge Organization and the Human Interface.

Proc. 1st Int. ISKO Conf., Darmstadt, 14-17 Aug.1990. Frankfurt/Main: INDEKS Verlag 1990/91, p.109-117

341 **285**
Rimbert, M.: **Problčmes posés par l'établissement automatique de réseaux de corrélations au sein de systčme documentaires en grandeur réelle.** Problems raised by the automatic elaboration of correlations networks within natural-size documentation systems. (Orig. fr). In: Probičmes de classification en documentation automatique, No.2, 1974. Also in In: Documentaliste Vol.11 , No.2 1974, p.89

342 **285**
Rindflesch, T.C.: Aronson, A.R.: **Ambiguity resolution while mapping free text to the UMLS Metathesaurus.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1994. p.240-244 ISSN 0195-4210

343 **286**
Rodriguez-Delgado, R.: **Towards the integration of decimal classifications and thesauri concepts.** (Orig. en). In: New Trends in Documentation and Information. FID Congress at Edinburgh, 25-28 Sept.1978 London, GB: Aslib 1978, 1p. Summary of paper only.

344 **287**
Rosenbaum, D.: **Development of compatible retrieval languages for the national and international information systems of Comecon member countries.** (Orig. de). In: Informatik Vol.21 , No.4 1974, p.28-30

345 **288-51/4**
Rosenberg, K.M.: Coultas, D.B.: **Acceptability of Unified Medical Language System terms as substitute for natural language general medicine clinic diagnoses.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1994. p.193-197 ISSN 0195-4210

346 **288-56**
Roulin, C.: **Towards the European Education Metathesaurus.** (Orig. en, fr). Brussels: Eurydice European Unit 1990 ISBN: 2-87116-150-X

347 **288-56**
Roulin, C.: **Sub-thesauri as part of a metathesaurus.** (Orig. en). In: In: Williamson,N. Hudon, M. (Eds.): Classification Research for Knowledge Representation and Organization. Proc.5th Int.Study Conf.on Classif.Res. 1991 Amsterdam, etc.: Elsevier 1992, p.329-336

348 **283**
Roulin, C.: Cooper, C.: **Bringing thesauri together: terminological problems and interests.** (Orig. en). In: In: Schmitz,K.-D.(Ed TKE'93: Terminology and Knowledge Engineering. Proc.3rd Int.Congress on Terminology and Knowledge Engineering, 25-27 Aug.1993, Cologne, Germany. Frankfurt/Main: INDEKS Verlag 1993, p.251-259

349 **283**
Ruchimskaya, E.M.: Tafipolskaya, M.Ya.: **The possibility of lexicon enrichment of an information retrieval thesaurus using the language of subject rubrics.** (Orig. en from ru). In: Autom.Doc.Math.Ling. Vol.20 , No.5 1986, p.76-87, 14 refs.
A comparative analysis of the subject heading language (SHL) and an information retrieval thesaurus (IRT) shows that the structures of these information languages are generally compatible and that an SHL can be used to expand an IRT vocabulary.

350 **288-6**
Sager, J.C.: Somers, H.L., McNaught, J.: **Thesaurus integration in the social sciences. Pt.1: Comparison of thesauri.** (Orig. en). In: Int.Classif. Vol.8 , No.2 1981, p.133-138, 45 refs.
The comparison considered 1) the knowledge structure (coverage, general descriptors, classification, hierarchies), 2) the linguistic structure (terms and descriptors, word categories, choice of descriptors and their form, relationships), and 3) formal features (typography, address codes, symbols, ordering, spelling variants).

351 **288-6**
Sager, J.C.: Somers, H.L., McNaught, J.: **Thesaurus integration in the social sciences: Pt.II: Stages towards integration.** (Orig. en). In: Int.Classif. Vol.9, No.1 1982, p.19-26
The methodology of the study is described and methods for a successful integration are proposed, involving a descriptor bank as an interim stage towards the objective of an integrated thesaurus. A checklist for the full linguistic, formal and technical description of thesauri is introduced.

352 **288-6**
Sager, J.C.: Somers, H.L., McNaught, J.: **Thesaurus integration in the social sciences. Pt.III: Guidelines for the integration of thesauri.** (Orig. en). In: Int.Classif. Vol.9 , No.2 1992, p.64-70
The guidelines - in conformity with international standards - take into account user requirements, the need for an exchange format of data and a suitable database management system.

353 **288-72**
Sakov, A.A.: Sonechkina, E.E.: **Progress with the common techno-economic information classification and coding system.** (Orig. ru). In: Problemy nauchnogo upravlenija socialisticheskim obshchestvom. Latvijskij Gos. Univ. 1977, p.32-43

354 **287**
Samuelson, K.: **Coordination of diversified data-bases and information networks through multiclassificatory concept notation.** (Orig. en). In: In: Mölgaard'H,R. Proc. 2nd Seminar on UDC and Mechanized Inform.Systems. Copenhagen, DK: Danish Centre for Doc. 1971, p.213-223

355 **288-56**
Sanllorenti, A.M.: **FID/CLA/LI: Compatibility experience of indexing languages in the education field.** (Orig. en). In: Int.Classif. Vol.11 , No.1 1984, p.28--30

356 **286**
Saraeva, G.A.: **On vocabulary compatibility between conventional and non-conventional indexing languages.** (Orig. ru). In: Predm.poisk.v tradicion. i netradicion.inform.-poiskov.sistemakh- Leningrad: 1980, p.86-103, 5 refs.
The conventional languages compared were the LBC and the subject catalogue language of the M.E.Saltykov-Shchedrin State Public Library and the non-conventional ones were natural language texts and key terms taken from document titles.

357 **286**
Schadlich, T.: **Changing from Sears to LC Subject Headings.** (Orig. en). In: Libr.Res. Techn.Serv. Vol.24 , No.4 1980, p.361-3, 1 ref.
The paper provides a quantitative evaluation of the compatibility of Sears and LC Headings.

358 **286**
Schubert, P.: **Probleme der Sacherschließung von Informationsquellen aus der Sicht einer Fachbibliothek.** Problems of indexing seen from the point of view of a special library. (Orig. de). In: Zentralblatt f.Biblwes. Vol.91 , No.7 1977, p.319-322, 15 refs.
It is recommended to use a classification system and a thesaurus for subject indexing in a special library of a company where the thesaurus ought to be made compatible with the classification system.

359 **288-51/4**
Schuyler, P.L.: Hole, W.T., Tuttle, M.S., Sherertz, D.D.: **The UMLS Metathesaurus: representing different views of biomedical concepts.** (Orig. en). In: Bull.Med.Libr.Assoc. Vol.81 , No.2 1993, p.217-222

360 **288-26**
Schreiber, H.: Bauer, C., Lange, B.: **Descriptor list with UDC- and IPC-notations added.** (Orig. de). In: Informatik Vol.1 22, p.24-26, refs.

361 **286**
Scibor, E.: **Methods of linking different indexing languages with particular attention to the UDC and descriptor languages.** (Orig. pl). Warszawa: IINTE 1974

362 **287**
Scibor, E.: **Plans concerning the establishment of Polish thesauri compatible with thesauri in other languages.** (Orig. en). In: In: Krumholz,W. Int.Org.for Standardization (ISO/TC46) and UNESCO Int.Scient.Symposium on Multilingual Thesauri. 8-10 Oct.1973, Berlin, DE, Proc. Berlin, DE: Leitstelle Politische Dokumentation 1973

363 **281**
Scibor, E.: **Methods of thesauri integration.** (Orig. pl). In: Aktual. Probl.Inform.Dok. Vol.30 , No.3 1985, p.9-14

364 **284**
Scibor, E.: **Indexing languages concordances: methodological issues.** (Orig. ru). In: Probl.Inform.Sistem , No.1 1986, p.45-63, 5 refs.
The compatibility of the ISTIS indexing languages with other indexing languages is analysed using the concordances. Presentation of structure and composition of the concordances.

365 **284**
Scibor, E.: Tomasik-Beck, J.: **On the establishment of concordances between indexing languages having a universal or interdisciplinary scope (Polish experience).** (Orig. en). In: Knowl.Org. Vol.21 , No.4 1994, p.203-212, 10 refs.
Report on investigations conducted at the Institute for Scientific, Technical and Economic Information (ISTEI) in Warsaw in 1992-1993 which aimed at a comparative analysis of four indexing languages developed and used in Poland and at the establishment o concordances between them.

366 **285**
Scott, M.L.: **Conversion tables: LC-Dewey, Dewey-LC.** Englewood, Colo.: Libraries Unlimited 1993. VIII,365p.

367 **288-6**
Sebestyen, G.: **The international integration effort and its impediment by the indexing languages used in social science IR.** (Orig. hu). In: Tud.es müsz.tajek. Vol.35 , No.8 1988, p.335-341, 8 refs.

368 **286**
Sechser, O.: **A note on the substitution of one document retrieval language into another document retrieval language.** (Orig. en). In: Prague Stud. Math. Linguist. Vol.4 1972, p.241-254

369 **288-53**
Seetharama, S.: Gopinath, M.A.: **Classification of books on disease in medicine: a comparative study of CC and UDC class numbers.** (Orig. en). In: DRTC Annual Seminar 9. Abstracting, indexing and reviewing periodicals. Pt 1: Papers Bangalore, IN: Doc.Res. and Training Centre 1971, p.446-466

370 **288-5332**
Seetharama, S.: **Compatibility among medical classification systems: a case study in the classification of cardio-vascular diseases.** (Orig. en). In: Int.Classif. Vol.12 , No.2 1985, p.80-86, 4 refs.
The study has been made at near-seminal level by the application of the Normative Principles of (Ranganathan's) General Theory of Classification and at the phenomenal level by classifying a set of a few hundred cardiovascular pathological conditions.

371 **282**
Shemakin, Yu.I.: Kulik, A.N.: **On information retrieval system compatibility and the role of a polytechnical thesaurus in a unified information network.** (Orig. ru). In: Nauch.-tekhn.inform.,Ser.2 , No.8 1970, p.30-34, 21 refs.
Includes a definition of IRS compatibility.

372 **288-51/4**
Shulova, I.E.: Lomova, N.F., Solov'eva, A.A. et al.: **Main principles underlying the reconciliation of the vocabulary of the Subject Headings List of the State Central Medico-Scientific Library.** with the descriptors of a thesaurus in medicine and health care. (Orig. ru). In: Med. Naukovedenie i

avtommatiz.inform.processov Moskva: 1984, p.176-177

373 **288-32**
Silvester, J.P.: Newton, R., Klingbiel, P.H.: **An operational system for subject switching between controlled vocabularies. A computational linguistics approach.** (Orig. en). McLean, VA: Planning Research Corp. 1984, 105p. ERIC ED 260 735
Description of a new automated process that pioneers full-scale operational use of subject switching by the NASA Scientificand Technical Information Facility. The switching is based on the NASA Lexical Dictionary (NLD).

374 **288-32**
Silvester, J.P.: Klingbiel, P.H.: **An operational system for subject switching between controlled vocabularies.** (Orig. en). In: Inform.Process.& Management Vol.29 , No.1 1993, p.477-59, 10 refs.

376 **286**
Sincai, A.: **Compatibility between traditional information systems and modern ones.** (Orig. ro). In: Probl.Inform.Doc. Vol.12 , No.3 1978, p.115-122, 26 refs.
The UDC is proposed as a common instrument for the realization of compatibility between information systems. Examples are given.

377 **286**
Singh, S.: Sobel, S.: **United Nations Bibliographic Information System (UNBIS).** (Orig. en). In: UNESCO J.Inform.Sci.Libr.Arch. Vol.2 , No.4 1980, p.220-226, 5 refs.
UNBIS was developed to provide expanded reference and information services, and to achieve comprehensive bibliographic control of United Nations documentary and externally acquired materials needed for the programmes and activities of the Organization. The primary objective of this online system development has been to achieve optimum compatibility with other international systems by following international standards for bibliographic description, by use of a vocabulary aligned with other major international lists and thesauri, and by achieving a one-to-one match of data elements with those used by the existing international information systems.

378 **287**
Sinyutina, S.V.: **All-Union classification schemes - the central component of the formalized data description tools.** (Orig. ru). In: Inform.yaz. i bazy dannikh v planir. Moskva, SU: 1980, p.87-94, 8 refs.

379 **287**
Skolek, J.: **The collection of retrieval languages of the International System for STI(MSNTI) - 3.** (Orig. cz). In: Cheskoslovenska Informatika Vol.21 , No.5 1979, p.136-141, 9 refs.

380 **281**
Slype, G. van: **Les langages documentaires dans la perspective utilisateur des reseaux internationaux d'IDIST.**
Documentary languages from the point of view of a user of international information systems. (Orig. fr). In: New Trends in Documentation and Information. Papers 39th FID Congress, Edinburgh, GB London, GB: Aslib 1978, 1 p.
Summary of paper with the proposals to harmonize documentary languages by the following measures: establishment of macrothesauri, 'languages de passage', thesaurus reconciliation, universal classification system, universal thesaurus, and natural languages.

381 **288-51/4**
Smith, L.C.: **Systematic searching of abstracts and indexes in interdisciplinary areas.** (Orig. en). In: J.Amer.Soc.Inform.Sci. Vol.25 , No.6 1974, p.343-353, 42 refs.
In order to permit systematic searching of multiple databases to satisfy user requests in interdisciplinary areas, a mapping of a portion of Medical Subject Headings (MeSh) to three other controlled vocabularies has been constructed. Report on the results of the mapping.

382 **281**
Smith, L.C.: **UNISIST revisited: Compatibility in the context of**

collaboratories. (Orig. en). In: In: Williamson,N. Hudon, M. (Eds.): Classification Research for Knowledge Representation and Organization. Proc.5th Int.Study Conf.Classif.Res., Toronto, 1991 Amsterdam, etc.: Elsevier 1992, p.337-344, 30 refs.

383 **284**
Smith, P.H.: Siberz,J.K., Devine,J.G., Well,J.C., Painter,J.A.: **Concordances.** (Orig. en). In: Literary Data Processing Conf. 1964 New York, NY: Materials Center, Mod.Language Assoc. 1964, p.113-170, 5 refs.

384 28.07
Soergel, D.: **Dokumentation und Organisation des Wissens - Versuch einer methodischen und theoretischen Grundlegung am Beispiel der Soziaiwissenschaften.** Documentation and Organization of Knowledge. An inquiry into the methodological and theoretical foundations with particular reference to the social sciences. Berlin: Duncker u. Humblot 1971. 380 p.
Contains in section B6. "comments and proposals" for the practical organization of cooperative information systems.

385 **282**
Soergel, D.: **A universal source thesaurus as a classification generator.** (Orig. en). In: J.Amer.Soc.Inform.Sci. Vol.23 , No.5 1972, p.299-305

386 **281**
Soergel, D.: **A general model for indexing languages: The basis for compatibility and integration.** (Orig. en). In: In: Wellisch, H.H. Subject Retrieval in the Seventies: New Directions. 1971. Westport, CT: Greenwood 1972, p.36-61, 10 refs.

387 **281**
Soergel, D.: **The idea of a Universal Source Thesaurus (UST).** (Orig. en). In: In: Soergel, D. Indexing languages and thesauri: Construction and maintenance Los Angeles, CA: Melville Publ.Co. 1974, p.516--520

388 **281**
Soergel, D.: **Compatibility of vocabularies.** (Orig. en). In: In: Riggs, F.W. The CONTA Conference. Proc.Conf.Conceptual & Terminol.Analysis in the Social Sciences, Bielefeld, May 24-27, 1981 Frankfurt: INDEKS Verlag 1982, p.209-223, 7 refs.

389 **289**
Sokolov, A.V.: **Compatibility of empirical microthesauri.** (Orig. ru). Moskva, SU: VINITI 1970, 23 p.
Based on an analysis of 9 microthesauri an attempt is made to evaluate the quantitative characteristics of their compatibility (ratio of common to unique terms) and the qualitative aspect of compatibility which is directly related to the problem of polysemy. Conclusions are given.

390 **287**
Sokolov, A.V.: **A possible approach to ensuring IRS compatibility (Universal 'fable' for information thesauri).** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.1 1977, p.19-24, 10 refs.
Proposal for a method for the construction of compatible thesauri based on a universal scheme of facets, referred to as the 'fable'. The scheme was used in the construction of thesauri in various fields of engineering. Its universal character was assessed in a comparison with the ERIC thesaurus.

391 **287**
Sokolov, A.V.: **Some problems in the type design of IR thesauri.** (Orig. ru). In: Struktur i priklad.lingv. , No.1 1978, p.172-181, 6 refs.
The type design should be such that it allows for compatibility among different thesauri. An IR thesaurus "plot" is introduced to mean the scheme of facets and lexical blocks, representing the thesaurus contents.

392 **287**
Sokolov, A.V.: Makevich, I.A., Fedotova, A.V.: **An experiment with the compiling of compatible, narrow-thematical information retrieval thesauri.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.10 1978, p.3-12, 8 refs.
Category-derived schemes permit the distribution of a descriptor vocabulary in

separate facets and blocks transferable from one thesaurus to another. The analysis of thesauri in the field of Navigation and Seaports confirmed the possibility to constructthesauri compatible in their facets and blocks. (English transl. in Autom.Doc.& Math.Linguist. 12(1978)No.4, p.8-19, 8 refs.).

393 **288-51/4**
Sperzel, W.D.: Tuttle, M.S., Olson, N.E., Erlbaum, M.S. et al.: **The Meta-1.2 engine: a refined strategy for linking biomedical vocabularies.** (Orig. en). In: Proc. Ann.Symp.Comput.Appl.Med.Care 1992 1992, p.304-308 ISBN: ISBN0195-4210

394 **286**
Spranger, U.: **Zusammenführung verschiedensprachiger und unterschiedlich strukturierter Thesauren unter einem gemeinsamen Code.** Joining multiple-language and differently structured thesauri under a common code. (Orig. de). In: Informatik Vol.20 , No.6 1973, p.15-17, 2 refs.

395 **286**
Shrejder, Yu.A.: **Interaction of classified and object languages.** (Orig. ru). In: Nauchno-tekhn. inform., Ser.2 , No.1 1978, p.20-23
The opposition of classified and object languages is discussed as well as the possibility of their combined use.

396 **288-51/4**
Stitt, F.W.: **The problem-oriented medical synopsis: coding, indexing, and classificatin sub-model.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1994. p.964 ISSN 0195-4210

397 **281**
Stys, T.: **Kompatibilitätsprobleme und Möglichkeiten ihrer Lösung.** Compatibility problems and possibilities for their solution. (Orig. de). In: Informatik Vol.32 , No.3, 4 1985, p.32-6, 27-29, 32 refs.
Syntactical, semantic and pragmatic problems of compatibility are treated.

398 **281**
Sukiasyan, E.R.: **Centralized classification: achievements and problems in regard to future development.** (Orig. en). In: UNESCO Bull.Libr. Vol.22 , No.4 1968, p.189-195

399 **283**
Svenonius, E.: **Translation between hierarchical structures: an exercise in abstract classification.** (Orig. en). In: In: Neelameghan, A. Ordering Systems for Global Information Networks. Proc.3rd Int.Study Conf.on Classif.Res., Bombay, Jan.6-11, 1975 Bangalore, IN: Sarada Ranganathan Endowment f.Libr.Sci. 1979, p.147-211
Consideration of difficulties arising from differences in classification schemes when attempting to translate from one scheme to another. Examples are taken froom the LCC and DDC schedules for architecture. The examples, the,are generalized and discussed in terms of a model (abstract) classification. The model indicated an approach toward measuring the descriptive power of classification. It also illustrated the point that a switching language switches by virtue of the fact that the ultimate semantic factors in the participating languages are similar.

400 **28.06.82-10-17**
Svenonius, E.: **Compatibility of retrieval languages. Introduction to a Forum.** (Orig. en). In: Int.Classif. Vol.10 , No.1 1983, p.2-4, 19 refs.
The 'Forum' was a conference held prior to the ASIS Annual Meeting at Columbus Ohio.

401 **281**
Svenonius, E.: **Use of classification in online retrieval.** (Orig. en). In: Libr.Resources Techn.Serv. Vol.27 , No.1 1983, p.76-80, 7 refs.
Points out a.o. that classification can be used to achieve compatibility of retrieval languages by serving as a mediating or switching language.

403 **286**
Swanson, G.L.: **Dewey to LC conversion tables.** (Orig. en). New York, NY: CCM Information Corp. 1972

404 **286**
Szczepanek, W.: **Choice of an association method of information retrieval languages exemplified by a keyword language and a thesaurus.** (Orig. pl). In: Aktual.Probl. Inform.Dok. Vol.26 , No.1 1981, p.7-10

406 **288-46**
Thomas, S.: **Unified Agricultural Thesaurus discussed.** (Orig. en). In: Agricult.Libr.Inform.Notes Vol.17 , No.1 1991, p.12-14

407 **288-51/4**
Tuttle, M.S.: Sherertz, D.D., Erlbaum, M.S. Sperzel, W.D. et al.: **Adding your terms and relationships to the UMLS Metathesaurus.** (Orig. en). In: Proc.Ann.Symp.Comput.Appl.Med.Care 1991. 1991, p.219-223 ISBN: ISSN0195-4210

408 **288-51/4**
Tuttle, M.S.: Sperzel, W.D., Olson, N.E., Erlbaum, M.S. et al.: **The homogenization of the Metathesaurus schema and distribution format.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1992. p.299-303 ISSN 0195-4210

409 **288-51/4**
Tuttle, M.S.: Nelson, S.J.: **The role of the UMLS in 'storing' and 'sharing' across systems.** (Orig. en). In: Int. J. Biomed. Comput. Vol.34 , No.1-4 1994, p.207-237

410 **281**
Ungvary, R.: **The structural framework of a thesaurus system.** (Orig. hu). In: Informatika, Budapest , No.3 1974, p.3-12, 19 refs.
A higher level thesaurus system is proposed which can be used to generate compatible and expandable working thesauri.

411 **28.06.74-09-10/12**
Unified IR Language System within the International Scientific and Technical Information System. (Orig. ru). Moscow: Mezhdunard. Cent. Nauchn. i Tekhn. Inform. 1974, 44p.
Program and abstracts of an international symposium held in Moscow, Sept.10-12, 1974. Section 2 of the symposium was devoted to IR language compatibility problems.

412a **281**
UNISIST Study Report on the feasibility of a World Science Information System. (Orig. en). Paris: UNESCO 1971, 161p.
Contains essential recommendations for the establishment of compatibility between order systems on the macro- and microlevel. Suggested the establishment of a Broad System of Ordering as a switching mechanism between universal and special classificationsystems as well as mechanisms to correlate the descriptors of different thesauri in one and the same field.

412b **281**
UNISIST. Etude sur la réalization d'un systčme mondial d'information scientifique. UNISIST. Study report on the feasibility of a World Science Information System. (Orig. en, fr). Paris: UNESCO 1971, 181p.

413 **281**
Ustimenko, A.I.: Grinina, R.F.: **Toward the problems of compatibility between the descriptor languages and the subject heading language.** (Orig. ru). In: Sostoyanie predmet.katalogov bibliotekakh SSSR 1978, p.193-211, 18 refs.
Grammatical devices were chosen for descriptor articulation as subject headings without additional reference to the contents of documents. Description of an experiment to test this methodology.

414 **2886**
Van Laer, C.J.P.: **UDC adapted to Library of Congress Classification: the compatibility of two universal classifications.** (Orig. nl). In: Open Vol.14 , No.4 1982, p.147-155, 6 refs.

415 **284**
Vickery, B.C.: **Coding for interconvertibility.** (Orig. en). In: In: Kent, A. (Ed.) Information Retrieval and Machine Translation. New York, NY: Interscience 1960/61, p.1159-1174

416 **281**
Vickery, B.C.: Slater, M., Pesanis, A., Reynolds, R.: **Classification in science information.** A comparative study undertaken by Aslib for the International Council of Scientific Unions. (Orig. en). Paris: UNESCO 1969
Presented as document UNISIST/CSI/5.8 at the 4th Session of the Central Committee to study the feasibility of a world science information system at UNESCO House, Paris, 15-17 Dec.1969.

417 **288-657**
Viet, J.: **Intégration de thésauri dans le champ du développement économique et social.** Thesaurus integration in the field of economic and social development. (Orig. fr). In: Documentaliste Vol.14 , No.5-6 1977, p.6-12
History of the OECD Macrothesaurus and revision plans taking care of the descriptors occurring in new thesauri of subfields of this Macrothesaurus and for integration methods for their reconciliation.

418 **282**
Vilenskaya, S.K.: **Compatibility of informtion retrieval languages (intermediary language and commutation language).** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.5 1977, p.16-21, 26 refs.

419 **282**
Vilenskaya, S.K.: **On the compatibility of different information retrieval languages within the integrated information system.** (Orig. en). In: In: Taylor,P.J.(Ed.) New trends in documentation and information. Proc.39th FID Congress. Univ.of Edinburgh, 25-28 Sept.1978. London: Aslib 1980, p.315-325, 15 refs.

420 **288-71/2**
Vilenskaya, S.K.: Bezrukov, A.: **The comparison of the MISON Rubricator and the UDC schedule for Economics. Economic Sciences.** (Orig. en). Moscow: USSR Acad.Sciences, INION 1981, 41p. 2680
Paper presented at the ECSSID WG3 Meeting, Prague, 9-11 Sept.1980

421 **287**
Vinogradov, S.N.: Lomakina, O.A., Nemchenko, V.N. et al.: **Integrating subsectoral thesauri into a sectoral thesaurus.** (Orig. ru). In: Nauchno-tekhn.inform.,Ser.2 , No.10 1986, p.16-18, 4 refs.

422 **285**
Virville, M.de: **Automatisation of correspondence tables between economic nomenclatures.** (Orig. fr). Paris: CNRS 1975, 334p., 38p.refs. DGRST-7371677

423 **288-55**
Vitukhnovskaya, A.A.: **The unit approach to thesaurus construction and the problem of compatibility.** (Orig. ru). In: Vopr. bibliotekoved. i bibliogr. Vol.4 Leningrad, SU: 1974, p.155-164, 7 refs.
The diagram of a fragment of the general structure of a psychology thesaurus is presented.

424 **286**
Vituchnovskaya, A.A.: **On the terminological linkage between a linguistics and a general psychology thesaurus.** (Orig. ru). In: Sborn.aspirantsk.rabot., Vyp.8 Leningrad: Institut Kultury 1976, p.82-98, 6 refs.

425 **281**
Vituchnovskaya, A.A.: **A contribution to the theory of information retrieval thesaurus compatibility.** (Orig. ru). In: Nauchn. i tekhn. biblioteki SSSR , No.11 1976, p.6-13, 20 refs.
A thesaurus development theory is advanced whereby a thesaurus is built from large semantic units. One unit may form a part of different specialized thesauri, whereby providing for vocabulary interchange and thesaurus compatibility.

426 **289**
Vladutz, G.: **Problems of compatibility between thesaurus and classification in information retrieval: advantages and disadvantages.** (Orig. it). In: In: Salimei, M. Vladutz, G.; Politica della documentazione. seminari sull'organizzazionee sviluppo dei

servizi d'informazione Roma: Ed.dell Ateneo & Bizarri 1978

427 **287**
Wahlin, E.: **The AR Complex. Adapted systems used together with a common reference system.** (Orig. en). In: Proc.Int. Symp. UDC in relation to other indexing languages, Herceg Novi, 28 June - 1 July 1971 Beograd: Yugoslav Center f. Techn.Scient.Doc. 1972, 24p.

428 **287**
Wahlin, E.: **The AR Complex. Adapted systems used in combination with a common reference system.** (Orig. en). In: In: Wojciechowski,JA Conceptual Basis of the Classif.of Knowledge. Proc.Ottawa Conf., Oct.1-5, 1971. München: Verlag Dokumentation 1974, p.416-449

429 **287**
Wahlin, E.: **Special classification systems used together with a common reference system. The AR Complex.** (Orig. en). In: In: Neelameghan,A.: Ordering Systems for Global Information Networks. Proc.3rd Int.Study Conf. on Classif.Res., Bombay, 6-11 Jan.1975. Bangalore, IN: Sarada Ranganathan Endowment f.Libr.Sci. 1979, p.407-22, 8 refs.

430 **282**
Wahlin, E.: **Der AR-Komplex - Spezial-klassifikationssysteme in Kombination mit einem gemeinsamen Referenz-system.** The AR-Complex - special classification systems in combination with a common reference system. (Orig. de). In: DK-Mitt. Vol.23 , No.1-4 1979, p.1-3, 5-9
Description of the AR-Complex and examples from difernt fields, such as building, commodity classification, classification of professions, a.o.

431 **283**
Wall, E.: Barnes, J.M.: **Intersystem compatibility and convertibility of subject vocabularies.** (Orig. en). Philadelphia, PA: Auerbach 1969, 138p. PB 184 144

432 **286**
Wall, R.A.: **Intelligent indexing and retrieval: A man-machine partnership.** (Orig. en). In: Inform.Process.& Management Vol.16 , No.2 1980, p.73-90, 40 refs.
Contains a.o. the proposal for a Universal Reference Vocabulary maintained online only and used to construct profiles before searching via natural language and/or class numbers.

433 **283**
Wall, E.: Landau, H.: **Interconvertibility of subject indexes.** (Orig. en). In: Soc.Impact Inform.Retr. Inform.Bazaar. 7th ann.Natl. Colloq. Philadelphia, PA: Coll.Physicians 1970, p.103-112, 2 refs.

434 **286**
Wall, R.A.: **Retrieval of multi-media information: simulation of a proposed system.** (Orig. en). Loughborough, GB: Univ.of Technology 1972, 350p. PhD Thesis
Part A is concerned with a reconciliation of a classification system (ZDC) with thesaurus terms.

435 **286**
Wall, R.A.: **Indexing language structure for automated retrieval.** (Orig. en). In: Inform.Storage Retrieval Vol.9 , No.11 1973, p.607-617
Reconciliation of a thesaurus with the UDC. The higher hierarchies are taken care of by the UDC and the lower ones by the thesaurus.

436 **284**
Wellisch, H.: **A concordance between UDC and Thesaurus of Engineering and Scientific Terms (TEST). Results of a pilot project.** (Orig. en). In: UDC in Relation to other Indexing Languages. Proc.Int.Symp., Herceg Novi 1971. Beograd, YU: Yugoslav Center f.Techn.Scient.Doc. 1972, 33p. Paper 20.

437 **286**
Wellisch, H.H.: **A concordance between the UDC and TEST (Thesaurus of Engineering and Scientific Terms).** (Orig. sh). In: Informatika, Yugosl. Vol.8 , No.2 1974, p.37-51
The study proved that it is possible to find UDC-numbers for the majority of TEST-descriptors.

439 **283**
Wersig, G.: **Experiences in compatibility research in documentary languages.** (Orig. en). In: In: Neelameghan, A. Ordering Systems for Global Information Networks. Proc.3rd Int.Study Conf.on Classif.Res., Bombay, 6-11 Jan.1975 Bangalore: Sarada Ranganathan Endowment f.Libr.Sci. 1979, p.423-430, 8 refs.
Presents theoretical considerations on the topic as well as results of a study of compatibility among the documentary languages used by the information systems of the federal agencies of Germany, based on an analysis at the syntactic, semantic, and pragmatic levels.

440 **282**
Wersig, G.: **Dokumentationssprachliche Probleme bei Verbundsystemen.** Problems for indexing languages in information networks. (Orig. de). In: Nachr.Dok. Vol.26 , No.6 19775, p.231-234, 12 refs.
Discussion of possibilities to overcome the problems of differing indexing languages in computerized information networks.

441 **288-6**
Whitelock, P.J.: **A descriptor bank of social science terms.** (Orig. en). In: Int.Classif. Vol.9 , No.3 1982, p.145-151

442 **288-652**
Wilson, P.T.: Spitzer, R.L.: **A comparison of three current classification systems for mental retardation.** (Orig. en). In: Amer.J.Ment.Deficiency Vol.74 , No.3 1969, p.428-435

443 **282**
Wolff-Terroine, M.: **A macrothesaurus - why? how?.** (Orig. en). In: In: Neelameghan,A. Ordering Systems for global Information Networks. Proc.3d Int.Study Conf.Classif.Res., Bombay, 6-11 Jan.1975. Bangalore, IN: Sarada Ranganathan Endowment Libr.Res. 1979, p.431-436

444 **281**
Wolff-Terroine, M.: **Multilingual systrems and thesaurus compatibility.** (Orig. en). In: Problems in the design and development of the Int.Scient.and Techn.Inform.System, Vol.1 Moskva, SU: VINITI 1975, p.145-149

445 **282**
Wolff-Terroine, M.: Plante, S., Gruson, C.: **Etude de faisabilité d'un macrothésaurus.** Feasibility study of a macrothesaurus. (Orig. fr). Paris: Bureau Natl.de l'Inform.Sci.& Techn. 1977, 62p. BNIST-291-758
Discussion of the methodology employed for the creation of a linguistic tool to allow compatibility between the documentary databases and a macroindexing of documents. Experimental indexing tests are included as well as the progamming for the processingand editing.

446 **288-51/4**
Yang, Y.: Chute, C.G.: **A schematic analysis of the Unified Medical Language System.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med. Care 1991. p.204-208 ISSN 0195-4210

447 **287**
Zeng, Lei: **Establishing a compatible general vocabulary in China: The capability.** (Orig. en). In: Int.Classif. Vol.17 , No.2 1990, p.91-98, 11 refs.
After a general analysis of the factors influencing compatibility between thesauri, the compatible capabilities between the special thesauri existing in China and the general ones as well as between Chinese thesauri and English ones are discussed.

448 **281**
Zeng, M.Lei: **Compatibility of indexing languages in an online access environment: A review of the approaches.** (Orig. en). In: In: Fidel, R. et al Proc.3rd ASIS SIG/CR Classification Research Workshop, Pittsburgh, PA, Oct.25, 1992 Medford, NJ: Learned Inform. 1993, p.161-181
Reprint from "Achieving compatibility of indexing languages", in Kent, A. (Ed.): Encyclopedia of Library and Information Science 50(1992) with some updating and revising. An excellent survey including a rtime schedule of works from 1957-1990.

449 **287**
Zonnevend, P.: Skripkin, W.: **Fragen der Entwicklung des Komplexes fon Informationsrecherchesprachen im Internationalen System der wissenschaftlichen und technischen Information.** Questions concerning the development of a complex of thesauri in the International System of Scientific and Technical Information. (Orig. de). In: TH Ilmenau. INER. 10.Kolloq. über Inform.u.Dok., Themenkreis 1 Ilmenau, GDR: INER 1977, p.25-32
Survey on the planning of a thesaurus complex for which an number of special compatible thesauri are elaborated which are not combined by a roof thesaurus.

450
288-546
Zielstorff, R.D.: Cimino,C., Barnett, G.O., Hassan, L., Blewett,D.R.: **Representation of nursing terminology in the UMLS Metathesaurus: a pilot study.** (Orig. en). In: Proc. Ann. Symp. Comput. Appl. Med.Care 1992. p.392-6 ISSN 0195-4210

RECOMMENDATIONS OF THE RESEARCH SEMINAR ON COMPATIBILITY AND INTEGRATION OF ORDER SYSTEMS

ORGANIZED BY THE INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO) AND THE SOCIETY FOR PROFESSIONAL INFORMATION (TIP) Warsaw, Poland, September 13-15, 1995

More than 60 information scientists and practitioners from 12 Western and Eastern Countries gathered in a joint study conference in Warsaw from September 13-15 to discuss matters of key concern to facilitating access to information across national and disciplinary boundaries. The conference organizers gratefully acknowledge support from the Foundation for German-Polish Cooperation, the British Council, and the European Commission.

The experts agreed that the global information society of the 21st century will rely increasingly on an information infrastructure which must have two essential components: the global telecommunication and electronic networks epitomized by the Internet; and, underpinning tit, a conceptual infrastructure reflecting the way knowledge and information are organized.

The experts further agreed on the following views and recommendations.

STATEMENT ON NEEDS AND POLICIES

It was recognized that

— The development and growth of international communication and information exchange are fundamental to a well-informed Information Society.

— There is a risk of a new division of people in the world between those who are able to communicate using new systems and technologies and those who are left out.

To prevent or at least mitigate such a negative development and its implications for social and economic disparity, and to further global understanding and European integration, the following goals and policies should be pursued.

— There is an urgent need for tools and functions that can overcome present barriers in the exchange of information between systems.

— These tools must also enable access to information in familiar languages.

— There is a need for tools that afford the user the opportunity for insight in the structure of an information system as well as guidance and orientation. Such functions must be optional in order to preserve the user~s freedom of decision in access and use of information.

There is a need for systems that give the user an overview of what information is available within a coherent structure and with descriptions that allow for the evaluation and validation of information sources.

All people need education in understanding and using information systems. Such education should cover not only issues but also social and ethical issues.

DETAILED RECOMMENDATIONS

1 Recommendations on Research, Education, and Information Exchange

1.1 Further investigations should be carried out into the principles and methodology of establishing concordances between order systems with particular attention to ordering systems of different structure.

1.2 Further investigations should be carried out into the benefits from and requirements for compatibility. Compatibility should not be pursued for its own sake but for specific objectives to be defined on a case by case basis. For instance, consider using existing thesauri and classifications as knowledge basis for machine-aided interactive processing of natural language queries.

1.3 In the education of information specialists knowledge organization (classification and indexing as well as knowledge representation and terminology) must be a central concern.

1.4 International exchange on the principles of knowledge organization with particular emphasis on cross-cultural comparison of ordering systems should be intensified through seminars and conferences, an INTERNET news group, faculty exchange, and publications. Specifically, an international inventory of software packages and other tools for the development of maintenance of order systems and correspondence among them.

2 Recommendations on system development and national and international collaborative efforts

2.1 Development of correspondence among order systems to support information exchange on national and international levels.

2.2 Long-range development of an open, multifunctional, multilingual, integrated knowledge base of concepts and terminology that preserves the integrity of the many sources on which it draws.

This open system should allow many contributors and be usable by many levels of users for improved information exchange, specifically for the following purposes:

— Tool for searching, particularly knowledge-based support for end-user searching on the Internet and other online services in multiple languages independent of the language used in each database.

— Tool for indexing, esp. basis for collaboration for more effective use of indexing effort.

— Conceptual basis for knowledge-based systems.

— Dictionary — in monolingual, bilingual, and multilingual mode – for human use.

— Dictionary/knowledge base for automated language processing – including natural language understanding and machine translation.

— Classification/ontology for data element standardization.

— Source for the extraction/development of specialized ordering systems, dictionaries, etc., both machine-readable and printed.

Such an open knowledge base would lead to savings in development effort and a potentiation of usefulness through the assembly of rich information from many sources that complement each other, and through establishing relationships among the concepts and terms from different sources.

2.3 Development of auxiliary thesauri for general use such as thesauri for geographical names, bibliographic forms, and languages.

2.4 International cooperation in the establishment and maintenance of authority files of names of persons, organizations, etc.

2.5 Development of criteria, methodologies, tools, and software for establishing, maintaining, converting, integrating, harmonizing, restructuring and translating of monolingual and multilingual order systems such as classification systems, thesauri and other special terminologies such as dictionaries, etc., with special attention to quality control.

POLISH LIBRARIANS ASSOCIATION

Polish Librarians Assiociation (PLA) is a self-governing non-profit organization of people who through their profession or research are involved in librarianship and information science in Poland .

OUR AIMS

Through its activity the Association seeks:

— to take part in developing library information and publishing policies ,

— to encourage and support research interesests and professional development of its members,

— to present the opinion of librarians to the authorities central administration, local government, and other organizations or associations,

— to protect humanist traditions of Polish culture,

— to contribute to the preservation and promotion of Polish and world literature,

— to undertake advocacy of the intergration of the entire library community.

A FEW FACTS FROM HISTORY

The predecessor of the Polish Librarians Association was the Commission for the History of Libraries and Library Science called to life in Warsaw in 1915. In 1917 members of the Commission founded the Polish Librarians Union, the first organization of librarians in the country as it regained independence after 120 years of being under partition.When Poland became free, the union engaged in a huge task to promote book reading and establish libraries. It helped to raise Polish culture from destruction caused by the partitions and to make up for the civilization gap. It sought to build up the prestige of the librarian's profession.The outbreak of World War II and the occupation of Polish lands by the hitlerite Third Reich caused that until 1944 the Union of Librarians was an underground organization committed to the protection of library collection against destruction or robbery by the Nazis. After the war, in 1946, the the organization was reborn as the Polish Librarians and Archivists Union and in 1953 changed to Polish Librarians Association .In spite of the change of name our organization has always served humanist ideas, books and reading. We have always supported these issues in the international arena. The Polish Librarians Union, our predecessor, by acceding in 1927 to the international Federation of Library Association founded at that time in Edinburgh, wonthe status of an IFLA founding member. Many Polish librarians have been members of IFLA's specialized agencies.

MEMBERSHIP

The Polish Librarians Association has 9000 members. They represent all types of libraries in the country. We are not an elitist organization — we live the life of the entire Polish library community. We possess — following the pattern of the administative division of the country — 48 voivodship branches represented by local boards. Subordinate to these boards are PLA circles which are basic segments of the Association. There are over 200 such circles. They are ussually based in major libraries, mainly public libraries.

KNOWLEDGE AND SPECIALIZATION

Definitely important in the structure of the Association is the position of expert teams called to life to deal with today's and future needs of librarians. The work of these teams serves to solve concrete problems related to library theory and practice.

The most important ones are:

— Section on Public Libraries,
— Section on Research Libraries,
— Section on Music Libraries,
— Section on Reading of the Physically Challenged Persons,
— Commission on Classification Schemes,
— Commission on Preservation and Conservation,
— Commission on Automation,
— Commission on Awards,
— Team on Regional Bibliography,
— Team on Parliamentary Act on Libraries.

The work of the teams results in the production of expert's reports to solve many important problems in various fields of librarianship. They promote development and disselmination of knowledge necessary for library services.

OUR PUBLICATION

The Polish Librarians Association is a publisher or co-publisher of 5 scientific journals related to Polish and foreign librarianship. Most of them have a many-years' history. Numerous articles on a wide range of topics were published in their pages by Polish and foreign authors.These journals are:

• „Bibliotekarz" (a monthly)
published since April 1919
Co-publisher: City of Warsaw Public Library
Editor-in-Chief: Jan Wołosz,

• „Przegląd Biblioteczny" (a quarterly)
published since 1927
Co-publisher Library of the Polish Academy of Sciences
Editor-in-Chief: Barbara Sordylowa,

• „Poradnik Bibliotekarza" (a monthly)
published since 1949
Editor-in-Chief: Władysława Wasilewska,
• Zagadnienia Informacji Naukowej (a semi-annual)
published since 1962
Co-publisher: Institute of Library and Information Science, University of Warsaw,
Editor-in-Chief: Bożenna Bojar,
• Biuletyn Informacyjny ZG SBP (PLA Information Bulletin)
Editor: Edward Jakimowicz.

PLA has a specialized publishing division preparing about a dozen books annually. The average impression ranges form 2 to 3 thousand copies. The Division caters mainly to the needs of librarians in the area of education and training, particularly library junior staff. The publications come out in three series:

— Science-Didactics-Practice,
— Proposals and Marterials,
— Polish Librarians in the Memories of Their Contemporaries.

The division publishes also bibliophilic prints, such as for example: "Chopin" by Ignacy Jan Paderewski. Furthermore,the Division provides libraries with leaflets, posters etc.

To share experience with others and learn about the latest achievements in librarianship. PLA organizes special symposia and conferences, also international ones. The PLA Publishing Division publishes and disseminates materials that were debated upon during these meetings.

For particular achievements in application of knowledge and skills, and for scientific achievements, each year PLA is awarding an annual Adam Łysakowski scientific award.

PLA AUTORITIES

The supreme authority of the Association is the Congress of Delegates convened every four years. The Congress elects the executive authorities: PLA President and members of the General Board. The General Board appoints the Presidium officers who direct the work of the Association between the meetings of the General Board. The General Board consists of the President and 14 members. The Presidium is composed of 7 persons.

For the present term of office (since June 15, 1993) the officers of the PLA Presidium are:

President: Stanisław Czajka
Secretary general: Janina Jagielska
Vice-President: Zofia Płatkiewicz
Vice-President: Andrzej Sroga
Treasurer: Dariusz Kuźmiński
Members: Andrzej Kempa, Marian Skomro

Once a year the PLA Forum is convened and the participants are the authorities of the Association, honorary members and representatives of all

divisions country. The Forum reviews current PLA activity and defines the tasks for the whole Association.

The accompanying event a each annual PLA Forum is a conference on current and most vital issues concerning Polish librarianship, such as library automation or library staff training according to world standarts.

The statutory activity of PLA is financed mainly by the income derived from. sales of publications and from promotion of books sponsored by publishers. The fees paid by individual members remain at the disposal of local PLA circles and division boards.

In order to pursue its statutory aims, PLA cooperates closely with the Ministry of Culture and Arts, National Library in Warsaw, research libraries, voivodship public libraries and associations that are active among librarians.

HONOURS AND AWARDS

To persons that are particularly worthy of ment for their contribution to Polish librarianship and PLA the title of an Honorary PLA Member is awarded. In all the years of the Association's existence this honour went to 45 persons.

The Association awards to its members also a special commemorative medal and an Honorary PLA Badge.These decorations go to persons who have distinguished themselves for their work for the Association and libraries .

LIBRARIAN'S DAY

May 8th was made Librarian's and Libraries Day. It is the holiday of our Association and it it serves to promote books and libraries in society. It is also a good occasion to consider the achievements and needs of people of our profession. It is a day to honour those who are the best and also to remind of or to promote librarians' successes and the importance of their profession in the development culture and achievement of our nation.

OUR ADDRESSES

Polish Librarians Association
02-103 Warszawa,
ul. St. K. Hankiewicza 1
Tel. 22-51-33
Fax 22-35-41
Poland
General Board Bureau
— Administrative Division
— Finance Division ,
— Distribution and Promotion Division

Polish Librarians Association
PLA Publishers
02-593 Warszawa,
ul. E. Konopczyńskiego 5/7
Tel. 27-52-96
Poland

Contens

Czytajcie czasopisma bibliotekarskie!

Miesięcz [illegible] otekarzy Polskich:

BIBLIOTE[illegible] się ważnymi i aktualnymi probl[illegible]

PORADNIK BIBLIOTEKARZA — pismo o charakterze instrukcyjno-metodycznym przeznaczone głównie dla bibliotekarzy bibliotek publicznych i szkolnych.

Oba pisma tylko w prenumeracie w „Ruchu", na poczcie i w siedzibie Wydawnictwa SBP. Szczegóły w obu pismach na stronie redakcyjnej.

ZIN (Zagadnienia Informacji Naukowej) — półrocznik poświęcony szeroko rozumianym zagadnieniom informacji.

Prenumerata i sprzedaż: Dział Promocji i Kolportażu SBP, ul. S. K. Hankiewicza 1, 02-103 Warszawa, tel. 22-43-45.

PRZEGLĄD BIBLIOTECZNY — kwartalnik o charakterze naukowym wydawany przez Bibliotekę PAN i SBP.

Prenumerata tylko w Redakcji: Pałac Kultury i Nauki VI p. 00-901 Warszawa, tel. 656-66-00, 20-33-02.

Jeśli masz kłopot z nabyciem

- druków bibliotecznych

(dziennik biblioteki, księga inwentarzowa, karty: katalogowe, książki, czytelnika, upomnienia, rewersy etc.),

- naszych czasopism (numery bieżące i zaległe),
- fachowej literatury bibliotekarskiej

napisz lub zadzwoń do:

DZIAŁ PROMOCJI I KOLPORTAŻU SBP,

02-103 Warszawa-Ochota, ul. S. K. Hankiewicza 1, tel. 22-43-45

ISBN 83-

WYDAWNICTWO
SBP