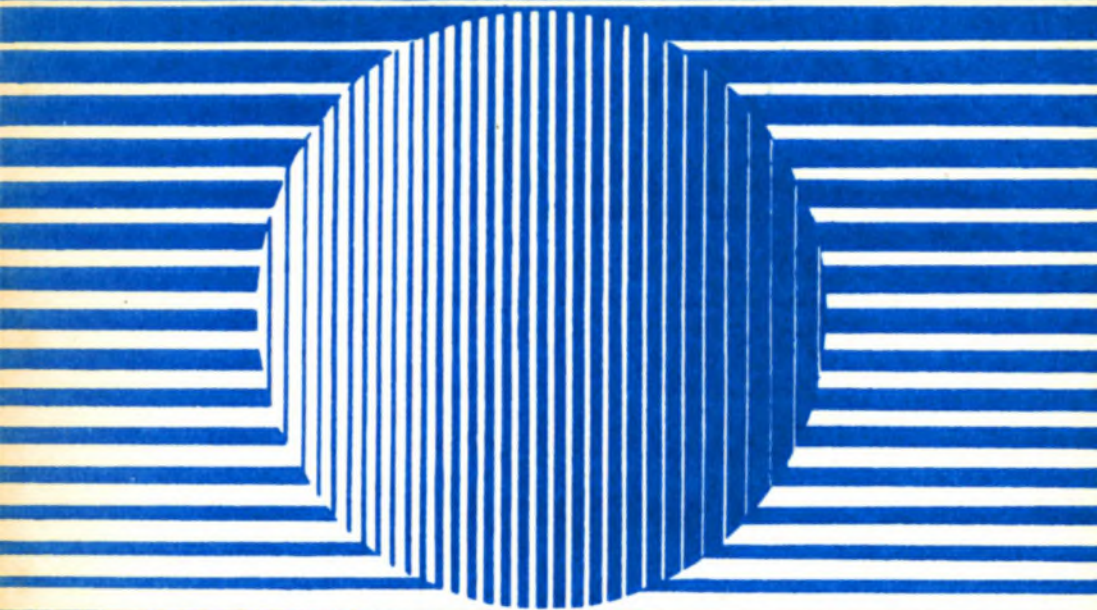


POLSKA AKADEMIA NAUK



OŚRODEK INFORMACJI NAUKOWEJ

PL ISSN 0324-8194

ZAGADNIENIA INFORMACJI NAUKOWEJ

1991

OSSOLINEUM

Nr 1 (58)

POLSKA AKADEMIA NAUK

OŚRODEK INFORMACJI NAUKOWEJ

ZAGADNIENIA
INFORMACJI
NAUKOWEJ

1991

Nr 1 - (58)

ZAKŁAD NARODOWY IMIENIA OSSOLIŃSKICH
WYDAWNICTWO POLSKIEJ AKADEMII NAUK

Rada Redakcyjna

Zdzisław Bobiatyński, Bożenna Bojar, Władysław M. Grabski,
Andrzej Gromek, Małgorzata Kłossowska, Barbara Krygier,
Juliusz L. Kulikowski, Anna Leśniewicz (sekretarz redakcji),
Bronisław Ługowski (redaktor naczelny), Elżbieta Malinowska,
Maria Szomańska, Olgierd A. Wojtasiewicz

Do roku 1971 czasopismo ukazywało się pod tytułem
„BIULETYN ODIiN PAN”

Adres Redakcji

Ośrodek Informacji Naukowej PAN
00-330 Warszawa, ul. Nowy Świat 72 (Pałac Staszica)

Maszynopis niniejszego numeru przekazano do Wydawnictwa we wrześniu 1991

Wykonano ze składuopisu dostarczonego przez OIN PAN

PL ISSN 0324-8194

Zakład Narodowy im. Ossolińskich, Wydawnictwo, Wrocław 1991
Zakład Graficzny Politechniki Wrocławskiej zam. 1133/91-510 egz.

JĘZYK WYSZUKIWAWCZY A POTRZEBY INFORMACYJNE UŻYTKOWNIKÓW

Kategorie użytkowników. Język indeksowania a język wyszukiwawczy. Język wyszukiwawczy jako język wyrażający potrzeby użytkowników. Rola biblioteki w tworzeniu profilu użytkownika. Elementy profilu użytkownika. Miejsce profilu użytkownika w systemie wyszukiwawczym. Profil użytkownika jako element systemu ekspertowego.

Informację coraz częściej traktuje się jak towar, który powinien być podporządkowany prawom rynku. Towar musi mieć odbiorcę i zaspokajać jego potrzeby. Tak więc podstawowym zadaniem projektantów systemów informacyjno-wyszukiwawczych (SIW) powinno być dokładne rozpoznanie potrzeb potencjalnych odbiorców i dostosowanie do nich tworzonych w systemie narzędzi informacyjnych (katalogów, bibliografii, baz danych). Znajomość potrzeb informacyjnych użytkowników ma więc istotne znaczenie dla efektywnego działania SIW. Od trafnego określenia tych potrzeb zależeć będzie znalezienie właściwych możliwości ich zaspokojenia. Wymaga to skoncentrowania w większym stopniu uwagi bibliotekarzy i pracowników informacji na użytkowniku, dla którego przeznaczony jest dany system i na jego potrzebach. Niestety, ze względu na trudności w sprecyzowaniu tych potrzeb, na ich niewymierny, ulotny i zmienny

charakter bardzo często zapominano, że osoba znajdująca się na końcu systemu jest ważniejsza od jego pozostałych elementów. O wiele łatwiej jest zmierzyć, przetestować i ocenić materialne, techniczne aspekty systemów informacyjnych niż trudne do zidentyfikowania aspekty intelektualne związane z potrzebami informacyjnymi. Dlatego biblioteki, ośrodki czy serwisy informacyjne funkcjonowały często w taki sposób, jakby centralnym punktem (podmiotem) ich działalności był nie użytkownik a dokument.

Pierwszym i zasadniczym krokiem w projektowaniu SIW lub w ocenie efektywności SIW już istniejącego jest określenie celów systemu w aspektach społeczności, jaką ma obsługiwać i usług, jakie ma świadczyć. Do tych elementów powinien być dostosowany zakres zbiorów, tworzone narzędzia informacyjne oraz sposoby ich prezentowania uwzględniające potrzeby obsługiwanej społeczności. Najistotniejszym i najtrudniejszym problemem jest zidentyfikowanie i określenie tych potrzeb. Wypracowane dotychczas metody badania potrzeb użytkowników okazały się nieskuteczne, a bliższa ich analiza ujawnia, że nie były one w zasadzie badaniami faktycznych potrzeb użytkowników, lecz badaniami ich zachowań i badaniami samych systemów informacyjnych //.

Wśród potrzeb użytkowników w zautomatyzowanych SIW możemy wyróżnić takie, które dotyczą sposobów wyszukiwania i takie, które dotyczą oczekiwanych wyników. Pierwsze związane są w znacznym stopniu z interfejsem, środkami komunikacji komputera z użytkownikiem, drugie - z zawartością bazy danych i dostępnymi środkami wyszukiwawczymi.

Podstawowe pytania, jakie powinien sobie postawić projektant SIW, to:

- w jaki sposób różnią się między sobą indywidualni użytkownicy i ich potrzeby?

- jak te różnice wpływają na wyniki wyszukiwania i stopień usatysfakcjonowania (ocenę uzyskanej informacji)?

-- jak te różnice wpływają na sam przebieg procesu wyszukiwania?

Użytkownicy danego SIW nie stanowią jednorodnej grupy ze standardowymi cechami, jednakowo reagującej w procesie wyszukiwania i jednakowo odbierającej jego wyniki; mają zindywidualizowane wymagania zarówno co do rodzaju i poziomu odpowiedzi, jakiej oczekują, jak i co do sposobów (metod, technik) jej wyszukiwania. Potrzeby informacyjne są wielkościami dynamicznymi, zależnymi od różnych czynników obiektywnych i subiektywnych. System może tylko wtedy funkcjonować efektywnie, jeśli potrafi dostosować się do tych zmiennych wielkości, do indywidualnych przyzwyczajeń, możliwości, preferencji użytkowników oraz do sposobów, w jaki prowadzą wyszukiwanie. Aby się do nich dostosować musi je najpierw zidentyfikować. Częściowo umożliwia mu to proces fizycznej interakcji, który w systemach online rozwinięty jest w wysokim stopniu. Interakcja między systemem a użytkownikiem zachodząca w trakcie procesu wyszukiwania wpływa w znacznym stopniu na modyfikowanie pracy systemu i jego reakcje oraz na zmiany w postępowaniu i decyzjach użytkownika, nie wystarcza jednak do pełnego dostosowania systemu do użytkowników. Aby to osiągnąć, należy najpierw ustalić kim są użytkownicy, dokonać ich rozróżnienia i podziału na różne kategorie i następnie określić potrzeby i oczekiwania poszczególnych kategorii.

KATEGORIE UŻYTKOWNIKÓW

Podziału użytkowników biblioteki i serwisów informacyjnych można dokonać ze względu na wiele kryteriów. Z punktu widzenia umiejętności korzystania z dokumentacyjnego systemu online istotne są dwa kryteria:

- znajomość obsługi systemów komputerowych;
- znajomość zasad wyszukiwania w systemach dokumentacyjnych.

Dokonując podziału użytkowników według obu tych kryteriów otrzymamy cztery grupy różniące się między sobą stopniem przygotowania do korzystania z systemu:

1) użytkownicy nie mający doświadczenia ani w wyszukiwaniu, ani w obsłudze komputera;

2) mający doświadczenie w wyszukiwaniu (umiejętność posługiwania się źródłami informacyjnymi, np. katalogiem, bibliografią, indeksem);

3) mający doświadczenie w obsłudze systemów komputerowych;

4) mający doświadczenie w wyszukiwaniu i w pracy z komputerem.

Każdej z wyodrębnionych grup należy zapewnić różne sposoby dochodzenia do tych samych (lub podobnych) rezultatów, każda ma różne potrzeby w zakresie środków dialogu i metody współpracy z systemem, dlatego dla każdej grupy należy przygotować inny interfejs, a raczej inną jego wersję (poziom).

Przy grupie pierwszej interfejs nie może zakładać żadnej uprzedniej wiedzy użytkownika ani w zakresie wyszukiwania, ani w zakresie systemów zautomatyzowanych. Wszelkie trudności i cała złożoność struktury tego typu systemu winna być przed użytkownikiem ukryta. Grupę czwartą stanowią głównie osoby odpowiednio przeszkolone, np. pracownicy biblioteki, choć i wśród użytkowników końco-

wych znajdą się przedstawiciele tej kategorii. Z drugiej strony część pracowników biblioteki może wchodzić w skład grup pozostałych. Niestety, najliczniejszą grupę użytkowników, przynajmniej w początkowym okresie funkcjonowania systemu online, stanowią reprezentanci grupy pierwszej, najmniej przygotowani do korzystania z systemu samodzielnie.

W artykule na temat systemów przyjaznych /1/ za najważniejszą cechę uznano adaptowalność SIW (lub interfejsu), tj. zdolność przystosowywania się systemu do różnego poziomu użytkowników, różnego stopnia znajomości obsługi komputerów i metod wyszukiwania w dokumentacyjnych zbiorach informacji, różnych przyzwyczajzeń, możliwości i upodobań. System powinien mieć wbudowany mechanizm dostosowywania się do wymienionych kategorii użytkowników, do ich doświadczenia, mechanizm umożliwiający wyszukiwanie w sposób tak sprawny i szybki, na ile pozwala sprawność użytkownika. Osoba korzystająca z systemu po raz pierwszy wymagać będzie szczegółowych wyjaśnień i instrukcji, prowadzenia "krok po kroku" przez zawiłości struktury bazy danych i strategii wyszukiwawczej, podpowiadania i korygowania. Specjalista natomiast całą procedurę wyszukiwania chce skrócić do minimum, dojść do wyników drogą najszybszą i najbardziej efektywną, nawet kosztem jej prostoty i przystępności. Użytkownicy korzystający z biblioteki regularnie powinni mieć pewne korzyści płynące z ich doświadczenia, obycia z systemem, poznania podstawowych operacji i metod skracania drogi poszukiwania, a nie być karani konieczności nużącego przechodzenia przez wszystkie kolejne etapy procesu i teksty informacji pomocniczych, nawet jeśli nie są one potrzebne i zamiast pomagać przeszkadzają. Dlatego obok dokonanego wcześniej podziału użytkowników ze względu na ich stopień przygotowania do korzystania z systemu powinniśmy

wprowadzić drugi podział, częściowo pokrywający się z pierwszym, choć ustalony według innego kryterium - kryterium częstości korzystania z systemu. W wyniku tego podziału otrzymamy trzy kategorie:

1) nowicjusz (korzystający z systemu po raz pierwszy);
2) użytkownik przypadkowy (korzystający z systemu rzadko, sporadycznie);

3) użytkownik stały (korzystający z systemu bardzo często).

W pierwszej grupie mogą znaleźć się użytkownicy, którzy korzystają z danego SIW po raz pierwszy, ale wcześniej korzystali już z innych SIW. Ich potrzeby są bardziej zbliżone do potrzeb drugiej grupy, która reprezentuje najczęstszy typ użytkownika. Do tej grupy należą czytelnicy biblioteki, którzy od czasu do czasu odczuwają potrzebę skorzystania z systemu wyszukiwawczego; znają oni ogólnie zasady korzystania (nie są nowicjuszami, kiedyś już z systemu korzystali), ale wymagają przypomnienia szczegółów, które od ostatniej sesji wyszukiwawczej mieli prawo zapomnieć, podpowiadania i korygowania ewentualnie popełnianych błędów.

Trzecią grupę stanowią głównie pracownicy biblioteki (ośrodka) oraz czytelnicy prowadzący bardzo często wyszukiwania, co może być związane z ich zawodem (np. dziennikarz, nauczyciel) lub zainteresowaniami. Korzystają oni z systemu na tyle często, że nie zdążą wyjść z wprawy, dlatego dla tej grupy może być przeznaczony krótki i efektywny, choć nieco trudny w stosowaniu język komend (por. artykuł /1/). Dla dwóch pierwszych grup najlepszą techniką dialogu jest system menu, choć opracowany na różnym poziomie. Dla użytkownika przypadkowego (grupa druga) - skrócone menu, przedstawiające dostępne opcje, ale bez wyjaśnień i przykładów (mogą być one wywoływane na żądanie na zasadzie tablic HELP); dla nowicjusza (grupa pierwsza) - rozbudowany i klarowny system

tablic menu, bardzo szczegółowe instrukcje, zrozumiałe i obszerne wyjaśnienia.

Dostosowanie się systemu do poziomu użytkownika musi być poprzedzone rozpoznaniem tego poziomu. Może to wskazać sam użytkownik, pośrednio lub bezpośrednio, trzema sposobami:

a) kiedy użytkownik rejestruje się w systemie i dostaje swój numer identyfikacyjny, wskazuje wybrany przez siebie poziom, który jest wpisany na stałe do systemu (w kartotece użytkowników); przy korzystaniu z systemu użytkownik podaje swój numer i system automatycznie ustawia dialog na wybranym poziomie; rozwiązanie to nie daje łatwości przechodzenia na poziom wyższy lub niższy (np. jeśli użytkownik nabierze wprawy lub jeśli ma trudności ze zrealizowaniem jakiegoś zadania);

b) przed każdorazowym rozpoczęciem pracy z systemem użytkownik wskazuje wymagany poziom, np. system wyświetla menu z dostępnymi wersjami dialogu i użytkownik wybiera jedną z nich, jego zdaniem najbardziej mu odpowiadającą; wadą tego rozwiązania jest przedłużenie każdej sesji wyszukiwawczej o jeden etap, o jedną dodatkową tablicę menu;

c) system ma wbudowany specjalny program, który automatycznie dopasowuje technikę dialogu do poziomu przygotowania użytkownika, np. jeśli użytkownik popełni na średnim poziomie elementarny błąd, system reaguje automatycznym przełączeniem się na poziom niższy, prostszy.

Ponieważ doświadczenie użytkownika rośnie przy częstym korzystaniu z systemu, dlatego powinno się zapewnić nie tylko możliwość wyboru różnego poziomu interfejsu, ale i możliwość przechodzenia z jednego poziomu na inny przed i w trakcie wyszukiwania, na każdym jego etapie.

Użytkownicy różnią się między sobą nie tylko stopniem doświadczenia, umiejętnością korzystania z systemu, różnią się także wiekiem, osobowością, uzdolnieniami, wykształceniem, zdolnościami przyswajania nowych rzeczy, zapamiętywania, logicznego myślenia, itp. Dlatego prawdziwie przyjazny interfejs powinien być tak zaprojektowany, aby uwzględniał także te subiektywne czynniki, umożliwiaił szybsze lub wolniejsze przechodzenie z jednego poziomu na drugi, szybsze lub wolniejsze przyswajanie niektórych pojęć, podawał mniej lub bardziej rozbudowane wyjaśnienia i instrukcje.

Wyróżnione dotychczas potrzeby użytkowników związane były głównie z samym procesem wyszukiwania, środkami dialogu, umiejętnością wprowadzania i wyprowadzania informacji, formułowania strategii wyszukiwawczej. Jak już jednak wspomniano, użytkownicy mają zindywidualizowane wymagania nie tylko co do sposobów prowadzenia wyszukiwania, ale i co do jego wyników, tj. rodzaju, zakresu i poziomu informacji, jakiej oczekują w odpowiedzi na postawione pytanie. Te potrzeby są znacznie trudniejsze do zidentyfikowania i sprecyzowania i związane są w większym stopniu z językiem wyszukiwawczym danego systemu.

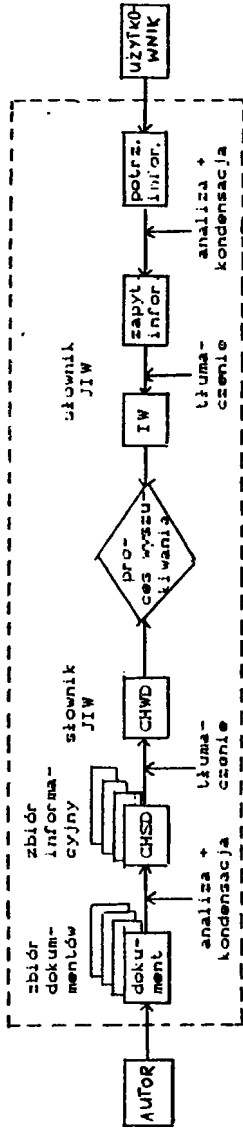
JĘZYK INDEKSOWANIA A JĘZYK WYSZUKIWAWCZY

Historię języków informacyjno-wyszukiwawczych (JIW) rozpatrywać można w aspekcie powolnego procesu oddzielania się języka wyszukiwawczego (czyli języka służącego do przedstawienia zapotrzebowania informacyjnego użytkowników) od języka indeksowania (czyli języka służącego do charakteryzowania dokumentów zawierających informację). Dla lepszego zrozumienia różnic w ich charakterze i budowie należy uświadomić sobie, jakie funkcje pełnią w procesie

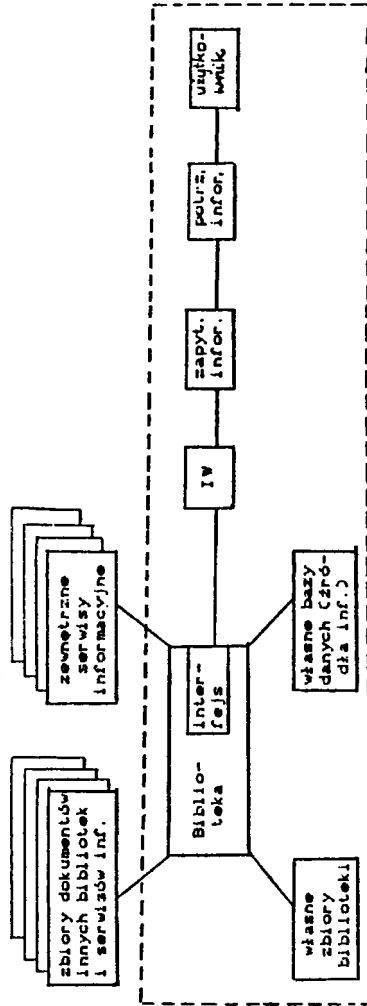
informacyjnym. Zilustruje to lepiej diagram, przedstawiający ten proces na przykładzie systemu informacyjnego jakim jest biblioteka (rys. 1).

Diagram ten pokazuje główne komponenty procesu wyszukiwania; po obu stronach widać dwa ciągi operacji, po lewej - ciąg operacji opracowywania dokumentu, po prawej - ciąg operacji formułowania instrukcji wyszukiwawczej (IW) służącej do wyrażenia potrzeby informacyjnej użytkownika. Twórca informacji pierwotnej - autor - tworzy dokument, który zakupiony przez bibliotekę staje się elementem systemu (oznaczonego tu linią przerywaną). Dokument zostaje poddany opracowaniu (indeksowaniu, katalogowaniu, klasyfikowaniu), czego rezultatem jest najpierw Charakterystyka Słowna Dokumentu (ChSD), a następnie-po przełożeniu przy pomocy słownika systemu (tezaurusu, tablic klasyfikacyjnych, słownika haseł przedmiotowych) na sformalizowany JIW - Charakterystyka Wyszukiwawcza Dokumentu (ChWD). ChWD charakteryzuje cechy treściowe dokumentu oraz jego cechy formalne związane z autorstwem, strukturą i nośnikiem. Opracowany dokument tworzy wraz z pozostałymi, wcześniej zakupionymi i opracowanymi dokumentami zbiór biblioteczny (księgozbiór), natomiast utworzona ChWD wraz z opisami pozostałych dokumentów - ustrukturalizowany zbiór wyszukiwawczy (katalog, spis bibliograficzny, indeks).

Po drugiej stronie schematu znajduje się użytkownik, który przychodzi do biblioteki z konkretnym zapotrzebowaniem informacyjnym wyrażonym w języku naturalnym w postaci zapytania. Zapytanie informacyjne poddawane jest podobnym transformacjom jak dokument biblioteczny, tj. przy pomocy słownika systemu tłumaczone jest na IW, a następnie porównywane pod kątem zgodności z elementami zbioru wyszukiwawczego, tj. z ChWD. Efektem tego porównania jest



Rys. 1. Schemat systemu informacyjno-wyszukiwawczego
(na przykładzie biblioteki)



Rys. 2. Model biblioteki podłączonej do sieci

wyselekcjonowanie na podstawie przyjętych dla danego systemu kryteriów odpowiedniości (relewancji technicznej) takiego podzbioru całego zbioru wyszukiwawczego, który spełniając warunek relewancji technicznej wydawany jest jako odpowiedź na zapytanie użytkownika. Na podstawie otrzymanych ChWD użytkownik trafi do dokumentów oryginalnych, w których powinna się znajdować relewantna informacja, czyli informacja zaspokajająca jego potrzebę informacyjną.

Podany diagram przedstawia tradycyjny model biblioteki, z jedną bazą danych. Przy bliższym przyjrzeniu się temu schematowi rzuca się w oczy podobieństwo po obu stronach procesu wyszukiwania; każdy element procesu opracowywania dokumentu ma swój odpowiednik po stronie procesu opracowywania potrzeby informacyjnej użytkownika. Wynikami obu procesów są odpowiednio ChWD i IW, które są następnie poddane porównaniu w celu ustalenia ich zgodności. Obie zapisane są w sformalizowanym języku przy wykorzystaniu słownika systemu, pytanie, jakie można tu postawić brzmi: czy jest to ten sam słownik i ten sam język?

W tradycyjnym systemie bibliotecznym zadaniem użytkownika przy wyszukiwaniu było odnalezienie w zbiorze informacyjnym (katalogu, indeksie, bibliografii) terminów (haseł, symboli) "przypisanych" dokumentowi przez bibliotekarza w procesie indeksowania, dopasowanie IW do istniejących już w zbiorze ChWD. Użytkownik odszukiwał po prostu w odpowiednim miejscu ciągu alfabetycznego lub systematycznego wybrane terminy. Konsekwencją takiego postępowania było dokładne odzwierciedlenie przez instrukcję wyszukiwawczą utworzonych wcześniej charakterystyk wyszukiwawczych. Instrukcja wyszukiwawcza była w zasadzie tworem abstrakcyjnym, a nie fizycznie istniejącym zapisem, a jej zmiana i modyfikacja w trakcie procesu wyszukiwania miała charakter ciągły i nieuchwytny (trudny do wyspecyfi-

kowania). Człowiek realizujący wyszukiwanie wykorzystywał w trakcie tego procesu swą wiedzę językową i pozajęzykową, a posiadając zdolność wykonywania kilku operacji myślowych jednocześnie nie rozgraniczał poszczególnych etapów i faz tego procesu. Dlatego w wyszukiwaniu prowadzonym metodami manualnymi nie są te etapy wyraźnie wyodrębnione i oznaczone, a ich wykładniki nie są jawne. Użytkownik, aby odszukać pożądaną ChWD w zbiorze, musiał stworzyć IW będącą jej wiernym odbiciem (obrazem), o tej samej formie i strukturze. Tak więc, w tradycyjnym systemie wyszukiwawczym nie było potrzeby oddzielania języka wyszukiwawczego od języka indeksowania, gdyż były one w zasadzie identyczne, choć pełniły różne funkcje. Język wyszukiwawczy starał się upodobnić, odzwierciedlić język indeksowania.

Wraz z rozwojem nowych technik wyszukiwawczych następuje stopniowe odseparowanie się obu języków. Wprowadzenie metod indeksowania współrzędnego, możliwe dzięki mechanizacji, a następnie automatyzacji procesów wyszukiwania, doprowadziło do zróżnicowania się obu języków, widocznego zarówno w ich słowniku jak i w gramatyce. Język wyszukiwawczy może być bogatszy lub uboższy o pewne środki gramatyczne, np. w tekstach języka wyszukiwawczego mogą być wprowadzone dodatkowo wskaźniki wagi (różnicujące wartość poszczególnych jednostek leksykalnych dla użytkownika), które nie występują w ChWD (czyli w języku indeksowania). Z drugiej strony w IW pomijane są wskaźniki relacji tekstowych typu wskaźników więzi, które wprowadzane bywają do wielozdaniowych ChWD w celu zapobieżenia fałszywej koordynacji. Różnice w gramatykach obu podjęzyków najbardziej widoczne są w języku deskryptorowym, w którym operatory Boole'a wykorzystywane są tylko do budowy IW.

Różnice w zasobach leksykalnych występują w mniejszym stopniu, ale zdarza się, że słownik wykorzystywany jest w danym systemie tylko do budowy IW, natomiast ChW0 powstaje na podstawie indeksowania swobodnego (niekontrolowane słowa kluczowe lub pełne teksty dokumentów) i wtedy struktura i zestaw obu zasobów leksykalnych jest bardzo różny.

Słowniki obu podjęzyków nie muszą się całkowicie pokrywać, choć pewne ich podzbiory muszą być identyczne, jeśli ma być w ogóle możliwy proces porównywania ich elementów. Porównywanie bowiem wciąż jeszcze dokonywane jest na podstawie formalnych cech jednostek leksykalnych, a nie cech semantycznych.

W systemach zautomatyzowanych konieczne jest wyraźne wyodrębnienie i oznaczenie poszczególnych etapów procesu wyszukiwania, użycie formalnych wykładników komend wyszukiwawczych, sprecyzowanie jednoznacznego kryterium relewancji technicznej, które, jak pamiętamy, nie były niezbędne w systemach manualnych obsługiwanych przez człowieka. Środki te składają się także na strukturę języka wyszukiwawczego, wzbogacając go o elementy nie występujące w języku indeksowania. Proces rozdzielania się obu języków posuwa się jeszcze dalej w związku ze zmianami w funkcji i modelu współczesnej biblioteki.

ROLA BIBLIOTEKI W PROCESIE WYSZUKIWANIA

Model biblioteki tradycyjnej pokazany na rys. 1 zawierał jeden zbiór biblioteczny i jeden zbiór wyszukiwawczy (katalog), nawet jeśli składał się on z kilku wyodrębnionych podzbiorów, np. w przypadku zbioru dokumentów - księgozbiór dydaktyczny, podręczny, bibliofilski, zbiory kartograficzne, regionalne, starych

druków, itp., a w przypadku zbioru wyszukiwawczego - katalog czasopism, katalog lektur, katalog alfabetyczny, przedmiotowy, systematyczny, itp.

Postęp w zakresie technologii informacyjnej w ostatnich latach wpłynął na zmianę funkcji biblioteki i sposobów jej wykorzystywania. Szczególne znaczenie odegrały tu:

1) wzrost zastosowań komputerów, spowodowany głównie spadkiem cen sprzętu i rozwojem specjalistycznego oprogramowania (do zastosowań bibliotecznych), coraz liczniejszego i doskonalszego; umożliwiło to wprowadzenie nowych sposobów operowania informacją, a w rezultacie znaczne podniesienie jakości usług;

2) rozwój sieci telekomunikacyjnych, który doprowadził do zwiększenia możliwości i dostępności sieci przy zmniejszeniu kosztów; połączenie telekomunikacji z techniką komputerową stworzyło nowe możliwości przekazu i rozpowszechniania informacji; problem odległości (fizycznego oddalenia) odbiorcy informacji od jej źródła i szybkość przekazu nie stwarza już (z pominięciem kosztów) żadnych problemów; wykorzystanie tych dwóch technologii w bibliotekach doprowadziło do powstawania komputerowych sieci bibliotecznych, dzięki którym mogło się urzeczywistnić odwieczne marzenie bibliotekarzy o wspólnym katalogowaniu, wspólnych zbiorach, wzajemnym korzystaniu ze swoich zasobów i katalogów;

3) powstanie i rozwój nowych nośników informacji (mikroform, nośników magnetycznych oraz optycznych), umożliwiających miniaturyzację zbiorów bibliotecznych, a w konsekwencji - ograniczenie przestrzenne biblioteki; nośniki te charakteryzują się niewielkimi rozmiarami i dużą pojemnością, niskimi kosztami wytwarzania i powielania, łatwością przesyłania; nośniki optyczne typu CD ROM umożliwiają stosunkowo tanim kosztem korzystanie lokalne (na

miejscu, bez pośrednictwa sieci telekomunikacyjnej) z ogromnych baz danych typu katalogowego, bibliograficznego lub encyklopedycznego.

Konsekwencją tych zmian w działalności bibliotek (gromadzeniu, opracowywaniu i udostępnianiu) z punktu widzenia użytkownika jest:

- zapewnienie lepszego serwisu, szczególnie w zakresie działalności informacyjnej;

- dostęp do większej liczby źródeł niż było to możliwe przy źródłach w formie dokumentów drukowanych; oprócz lokalnych źródeł biblioteki (baza danych katalogowa) także źródła zewnętrzne (bibliograficzne i katalogowe bazy danych zewnętrznych serwisów informacyjnych i innych bibliotek),

a z punktu widzenia bibliotekarza:

- zmniejszenie lub wyeliminowanie mechanicznych czynności bibliotekarskich (np. sporządzania kart katalogowych, szeregowania);

- oszczędności wynikające z odciążenia personelu od wielu prac i z posiadania wspólnych zasobów bibliotecznych;

- miniaturyzacja zbiorów wynikająca z mniejszych rozmiarów niektórych nośników i korzystania ze zbiorów spoza biblioteki.

Obecnie komputer (terminal) umieszczony w bibliotece może udostępniać nie tylko własne źródła informacji (jest źródłem informacji o własnych zasobach), ale i inne bazy danych przechowywane w bibliotece (na dyskietkach, taśmach lub CD ROM) lub poza nią, a dostępne online za pomocą sieci telekomunikacyjnych. Biblioteka udostępnia nie tylko własne zbiory, lecz dzięki sieciom komputerowym, nowym technikom przekazu i powielania informacji oraz wypożyczaniu międzybibliotecznemu także zbiory innych bibliotek i systemów. Zmiany te prowadzą do poszerzenia pierwotnych funkcji i samego pojęcia "biblioteki".

w tradycyjnym modelu biblioteki (por. rys.1) użytkownik korzysta z jednego zbioru dokumentów i jednego zbioru wyszukiwanego, w modelu przyszłościowy: użytkownik korzysta z różnych zasobów dokumentacyjnych i informacyjnych. Następuje odejście od tradycyjnej biblioteki bazującej na własnych zbiorach i własnych katalogach do biblioteki mającej dodatkowo dostęp do zbiorów i źródeł innych bibliotek i serwisów informacyjnych, co zilustrować można uproszczonym schematem (rys.2).

W modelu tym zmianie uległa struktura i niektóre procesy biblioteczne, dalej jednak biblioteki służą tym samym celom co zawsze, ich zadaniem wciąż jest pośredniczenie i ułatwianie fizycznego kontaktu między użytkownikiem a dokumentem, który chciałby przeczytać. Różnica dotyczy tylko zakresu i zasięgu tego pośrednictwa: w nowym układzie biblioteki pośredniczą między użytkownikiem i nie tylko swoimi zbiorami, lecz zbiorami i źródłami informacji o nich sporządzonymi i przechowywanymi poza biblioteką. Wylaniają się tu dwa nowe problemy:

1) problem dostarczania użytkownikowi oryginalnych dokumentów (zapewnienie serwisu sprawniej działającego niż tradycyjne wypożyczanie międzybiblioteczne);

2) problem opracowania narzędzi (programów) umożliwiających poruszanie się po tym złożonym systemie zbiorów i źródeł. Korzystanie bowiem z zewnętrznych serwisów i zasobów nie jest wcale proste (znajomość protokołów, języków, wybór bazy danych itp., por. artykuł /1/) i wymaga specjalnego przygotowania (przeszkolenia) albo opracowania odpowiedniego interfejsu dla użytkowników końcowych, który umożliwi im swobodne poruszanie się po tych różnorodnych serwisach, bazach danych i ich środkach wyszukiwania.

Funkcja pośrednicząca biblioteki będzie jeszcze bardziej widoczna w przyszłości, kiedy to dalsze rozpowszechnianie technologii komputerowej umożliwi dostęp do zasobów i źródeł bez konieczności przychodzenia do biblioteki, za pośrednictwem terminala umieszczonego w szkole, w biurze, w akademiku, w domu. Prowadzi to do innego rozumienia idei biblioteki, należy przestać myśleć o niej w kategoriach obiektu (jednostki fizycznej), a zacząć myśleć o bibliotece jako organizacji, jako sumie źródeł informacyjnych, do których umożliwia dostęp. Ewolucja technologii informacyjnej przyspieszyła postrzeganie informacji, niezależnie od nośnika, na którym jest utrwalona i od zewnętrznego otoczenia fizycznego, w którym jest przechowywana (np. biblioteki).

Jednym ze sposobów patrzenia na rozwój bibliotekarstwa jest zaobserwowanie prób fizycznego zolizowania źródeł informacji do użytkownika, ułatwienia korzystania z nich, zwiększenia ich dostępności (wolny dostęp do półek, księgozbiór podręczny, publiczne katalogi, itp.). Być może przyszłość całkowicie wyeliminuje bariery między użytkownikiem a źródłem; po to, aby uzyskać informację, skorzystać z dokumentów nie będzie już konieczne fizyczne przemieszczenie się użytkownika, to informacja zostanie przemieszczona do miejsca, w którym się on znajduje.

JĘZYK WYSZUKIWAWCZY JAKO JĘZYK WYRAŻAJĄCY POTRZEBY UŻYTKOWNIKÓW

W nowym modelu biblioteki udostępniają informacje nie tylko o swoich zbiorach, zapewniają dostęp nie tylko do własnej katalogowej bazy danych, ale do baz danych innych bibliotek i serwisów bibliograficznych. Język wyszukiwawczy, za pomocą którego możliwe byłoby wyszukiwanie we wszystkich tych systemach, powinien być niezależny od jednej bazy danych, nie powinien, jak to bywało

w systemach tradycyjnych, odzwierciedlać jedną bazę danych i stosowan w niej język indeksowania, lecz służyć do wyrażenia zapytania i potrzeb użytkowników.

Przegląd istniejących JIW wykazuje, że najczęściej odzwierciedlały one albo strukturę księgozbioru, dla którego były tworzone, albo ogólny stan piśmiennictwa w danej dziedzinie na podstawie klasyfikacji nauk. Sytuacja pierwsza miała miejsce, jeśli JIW tworzony był dla konkretnej biblioteki (np. Klasyfikacja Biblioteki Kongresu), druga, jeśli tworzony był niezależnie od księgozbioru jednej biblioteki, dla większej ich liczby (np. Klasyfikacja Dziesiąta Dewey'a lub Klasyfikacja Bibliograficzna Blissa). Z zadań biblioteki wynika jednak, że JIW powinien odzwierciedlać strukturę nie zbioru i nie nauki, ale strukturę potrzeb użytkowników. Biblioteki jako instytucje usługowe powinny dbać o dostosowanie się do potrzeb swoich klientów.

W systemie manualnym użytkownik budował (czasami przy pomocy bibliotekarza) instrukcję wyszukiwawczą zgodnie z wzorcem: "Jak książka na interesujący mnie temat mogła być zaindeksowana? Jakie hasła przydzielono jej w procesie indeksowania?". Nic więc dziwnego, że przy takim podejściu punktem odniesienia, głównym obiektem działalności, był dokument, a nie użytkownik. Użytkownik, podobnie jak twórca dokumentu (autor), znajdował się poza systemem bibliotecznym rozpatrywanym w funkcji wyszukiwawczej (por. rvs.1). Obecnie, przy ześrodkowaniu uwagi na użytkowniku, bibliotekarz indeksujący dokument powinien się zastanawiać przy przydzielaniu haseł: "Dla jakiego użytkownika może być ta książka interesująca? Jakie pytanie może postawić? Jakich terminów użyje?" Tak więc, nie język wyszukiwawczy powinien odzwierciedlać strukturę zbioru ChWD, lecz język indeksowania powinien być bardziej dostosowany

do zbioru IW wyrażających potrzeby informacyjne użytkowników. Język wyszukiwawczy powinien być skierowany nie na dokument, nie powinien być odbiciem struktury zbioru informacyjnego i pośrednio formalnych i treściowych cech zbioru bibliotecznego, lecz odbiciem potrzeb użytkowników. Do tej pory, nawet jeśli nastawiony był na użytkownika, to głównie w celu jak najwierniejszego odzwierciedlenia w postaci IW zawartości jego zapytania (np. rozbudowane profile użytkowników w systemie typu SDI) - potrzeby informacyjnej już nie uwzględniał (nie analizował).

Potrzeby użytkowników nie są czymś stałym, zależą od wielu czynników obiektywnych i subiektywnych. Wiadomo, że zadowolenie z dokumentów otrzymanych w odpowiedzi na pozornie to samo pytanie (formalnie identyczne) może być różne u różnych użytkowników. Przykładem czynników obiektywnych mających wpływ na charakter potrzeb informacyjnych może być rodzaj działalności uprawianej przez użytkownika (zawód, funkcja społeczna); zupełnie inne są potrzeby naukowca badającego jakieś zjawisko niż konstruktora projektującego jakieś urządzenie wykorzystujące to zjawisko lub ucznia przygotowującego na ten temat referat do szkół. Czynniki takie, jak poziom i rodzaj wykształcenia użytkownika, umiejętność korzystania z literatury fachowej, znajomość języków obcych, jego cechy osobnicze, ambicje zawodowe, system kulturowy, polityczny i ekonomiczny, do którego należy itp. jeśli nie decydują, to przynajmniej wpływają w znacznym stopniu na decyzję wyboru odpowiednich dokumentów i ocenę ich przydatności. Wiedzą o tym pracownicy oddziałów informacyjnych bibliotek, którzy intuicyjnie oceniali te elementy na podstawie wywiadu przeprowadzonego z czytelnikiem przed przystąpieniem do wyszukiwania. Obecnie w systemach zautomatyzowanych obsługujących użytkowników końcowych

musi ich w tym zastąpić odpowiednio przygotowany interfejs.

Aby system mógł zidentyfikować i określić potrzeby wpływające z wymienionych czynników musi posiadać pewne dane, z których przy wykorzystaniu wbudowanych reguł wnioskowania może je wyprowadzić. Tak więc, do języka wyszukiwawczego należy wprowadzić elementy opisujące samego użytkownika (jego wykształcenie, zawód, znajomość języków obcych, itp.), umożliwiające utworzenie profilu użytkownika. Profil ten, przez analogię do ChW0, można nazwać **Charakterystyką Wyszukiwawczą Użytkownika (ChWU)**, a jego elementy powinny składać się na słownik języka wyszukiwawczego, w takim samym stopniu, jak elementy ChW0 tworzą słownik języka indeksowania. Profil powinien zawierać te cechy charakteryzujące użytkownika, które mogą okazać się przydatne przy ustalaniu jego faktycznych potrzeb informacyjnych w poszczególnych sesjach wyszukiwawczych.

Do tej pory elementy charakteryzujące użytkownika nie owiły z reguły wykorzystywane przy wyszukiwaniu, chociaż bibliotekarz pośredniczący w wyszukiwaniu prawdopodobnie niektóre z tych cech uwzględniał przy formułowaniu strategii wyszukiwawczej i ocenie wyników (często nawet nieświadomie). Obecnie, przy coraz częstszym korzystaniu z systemu przez użytkowników końcowych, ich profile powinny być wprowadzone na stałe do systemu (tworząc oddzielną bazę danych użytkowników), a odpowiednio opracowane interfejsy powinny zapewniać mechanizmy wykorzystywania tych danych do wyboru odpowiedniej formy dialogu i strategii wyszukiwawczej.

FUNKCJA BIBLIOTEKI W TWORZENIU PROFILU UŻYTKOWNIKA

Proces oddzielania się języka wyszukiwawczego od języka indeksowania spowodowany wprowadzeniem nowych technologii informacyjnych jest jeszcze bardziej wyraźny przy wspomnianych już zmia-

nach modelu biblioteki, pewnych jej funkcji i komponentów. Pomimo wżenia się gotowych serwisów informacyjnych i ich dostępności online za pośrednictwem sieci telekomunikacyjnej bezpośrednio korzystanie z nich przez osoby indywidualne nie jest sprawą prostą. Obecnie istnieje ogromne bogactwo informacji potencjalnie (!) dostępnej na rynku, jednak sama jej objętość, złożoność i różnorodność są przytłaczające dla przeciętnego odbiorcy. Trudności, z jakimi spotyka się użytkownik indywidualny przy próbach korzystania z systemów online są liczne: gdzie szukać (który serwis, która baza danych, jaki adres), skomplikowana procedura rejestracji i łączenia z serwisem, zróżnicowane języki dialogu i zestaw środków wyszukiwawczych, liczne bazy danych o różnej strukturze, zawartości, zakresach, zasięgach i aktualności, to wszystko najczęściej przekracza możliwości użytkownika. Aby skorzystać z systemu, musi on dysponować odpowiednim sprzętem, musi mieć umowę z dostawcą informacji i własne konto, musi mieć przydzielony numer identyfikacyjny, hasło, musi znać procedurę korzystania z sieci i z systemu, a przede wszystkim musi dysponować funduszami na opłacenie tych usług. Konsekwencją tych trudności dla indywidualnego użytkownika jest nieopłacalność korzystania z systemów tego typu z punktu widzenia nakładów czasu i pracy poświęconych na dostęp oraz potrzebnych środków finansowych. Biblioteka może być tą organizacją, która mu to ułatwi, zapewni niezbędny sprzęt, kontrakt, hasło, a może nawet i pieniądze, pomoże przebrnąć przez "dżunglę" serwisów informacyjnych i baz danych oraz przepisów i zasad korzystania z nich. Spowoduje to pewną zmianę relacji między użytkownikiem i biblioteką.

Ewolucja funkcji biblioteki prowadzi do włączenia się modelu biblioteki-pośrednika, łącznika zapewniającego dostęp do własnych

i zewnętrznych źródeł informacji, umożliwiającemu tym, którzy nie posiadają własnego komputera korzystanie z nich na miejscu (w bibliotece), a tym, którzy go posiadają - korzystanie zdalne (spoza biblioteki). Obok funkcji tradycyjnych i wspomnianych funkcji pośredniczących zadaniem biblioteki będzie przetwarzanie i dostosowywanie informacji do potrzeb własnych odbiorców. Bibliotekarze, zwolnieni z wielu rutynowych i mechanicznych prac mogą więcej czasu poświęcić na działalność intelektualną, dokładniejszą selekcję zbiorów, działalność informacyjną, rzeczowe opracowanie zbiorów pod kątem potrzeb i poziomu własnych użytkowników. Aby móc dobrze wypełniać te zadania biblioteki muszą być w większym stopniu niż dotychczas zorientowane na użytkownika, co z kolei wymaga wypracowania mechanizmów lepszego rozpoznawania i przedstawiania ich cech charakterystycznych, potrzeb, przzwyczajzeń i zainteresowań (dlatego w modelu biblioteki na rys. 2 użytkownik znajduje się w obrębie przerywanej linii, staje się elementem biblioteki jako organizacji). Biblioteka powinna zbierać i utrzymywać, podobnie jak informacje o swych zbiorach, tak też informacje o swych użytkownikach i w większym stopniu wykorzystywać je później przy wyszukiwaniu.

Zadaniem biblioteki jest przekazywanie informacji do potrzebujących jej w sposób i w formie dla nich najwygodniejszej. Aby system mógł obsługiwać użytkowników zgodnie z ich faktycznymi potrzebami, powinien potrafić dostosować się do ich wymagań, poziomu i doświadczenia. Zdolność adaptacji systemu do użytkownika jest szczególnie ważna, kiedy system obsługuje bardzo zróżnicowaną populację, jak to ma miejsce w bibliotekach publicznych. Ułatwioną sytuację mają biblioteki specjalistyczne lub ośrodki informacji, które bezpośrednio współpracują ze swymi użytkownikami - specjalis-

tami z danej dziedziny, które znają swoich odbiorców, są z nimi w stałym kontakcie, mogą lepiej i szybciej reagować na ich zmieniające się potrzeby. W bibliotece publicznej czytelnik jest w dużym stopniu anonimowy. Owszem, biblioteki mają pewne dane o swych czytelnikach (nazwisko, wiek, miejsce pracy, adres), ale po pierwsze, są one niewystarczające, a po drugie, i tak nie były wykorzystywane do oceny potrzeb i przygotowania użytkownika w procesie wyszukiwania. Wynikało to ze zróżnicowanych funkcji dotychczasowych bibliotek. Obecnie funkcje te ulegną zmianie, nie będą się tak różnić w zależności od typu biblioteki. W sytuacji, kiedy wszystkie serwisy informacyjne będą osiągalne z każdej biblioteki, także z publicznej, można przypuszczać, że użytkownicy zamiast korzystać z wielu wyspecjalizowanych bibliotek będą korzystać z jednej, najbliższej zlokalizowanej. Badania wykazały, że użytkownika cechuje tendencja do wykorzystywania wszelkich dostępnych mu lokalnych instytucjonalnych źródeł informacji, zanim podejmie trud zidentyfikowania i skontaktowania się z wyspecjalizowanymi ośrodkami informacji /6/. Oznacza to, że bliskość i znajomość biblioteki może odgrywać coraz większe znaczenie. Dotąd użytkownik skazany był często na korzystanie z systemów specjalistycznych, gdyż biblioteki publiczne nie dysponowały tego typu informacjami. W przyszłości funkcje i usługi poszczególnych bibliotek nie będą tak zróżnicowane, bardziej zróżnicowani natomiast będą ich użytkownicy i ich potrzeby. Dla ustalenia tych potrzeb biblioteki powinny zacząć tworzyć profile swoich stałych (częstych) użytkowników. Oczywiście, pojawi się tu problem ochrony danych, ich prywatności, tj. kto ma prawo wglądu w te dane, kto może z tych profilów korzystać i w jakim celu, kto może je zmieniać, itp.

Suma wszystkich profili użytkowników (baza danych użytkowników) jest kolektywnym zbiorem zainteresowań i potrzeb informacyjnych czytelników biblioteki. Pozwala to budować "lokalne" bazy danych odzwierciedlające indywidualne i grupowe zainteresowania i potrzeby użytkowników, pozwala je badać i dostosowywać do nich pozostałe elementy systemu bibliotecznego, np. zbiory. Bibliotekarz wreszcie dostał do ręki sprawne narzędzie rzeczowego badania potrzeb użytkowników.

ELEMENTY PROFILU UŻYTKOWNIKA

Aby efektywnie wypełniać funkcje pośredniczące, pomagać w wyborze bazy danych do przeszukiwania, w wypracowaniu najlepszej strategii i jej modyfikowaniu, w ocenie i selekcji najbardziej relewantnych dokumentów itp., system powinien dysponować profilami użytkowników wbudowanymi na stałe w interfejs komputera.

Co powinno składać się na taki profil? Wśród informacji charakteryzujących użytkownika należy wyróżnić takie, które są informacjami stałymi, niezmiennymi lub przynajmniej nie podatnymi na ciągłe zmiany (np. wykształcenie, znajomość języków obcych) oraz takie, które mogą ulegać zmianom w poszczególnych sesjach wyszukiwawczych (np. cel wyszukiwania).

Profile wbudowane na stałe w interfejs komputera (tworzące bazę danych użytkowników) powinny zawierać tylko te elementy, które są względnie stałe oraz te elementy zmienne, które system sam może modyfikować na podstawie raz podanych danych (np. wiek użytkownika). Elementy różniące się w poszczególnych sesjach muszą być uzyskiwane każdorazowo od użytkownika na zasadzie wywiadu.

Informacjami, które posłużyć mogą do oceny zdolności użytkownika do korzystania z dokumentów, tj. ocenić poziom, zawartość

i charakter dokumentów wydawanych w odpowiedzi pod kątem potrzeb, mogą być:

- 1) wiek (inny zestaw dokumentów dla dzieci, inny dla dorosłych);
- 2) poziom i profil wykształcenia;
- 3) zawód wykonywany, rozumiany także jako środowisko (inny typ informacji odpowiedni będzie dla różnych grup środowiskowych, np. dla pracownika naukowego, studenta, ucznia, dziennikarza, konstruktora, pracownika administracyjnego, robotnika);
- 4) pełnione funkcje i stanowisko służbowe (np. kadra kierownicza);
- 5) dodatkowe zainteresowania.

Na podstawie tych danych system powinien ustalić pewne kryteria, które pozwolą mu ocenić stopień relewantności wyszukanych dokumentów, a wcześniej sformułować odpowiednią strategię wyszukiwawczą. Jednym z takich kryteriów jest znajomość przez użytkownika dziedziny (tematu), której dotyczy wyszukiwanie. Wiadomo, że innych dokumentów oczekuje specjalista z danej dziedziny, znający dobrze problematykę, a innych laik, który często nie potrafi dobrze sprecyzować tematu wyszukiwania. Specjalista z reguły pragnie uzupełnić, podwyższyć swoją wiedzę, dowiedzieć się o aktualnym stanie i kierunkach rozwoju reprezentowanej przez siebie dziedziny, laika zadowolą często podstawowe, ogólne informacje typu encyklopedycznego. Znajomość problematyki inna jest z reguły, jeśli jest to dziedzina działalności zawodowej użytkownika, a inna jeśli jest to dziedzina dodatkowych zainteresowań (hobby). Wśród użytkowników wymagających wysokiego poziomu naukowego, fachowego, rozróżnić należy specjalistów zajmujących się daną problematyką stale (zawodowo, naukowo) i specjalistów z dziedzin

pokrewnych. Innym rozróżnieniem wartym odnotowania jest podział na "teoretyków" (np. pracowników naukowych) i praktyków (lekarz, inżynier budowlany).

Wyróżnić można więc co najmniej cztery grupy użytkowników, biorąc pod uwagę ich znajomość tematu, którego dotyczy pytanie:

- 1) specjalista zajmujący się tym tematem i obeznany z literaturą;
- 2) specjalista z dziedzin pokrewnych;
- 3) praktyk (z praktyczną znajomością tematu, ale nie obeznany z literaturą);
- 4) laik.

Do której z tych grup należy zaliczyć użytkownika system może ustalić na podstawie stałych elementów charakterystyki użytkownika (zawód, poziom i profil wykształcenia, zainteresowania), jak również na podstawie elementów zmiennych (dodatkowych pytań). Dobrze opracowany "inteligentny" interfejs powinien wykorzystywać do tej oceny dodatkowe elementy, np. terminologia użyta przy formułowaniu pytania (terminy potoczne, specjalistyczne, naukowe) świadczyć może w pewnym stopniu o jego znajomości problemu i typie oczekiwanej odpowiedzi. Należy też pamiętać, że znajomość tematu może ulec zmianie po przeczytaniu pierwszych dostarczonych dokumentów, dlatego system powinien zapamiętać prowadzone już wyszukiwania i przy kolejnych sesjach proponować uzupełniający zestaw dokumentów.

Nie wszystkie informacje potrzebne do ustalenia odpowiedniej strategii wyszukiwawczej i kryterium relewancji można zdobyć na podstawie stałych elementów znajdujących się w "karcie" użytkownika. Niektóre elementy mogą się zmieniać przy poszczególnych sesjach. Informacje te system powinien uzyskiwać każdorazowo, na zasadzie wywiadu z użytkownikiem (tak jak to robił bibliotekarz).

Wśród pytań powinny się znaleźć przykładowo pytania o:

1. cel wyszukiwania;

2. liczbę oczekiwanych pozycji zależną od tego, ile użytkownik planuje przeczytać (informacje te można czasami wywnioskować na podstawie odpowiedzi na pytanie pierwsze, o cel);

3. słowa kluczowe lub frazy najlepiej opisujące zapytanie użytkownika, wraz z synonimami, terminami węższymi, szerszymi, pokrewnymi; jeśli użytkownik nie jest w stanie ich wymienić system mu w tym pomaga na podstawie wbudowanego słownika (tezaurusu); charakter i ilość terminów podanych przez użytkownika oraz ich poprawność może, jak już wspomniano, dać obraz znajomości tematu przez użytkownika;

4. ograniczenie co do daty wydania dokumentu (informacje na temat ograniczenia pod względem języka dokumentu system posiada na podstawie charakterystyki użytkownika);

5. pożądany format i sposób podania wyników (np. na monitorze czy wydruk).

Oczywiście, nie wszystkie te pytania podane są na raz, będą się one sukcesywnie pojawiać w odpowiednich momentach na poszczególnych etapach wyszukiwania i nie zawsze - np. pytanie dotyczące ograniczenia daty wydania dokumentów pojawi się tylko w przypadku zbyt dużej liczby pozycji wyszukanych.

Niedocenianym często elementem mającym ogromny wpływ na dobór dokumentów jest cel wyszukiwania. Określa to w pewnym stopniu wagę, jaką użytkownik przywiązuje do odnalezienia informacji, jej szczegółowość, poziom. Celem wyszukiwania może być chęć zdobycia ogólnych (elementarnych) informacji na jakiś temat, uzupełnienie (aktualizacja) wiadomości już posiadanych, zdobycie pełnego zestawu dokumentów w celu utworzenia bibliografii, sprawdzenie

w pierwszej lepszej pracy brakujących danych (np. daty wydarzenia historycznego) lub sprawdzenie, czy ktoś już zajmował się danym problemem. Znajomość celu pozwoli ustalić szerokość wyszukiwania, liczbę pozycji wydanych i ich poziom.

Dwie skrajne sytuacje występują, kiedy użytkownik prosi o wszelką informację dotyczącą danego problemu lub kiedy szuka konkretnej odpowiedzi na konkretne pytanie. W drugim wypadku, kiedy użytkownik chce zdobyć informacje dotyczące np. daty wydarzenia, parametrów materiału, struktury chemicznej związku, działania leku, danych statystycznych itp. często pierwszy dokument, który udzieli mu tych informacji całkowicie zaspokoi jego potrzebę informacyjną. Jeśli jednak prowadzi badania, pisze pracę doktorską lub tworzy bibliografię—zależy mu na wyszukaniu jak największej liczby dokumentów, a fakt, że wiele z nich będzie "powtarzać" pewne informacje nie oznacza, że nie powinien ich przejrzeć.

Ustalenie pełnego zestawu elementów charakterystyki użytkownika składających się na jego profil i zestawu pytań prowadzących do zidentyfikowania jego potrzeby informacyjnej powinno być poprzedzone dokładnymi badaniami zachowań użytkowników, ich potrzeb i sposobów oceny przez nich stopnia relewancji wyszukanych dokumentów w zależności od różnych sytuacji wyszukiwawczych.

MIEJSCE PROFILU UŻYTKOWNIKA W SYSTEMIE WYSZUKIWAWCZYM

Dla wypracowania najefektywniejszej strategii wyszukiwawczej i wybrania najodpowiedniejszych środków dialogu oraz dokumentów system powinien dysponować jak największą ilością danych o użytkowniku i o jego aktualnej potrzebie informacyjnej. Może te informacje uzyskiwać każdorazowo, przy kolejnych sesjach wyszukiwawczych poprzez zadawanie pytań "Czy korzystałeś już z systemu?", "Ile masz

lat?, "Czy jesteś uczniem, studentem, nauczycielem...." itd. Metoda ta byłaby jednak czasochłonna i irytująca dla użytkownika, tym bardziej że część z tych danych powtarzana byłaby przy każdym wyszukiwaniu. Dlatego, jak już stwierdziliśmy, lepszym rozwiązaniem jest umieszczenie tych danych na stałe w systemie i wywoływanie ich przez każdą sesję wyszukiwawczej użytkownika (po podaniu przez niego swojego numeru identyfikacyjnego). W takim wypadku użytkownik musiałby odpowiedzieć tylko na te pytania, które związane są ze zmiennymi cechami relewantnymi dla danego wyszukiwania.

Profil użytkownika powinien być integralnym elementem interfejsu. Wyłania się teraz pytanie, gdzie taki interfejs umieścić. Może się on znajdować w różnych miejscach między użytkownikiem a systemem docelowym, czyli tym, w którym następuje faktyczne wyszukiwanie, np. bazą danych, którą użytkownik zamierza przeszukiwać za pośrednictwem biblioteki i sieci telekomunikacyjnej. Każde z rozwiązań ma swoje plusy i minusy, wydaje się jednak, że skoro interfejs wymaga stałej interakcji (uzupełniania danych) lepszym rozwiązaniem jest umieszczenie go jak najbliżej użytkownika, np. w jego komputerze domowym. Rozwiązanie to ma jednak i złe strony. Po pierwsze mikrokomputer średniej klasy, który w najlepszym wypadku jest "terminalem" użytkownika ma swoje ograniczenia co do pojemności pamięci i szybkości działania i trudne lub wręcz niemożliwe jest wprowadzenie do niego na stałe programów wymaganych w tego typu interfejsach, tj. zawierających opisy wszystkich serwisów informacyjnych, charakterystyki dostępnych w nich baz danych, stosowanych w nich słowników i środków wyszukiwawczych, programu typu ekspertowego z wbudowanymi regułami wnioskowania, itp. Po drugie, dane te ulegają stałym zmianom (powstają nowe serwisy, nowe bazy danych, stare ulegają likwidacji, wprowadzane

są nowe środki wyszukiwania, zakresy i zasięgi istniejących baz danych ulegają zmianie) i w związku z tym systemy tego typu wymagają stałej aktualizacji. Po trzecie, zakup takich programów i ich utrzymywanie byłyby zbyt kosztowne dla indywidualnych użytkowników, szczególnie dla tych, którzy korzystają z nich stosunkowo rzadko. Dlatego optymalnym rozwiązaniem, przynajmniej w najbliższym czasie, wydaje się być opracowanie i przechowywanie takiego interfejsu w instytucji pośredniczącej w wyszukiwaniu, tj. w bibliotece lub ośrodku informacji. Interfejs umieszczony byłby między użytkownikiem a pakietem programów realizujących wyszukiwanie w konkretnej bazie danych. Biblioteka lub ośrodek odpowiedzialne byłoby za opracowanie takiego programu lub jego zakup, za jego doskonalenie, przechowywanie, zabezpieczanie i aktualizację. Biblioteka troszczy się o to, aby zapewnić dostęp do jak największej liczby serwisów informacyjnych załatwiając wszelkie formalności z tym związane, wprowadza te dane do systemu, ponosi koszty związane z zakupem i rejestracją, a użytkownik opłaca tylko koszty związane z faktycznym wyszukiwaniem (połączenie telefoniczne i korzystanie z bazy danych). Obowiązkiem biblioteki jest zbieranie danych o użytkowniku i jego potrzebach, przechowywanie ich, wywoływanie do interfejsu w celu wykorzystania przy wyszukiwaniu. Większość z czasochłonnych prac związanych z wyszukiwaniem może być wtedy wykonana na miejscu, w bibliotece, np. ustalenie strategii wyszukiwawczej, co prowadzi do krótszego czasu wyszukiwania online a tym samym do zmniejszenia kosztów.

Stawia to przed biblioteką nowe zadania, ale i nowe możliwości, służyć może zwiększeniu jej popularności wśród użytkowników, dla których stanie się prawdziwym "oknem na świat".

PROFIL UŻYTKOWNIKA JAKO ELEMENT SYSTEMU EKSPERTOWEGO

Profil użytkownika, będący integralnym elementem interfejsu, ma być podstawą do ustalenia faktycznej potrzeby użytkownika, ma być pomocny w intelektualnych czynnościach procesu wyszukiwania, w wyborze odpowiedniej bazy danych, sformułowaniu strategii wyszukiwawczej, w ocenie i selekcji wyników wyszukiwania. Profil taki powinien rzeczywiście opisywać użytkownika, a nie cechy dokumentu, który ma być dla niego wyszukany, jak to miało miejsce w systemach typu SDI.

Przy udostępnianiu zbiorów rejestrowane są w bibliotece zawsze dwa rodzaje informacji: o dokumencie i o wypożyczającym go użytkowniku. Tak było w systemach tradycyjnych i tak jest w zautomatyzowanych. Niestety, w obu typach systemów wiadomości o użytkowniku nie były wyczerpujące i sporządzane były głównie dla celów administracyjnych (kontroli), a nie do kształtowania profili mających wpływ na strategię i przebieg wyszukiwania. Postęp w zakresie technologii informacyjnej pozwala tworzyć bazy danych użytkowników i wykorzystywać je efektywnie do wyboru sposobów wyszukiwania i jego rezultatów, tj. tworzyć zindywidualizowane profile użytkowników i dostosowywać do nich środki dialogu i strategii wyszukiwawczej.

Efektywne wykorzystywanie tego typu baz danych wymaga pewnych zmian w interfejsie i w samym systemie wyszukiwawczym, a konkretnie w jego języku informacyjno-wyszukiwawczym. Powiedzieliśmy już, że użytkownik dopiero teraz zaczyna być traktowany jako podmiot działalności informacyjnej, do niedawna centralnym punktem systemu był dokument. Wyszukiwanie realizowano według zasady: Jaka może być ChW dokumentu, który zaspokoi potrzebę użytkownika, odpowie na jego pytanie? Obecnie zasada ta powinna być zmieniona, w proce-

sie indeksowania należy zadać pytanie: Jakie potrzeby informacyjne użytkownika dany dokument zaspokoi, na jakie zapytanie będzie odpowiadać? Zmiana zasad wyszukiwania implikuje zmianę zasad indeksowania, wprowadza konieczność modyfikacji bazy danych opisów dokumentów, charakterystyk wyszukiwawczych. Muszą się w nich pojawić elementy, które do tej pory nie występowały, elementy umożliwiające porównanie CHWD z opisami potrzeb użytkowników w celu wyselekcjonowania tych, które te potrzeby najlepiej zaspokoją. Elementami takimi będą np. informacje dotyczące poziomu pracy, jej aktualności, przeznaczenia czytelniczego, itp.

Błędem byłoby przypuszczać, że suche dane zgromadzone w bazie danych użytkowników wystarczą do ustalenia jego stopnia przygotowania do korzystania z informacji. W rzeczywistości dopiero na ich podstawie można zbudować faktyczny profil użytkownika, tak więc dane te nie stanowią profilu lecz służą do jego utworzenia. Jeśli interfejs tego typu zastąpić ma specjalistę - człowieka pośredniczącego w wyszukiwaniu, musi sam na podstawie podanych cech wyciągać wnioski, oceniać i podejmować decyzje. Tak więc musi być zbudowany na zasadach systemu ekspertowego. System tego typu powinien dostosowywać się do poziomu użytkownika na podstawie wprowadzonych danych (z bazy danych użytkowników), na podstawie uzyskanych w danych sesji odpowiedzi na postawione pytania oraz na podstawie doświadczeń z pracy z użytkownikiem w sesjach poprzednich, które powinien zapamiętywać. W artykule /1/ określono system przyjazny, odwołując się do potocznego rozumienia tego wyrazu, jako system lubiany przez użytkowników, do którego chętnie wracają. Teraz można dodać jeszcze jedną cechę takiego systemu - system przyjazny to taki, który dobrze zna użytkownika, pamięta jego zainteresowania, słabości, przyzwyczajenia i upodobania.

Interfejs tego typu ma być systemem "inteligentnym", oprócz umiejętności wyciągania wniosków z wprowadzonych danych powinien być systemem samouczącym, powinien wykorzystywać dotychczasowe doświadczenia w kontaktach z użytkownikiem na tej samej zasadzie, jak pracownicy informacji pamiętali swoich stałych klientów. Ponadto, tak jak pracownik informacji potrafi wyrobić sobie obraz użytkownika i jego umiejętności po rozmowie z nim przeprowadzonej przed przystąpieniem do wyszukiwania (na podstawie jego wypowiedzi, zachowania, reakcji), tak też system ekspertowy powinien wykorzystywać te dane do oceny jego potrzeb. Na przykład zbyt powolne "wstukiwanie" odpowiedzi może być podstawą decyzji o zamianie klawiatury na ekran dotykowy. Na podstawie elementarnych lub zbyt częstych błędów może ocenić czy jest to użytkownik przypadkowy, czy nowicjusz. Błędy w pisowni mogą wskazywać na to, że mamy do czynienia z użytkownikiem o niskim poziomie wykształcenia i odpowiednio do tego poziomu selekcjonować wyszukiwane prace. Jeśli terminy użyte do sformułowania zapytania są niepoprawne (nie są terminami fachowymi, naukowymi) lub jeśli użytkownik ma trudności z ustaleniem synonimów oznacza to, że prawdopodobnie temat jest mu obcy lub mało znany. Oczywiście, czasami wnioski te mogą być błędne, ale czy pośredniczący specjalista nigdy się nie mylił?

Jednym słowem: system powinien być na tyle "inteligentny", by sam ustalił o co właściwie chodzi użytkownikowi, który czasami sam nie potrafi dostatecznie jasno i precyzyjnie sformułować zapytania. Zaprojektowanie takich interfejsów wymaga wypracowania specjalnej logiki pytań, metod strukturalizacji i reprezentacji wiedzy, reguł wnioskowania. Następną generacją interfejsów wykorzystująca rozbudowane środki sztucznej inteligencji uwolni użyt-

kownika od wszelkich trudności związanych z wyszukiwaniem informacji. Interfejsy ewoluować będą w kierunku bardziej naturalnych, tj. prowadzących rozmowę w języku użytkownika (łącznie ze stosowaną przez niego terminologią zależną od jego wykształcenia, zawodu i środowiska, z którego pochodzi), bardziej osobistych (indywidualizowanych), tolerancyjnych i prostych. Przy ich pomocy użytkownik może oczekiwać takiej łatwości poruszania się po różnych serwisach informacyjnych i ich bazach danych, jak obecnie z łatwością przechodzi od bibliografii do katalogu, a z katalogu do półek z książkami. Interfejs tego typu pomoże sformułować zapytanie, wybrać część danych, pasujące do niej słowa kluczowe, doradzi przy wyborze terminów, utworzy instrukcję wyszukiwawczą, oceni wyniki, dokona selekcji, jeśli uzna rezultaty za niesatysfakcjonujące - określi przyczynę niepowodzeń i zaproponuje modyfikację strategii wyszukiwawczej, sprawdzi poprawność zapisów, skoryguje błędy, pomoże w przesłaniu i wydrukowaniu wyników w najbardziej odpowiedniej formie, a to wszystko na zasadzie "inteligentnej" rozmowy. Interfejs tego typu pozwoli zbudować profil użytkownika przedstawiający jego własną wizję informacji w systemie, własny zbiór komend i sposoby wyszukiwania.

Literatura

1. CHMIELEWSKA -GORCZYCA E. O tak zwanych systemach przyjaznych użytkownikowi. Zag.Inf.Nauk. 1990 nr 2(57) s.45-82.
2. CHMIELEWSKA-GORCZYCA E. Struktura instrukcji wyszukiwawczej. Zag.Inf.Nauk. 1984 nr 1(44), s.49-63.
3. GORMAN M. The Impact of Technology on the Organization of Libraries. CLSI Publications 1985 London.

4. HIERPPE R. What artificial intelligence can, could, and can't do for libraries and information services. W: 7th ONLINE Information Meeting. London 6-8 December 1983. Learned Information 1984 Oxford, s.7-25.
5. POMYKALSKI A. System informacji w badaniach naukowych. PWE 1980 Warszawa.
6. TURNER Ch. Organizing Information. Principles and practice. Clive Bingley 1987 Oxford.
7. WOŹNIAK J. O tak zwanych potrzebach informacyjnych. Zag.Inf.Nauk. 1989 nr 1(54), s.39-59.

25.02.1991 r.

INFORMATION RETRIEVAL LANGUAGE AND INFORMATION USERS' NEEDS

Summary

There is pointed out, in the article, the necessity of adjusting the elements of the information retrieval system to the individualized information needs of the users. There are distinguished several categories of users view of their differentiated experience in using the automatized information retrieval systems and of their expectations as far as the results of the retrieval are concerned (kind of the results, their scope and level). The features of the information retrieval system needed for each category of users are qualified as well as the methods of ensuring of these features. There is discussed the interdependence between the indexing language and the retrieval language in the traditional and automatized information retrieval systems. There are shown greater and greater differences between these languages as the new tools have been introduced (the new information technologies) and as the informational and library

functions have evaluated. There is presented the information retrieval language as a language which serves expressing the users' real needs, depending on such elements as: age, level and profile of education, profession etc. There is indicated the necessity to establish users' profiles (users data bases) and to use these profiles when formulating the search strategy. There are distinguished the permanent and varying elements of which the user profile is composed of (among others the aim of searching, time limits regarding the date publication of documents, command of foreing languages and so on) There is presented the role of the library in building of interfaces which meet the individualized and varyig needs of its users.

Информационно-поисковый язык а информационные потребности потребителей

Р е з ю м е

В статье указана необходимость приспособления элементов информационно-поисковой системы к индивидуализированным информационным потребностям потребителей. Выделены разные категории потребителей ввиду их дифференцированной подготовки к использованию автоматизированных ИПС, а также ввиду их ожиданий относительно результатов поиска (вида, объема, уровня итп.). Для каждой категории потребителей определены требуемые признаки ИПС и методы их обеспечения. Обсуждена зависимость между языком индексирования данной ИПС и её поисковым языком в традиционной и автоматизированной системе. Указана всё большая дифференциация этих языков по мере введения новых информационных инструментов (информационных технологий) и эволюции информационных и библиотечных функций. Представлен информационно-поисковый язык как язык служащий выражению действительных информационных потребностей, зависящих от таких элементов как : возраст, уровень

и профиль образования, профессия и тп. Указана необходимость создания профилей (баз данных) потребителей и их использования в процессе формулирования поисковой стратегии. Выделены постоянные и изменяющиеся элементы из которых состоит профиль потребителя (среди прочин цель поиска, ограничение относительно даты опубликования документов, знание иностранных языков). Представлена роль библиотеки в создании интерфейсов приспособленных к индивидуализированным и изменяющимся потребностям своих потребителей.

BARBARA SOSIŃSKA-KALATA

Instytut Bibliotekoznawstwa
i Informacji Naukowej UW

TERMINOLOGICZNA BAZA DANYCH Z ZAKRESU TEORII JĘZYKÓW INFORMACYJNO-WYSZUKIWAWCZYCH

Organizacja prac nad projektem wersji wyjściowej bazy, jej ogólna charakterystyka, zakres tematyczny, ustalenie zbioru leksykalnego. Odpowiedniki obcojęzyczne terminów polskich, definicje terminologiczne. Pola uwag i materiałów ilustracyjnych. Kod logiczny terminu, kod autora, data wprowadzenia terminu. Stan prac po zakończeniu pierwszego etapu realizacji bazy.

Komentowany już od wielu lat przez wielu specjalistów niezadowolający stan źródeł terminologicznych z zakresu informacji naukowej, a także bibliotekoznawstwa, bibliografii, księgoznawstwa i innych nauk związanych z badaniem książki, informacji i komunikacji społecznej, powoduje ogromne utrudnienie zarówno procesu kształcenia studentów kierunków bibliotekoznawstwa i informacji naukowej, jak również popularyzację wśród bibliotekarzy i dokumentalistów wyników badań prowadzonych w tym zakresie. Warunkuje to przecież ewolucję środków, metod i technik ich pracy, a od tego w największym stopniu zależy możliwość usprawnienia obsługi informacyjnej użytkowników. W ostatnich latach zespół rekrutujący się spośród pracowników Instytutu Bibliotekoznawstwa i Informacji

Naukowej Uniwersytetu Warszawskiego, specjalizujących się w problematyce języków informacyjno-wyszukiwawczych, pod kierunkiem doc.dr nau. Bożenny Bojar, opracował słownik encyklopedyczny tej dziedziny, .jednakże tradycyjne techniki jego publikacji nie gwarantują szybkiego przekazania tego opracowania zainteresowanemu środowisku. Przyspieszenie procesu przygotowania i udostępniania tak potrzebnych źródeł terminologicznych można osiągnąć odchodząc od tradycyjnych ich form książkowych i wykorzystując techniki komputerowe. Działania takie znane są już na świecie od wielu lat, czego przykładem może być choćby znany system informacji o terminologii naukowej INFOTERM, jednakże w Polsce dotychczas nie podejmowano badań nad zautomatyzowanymi słownikami w postaci terminologicznych baz danych. Jesienią 1990 r. podjęto więc w Zakładzie Systemów Informacyjnych IBIN UW badania nad utworzeniem takiego automatycznego słownika, nazwanego terminologiczną bazą danych BIS, która w wersji wyjściowej ma obejmować terminologię teorii języków i systemów informacyjno-wyszukiwawczych, natomiast w wersji docelowej stać ma się dostępnym na sprzęcie mikrokomputerowym słownikiem z zakresu nauki o informacji, książce i komunikacji społecznej.

Organizacja prac nad projektem wersji wyjściowej bazy BIS-1

Prace nad projektem i przygotowaniem wersji eksploatacyjnej terminologicznej bazy danych z zakresu teorii języków informacyjno-wyszukiwawczych "BIS-1" przewidziano wstępnie w trzech etapach, które obejmują następujące zadania szczegółowe:

- etap pierwszy: "Przygotowanie roboczej wersji bazy danych"

1. Opracowanie ogólnej koncepcji terminologicznej bazy danych, na podstawie której przygotowany zostanie projekt techniczny;

2. Ustalenie zakresu tematycznego i struktury pola semantycznego reprezentowanego przez terminologię wprowadzoną do bazy;

3. Zgromadzenie podstawowego zbioru leksykalnego i określenie relacji synonimii między wybranymi terminami;

4. Przygotowanie aparatury i środowiska systemowego dla projektowanej bazy;

5. Opracowanie projektu technicznego terminologicznej bazy danych;

6. Opracowanie definicji terminologicznych i artykułów encyklopedycznych pilotażowego zbioru terminów;

7. Wprowadzenie do bazy i korekta rekordów pilotażowego zbioru terminów.

- etap drugi: "Przygotowanie wersji eksploatacyjnej bazy danych "8IS-1"

1. Opracowanie projektu technicznego wersji eksploatacyjnej terminologicznej bazy danych;

2. Opracowanie definicji terminologicznych i artykułów encyklopedycznych ok. 1250 terminów;

3. Opracowanie angielskich i rosyjskich odpowiedników terminów polskich;

4. Opracowanie struktury logicznej systemu terminologicznego reprezentowanego przez zbiór leksykalny bazy oraz kodów logicznych terminów;

5. Wprowadzenie do bazy ok. 1250 nowych rekordów i modyfikacja 50 rekordów wcześniej wprowadzonych;

6. Opracowanie roboczej wersji środków humanizacji języka komunikacji użytkownika z systemem;

7. Opracowanie technik i instrukcji wyszukiwania danych.

- etap trzeci: "Aktualizacja wersji eksploracyjnej bazy "BIS-I" i opracowanie użytkowej wersji języka komunikacji".

1. Opracowanie francuskich i niemieckich odpowiedników terminów polskich;
2. Weryfikacja i modyfikacja ok. 1300 rekordów;
3. Aktualizacja zbioru leksykalnego;
4. Opracowanie użytkowej wersji języka komunikacji użytkownika z systemem.

Realizacja pierwszego etapu prac przewidziana została na okres czterech końcowych miesięcy 1990 r. Etap drugi planuje się na rok 1991/92, natomiast realizację zadań składających się na etap trzeci i zakończenie prac nad projektowaną bazą przewiduje się w 1993 r.

Ogólna charakterystyka bazy "BIS-1"

Jak powiedziano już wcześniej, problematyka terminologicznych baz danych, a zwłaszcza projektowanie, instalowanie, praktyczne zarządzanie i wykorzystanie terminologicznych baz danych w Polsce jest dziedziną, w której dotychczas nie podejmowano badań naukowych. Należy przy tym podkreślić, że tego typu bazy danych stanowiącą istotne wzbogacenie warsztatu naukowego zapewniając stosunkowo szybką i taną edycję łatwo i systematycznie aktualizowanych słowników terminologicznych, oferujących - w porównaniu ze słownikami tradycyjnymi - znacznie więcej możliwości wyszukiwania znormalizowanego i nieznormalizowanego słownictwa specjalistycznego, a także obcojęzycznych odpowiedników terminów polskich. Terminologiczna baza danych pełni więc funkcje dwóch rodzajów słowników specjalistycznych - terminologicznego słownika objaśniającego i przekładowego słownika terminologii specjalistycznej.

Istniejące zagraniczne bazy danych, udostępniające informację terminologiczną mają charakter prostych systemów faktograficznych, zasadniczo zapewniających dostęp do informacji jedynie poprzez wybranie klucza wyszukiwawczego równokształtnego z terminem stanowiącym przedmiot poszukiwań lub z jego synonimem, względnie poprzez wybranie klucza wyszukiwawczego o kształcie częściowo pokrywającym się z kształtem pewnego terminu czy jego synonimu ogólnojęzykowego. Projektowana baza danych BIS pomyślana jest jako specyficzny rodzaj systemu, który do pewnego stopnia funkcjonuje podobnie jak system ekspertowy, operujący szczegółową informacją terminologiczną i rozbudowanymi technikami organizacji danych, ich wyszukiwania oraz językiem komunikacji użytkownika nieprofesjonalnego z systemem o stosunkowo wysokim stopniu tzw. "humanizacji".

Dostęp do informacji w bazie BIS umożliwiać będzie zróżnicowana grupa kluczy wyszukiwawczych, obejmująca zarówno wyrażenia równokształtne ze zdefiniowanymi terminami, wyrażenia tzw. nieznormalizowane, czyli synonimy zdefiniowanych terminów, obcojęzyczne odpowiedniki terminów polskich, jak i wyrażenia w pewien sposób spokrewnione z terminami zdefiniowanymi i cytowane w ich definicjach lub w artykułach encyklopedycznych oraz kody logiczne terminów zdefiniowanych i pól tematycznych (pól semantycznych), do których terminy te należą. Te dwa ostatnie rodzaje kluczy wyszukiwawczych w pewien sposób pozwalają symulować wyszukiwanie logiczne, dzięki czemu wyszukiwanie realizowane w bazie BIS zapewniać będzie nie tylko odnalezienie wyjaśnienia znaczenia określonego terminu i podstawowych informacji o jego desygnatach, ale również pozwoli ustalić zespoły terminów związanych z pewnym tematem, a zarazem encyklopedyczne noty o podstawowych elementach tego tematu.

Ogólna struktura bazy BIS obejmować będzie dwa podstawowe zbiory informacyjne: zbiór główny i tzw. zbiór wyszukiwawczy^{1/}. Zbiór główny zawierać będzie materiał leksykograficzny zorganizowany w rekordy obejmujące dziewięć rodzajów danych:

- A. termin standaryzowany przyjęty jako jednostka definiowana,
- B. wyrażenia synonimiczne wobec terminu zdefiniowanego,
- C. odpowiedniki obcojęzyczne terminu zdefiniowanego (w językach: angielskim, francuskim, niemieckim i rosyjskim),
- D. definicje terminologiczne,
- E. uwagi o charakterze komentarza do definicji terminu i artykułu encyklopedycznego zawierającego najistotniejsze informacje o desygnatach terminów zdefiniowanych,
- F. materiały ilustracyjne (schematy graficzne wspomagające eksplikację słowną),
- G. kod logiczny terminu,
- H. kod autora definicji i uwag,
- I. data wprowadzenia rekordu.

Zbiór wyszukiwawczy generalnie zawierać będzie wszystkie takie wyrażenia, które mogą być wykorzystywane jako klucze wyszukiwawcze. Zbiór ten podzielony jest na siedem głównych podzbiorów:

- 1. terminy polskie cytowane w tekście definicji i uwag encyklopedyczno-komentujących,
- 2. terminy polskie zdefiniowane jako główne hasła definicji,
- 3. terminy angielskie, wymienione jako odpowiedniki zdefiniowanych terminów polskich,

^{1/} Struktura bazy danych realizowanej za pomocą oprogramowania COS/ISIS zdefiniowana jest przez logiczną organizację baz danych budowanych w oparciu o to oprogramowanie. Szczegółową charakterystykę tej struktury zawarto w opracowaniu "Projekt techniczny terminologicznej bazy danych z zakresu teorii języków informacyjno-wyszukiwawczych" przygotowanym przez U. Ohnsorge - omówiony w bieżącym numerze ZIN s. 73-90.

4. **terminy francuskie**, wymienione jako odpowiedniki zdefiniowanych terminów polskich.
5. **terminy niemieckie**, wymienione jako odpowiedniki zdefiniowanych terminów polskich,
6. **terminy rosyjskie**, wymienione jako odpowiedniki zdefiniowanych terminów polskich,
7. **kody logiczne terminów zdefiniowanych**.

W podzbiorze (2) wydzielono ponadto podzbiór terminów standaryzowanych, które przyjęto jako hasło definicji oraz podzbiór terminów będących ich synonimami.

Roboczą bazą danych wspomagającą budowę terminologicznej bazy "BIS" jest utworzona strukturalnie niezależnie, w oparciu o oprogramowanie dBase III plus, prosta baza "KARTOTEKA", w której gromadzony jest wstępny materiał leksykalny. Baza ta utworzona i wykorzystywana jest wyłącznie w celach kontroli doboru i stanu opracowania słownictwa oraz organizacji prac nad przygotowaniem danych dla bazy "BIS", toteż nie przewiduje się udostępniania jej użytkownikom bazy terminologicznej.

Określenie zakresu tematycznego bazy "BIS-1"

Zakres tematyczny projektowanej terminologicznej bazy danych określono wstępnie jako pole tematyczne teorii języków informacyjno-wyszukiwawczych. Dziedzina ta jest jednym z ważniejszych, a zarazem, jednym z trudniejszych i znacznie rozbudowanych elementów nauki o informacji naukowej. Dobrze zbudowany język informacyjno-wyszukiwawczy (JIW) warunkuje właściwe funkcjonowanie każdego systemu informacyjno-wyszukiwawczego (SIW) stanowiącego obecnie główny przedmiot badań informacji naukowej. JIW jest narzędziem przeznaczonym do reprezentacji informacji przechowywanej w SIW i będącej obiektem poszukiwań użytkowników tego systemu. Jeżeli

narzędzie to jest wadliwie zbudowane, czy też należy do niskiej klasy JIW powodujących znaczne uogólnienie reprezentowanej informacji i nieprecyzyjne jej wyrażanie, nawet technicznie niezawodny SIW wykorzystujący tego typu język, nie będzie miał dużej wartości użytkowej. Warto przy tym podkreślić, że ta właśnie wartość użytkowa jest rzeczywistą wartością każdego serwisu informacyjnego, zaś techniki komputerowe powinny umożliwić jej podwyższenie, a nie - jak niekiedy się dzieje - powodować faktyczne obniżenie wartości użytkowej serwisu informacyjnego mimo nowoczesnej formy realizacyjnej.

Języki informacyjno-wyszukiwawcze są sztucznymi systemami semiotycznymi tworzonymi dla realizacji ściśle określonych rodzajów procesów komunikacyjnych, toteż zdawanie sobie sprawy z reguł budowy i funkcjonowania, jakim one podlegają konieczne jest dla prawidłowego i pełnego ich wykorzystywania, a także dla właściwego ich konstruowania oraz aktualizacji JIW już istniejących. Trudności związane z poznawaniem JIW, ich teorii i zasad użytkowania związane są m.in. ze znacznie rozbudowanym i dosyć skomplikowanym systemem terminologicznym tej dyscypliny. W dużym stopniu stąd właśnie wynika niezadowalająca jego znajomość, niezbędna przecież dla zrozumienia procesów związanych z tworzeniem i stosowaniem w działalności praktycznej JIW, ich specyficznych własności, będących konsekwencją podjęcia określonych decyzji w fazie konstruowania tych sztucznych systemów językowych. Wśród materiałów wykorzystywanych w nauczaniu JIW brak jakichkolwiek - polskich czy zagranicznych - źródeł leksykograficznych lub encyklopedycznych, które zapewniłyby optymalnie szczegółową prezentację systemu terminologicznego JIW. Źródła istniejące - bibliologiczne, bibliotekoznawcze i z zakresu nauki o informacji naukowej - dziedzinę tę z koniecz-

ność przedstawiają sposób wysoce niekompletny, przy czym sposób dotychczas opublikowany, w źródłach polskich najmłodsze ukazało się lat temu 30 w przypadku tak szybko rozwijającej się dziedziny, jaką jest teoria JIW, przesądza o małej już obecnie ich użyteczności. Jak powiedziano we wstępie do tego artykułu, w drugiej połowie lat 80-tych zespół pod kierunkiem doc.dr hab. Bożenny Bojar opracował dotychczas jeszcze nieopublikowany słownik encyklopedyczny obejmujący zagadnienia JIW i SIW, który dla powstającej bazy terminologicznej jest jednym z najważniejszych źródeł. Ze względu na hermetyczny metafizyk i stylistykę opisu znaczenia, zastosowane przy formułowaniu definicji, słownik ten jest jednak źródłem trudnym w percepcji, zwłaszcza przez tego typu użytkowników, dla których przeznaczona jest baza "BIS-1", tj. przede wszystkim dla osób praktycznie zajmujących się działalnością informacyjną w bibliotekach i ośrodkach informacyjnych oraz studentów kierunków bibliotekoznawczych. Warto przy tym zwrócić uwagę także na fakt, iż słownik ten praktycznie nie jest dostępny.

Poza brakiem ogólnodostępnych źródeł leksykograficznych dodatkową trudność w poznawaniu systemu terminologicznego JIW stwarza jego specyficzny charakter: terminologia tej dziedziny generuje się w dużej mierze z systemów terminologicznych innych nauk, takich jak semiotyka, logika, językoznawstwo, ogólna teoria systemów, cybernetyka, prakseologia, matematyka, informatyka. Trzeba jednak pamiętać, że wiele terminów przejętych z systemów terminologicznych innych dyscyplin naukowych, w teorii JIW ma przypisaną specyficzną dla tej właśnie dziedziny interpretację semantyczną - generalnie zgodną logicznie ze znaczeniem przyjętym dla tych terminów w dziedzinie macierzystej, jednakże uzupełnioną o cechy specyficzne właściwe JIW. Na przykład, choć tzw. Elementar-

ny Znak Graficzny JIW można zdefiniować jako literę JIW, to jednak pomija się w ten sposób istotną cechę Elementarnego Znak Graficznego, jaką jest posiadanie przez niego stałej wartości porządkowej, gdyż litera - zgodnie z językoznawczym znaczeniem tego terminu - nie musi cechy takiej posiadać.

Słuszne wydaje się więc przyjęcie założenia, iż teoria JIW ma już własny system terminologiczny, w części derywowany wprawdzie z systemów terminologicznych kilku innych od dawna ukonstytuowanych dyscyplin wiedzy, wykazujący jednak już swoją odrębność dość wyraźnie, aby można było podjąć prace nad jego prezentacją. Prezentacja ta powinna być przy tym możliwie najbardziej wyczerpująca, co oznacza, że powinna zawierać poza informacją stricte terminologiczną w postaci sprawozdawczych definicji znaczenia terminów, także obszerny komentarz dotyczący zarówno użycia poszczególnych terminów w literaturze przedmiotu, jak i przedstawiający najistotniejsze informacje encyklopedyczne o desygnatach tych terminów. Założenie to odbiega nieco od norm przyjmowanych zwykle w pracach terminologicznych, uznano jednak, że ze względu zarówno na brak kompleksowych opracowań encyklopedycznych lub monograficznych teorii JIW, jak i ze względu na omówione wyżej szczególne cechy jej systemu terminologicznego, przyjęcie takiej właśnie, wypróbowanej już w słowniku pod redakcją B. Bojar, a w projektowanej bazie jeszcze bardziej rozwiniętej terminologiczno-encyklopedycznej - formuły opracowania materiału leksykalnego będzie najbardziej użyteczne dla osób, do których jest ta baza adresowana.

Zakres tematyczny projektowanej terminologicznej bazy danych "BIS-1" zdefiniowano więc jako terminologię teorii JIW, czyli obejmuje on słownictwo związane z budową, opisem i wykorzystaniem tych systemów językowych. Słownictwo to wypełnia tzw. centralne pole tematyczne projektowanej bazy, natomiast pola uznane za

50

peryferyjne lecz niezbędne dla precyzyjnej eksplikacji znaczenia terminów z pola centralnego, wypełniają terminy związane przede wszystkim z rodzajami i budową źródeł informacyjnych i systemów informacyjno-wyszukiwawczych oraz terminy należące do tzw. dziedzin pokrewnych informacji naukowej szczególnie istotnych w badaniach nad JIW, takich jak językoznawstwo (szczególnie językoznawstwo strukturalne), semiotyka, logika, ogólna teoria systemów i cybernetyka.

Ustalenie zbioru leksykalnego bazy "BIS-1"

Podstawowy zbiór leksykalny dla bazy "BIS" zgromadzony w roboczej bazie "KARTOTEKA" obejmuje 1300 wyrazów należących do wyznaczonego zakresu tematycznego bazy. W bazie "KARTOTEKA" słownictwo to uporządkowane jest generalnie alfabetycznie (zgodnie z wartościami przyjętymi dla znaków alfabetu polskiego w kodzie ASCII), przy czym wyrażenia nie będące terminami, lecz istotnymi dla zakresu tematycznego bazy nazwami własnymi, (np. nazwy własne wybranych systemów informacyjno-wyszukiwawczych lub poszczególnych języków), umieszczone zostały na początku ciągu alfabetycznego. Dla każdego wyrażenia ustalono ciąg wyrazów wobec niego synonimicznych, nie decydując jednak, które z wyrazów synonimicznych powinno być uznane za termin standaryzowany, funkcjonujący jako główne hasło definicji terminologicznej. Należy podkreślić, że selekcionując słownictwo przeznaczone do definiowania w projektowanej bazie uwzględniono zarówno wyrażenia o charakterze terminów sensu stricto, jak też wyrażenia złożone, których znaczenie określone jest przez znaczenie elementów składowych, jednakże w przypadku których podanie definicji oraz encyklopedycznej charakterystyki desygnatów uznano za szczególnie wartościowe ze względów dydaktycznych

(chodzi tu przede wszystkim o wyrażenia stanowiące nazwy pewnych istotnych tematów i obiektów badawczych, np. gramatyka języka informacyjno-wyszukiawczego, słownik języka informacyjno-wyszukiawczego, produktywność klasyfikacji).

Tworząc zbiór leksykalny projektowanej bazy "GIS" zastosowano mieszaną, tzw. dedukcyjno-indukcyjną, metodę gromadzenia słownictwa, zbliżoną do tego typu metody gromadzenia materiału językowego stosowanej przy budowie tezaursów. Metodę tę charakteryzuje przede wszystkim założenie, iż materiał leksykalny winien być ustalony zarówno w oparciu o istniejące autorytatywne (tzw. normatywne) źródła encyklopedyczne i leksykograficzne, jak i na podstawie bieżącej literatury przedmiotu, zapewniającej przede wszystkim aktualizację tworzonego systemu terminologicznego i gwarantującej uwzględnienie wyrażen stosowanych współcześnie przez specjalistów.

Poddano analizie następujące rodzaje normatywnych źródeł informacji:

1. polskie słowniki z zakresu informacji naukowej, bibliotekoznawstwa i nauki o książce,
2. polskie encyklopedie z zakresu informacji naukowej bibliotekoznawstwa i nauki o książce,
3. podstawowe polskie i zagraniczne podręczniki z zakresu teorii i praktyki JIW,
4. zagraniczne wydawnictwa leksykograficzne i encyklopedyczne z zakresu bibliotekoznawstwa i informacji naukowej,
5. polskie i międzynarodowe normy terminologiczne z zakresu informacji naukowej i dokumentacji, bibliografii i bibliotekoznawstwa,
6. polskie wydawnictwa informacyjne z zakresu logiki, językoznawstwa i informatyki.

Selekcjonując słownictwo na podstawie źródeł normatywnych, z których zdecydowana większość ma charakter opracowań o zakresie szerszym, aniżeli przyjęty zakres tematyczny projektowanej bazy, posłużono się następującym ogólnym schematem podziału tematycznych, których słownictwo powinno znaleźć się w bazie:

1. Ogólna teoria informacji i komunikacji
 - 1.1. Informacja i jej rodzaje
 - 1.2. Układ informacyjny i jego elementy
 - 1.3. Procesy informacyjne
 - 1.4. Zjawiska informacyjne
 - 1.5. Własności informacji
2. Język informacyjno-wyszukiwawczy
 - 2.1. JIW - funkcje i własności
 - 2.2. JIW - typologie
 - 2.3. JIW - struktura
 - 2.4. JIW - kodowanie informacji
 - 2.5. JIW - budowa poszczególnych rodzajów JIW
3. Opracowanie dokumentów i informacji
 - 3.1. Opracowanie formalne i języki opracowania formalnego
 - 3.2. Opracowanie rzeczowe i analiza informacyjno-logiczna
4. Zbiory informacji
 - 4.1. Rodzaje zbiorów informacji
 - 4.2. Organizacja zbiorów informacji
5. Źródła informacji
 - 5.1. Dokumenty
 - 5.2. Źródła instytucjonalne
6. System informacyjno-wyszukiwawczy
 - 6.1. Rodzaje SIW
 - 6.2. Struktura SIW

6.3. Własności SIW

7. System informatyczny

7.1. Struktura systemu informatycznego

7.2. Własności systemu informatycznego

7.3. Dane i ich organizacja

7.4. Język opisu danych

7.5. Język manipulacji danymi

7.6. Oprogramowanie

7.7. Języki programowania

7.8. Sieci przekazu informacji

8. Logika

8.1. Klasyfikacja

8.2. Algebra Boole'a

8.3. Rachunek zdań i rachunek predykatów

8.4. Logiczna kategoryzacja wyrażeń

8.5. Definicje

9. Językoznawstwo

9.1. Typologia systemów językowych

9.2. System języka

9.3. Funkcje językowe

9.4. Ogólna struktura języka

9.5. Plany językowe

9.6. Kategorie językowe

9.7. Słownictwo

9.8. Gramatyka

9.9. Semantyka

9.10. Pragmatyka

Realizując indukcyjną część metody gromadzenia zbioru leksykalnego wykorzystano technikę wyłaniania słów kluczowych z tekstów spe-

cyjalistycznych poświęconych tematyce JIW. Poddano analizie artykuły zawarte w 10 ostatnich rocznikach dwóch polskich czasopism specjalistycznych: "Zagadnień Informacji Naukowej" i "Aktualnych Problemów Informacji Naukowej".

W trakcie gromadzenia słownictwa wstępnie ustalono związki synonimiczne między wyrażeniami wprowadzonymi do bazy roboczej "KARTOTEKA". Za wyrażenia synonimiczne uznano:

1. terminy specjalistyczne i ich odpowiedniki w języku ogólnym i lub innym systemie terminologicznym (np. fonem - głoska, sylaba - zgłoska, układ - porządek pionowy),
2. wyrażenia równoznaczne o różnej pisowni lub formie gramatycznej (np. notacja mnemoniczna - notacja mnemotechniczna, katalog online - katalog on-line, tezaurofaseta - tezaurus fasetowy),
3. wyrażenia równoznaczne o różnej etymologii (np. notacja liniowa - notacja linearna, notacja cyfrowa - notacja numeryczna),
4. synonimy chronologiczne (np. indeks - skorowidz),
5. wyrażenia bliskoznaczne funkcjonujące jako synonimy kontekstowe (np. klasyfikacja fasetowa - klasyfikacja polichierarchiczna, biblioteka specjalna - biblioteka fachowa),
6. pełne formy wyrazów i ich formy skrócone (najczęściej dotyczy to nazw własnych, np. Uniwersalna Klasyfikacja Dziesiętna - UKD),
7. przyjęte w polskim piśmiennictwie wyrażenia obce i ich odpowiedniki polskie (np. UBC /Universal Bibliographic Control/ - Uniwersalna Rejestracja Bibliograficzna, system online - system dialogowy).

Odpowiedniki obcojęzyczne

Jednym z istotnych elementów charakterystyki terminów zawartych w bazie "BIS" są ich odpowiedniki w czterech językach dominujących w piśmiennictwie naukowym: angielskim, francuskim, niemieckim i rosyjskim. Poza przekonaniem, iż każdy dobrze opracowany specjalistyczny serwis informacji terminologicznej powinien spełniać poza funkcją wyjaśniającą, także funkcję specjalistycznego słownika przekładowego, decyzja wprowadzenia do bazy "BIS" odpowiedników obcojęzycznych polskich terminów motywowana była również niezadowalającym stanem opracowania terminologii obcojęzycznej w dziedzinie teorii JIW. Problematyka JIW, podobnie jak w piśmiennictwie polskim, jest szeroko reprezentowana także w literaturze zagranicznej, przy czym - również podobnie jak terminologia polska obcojęzyczna terminologia tej dziedziny jest nieuorządkowana i bardzo zróżnicowana. Poza wydanym przed 11 laty, znacznie zaktualizowanym już "Słownikiem terminologicznym informacji naukowej" (Wrocław, Ossolineum 1979), uwzględniającym tematykę JIW w sposób bardzo skromny, brak jakichkolwiek względnie nowych i kompletnych źródeł zapewniających poprawny przekład specjalistycznych tekstów w tej dziedzinie.

Opracowanie odpowiedników obcojęzycznych stanowi odrębne zadanie w ramach realizowanego projektu badawczego, które wymagać będzie uzupełnienia dwuosobowego obecnie zespołu opracowującego bazę "BIS" o specjalistów zajmujących się terminologią informacji naukowej w językach obcych. W obecnej wersji bazy "BIS-1" zapewniono możliwość wprowadzenia odpowiedników obcojęzycznych w zaprojektowanym formacie danych oraz możliwość wyszukiwania za ich pomocą, przeglądania wyrazów terminologii obcojęzycznej i wyświetlania danych według formatów przewidzianych dla identyfikacji

odpowiedników w poszczególnych językach oraz formatów przewidzianych dla identyfikacji znaczenia terminów obcojęzycznych. Prace nad ustaleniem odpowiedników obcojęzycznych przewidziano natomiast w drugim i trzecim etapie realizacji projektu badawczego.

Definicje terminologiczne

Wyjaśniająca funkcja terminologicznej bazy danych "BIS" realizowana jest przez umieszczenie w rekordzie każdego terminu dwóch pól powtarzalnych (o praktycznie nieograniczonej długości), w których zawarte są teksty definicji terminologicznych i uwag o charakterze komentarza do definicji i/lub artykułu encyklopedycznego.

Większość definicji ma charakter definicji sprawozdawczych, a więc zadaniem ich jest przedstawienie znaczenia w jakim dany termin funkcjonuje w piśmiennictwie. Ze względu na to, że pewna część terminologii teorii JIW derywowana jest z systemów terminologicznych innych dyscyplin naukowych, a przy tym w teorii JIW używana jest w specyficznym znaczeniu, w wielu przypadkach terminy wyposażone są w kilka definicji przedstawiających różne interpretacje semantyczne tego samego wyrażenia. Znaczenie pochodzące z dziedzin macierzystych terminów wyrażone jest definicją umieszczoną jako pierwsza, zmodyfikowane interpretacje przyjęte w teorii JIW referują definicje umieszczone na końcu bloku definicyjnego, np.:

KLASYFIKACJA 1. dowolny PODZIAŁ LOGICZNY pewnego zbioru; 2 PODZIAŁ LOGICZNY rozgałęziony, czyli taki, w którym KLASY uzyskane z podziału ulegają dalszemu dzieleniu; klasyfikacja w przeciwieństwie do TYPOLOGII oparta jest na zasadzie OPOZYCJI PRYWATYWNEJ; 3. JĘZYK INFORMACYJNO-WYSZUKIWAWCZY grupujący ELEMENTARNE JEDNOSTKI LEKSYKALNE

w system KLAS, uporządkowanych zwykle hierarchicznie i znakowanych za pomocą pewnej NOTACJI SZTUCZNEJ.

Definicje terminów przyjęte w bazie "BIS-1" przedstawiają ich znaczenie konotacyjne, czyli wyliczają wszystkie istotne cechy desygnatów terminów, pozwalające w sposób jednoznaczny wskazać takie elementy rzeczywistości, które zgodnie z regułami języka można nazwać za pomocą definiowanego terminu. Określenie znaczenia konotacyjnego wyrażenia językowego w logice i językoznawstwie bywa różnie interpretowane. Zespoli cech desygnatów składających się na konotację ustala się albo jako tzw. treść konstytutywną, czyli zbiór konieczny i wystarczający dla jednoznacznego zidentyfikowania desygnatu, albo jako treść charakterystyczną (zwaną też treścią językową) o charakterze konsekwentnym, czyli jako zbiór zawierający poza cechami koniecznymi i wystarczającymi dla zidentyfikowania desygnatów, również inne, redundantne cechy charakterystyczne desygnatów, które na mocy reguł semantycznych języka uważane są przez użytkowników tego języka za elementy znaczenia określonych wyrażań.

W terminologicznej bazie danych, która ma spełniać rolę swoistego systemu informacji faktograficznej, w której możliwe będzie wyszukiwanie pełnotekstowe, jedynym konstruktywnym rozwiązaniem jest przyjęcie drugiej metody określania znaczenia konotacyjnego. Podejście takie zapewnia, że teksty definicji zawierają znaczną liczbę innych - z reguły semantycznie prostszych - terminów, pokrewnych znaczeniowo terminowi definiowanemu. Terminy te należą więc do tego samego pola semantycznego, co z reguły oznacza, iż związane są z pewnym szczegółowym tematem badawczym. Weryfikacja powyższej tezy na przykładzie wyszukiwania w bazie terminologicznej pozwoliłaby stwierdzić, iż technika wyszukiwania asocjacyjnego

(za pomocą ciągu terminów pozostających wobec siebie w pewnej relacji kojarzeniowej) prowadzi do podwyższenia wskaźników efektywności wyszukiwania informacji według kryteriów rzeczowych.

W projektowanej bazie, poza wyrażeniami będącymi terminami, umieszczono także pewną liczbę nazw własnych, nazw języków i systemów informacyjno-wyszukiwawczych, nazw systemów notacyjnych, nazw organizacji związanych z badaniami w zakresie teorii i praktyki JIW. Wyrażenia te nie mogą być oczywiście definiowane konotacyjnie, toteż w ich przypadku zamiast definicji sprawozdawczej wprowadzono syntetyczne artykuły encyklopedyczne, których zadaniem jest przedstawienie podstawowych informacji o ich desygnatach.

Uwagi

Pole uwag, podobnie jak pole definicji, jest polem powtarzalnym, w którym umieszczony jest tekst komentarza uzupełniającego definicję terminologiczną oraz tekst skondensowanego artykułu encyklopedycznego, w którym przedstawione są podstawowe informacje na temat obiektów czy zagadnień reprezentowanych przez termin. Dane zawarte w tym polu mają za zadanie spełniać rolę informacji faktograficznej charakterystycznej dla specjalistycznego podręcznika i adresowane są przede wszystkim do studentów poznających problematykę teorii JIW. Dlatego też treść uwag przede wszystkim skoncentrowana jest na problematyce reprezentowanej przez zdefiniowany termin w teorii JIW, np.:

GRAMATYKA POZYCYJNA - GRAMATYKA, w której rolę wykładnika RELACJI SYNTAGMATYCZNEJ pełni szereg WYRAZEŃ PROSTYCH w strukturze WYRAŻENIA ZŁOŻONEGO; oznacza to, iż zmiana miejsca elementów pewnego wyrażenia powoduje zmianę jego ZNACZENIA lub prowadzi do utworzenia wyrażenia nie należącego do danego języka.

(Uwagi:) W teorii JĘZYKÓW INFORMACYJNO-WYSZUKIWAWCZYCH termi-
n gramatyka pozycyjna odnosi się przede wszystkim do takich języków,
jak JĘZYK HASEŁ PRZEDMIOTOWYCH- KLASYFIKACJE FASETOWE. Wyrażenie
złożone w obu tych językach ma ściśle określoną strukturę, w
której kolejne miejsca są zarezerwowane dla wyrazów prostych,
reprezentujących określone treści dokumentu, np. HASŁO PRZEDMIOTOWE
składa się z ciągu ELEMENTARNYCH JEDNOSTEK LEKSYKALNYCH JĘZYKA
HASEŁ PRZEDMIOTOWYCH uszeregowanego tak, iż jednostka umieszczona
na pierwszym miejscu w hasle jest TEMATEM reprezentującym PRZEDMIOT
lub formę DOKUMENTU, a jednostki umieszczone na następnych miejs-
cach - różne kategorie OKREŚLNIKÓW, których następstwo jest także
ściśle określone. W klasyfikacjach fasetowych gramatyka pozycyjna
przyjmuje postać FORMUŁY FASETOWEJ.

Z drugiej strony specyficzna budowa tekstu uwag, wypełnionego
licznymi terminami o znaczeniu pokrewnym, zwykle oaziej szczegóło-
wymi aniżeli termin hasłowy, pozwala wzbogacić możliwości asocjacyj-
nego wyszukiwania pełnotekstowego.

Zarówno w tekście definicji, jak i w tekście uwag wszystkie
terminy zdefiniowane jako odrębne hasła zostały wyróżnione graficz-
nie poprzez zapis wersalikami. W zbiorze wyszukiwawczym występują
one w zróżnicowanych wariantach gramatycznych, zgodnie z formą
 użytą w poszczególnych tekstach, toteż posługiwanie się nimi
w celu wyszukania wszystkich informacji związanych z reprezentowa-
nym przez te terminy zagadnieniem, wymaga maskowania sufiksów
gramatycznych.

Materiały ilustracyjne

Pole materiałów ilustracyjnych jest swoistym przedłużeniem
pola uwag, w którym umieszcza się proste schematy graficzne
(np. drzewa klasyfikacyjne, graficzne przedstawienie operacji

rachunku mnogości imitujące diagramy Venna) oraz wykazy tabelaryczne (np. tablicową prezentację układu klasyfikacyjnego). Konieczność wprowadzenia takiego pola stwierdzono w wyniku testowania efektów wyświetlania według różnych formatów tych rekordów zbioru głównego, w których w polu uwag umieszczone były materiały o ściśle określonym układzie graficznym.

Wprowadzenie pola materiałów ilustracyjnych przewidziane jest w drugim etapie prac, w którym wszystkie przewidziane formaty wyświetlania danych poddane zostaną weryfikacji na próbie zwiększonej do kilkuset rekordów zbioru głównego.

Kod logiczny terminu

Kod logiczny terminu jest podstawowym narzędziem, które zapewnić ma symulację wyszukiwania logicznego w bazie terminologicznej oraz konieczności wcześniejszego ustalenia formy terminów logicznie (i semantycznie) pokrewnych wobec terminu wyjściowego. Kod ten ma więc w założeniu odwzorowywać zależności zachodzące między znaczeniem poszczególnych terminów poprzez wyrażanie ich wspólnych składników semantycznych tymi samymi znakami w strukturze kodu. Oznacza to, iż planuje się opracowanie dla całego systemu terminologicznego uwzględnionego w bazie tzw. notacji komponentalnej. Prace nad tą notacją można będzie rozpocząć po opracowaniu definicji dla podstawowego zbioru leksykalnego bazy, stąd przewidziane są one w drugim etapie realizacji projektu.

Kod autora i data wprowadzenia terminu

Poza polami przewidzianymi dla informacji merytorycznych w rekordzie każdego terminu umieszczono dwa pola informacji formalnych - dotyczących autora charakterystyki terminu i daty wprowadzenia.

BAZA TERMINOLOGICZNA JEZYKOW INFORMACYJNO - WYSLIWIWACZYCH

Wydru: w formacie kontrolnym

Rekord 000026

NOTACJA HIERARCHICZNA

Synonimy: znakowanie hierarchiczne:

NOTACJA STRUKTURALNA generująca WYRAZENIA o strukturze odwzorowującej RELACJE HIERARCHICZNE między ich ZAKRESAMI ZNACZENIOWYMI.

Uwagi: Najpopularniejszym rodzajem notacji hierarchicznej jest NOTACJA DZIESIĘTNA, należy jednak pamiętać, że istnieje wiele innych wariantów notacji hierarchicznej będących zarówno NOTACJAMI NUMERYCZNYMI (np. NOTACJA SETNA), jak i NOTACJAMI ALFABETYCZNYMI lub NOTACJAMI ALFANUMERYCZNYMI; ze względu na prostotę budowy, zdolność wskazywania związków znaczeniowych między wyrażeniami, a przy tym zapewnienie wysokiej PRODUKTYWNOŚCI JIW notacja hierarchiczna jest najczęściej zalecanym dla JIW systemem znakowania.

[BS] 28/11/90.

Rekord 000030

NOTACJA MIESZANA

Synonimy: notacja niejednorodna; znakowanie mieszane; znakowanie niejednorodne; notacja niejednolita; znakowanie niejednolite; notacja wielorodna; znakowanie wielorodne;

NOTACJA, której BAZE NOTACYJNE stanowią ZNAKI różnego typu, np. NOTACJA ALFANUMERYCZNA jest notacją mieszaną, której baza notacyjna zawiera literowy pewnego ALFABETU JEZYKA NATURALNEGO oraz cyfry arabskie; notacja alfanumeryczna jest najczęściej stosowanym rodzajem notacji mieszanej.

Uwagi: Określenie NOTACJA MIESZANA stosowane jest wówczas, gdy ELEMENTARNE ZNAKI GRAFICZNE, należące do różnego typu znaków, wykorzystywane są do ZNAKOWANIA PODSTAWOWYCH JEDNOSTEK LEKSYKALNYCH danego JIW, np. SYMBOLI GŁÓWNYCH pewnej KLASYFIKACJI; jeśli ZNAKI SPECJALNE, np. znaki matematyczne czy interpunkcyjne, wykorzystywane są jako różnego rodzaju WSIĄZNIKI, a symbole stanowiące podstawowe słownictwo JIW oznakowane są jednorodnymi ciągami alfabetycznymi czy jednorodnymi ciągami numerycznymi - tego typu notacji nie traktuje się jako notacji mieszanej (por. UNIWERSALNA KLASYFIKACJA DZIESIĘTNA).

[BS] 28/11/90.

BAZA TERMINOLOGICZNA JEZYKÓW INFORMACYJNYCH - WYSZUKIWAWCZYCH

Wyodrębni w formacie wyrocznym

NOTACJA HIERARCHICZNA [znakowanie hierarchiczne] - **NOTACJA STRUKTURALNA**
generująca WYRAŻENIA o strukturze odwzorowujące RELACJE HIERARCHICZNE
między ich ZAKRESAMI ZNACZENIOWYMI.

UWAGI: Najpopularniejszym rodzajem notacji hierarchicznej jest **NOTACJA
DZIESIĘTNA**, należy jednak pamiętać, że istnieje wiele innych wariantów
notacji hierarchicznych będących zarówno **NOTACJAMI NUMERYCZNYMI** (np. **NOTACJA
SETNA**), jak i **NOTACJAMI ALFABETYCZNYMI** lub **NOTACJAMI ALFANUMERYCZNYMI**; ze
względu na prostotę budowy, zdolność wskazywania związków znaczeniowych
między wyrażeniami, a przy tym zapewnienie wysokiej **PRODUKTYWNOŚCI** JIW
notacja hierarchiczna jest najczęściej zalecanym dla JIW systemem
znakowania.

NOTACJA MIESZANA [notacja niejednorodna; znakowanie mieszane; znakowanie
niejednorodne; notacja niejednorodna; znakowanie niejednorodne; notacja
wielorodna; znakowanie wielorodne] - **NOTACJA**, której **BAZE NOTACYJNE**
stanowią **ZNAKI** różnego typu, np. **NOTACJA ALFANUMERYCZNA** jest notacją
mieszaną, której baza notacyjna zawiera litery pewnego **ALFABETU JEZYKA
NATURALNEGO** oraz cyfry arabskie; notacja alfanumeryczna jest najczęściej
stosowanym rodzajem notacji mieszanej.
UWAGI: Określenie **NOTACJA MIESZANA** stosowane jest wówczas, gdy **ELEMENTARNE
ZNAKI GRAFICZNE**, należące do różnego typu znaków, wykorzystywane są do
ZNAKOWANIA PODSTAWOWYCH JEDNOSTEK LEKSYKALNYCH danego JIW, np. **SYMBOL:
GŁÓWNYCH** pewnej **KLASYFIKACJI**; jeśli **ZNAKI SPECJALNE**, np. znaki matematyczne
czy interpunkcyjne, wykorzystywane są jako różnego rodzaju **WSKAZNIKI**, a
symbole stanowiące podstawowe słownictwo JIW oznakowane są jednorodnymi;
ciągami alfabetycznymi czy jednorodnymi ciągami numerycznymi - tego typu
notacji nie traktuje się jako notacji mieszanej (por. **UNIWERSALNA
KLASYFIKACJA DZIESIĘTNA**).

NOTACJA RETROAKTYWNA [znakowanie retroaktywne] - **NOTACJA** z wyrocznionymi w
BAZIE NOTACYJNEJ dla każdej **GALEJI KLSYFIKACYJNEJ** lub **FASETY** podziorami
znaków o wartości porządkowej wyższej od wartości porządkowej **WYRAŻENIA**
reprezentującego **KLASE WELDWA** wobec rozpoczynającą daną galeję lub fasetę
oraz z regułą inwersji **PORZĄDKU POZIOMEGO** i wobec **PORZĄDKU PIONOWEGO**.

UWAGI: Np. w pewnej **KLASYFIKACJI fasetowej**, której **SŁOWNICTWO** zostało
pogrupowane w trzech fasetach "Biblioteki", "Materiały biblioteczne" i
"Operacje biblioteczne" zgodnie z zasadą notacji retroaktywnej poszczególne
KLASY otrzymany następujące symbole:

D	Operacje biblioteczne
DE	Gromadzenie
DF	Opracowanie
DFE	Opracowanie rzeczowe
DFF	Opracowanie formalne
DG	Udostępnianie
DH	Przechowywanie
E	Materiały biblioteczne
EF	Dokumenty piśmiennicze
EFE	Rękopisy
EFF	Dokumenty drukowane
EG	Dokumenty niepiśmiennicze
EGE	Materiały audialne
EGF	Materiały wizualne
F	Biblioteka

dzenia tej charakterystyki do bazy "BIS". Dane te przeznaczone są dla kontroli administracyjnej (np. korekta danych, planowanie cykli aktualizacji), stąd umieszczono je wyłącznie w tzw. formacie kontrolnym, dostępnym jedynie dla administratora bazy.

Informacje formalne w roboczej wersji bazy danych "BIS" wprowadzone są w postaci zakodowanej. Dla danych autorskich obecnie przyjęto tradycyjną w wydawnictwach encyklopedycznych i leksyko-graficznych postać inicjałów pierwszego imienia i nazwiska autora opracowania. Ponieważ w tej wersji bazy nie wprowadzono odpowiedników obcojęzycznych (przewidzianych w wersji drugiej), nie zachodzi potrzeba różnicowania kodów autora opracowania artykułu definicyjnego i encyklopedycznego oraz autora lub autorów opracowania części przekładowej charakterystyki terminów. W wersji drugiej - w razie konieczności - kody te zostaną zmodyfikowane.

Informacje o dacie wprowadzenia danych mają tradycyjną postać kodu sześciocyfrowego tworzonego według formuły: dd/mm/rr (dzień/miesiąc/rok). Kody te mają wskazywać datę ostatniej modyfikacji zawartości rekordu.

Stan prac po pierwszym etapie realizacji bazy "BIS"

Zamierzonym celem realizacji zadań składających się na pierwszy etap opracowania terminologicznej bazy danych z zakresu teorii JIW było przygotowanie tzw. roboczej wersji tej bazy oraz technicznych warunków jej realizacji. Przez wersję roboczą rozumie się opracowanie wyjściowego projektu ideowego bazy i jej podstawowego oprogramowania, zapewniającego techniczną realizację zamierzonych funkcji. W projekcie ideowym, którego charakterystykę przedstawiono powyżej, określono rodzaj i funkcje poszczególnych elementów charakterystyki terminu (a więc generalnie - schemat logicz-

ny rekordu terminu) oraz ustalono metody opracowania danych. W projekcie technicznym ustalono rodzaj optymalnego dla bazy oprogramowania oraz opracowano wyjściowy wariant tzw. systemu STJIW, stanowiącego modyfikację uniwersalnego systemu gromadzenia i wyszukiwania informacji CDS/ISIS w wersji na minikomputer typu IBM PC.

W roboczej wersji bazy "BIS-1" utworzono pilotażowy zbiór główny zawierający 50 rekordów i wygenerowano na jego podstawie zbiór wyszukiwawczy zawierający ok. 380 terminów wyszukiwawczych, co pozwoliło przeprowadzić testy sprawdzające użyteczność przyjętych rozwiązań zarówno logicznych, jak i technicznych. W tej wersji bazy zastosowano docelowy formularz wejściowy, jednakże ze względu na zbyt wysokie - wobec środków, którymi dysponował wykonawca - koszty opracowania pełnych charakterystyk terminów odzwierciedlonych w tym formularzu, zbiór pilotażowy opracowany został w zakresie zawężonym, nie obejmującym kodów logicznych (opracowanie ich przewidziane jest dla całego zbioru leksykalnego w drugim etapie prac) ani odpowiedników odczołowych. W wersji roboczej funkcjonuje więc jedynie część wyjaśniająca bazy (nie funkcjonuje jeszcze przewidziana w następnych etapach realizacji projektu część przekładowa) oraz standardowe, wynikające z możliwości systemu CDS/ISIS, techniki wyszukiwania (nie funkcjonuje, również przewidziane w wersji eksploatacyjnej, wyszukiwanie asocjacyjne i logiczne).

Poza celem zasadniczym, jakim jest przygotowanie wersji roboczej bazy, w pierwszym etapie prac zaplanowano i zrealizowano także prace wstępne związane z przygotowaniem podstawowego zbioru leksykalnego - ustalenie zakresu tematycznego, wytypowanie i zgromadzenie materiałów źródłowych oraz zgromadzenie i wstępne opracowanie podstawowego zbioru leksykalnego (baza "KARTOTEKA").

Tak więc wszystkie zadania pierwszego etapu realizacji bazy "BIS" zostały wykonane, przygotowano także materiały dla realizacji etapu drugiego, którego ukończenie planowane jest w roku 1992.

Literatura wykorzystana przy gromadzeniu zasobu leksykalnego bazy "BIS-1"

1. BIELICKA L.A. Języki informacyjne - przegląd analityczny badań efektywności. Warszawa: IINTE 1977
2. BIELICKA L.A. Konwencjonalne metody indeksowania współrzędnego. Warszawa: IINTE 1979
3. BIELICKA L.A. Metodyka i organizacja opracowania tezaury. Warszawa: CİNTE 1988
4. BIELICKA L.A. Zasady gramatyczne w językach deskryptorowych. Warszawa: IINTE 1985
5. BIELICKA L.A., PACIEJEWSKI J., ŚCIBOR E. Pojęcie spójności i metody jej osiągania w językach informacyjnych. Warszawa: IINTE 1984
6. BIELICKA L.A., ŚCIBOR E. Języki informacyjne. Rodzaje i zastosowanie w działalności informacyjnej. Warszawa: CİNTE 1982
7. BIELICKA L.A., ŚCIBOR E. Wprowadzenie do teorii języków informacyjnych. Warszawa: CİNTE 1981
8. BOBROWSKI J. Wybrane zagadnienia efektywności zautomatyzowanych systemów wyszukiwania dokumentów. Warszawa: CİNTE 1981
9. BOJAR B. Zarys językoznawstwa dla informatyków. Warszawa: Wydaw. UW 1976
10. BUCHANAN B. Theory of library classification. London: Clive Bingley 1979
11. CHMIELEWSKA-GORCZYCA E. Indeksy permutacyjne. Warszawa: CİNTE 1977
12. CHMIELEWSKA-GORCZYCA E. Relacje syntagmatyczne w językach informacyjno-wyszukiwawczych (rozprawa doktorska) Warszawa: Uniwersytet Warszawski 1981 (maszynopis)

13. CHMIELEWSKA-GORCZYCA E., SOSIŃSKA-KALATA B. Informacja naukowa z elementami naukoznawstwa. Warszawa: Wydaw. Szkolne i Pedagogiczne (w druku)
14. CLASON W.E. Elsevier's dictionary of library science, information and documentation. Amsterdam: Elsevier Scientific Publ. Company 1973
15. CZERNY A.I. Syntagmatyczne stosunki między deskryptorami. Metodyka budowy tezaurusów. Warszawa: CİNTE 1969
16. CZERNY A.I. Wstęp do teorii wyszukiwania informacji. Warszawa: OIN PAN 1978
17. ĆWIEKOWA J. Opracowanie tematyczne piśmiennictwa. Warszawa: PWN 1988
18. DOBROWOLSKI Z. Budowa klasyfikacji. Warszawa: PWT 1956
19. ENCYKLOPAEDIA of library and information science. Ed. A. Kent, H. Lancour. New York: M. Dekker 1968-1975
20. ENCYKLOPEDIA wiedzy o książce. Red. A. Birkenmajer, B. Kocowski, J. Trzynadłowski. Wrocław: Zakł. Narod. im. Ossolińskich 1971
21. ENCYKLOPEDIA wiedzy o prasie. Red. J. Maślanko. Wrocław: Zakł. Narod. im. Ossolińskich 1976
22. ENCYKLOPEDIA współczesnego bibliotekarstwa polskiego. Red. K. Głombowski, B. Świderski, H. Więckowska. Wrocław: Zakł. Narod. im. Ossolińskich 1976
23. FOSKETT A.C. The subject approach to information. 4th ed. London: Clive Bingley - Linnet Books 1981
24. GŁOWACKA T. Katalogi rzeczowe. Warszawa: PWN 1975
25. GRABOWSKA M. Streszczenia dokumentacyjne (Wybrane problemy). Warszawa: CİNTE 1979
26. GRAMATYKA języka deskryptorowego na tle gramatyk innych języków informacyjnych. Red. M. Poletyło. Warszawa: CİNTE 1975
27. HUTCHINS W.J. Languages of indexing and classification. London: Peter Peregrinus 1979
28. INFORMACIONNO-poiskovyj tezaurus po informatykie. Red. A.I. Čer-nyj. Moskva: VINITI 1974

29. JABRZEMSKA E. ŚCIBOR E. Zasady opracowania i stosowania dziedziczo-gałęziowych klasyfikacji tematycznych. Warszawa: CİNTE 1986
30. LANCASTER F.W. Vocabulary control for information retrieval. Washington: Information Resources Press 1972
31. LANGRIDGE D. Approach to classification for students of librarianship. London: Clive Bingley 1973
32. LESKA M., LESKI K. Tezaurus informacji naukowej. Warszawa: IINTE 1972
33. LESKI K. Zasady budowy tezaurusów. Warszawa: OIN PAN 1978
34. ŁYSAKOWSKI A. Katalog przedmiotowy. Cz.1 Teoria. Włino 1920
35. ŁYSAKOWSKI A. Katalog przedmiotowy. Cz.2 Podręcznik. Warszawa: Biblioteka Narodowa 1946
36. MALTBY A. Sayers' manual of classification for librarians. 5th ed. London: Andre Deutsch - A Grafton Book 1975
37. MAŁA encyklopedia logiki. Red. W. Marciszewski. Wrocław: Zakł. Narod. im. Ossolińskich 1970
38. MARCISZEWSKI W. Metody analizy tekstu naukowego. Warszawa: PWN 1981
39. MORYK A. Klasyfikacja patentowa jako język wyszukiwawczy w informacji technicznej. Warszawa: CİNTE 1988
40. MOSKOVIC V.A. Informacionnye jazyki. Moskva: Nauki 1971
41. RANGANATHAN S.R. Prolegomena to library classification. 3rd ed. London: Asia Publ. House 1967
42. ROBOWSKI J. Języki deskryptorowe. Analiza porównawcza języków deskryptorowych i innych języków informacyjnych. Warszawa: IINTE 1974
43. RUONICKI P. Języki informacyjno-logiczne w informacji naukowej. Warszawa: IINTE 1980
44. SAOOWSKA J., TUROWSKA T. Języki informacyjno-wyszukiwawcze. Katalogi rzeczowe. Warszawa: CUKB 1991
45. SLOVAR terminov po informatike na russskom i anglijskom jazykam. Moskva: Nauka 1971

46. SŁOWNIK encyklopedyczny terminologii języków i systemów informacyjno-wyszukiwawczych. Red. B. Bojar. Oprac. E. Artowicz, B. Bojar, E. Chmielewska-Gorczyca, B. Sosińska, J. Woźniak. Warszawa 1987 (maszynopis)
47. SŁOWNIK terminologiczny informacji naukowej. Red. M. Dembowska. Wrocław: Zakł. Narod. im. Ossolińskich 1979
48. SŁOWNIK terminologii językoznawczej. Red. Z. Gołąb, A. Heinz, K. Polański. Warszawa: PWN 1968
49. SOSIŃSKA-KALATA B. Relacje między planem treści i planem wyrażenia w językach informacyjno-wyszukiwawczych. Warszawa: IINTE 1989
50. SZULC A. Podręczny słownik językoznawstwa stosowanego. Warszawa: PWN 1984
51. ŚCIBOR E. Biblioteczno-bibliograficzne systemy porządkowania dokumentów. Warszawa: CİNTE 1976
52. ŚCIBOR E. Typologia strukturalna języków informacyjnych. Warszawa: IINTE 1982
53. ŚCIBOR E. Zasady stosowania Polskiej Klasyfikacji Tematycznej. Warszawa: CİNTE 1982
54. ŚCIBOR E., TOMASIK-BECK J. Analiza porównawcza języków informacyjnych. (Wybrane zagadnienia). Warszawa: CİNTE 1970
55. TERMINOLOGİČESKIJ slovar po informatike. Moskva: MCNTI 1975
56. TITTENBRUN K. Polskie piśmiennictwo terminologiczne z zakresu informacji naukowej i dziedzin pokrewnych. Bibliografia adnotowana 1945-1978. Warszawa: IINTE 1981
57. THOMPSON A. Vocabularium bibliothecarii. 2nd ed. Bruges: UNESCO 1962
58. TOMASIK-BECK J. Zasady budowy tezaursów wielojęzycznych. Warszawa: IINTE 1977
59. TOPULOS A. Problemy terminologii naukowej i technicznej. Warszawa: CİNTE 1979
60. UNGURIAN O. Elementy teorii języków informacyjnych. Warszawa: DIN PAN 1976
61. UNGURIAN O. Teoria i praktyka klasyfikacji fasetowej S.R. Ranganathana. Warszawa: CİNTE 1975

62. UNGURIAN O. Wprowadzenie do Uniwersalnej Klasyfikacji Dziesiątnej. Wyd.3 Warszawa: Wydaw. łącz. i Komunikacji 1978
63. WERSIG G., Noveling U. Terminology of documentation. Paris: The UNESCO Press 1976
64. WILLIAMS H.L. Computerised systems in library and information services. Lon don: Aslib 1982
65. WÓJCIK T. Zarys teorii klasyfikacji. Zagadnienia formalne. Warszawa: PWN 1965.

THE TERMINOLOGICAL DATA BASE ON THE THEORY OF
INFORMATION RETRIEVAL LANGUAGES

Summary

There is presented in the article the characteristics of the structure of the BIS data which has ben worked out at the Institute of Librarianship and Information Science and which gives the access to the information on the meanings of terms in the field of theory information retrieval languages (IRL) and on their equivalentents in the four foreign languages, mostly used in literature. The organization of works on establishing the data base as well as the general technological structure of this base which is determined by the use of the standard software CDS/ISIS in the version adopted for IBM PC are discussed. There are characterized furthermore: the logical structure of records in the master file and in the index file, rules of defining the thematic scope of the base and methods of acquiring the lexical resources, rule of establishing the synonymy relationships between words input into the files, the adopted methods of working out of the data for the respective fields of the record (foreign language

equivalents fields, terminological definition fields, the encyclopaedic remarks fields, the illustrative materials' fields, fields for the logical code of the term, author and the input data fields). The state of the art after finishing the first stage of the data base implementation is presented. The list of publications used within the works on the acquiring the lexical sources is enclosed to the article (65 items).

Терминологическая база данных в области
теории информационно-поисковых языков

Р е з ю м е

В статье представлена характеристика баз данных БИБ разрабатываемой в Институте библиотековедения и научной информации Варшавского университета обеспечивающей доступ к информации о значении терминов в области теории информационно-поисковых языков, а также об их эквивалентах на четырех иностранных языках наиболее популярных в научной литературе. Представлена организация разработок по созданию базы, ее общая техническая структура, предопределенная использованием стандартного программного обеспечения CDB/ISIS во версии на языке эвм типа IBM 30, логическая структура рекода в основном файле и в файле поисковых терминов, принципы определения тематического охвата базы, методы накопления лексических ресурсов, принципы определения синонимических связей между выражениями вводимыми в базу, принятые методы обработки данных для отдельных полей рекода (поля эквивалентов на иностранных языках, поля терминологической дефиниции, поля энциклопедических примечаний, поля иллюстрирующих материалов, поля логического кода термина, а также поля кодов автора и даты ввода). Обсуждено состояние работ после окончания первого этапа реализации базы. Приложен список публикаций использованных в работах по накоплению лексических ресурсов базы (65 позиций).

PROJEKT TECHNICZNY
TERMINOLOGICZNEJ BAZY DANYCH Z ZAKRESU TEORII JĘZYKÓW
INFORMACYJNO-WYSZUKIWAWCZYCH

Założenia merytoryczne. Struktura logiczna i fizyczna bazy. Budowa rekordów zbioru głównego. Formularz wprowadzania danych. Formaty wyświetlania. Słownik terminów wyszukiwawczych. Produkowanie wyciągów z bazy. Modyfikacje systemu ISIS ukierunkowane na użytkownika nieprofesjonalnego. Perspektyw rozwoju.

Terminologiczna baza danych BIS^y z zakresu teorii języków informacyjno-wyszukiwawczych, pełniąca rolę swoistego słownika terminologicznego na nośniku komputerowym, została zrealizowana w Zakładzie Systemów Informacyjnych Instytutu Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu Warszawskiego.

Po rozpoznaniu potrzeb użytkownika zdecydowano, że do realizacji pakietu zostanie użyty uniwersalny system gromadzenia i wyszukiwania informacji CDS/ISIS w wersji na mikrokomputer typu IBM PC. Za takim rozwiązaniem przemawiały następujące fakty:

1. Mikrokomputery zgodne ze standardem IBM PC należą do najczęściej używanych obecnie w kraju;

^y Zob omówienie Bazy: B.Sosińska-Kalata: Terminologiczna baza danych z zakresu teorii języków informacyjno-wyszukiwawczych. ZIN 1991 nr 1(58) s.41-71.

2. System CDS/ISIS jest jedynym bezpłatnym legalnie rozprowadzonym w kraju systemem gromadzenia i wyszukiwania informacji;

3. System ten jest ponadto szczególnie przydatny do tworzenia terminologicznych baz danych;

4. Dzięki odpowiednim narzędziom wchodzącym w skład pakietu ISIS (menu i formularze w języku polskim) zdefiniowane przy jego pomocy bazy danych mogą być eksploatowane przez użytkowników będących laikami w dziedzinie informatyki.

Baza danych BIS wraz z odpowiednio skonfigurowanym i zmodyfikowanym systemem mikroISIS zwane będą w dalszej części niniejszej dokumentacji systemem BIS. W kolejnych rozdziałach omówiona zostanie zarówno logiczna jak i fizyczna budowa systemu, struktura rekordów wchodzących w skład bazy danych, formularz wprowadzania danych, formaty wyświetlania, budowę słownika terminów wyszukiwawczych, możliwości i rodzaje generowania raportów z bazy, a także sposób łączenia podbaz w jedną spójną bazę danych.

Założenia merytoryczne

Przedmiotem projektu jest baza danych zawierająca elementy słownika terminologicznego. Przez elementy słownika rozumie się opisy haseł słownika. Składają się one z terminu definiowanego, jego synonimów, definicji, odpowiedników obcojęzycznych i ewentualnych dodatkowych wyjaśnień. Dziedziczne powiązania poszczególnych terminów odwzorowane zostaną przy pomocy specjalnie opracowanych kodów. Dla ułatwienia prowadzenia dokumentacji wprowadzono dodatkowe informacje takie jak kod autora definicji i datę wprowadzania.

Zadaniem systemu BIS jest przechowywanie na nośniku magnetycznym (dysku twardym mikrokomputera) oraz szybkie i wygodne udostępnianie informacji dotyczących zdefiniowanych terminów. Naj-

większy nacisk położono na stworzenie takiego środowiska programowego, aby mogły z tej bazy korzystać osoby nie będące fachowcami w dziedzinie tworzenia i eksploatacji baz danych. Będą temu sprzyjały zarówno sposób komunikacji z użytkownikiem (przez system menu w języku polskim) jak i odpowiednio przygotowane i wypełnione formularze ekranowe, zawierające parametry drukowania i umożliwiające uzyskanie potrzebnych wyciągów z bazy w sposób automatyczny.

Z uwagi na pracochłonność procesu wprowadzania danych ważną jest możliwość równoczesnego tworzenia na wielu maszynach podzbiorów docelowej bazy i łączenie ich w jedną całość przy wykorzystaniu mechanizmów importowania i eksportowania danych.

Ważnym problemem związanym z eksploatacją słownika terminologicznego w języku polskim jest obsługa polskich znaków diakrytycznych. Problem ten został rozwiązany przez użycie odpowiedniego generatora znaków (w standardzie Mazowia). Rozwiązanie to pociąga za sobą pewne ograniczenia, niestety w tym wypadku konieczny był kompromis. Naturalną konsekwencją jest możliwość przenoszenia pakietu tylko na instalacje wyposażone w generator znaków Mazowia (zarówno na karcie grafiki jak i w drukarce). Ponadto sortowanie terminów uwarunkowane kodami znaków jest niedoskonałe (kody znaków diakrytycznych są wyższe niż kody wszystkich pozostałych znaków alfanumerycznych). Wydaje się jednak, że w obecnych warunkach było to jedyne możliwe rozwiązanie.

System BIS został zaimplementowany na komputerze typu IBM PC pod systemem operacyjnym MS DOS ver. 4.0. Dla prawidłowej pracy pakietu BIS w konfiguracji systemu komputerowego niezbędne są następujące elementy:

- dysk twardy o pojemności min. 20 MB;
- minimum jedna stacja dysków miękkich;

- karta grafiki z generatorem znaków w standardzie Mazowia;
- minimum 512 K RAM;
- drukarka z zainstalowanym generatorem znaków w standardzie Mazowia.

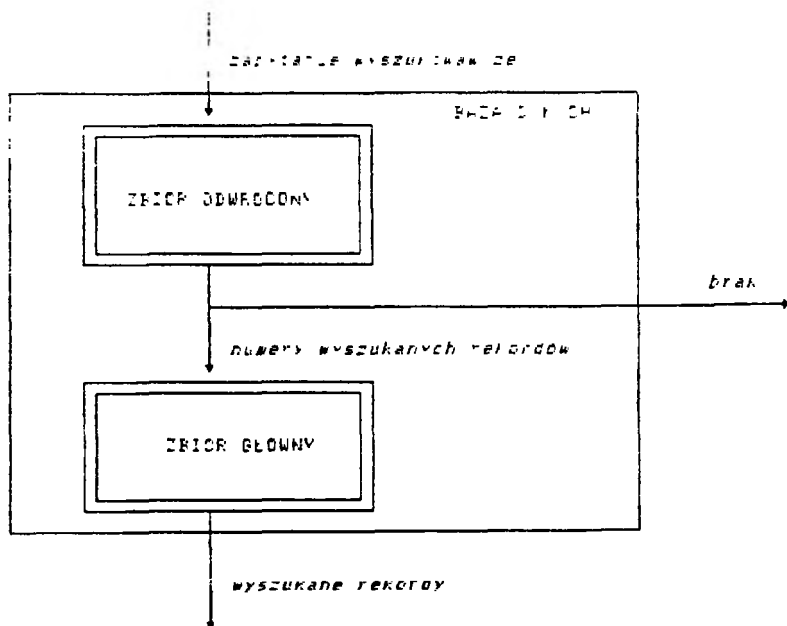
Struktura logiczna bazy BIS

Z punktu widzenia użytkownika baza danych BIS składa się ze zbioru rekordów opisujących hasła słownika terminologicznego. Każde hasło opisane jest za pomocą jednego rekordu. W rekordzie można wyodrębnić mniejsze jednostki informacji zwane polami i podpolami. Poszczególne pola zawierają odrębne części opisu terminu tj. sam termin definiowany (w pełnym brzmieniu), jego synonimy, definicje poszczególnych znaczeń, uwagi umożliwiające pełniejsze zrozumienie znaczenia terminu, odpowiedniki obcojęzyczne. Polom nadano odpowiednie mnemotechniczne nazwy oraz identyfikatory numeryczne zgodnie ze standardem systemu mikroISIS.

W systemie BIS można wyszukiwać opisy jedynie na podstawie tych elementów rekordów, które zostały wybrane do słownika terminów wyszukiwawczych (tzw. zbioru odwróconego). Zbiór odwrócony składa się z uporządkowanych alfabetycznie terminów wyszukiwawczych, a każdemu z tych terminów przyporządkowany jest wykaz numerów rekordów, w których dany termin występuje (a także inne dane umożliwiające lokalizację terminu). Dzięki takiej organizacji bazy wyszukiwanie jest natychmiastowe.

W pierwszym przybliżeniu bazę danych w systemie BIS można postrzegać jako dwa powiązane ze sobą zbiory: zbiór główny i zbiór odwrócony. Zbiór odwrócony reprezentuje jedynie podzbiór danych zbioru głównego -- tylko te dane, które zostały wybrane do słownika terminów wyszukiwawczych, ale są one zorganizowane w sposób umożli-

włażący najszvbsze wyszukiwanie. Schemat wyszukiwania przedstawia rysunek:



Dane stanowiące zawartość rekordów pliku prostego mogą zostać wyświetlone na ekranie monitora lub wydrukowane na drukarce. System BIS umożliwia wyświetlanie i drukowanie rekordów w wielu rozmaitych formatach w zależności od potrzeb użytkownika. Procesy sortowania rekordów bazy i ich drukowania są sterowane za pomocą specjalnie przygotowanych formularzy. Zawierają one parametry tych funkcji i są dedykowane przede wszystkim użytkownikom nieobeznanym z systemem CDS/ISIS.

Wprowadzanie danych do bazy i ich modyfikacja polega na wypełnianiu odpowiednich formularzy. Dzięki temu przygotowanie osób

wprowadzających dane nie wymaga specjalnego przeszkolenia. W systemie BIS zdefiniowano przejrzysty i łatwy w użyciu formularz wprowadzania danych opisany w dalszej części niniejszego artykułu.

Struktura fizyczna bazy BIS

Zgodnie z zasadą funkcjonowania systemu mikroISIS poszczególne zbiory bazy danych na poziomie logicznym składają się z wielu plików na poziomie fizycznym.

Strukturę bazy danych określają następujące pliki:

- BIS.FDT - tablica definicji pól rekordów bazy;
- BIS.FST - tablica selekcji pól do zbioru odwróconego;
- BIS.PFT - standardowy format wyświetlania rekordów;
- ABIS.FMT, BBIS.FMT i CBIS.FMT - definicje kolejnych stron formularza wejściowego.

Zbiór główny składa się z dwóch plików:

- BIS.MST - zawiera rekordy zmiennej długości;
- BIS.XRF - zawiera skorowidz zbioru głównego, który umożliwia szybki dostęp do rekordów zbioru głównego.

Zbiór odwrócony składa się z następujących plików:

- BIS.L01 - słownik terminów krótkich (do 10-ciu znaków);
- BIS.L02 - słownik terminów długich (do 30-tu znaków);
- BIS.N01 - skorowidz słownika terminów krótkich;
- BIS.N02 - skorowidz słownika terminów długich;
- BIS.CNT - plik kontrolny;
- BIS.IFP - lista odsyłaczy do zbioru głównego.

Do sterowania sortowaniem i drukowaniem wyciągów z bazy stworzono następujące formularze:

- AC00.FRM - do sortowania wg kodów terminów;

ATERM.FRM - do sortowania terminów w porządku leksykograficznym;

ASTS.FRM i APSTS.FRM - do drukowania zawartości bazy terminologicznej w formacie skróconym;

ASTP.FRM i APSTP.FRM - do drukowania wyciągów z bazy w formacie pełnym;

ALTZ.FRM - do drukowania list terminów wyszukiwawczych w porządku alfabetycznym;

ALTZC.FRM, APLTZC.FRM - do drukowania list terminów wyszukiwawczych w porządku wynikającym z ich kodów.

W trakcie pracy systemu BIS tworzone są dynamiczne rozmaite pliki robocze. Zawierają one między innymi: tymczasowe wyniki wyszukiwania, wyrażenia wyszukiwawcze bieżącej sesji, pośrednie wyniki sortowania itp. Budowa tych plików nie jest istotna z punktu widzenia użytkownika.

Budowa rekordu zbioru głównego

Rekordy zbioru głównego bazy BIS składają się z następujących pól zmiennej długości:

1. TERMIN - pole zawierające termin definiowany w postaci łańcucha znaków alfanumerycznych. Maksymalna liczba znaków: 70. Pole nie jest polem powtarzalnym, nie zawiera podpól.
2. SYNONIMY - pole powtarzalne przeznaczone na synonimy terminu definiowanego. Każde wystąpienie pola zawiera jeden synonim w postaci łańcucha znaków alfanumerycznych. Jeżeli dany synonim dotyczy tylko niektórych znaczeń hasła definiowanego, numer (numery) tych znaczeń należy umieścić w podpolu¹ z. Każdy synonim może mieć długość do 70 znaków. Pole nie jest obowiązkowe - może nie wystąpić w rekordzie.

3. KOD TERMINU - pole przeznaczone na kod odwzorowujący dziedzinowe powiązania z innymi terminami zdefiniowanymi w bazie. Długość 25 znaków alfanumerycznych, pole nie jest powtarzalne.
4. DEFINICJA - pole powtarzalne przeznaczone na definicje haseł słownika. Dla każdego znaczenia terminu definiowanego przeznaczono jedno wystąpienie pola o długości do 1650. znaków alfanumerycznych. Nie zawiera podpól.
5. UWAGI OGÓLNE - pole przeznaczone na ogólne uwagi umożliwiające szersze objaśnienie znaczenia hasła. Z uwagi na systemowe ograniczenie maksymalnej długości jednego wystąpienia pola do 1650. znaków zdefiniowano to pole jako powtarzalne. Pociąga to za sobą konieczność dzielenia dłuższych uwag na podrozdziały umieszczane w poszczególnych wystąpieniach pola. Nie zawiera podpól i nie jest obowiązkowe.
6. ODPOWIEDNIKI ANGIELSKIE - pole powtarzalne. Każde wystąpienie tego pola zawiera jeden odpowiednik angielski definiowanego terminu w postaci łańcucha znaków o długości nie większej niż 60. Nie zawiera podpól.
7. ODPOWIEDNIKI FRANCUSKIE - pole powtarzalne. Każde wystąpienie tego pola zawiera jeden odpowiednik francuski definiowanego terminu w postaci łańcucha znaków o długości nie większej niż 60. Nie zawiera podpól.
8. ODPOWIEDNIKI NIEMIECKIE - pole powtarzalne. Każde wystąpienie tego pola zawiera jeden odpowiednik niemiecki definiowanego terminu w postaci łańcucha znaków o długości nie większej niż 60. Nie zawiera podpól.
9. ODPOWIEDNIKI ROSYJSKIE - pole powtarzalne. Każde wystąpienie tego pola zawiera jeden odpowiednik rosyjski definiowanego terminu

zapisany w transkrypcji fonetycznej w postaci łańcucha znaków o długości nie większej niż 60. Nie zawiera podpól.

10. KOD AUTORA DEFINICJI - pole umożliwia zarejestrowanie kodu autora definicji w postaci maksimum 6. znaków. Nie jest powtarzalne, nie zawiera podpól, służy jedynie celom dokumentacyjnym.

11. DATA WPROWADZANIA - pole pojedyncze zawierające datę utworzenia danego rekordu według wzorca: DD/MM/RR.

Formularz wprowadzania danych

Wprowadzanie danych w systemie BIS polega na wypełnianiu odpowiednio przygotowanych formularzy wejściowych. Dla pierwszej - roboczej wersji systemu zdefiniowano jeden standardowy formularz BIS. Formularz składa się z trzech stron.

Na pierwszej stronie znajdują się pola do wprowadzania wszystkich danych poza definicją i uwagami. Do wprowadzania synonimów i odpowiedników obcojęzycznych zastosowano pola przesuwne - tj. pola umożliwiające wprowadzenie większej liczby znaków niż objętość tworzonego przez nie okna na ekranie. Przy wprowadzaniu synonimów, odwołania do konkretnych znaczeń (w postaci numerów tych znaczeń) należy poprzedzać identyfikatorem podpola [^] z, a poszczególne synonimy oddzielać separatorem pól powtarzalnych zdefiniowanym w tym systemie jako znak | . Również odpowiedniki obcojęzyczne w obrębie jednego języka należy między sobą separować znakiem | . Kod autora definicji można wprowadzać w postaci jego inicjałów. Przy wprowadzaniu danych możliwe jest zdefiniowanie w niektórych rubrykach formularza wejściowego wartości domyślnych. Będzie to przydatne w polu DATA WPROWADZANIA oraz przy wprowadzaniu wielu definicji jednego autora.

Cała druga strona formularza stanowi okno (pole przesuwne) służące do wprowadzenia definicji terminu. Poszczególne definicje należy oddzielać separatorem | i ich objętość nie może przekroczyć 1650 znaków. Definicje terminów o wielu znaczeniach powinny być numerowane, pojedyncze nienumerowane definicje dobrze jest rozpoczynać trzema spacjami dla ujednolicenia wydruków. Terminy cytowane w obrębie definicji wprowadzane będą dużymi literami (dla większego wyróżnienia ich w tekście) i ujmowane w nawiasy kątowe (co spowoduje skierowanie ich do zbioru odwróconego).

Na trzeciej stronie formularza zdefiniowano pole przesuwne służące do wprowadzenia uwag do terminu. Wynikającą z ograniczeń systemu COS/ISIS fizyczną maksymalną długość pola (1650 znaków) powiększono sztucznie definiując to pole jako powtarzalne. Umożliwia to wprowadzenie do pola UWAGI tekstu dłuższego niż 1650 znaków w postaci np. paragrafów oddzielanych między sobą znakiem | Terminy cytowane w obrębie pola UWAGI wprowadzane będą dużymi literami (dla większego wyróżnienia ich w tekście) i ujmowane w nawiasy kątowe (co spowoduje skierowanie ich do zbioru odwróconego).

Wypełnianie każdego pola formularza polega na wprowadzeniu łańcucha znaków stanowiącego treść pola rekordu. Odpowiednie wzorce zostają wyświetlone po wciśnięciu klawisza F1.

Formaty wyświetlania danych

Rekordy bazy BIS mogą zostać wydrukowane lub wyświetlone w rozmaity sposób, zgodnie z wybranym formatem. Dla pierwszej - roboczej wersji systemu zdefiniowano następujące formaty:

BIS - format standardowy;

KONTR - format kontrolny;

SKROT - skondensowany format typowy dla słowników;

SŁOWNIK - format typu słownik przekładowy wszystkich czterech języków;

ANG - format typu słownik przekładowy języka angielskiego;

NIEM - format typu słownik przekładowy języka niemieckiego;

FRANC - format typu słownik przekładowy języka francuskiego;

ROS - format typu słownik przekładowy języka rosyjskiego.

Standardowy format BIS służy do wyświetlania treści rekordu w sposób przejrzysty, z pominięciem pól KOD AUTORA DEFINICJI i DATA WPROWADZANIA. Poszczególne części opisu zaczynają się w nowej linii i oddzielane są jedną linią pustą. Termin definiowany wprowadzany jest dużymi literami. Kod terminu ujmowany jest w nawiasy kwadratowe. Następnie wyprowadzane są synonimy poprzedzone etykietą "Synonimy:". Jeżeli dany synonim odnosi się tylko do konkretnej definicji hasła, to numer tej definicji wyprowadzany jest w nawiasach okrągłych. Poszczególne synonimy oddzielane są średnikiem. Zarówno definicje terminu jak i uwagi wyprowadzane są z pominięciem nawiasów kątowych wyodrębniających słowa kluczowe. Ostatnią część opisu stanowią odpowiedniki obcojęzyczne. Każda grupa wyprowadzana jest w nowej linii i poprzedzana etykietą określającą język. Odpowiedniki jednego języka oddzielane są średnikiem. Rekordy oddzielane są optycznie linią przerywaną.

Zmianę formatu wyświetlania umożliwia opcja F w menu PXGEN - "Wyszukiwanie danych". Po wybraniu tej opcji wyświetlana jest definicja aktualnego formatu wyświetlania. Zmianę na jeden ze zdefiniowanych formatów uzyskuje się przez wprowadzenie na początek tej definicji nazwy nowego formatu poprzedzonej znakiem @ i zakończonej minimum dwiema spacjami.

Format kontrolny umożliwia wyprowadzenie dokładnie całej zawartości rekordu. Stanowi niejako rozszerzenie formatu standardowe-

go. Dodatkowo wyprowadzane są: numer fizyczny rekordu, kod autora definicji w nawiasach kwadratowych i data wprowadzania danych. Kod terminu nie jest umieszczany w nawiasach, lecz specjalnie wyróżniony w prawym górnym rogu ekranu dla większej czytelności.

Format skondensowany SKROT jest typowym formatem występującym w słownikach wydanych na nośnikach papierowych. Termin wyprowadzany jest dużymi literami, synonimy ujęte zostają w nawiasy kwadratowe. Po myślniku następują definicje, następnie (od nowej linii) uwagi. W kolejnej linii wyprowadzane są odpowiedniki obcojęzyczne a na koniec kod terminu w nawiasach kwadratowych. Treści rekordów oddzielane są jedynie dwiema liniami odstępu.

Format SLOWNIK umożliwia korzystanie z zawartości bazy jako słownika przekładowego. Wyprowadza termin definiowany oraz jego synonimy ujęte w nawiasy kwadratowe, a następnie (od nowych linii) odpowiedniki obcojęzyczne. Jako ostatni wyprowadzany jest kod terminu w nawiasach kwadratowych.

Traktowanie bazy jako prostego słownika przekładowego określonego języka (angielskiego, francuskiego, niemieckiego czy rosyjskiego^{2/}) umożliwiają odpowiednio formaty ANG, FRANC, NIEM, ROS. Ich działanie jest analogiczne do formatu SLOWNIK z tym, że nie są wyprowadzane kody terminów.

Słownik terminów wyszukiwawczych

Jak już wspomniano, na podstawie zawartości zbioru głównego tworzony jest zbiór odwrócony, stanowiący słownik terminów wyszukiwawczych. Wybrane pola rekordów zbioru głównego są pobierane zgodnie ze zdefiniowanym formatem, a następnie w oparciu o zdefiniowaną dlań technikę selekcji mieszczane w słowniku.

^{2/} tylko w transkrypcji fonetycznej

Terminy definiowane są umieszczane w nawiasach kwadratowych i zakańczane wykrzyknikiem. Terminy dłuższe niż 27 znaków są prawostronnie obcinane do tej długości. Każdemu terminowi definiowanemu w zbiorze głównym odpowiada jeden element słownika terminów wyszukiwawczych.

Synonimy terminów definiowanych są również umieszczane w nawiasach kwadratowych, z tą jednak różnicą, że nie są zakańczane wykrzyknikiem. Synonimy dłuższe niż 28 znaków są prawostronnie obcinane do tej długości. Każdemu synonimowi w zbiorze głównym odpowiada jeden element słownika terminów wyszukiwawczych.

Kody terminów definiowanych wyprowadzane są z rekordów zbioru głównego bez żadnych modyfikacji. Każdy unikalny kod stanowi pojedynczy termin wyszukiwawczy.

Słowa kluczowe występujące w polach: DEFINICJA i UWAGI OGÓLNE a wprowadzone w nawiasach kątowych zostają wprowadzone do słownika z pominięciem tych nawiasów. Każdy z nich stanowi indywidualny element wyszukiwawczy. Terminy dłuższe są obcinane prawostronnie do 30 znaków.

Odpowiedniki obcojęzyczne są poprzedzane przedrostkami określającymi ich pochodzenie. I tak odpowiedniki angielskie poprzedzane są prefiksem ~A~, francuskie ~F~, niemieckie ~N~ a rosyjskie ~R~. Każdy odpowiednik obcojęzyczny obcinany jest prawostronnie do 27 znaków i stanowi odrębny element słownika terminów wyszukiwawczych.

Dodanie przedrostków do terminów wyszukiwawczych umożliwiło automatyczne podzielenie słownika na następujące rozdziały:

- terminy zdefiniowane i ich synonimy - klucz E
- słowa kluczowe wymieniane w definicjach i uwagach - klucz A
- kody terminów definiowanych - klucz I
- odpowiedniki angielskie - klucz ~A~

- odpowiedniki niemieckie - klucz ~N ~
- odpowiedniki francuskie - klucz ~F ~
- odpowiedniki rosyjskie - klucz ~R ~

Odpowiedni rozdział słownika można wyświetlić na ekranie podając odpowiedni klucz. Klucz ten ma charakter hierarchiczny a terminy wyszukiwawcze uporządkowane są w słowniku w porządku zgodnym z wartościami kodów znaków ASCII. W związku z tym dostęp do odpowiedniego podrozdziału osiąga się przez prawostronne dołączenie do klucza podstawowego odpowiedniego znaku bądź ciągu znaków.

Produkowanie wyciągów z bazy

Proces generowania raportów z bazy sterowany jest przez podanie parametrów drukowania w specjalnie przygotowanym formularzu. W obecnej - roboczej wersji systemu BIS opracowano formularze umożliwiające wyprowadzanie następujących raportów z bazy:

- listy terminów zdefiniowanych w porządku quasi - alfabetycznym^{3/};
- listy terminów zdefiniowanych w porządku wynikającym z kodów tych terminów;
- raportu w postaci zbliżonej do słownika terminologicznego w formacie standardowym systemu BIS;
- wyciągu z bazy w postaci słownika terminologicznego w formacie skondensowanym, typowym dla wydawnictw słownikowych.

Do generowania raportów służą funkcje dostępne w menu systemowym PXPRT. Pojawia się ono po wybraniu opcji P z menu głównego.

^{3/} czyli w porządku wynikającym z kodów znaków w generatorze znaków. Polskie znaki diakrytyczne mają niestety kody wyższe niż wszystkie podstawowe znaki alfabetu.

Wykorzystanie specjalnie zdefiniowanych dla systemu BIS specyfikacji umożliwia funkcja S - Wydruk wg własnej specyfikacji. Po wybraniu tej opcji użytkownik dokonuje wyboru rodzaju raportu przez podanie nazwy odpowiedniego formularza sterującego wydrukiem. I tak:

- lista terminów zdefiniowanych w porządku quasi - alfabetycznym sterowana jest przez formularz LIZ (gdy rezultat ma być wyprowadzany do pliku tekstowego) oraz PLIZ (gdy rezultat ma być drukowany na drukarce) ;

- lista terminów zdefiniowanych w porządku wynikającym z kodów tych terminów produkowana jest według parametrów zawartych w formularzu PLIZC (ze skierowaniem na drukarkę) oraz LIZC (gdy rezultat ma być wyprowadzany do pliku tekstowego);

- raport w postaci zbliżonej do słownika terminologicznego w formacie standardowym systemu BIS generowany jest zgodnie z parametrami formularzy STP (do pliku) i PSTP (na drukarkę);

- wyciąg z bazy w postaci słownika terminologicznego w formacie skondensowanym wyprowadzany jest według formularzy STS (do pliku tekstowego) i PSTS (na drukarkę).

Po dokonaniu wyboru odpowiedniego formularza użytkownik ma możliwość zaakceptowania bądź zmodyfikowania proponowanych parametrów wydruku. W przypadku zmiany formatu wydruku na jeden ze zdefiniowanych w systemie BIS nazwa formatu musi być poprzedzona znakiem @ . Po formularzu sterującym wydrukiem pojawia się formularz sterujący sortowaniem. Z uwagi na specyficzny język w jakim zdefiniowano parametry sortowania nie zaleca się ich modyfikacji. W obecnej wersji systemu zdefiniowano dwa formularze sortowania: IERM - sterujący sortowaniem quasi-alfabetycznym terminów i COD - sterujący sortowaniem według kodów terminów zdefiniowanych. Z uwagi na to, że w treści rekordów występują polskie znaki diakrytyczne zakodowane według standardu Mazovia konieczne jest aby

wydruki z bazy były drukowane na drukarce wyposażonej w generator znaków w tym standardzie.

Modyfikacje systemu ukierunkowane na użytkownika nieprofesjonalnego

W celu ułatwienia pracy użytkownikom systemu BIS wprowadzono kilka modyfikacji systemu mikroISIS. Podstawowe zmiany dotyczą menu. Użytkowników systemu BIS podzielono na dwie kategorie: modyfikujących bazę i wyszukujących informacje. Przy wywołaniu systemu (poleceniem SŁOWNIK) należy zadeklarować swoją przynależność do jednej z tych grup co spowoduje przyłączenie odpowiedniego zbioru menu (odpowiednio z katalogu SŁOWNIK MENUM albo SŁOWNIK MENU). Dla wersji roboczej skoncentrowano się na odcięciu dostępu do tych funkcji systemu mikroISIS, których używanie nie jest konieczne w kontekście słownika terminologicznego, a których nieumiejętne użycie mogłoby spowodować uszkodzenie bazy. Opracowano ogólną koncepcję modyfikacji menu systemu dla wersji ostatecznej. Rozszerzono menu sterujące generowaniem raportów (menu PXPRT) o informacje na temat dostępnych definicji raportów. Usunięto opcje umożliwiające zmianę języka konwersacji jako nieprzydatną w przypadku słownika w języku polskim dedykowanego polskiemu użytkownikowi.

Kolejną modyfikacją systemu jest przedefiniowanie znaczenia klawisza funkcyjnego F2. W oryginalnej wersji systemu powodował on podczas wprowadzania danych usunięcie zawartości pola i zakończenie jego edycji. Ponieważ osoby wprowadzające dane są przyzwyczajone do używania tego klawisza przy wprowadzaniu polskich znaków diakrytycznych (przyzwyczajenie z programu CHI-WRITER) - usunięto jego oryginalne znaczenie. Oryginalnym separatorem pól powtarzalnych systemu CDS/ISIS jest znak $\%.$ W systemie BIS zmieniono ten separator na symbol $\{$, uznano go bowiem za bardziej mnemotechniczny symbol separatora.

Dla potrzeb systemu BIS nieznacznie zmodyfikowano oryginalne komunikaty systemu ISIS. Chodziło głównie o to, aby wszystkie pojawiające się napisy były zrozumiałe dla użytkownika.

Perspektywy rozwoju systemu BIS

W dalszym ciągu prac nad systemem BIS przewidziane są następujące działania:

- opracowanie szerszego wachlarza formatów wyświetlania;
- opracowanie szeregu dalszych, w miarę możliwości najbardziej wyrafinowanych raportów z bazy;
- zrealizowanie powiązań dziedzinowych między zdefiniowanymi terminami (za pomocą pola KOD TERMINU);
- opracowanie pliku askryptorów skojarzeniowych ANY;
- prace nad umożliwieniem wprowadzenia do definicji prostych schematów i ilustracji;
- modyfikacje menu przeznaczonych dla użytkowników nieprofesjonalnych i zabezpieczenie systemu przed skutkami nieumiejętnego posługiwania się nim;
- prace nad humanizacją dialogu użytkownik - system.

Za szczególnie istotne uważa się zhumanizowanie procesu komunikacji między użytkownikiem a systemem i temu celowi podporządkowane zostaną wszystkie inne prace.

20.02.1991 r.

THE TECHNICAL PROJECT OF THE TERMINOLOGICAL DATA BASE
ON THE THEORY OF INFORMATION RETRIEVAL LANGUAGES

Summary

The article is devoted to characteristics of the technical project of the terminological data base BIS which has been worked out at the Institute of Librarianship and Information Science of the Warsaw University. The discussed data base was implemented with using the universal software package CDS/ISIS in the version adopted for IBM PC. There are presented: the physical and logical structure of the system, the structure of the data base's records, formats of displaying, structure of the vocabulary of the searching terms, possibilities of generating of the reports from the base, kinds of the reports available and ways of merging sub-bases into one compatible data base. The directions of the further works on the system also are shown.

Технический проект терминологической базы данных в области
теории информационно-поисковых языков

Резюме

Статья посвящена обсуждению технического проекта терминологической базы данных BIS в области теории информационно-поисковых языков разработанной в Институте библиотековедения и научной информации Варшавского университета. Для ее реализации использовался универсальный программный пакет CDS/ISIS во версии на микро эви типа IBM PC. Обсуждени: логическое строение системы, структура рекордов входящих в состав базы данных, формат дисплея, структура словаря поисковых терминов, возможности и виды генерирования рапортов из базы, а также способ соединения подбаз в одну совместимую базу данных. Представлены тоже направления дальнейших работ по развитию системы.

HANNA POPOWSKA

Instytut Informacji Naukowej,
Technicznej i Ekonomicznej

**WYBÓR BAZ DANYCH I SERWISÓW ONLINE
(KILKA WSKAZÓWEK METODYCZNYCH)**

Rozwój zachodniego rynku baz danych - trudności związane z wyborem potrzebnych baz danych i odpowiednich serwisów online. Wstępne założenia wykorzystania baz danych przez konkretnego użytkownika (własna, wewnętrzna eksploatacja baz danych lub korzystanie z baz danych za pośrednictwem serwisów online). Wybór baz danych: wybór źródeł informacji o bazach danych; kryteria przydatności baz danych; wieloetapowy proces wyboru baz danych. Wybór serwisów online.

Na Zachodzie jednym z podstawowych źródeł zaspokajania potrzeb informacyjnych są bazy danych, wykorzystywane w trybie bezpośrednim (online) - za pośrednictwem dystrybutorów baz danych (producentów baz danych lub wyspecjalizowanych pośredników, tzw. serwisów online) albo pozyskiwane na nośnikach maszynowych - do własnej eksploatacji.

Rozwój tej dziedziny usług informacyjnych jest szybki: w 1965 r. zachodnim użytkownikom udostępniano (wówczas jeszcze nie online) ok. 12-20 baz danych /54/, a obecnie można korzystać z kilku tysięcy baz danych, produkowanych i udostępnianych przez wiele instytucji. Na przykład, w katalogu /17/ Cuadra Associates, opublikowanym w połowie 1989 r., przedstawiono: 1870 producentów baz danych, 622 serwisy online i 4245 baz danych (zob. tab.1). Należy jednak zana-

"Zagadnienia Informacji Naukowej" 1991 nr 1(58)

czyć, że we wspomnianym katalogu nie są uwzględniane wszystkie powszechnie dostępne bazy danych, lecz jedynie bazy danych, które są powszechnie dostępne line poprzez sieci telekomunikacyjne o zasięgu międzynarodowym lub obejmujące swym zasięgiem jeden cały kraj lub kilka krajów. Innymi słowy, liczba udostępnianych na świecie baz danych znacznie przewyższa liczbę baz danych, spełniających kryteria przyjęte na użytek katalogu Cuadra Associates.

Tak więc, oferta światowego rynku baz danych jest bogata i stale się poszerza, co sprawia, że wybór baz danych (potrzebnych konkretnemu użytkownikowi) staje się coraz trudniejszy.

Od pewnego czasu daje się zauważyć wzrastające zainteresowanie metodyką oceny, porównywania i wyboru baz danych oraz serwisów online, o czym świadczy rosnąca liczba publikacji na ten temat, np. /3,33,34,37,39,40,49,50,51,52,53,56,58,59,60,73,78,83/.

Dotychczasowe wykorzystanie w Polsce zachodnich baz danych /68,86/ uznać należy za znikome, w stosunku do światowej oferty i naszych potrzeb. Tym samym, także i doświadczenia w wyborze potrzebnych baz danych są niewielkie, może się więc zdarzyć, że - na podstawie pochopnego wyboru - zamiast bazy danych, bardzo potrzebnej polskiemu użytkownikowi, zostanie sprowadzona do Polski baza mało przydatna lub całkowicie nieprzydatna. Koszty takiej błędnej decyzji są, oczywiście, bardzo duże, wymierne i niewymierne.

Ostatnio jednak także i w naszym kraju stopniowo wzrasta zainteresowanie możliwościami zachodniego rynku baz danych. Specjaliści z różnych polskich instytucji zastanawiają się nad wyborem zestawu najbardziej przydatnych baz danych. Problem formułowany bywa różnie, np.:

- Dyrekcja mojej firmy zdecydowała, że powinniśmy zacząć korzystać z najlepszych światowych baz danych z naszej dziedziny. Ile jest

takich baz, jak to wszystko zorganizować i ile takie przedsięwzięcie będzie kosztowało?

- Możemy kupić czytnik CD-ROM i przeznaczyć rocznie na zachodnie bazy danych ok. pięciu tysięcy USD. Jakie bazy danych warto sprawdzić?
- Ustalono, że nasz instytut może kupić jedną zachodnią bazę danych. Którą należy wybrać?
- Interesujemy się kilkoma bazami danych i chcemy z nich korzystać za pośrednictwem serwisu online. Wiemy, że te bazy są udostępniane przez dwa serwisy: DIALOG i ESA-IRS. Który z tych serwisów oferuje korzystniejsze warunki?
- Uznaliśmy, że w naszej instytucji bardzo często będziemy korzystać z dwóch baz danych: ABI/INFORM i MANAGEMENT CONTENTS, ale sporadycznie wykorzystamy także kilka baz z dziedziny statystyki i finansów. Czy najlepszym dla nas rozwiązaniem będzie nawiązanie współpracy z serwisem DIALOG?

Tak więc, wydaje się celowe zaproponowanie kilku wskazówek metodycznych, które mogą się okazać przydatne osobom, stojącym wobec niełatwego zadania wyboru kilku, kilkunastu czy kilkudziesięciu baz danych dla określonego grona polskich użytkowników.

To opracowanie przygotowałam na podstawie /67/ - jednego z napisanych przeze mnie rozdziałów pracy planowej /69/, wykonanej w 1990 r. w Instytucie INTE przez Barbarę Stefaniak, Adama Wysockiego i Hannę Popowską.

Niżej zamiast baza danych będzie stosowany skrót BD.

1. OKREŚLENIE WSTĘPNYCH ZAŁOŻEŃ WYKORZYSTANIA BAZ DANYCH

Proces wybierania BD i serwisów online musi zostać poprzedzony analizą potrzeb informacyjnych grona użytkowników, dla których dokonujemy tego wyboru, oraz wnikliwą oceną sytuacji informatycznej i finansowej. Tak więc, przede wszystkim trzeba dokładnie określić potrzeby informacyjne grona użytkowników, którym mają służyć wybrane BD. W ramach działań wstępnych, przygotowawczych, należy dokonać oceny posiadanych środków technicznych (sprzętowych), organizacyjnych i finansowych, a także ustalić, czy są możliwe jakieś (jakie?) zmiany w tym zakresie. Warto z góry dokładnie wiedzieć, jakimi kwotami rozporządzamy i ile będziemy mogli rocznie przeznaczać na BD.

Przed wyborem BD i serwisów online warto także określić, do jakich celów (np. SOI, RWI) i jak wybrane BD będą u nas (tj. w naszej instytucji) wykorzystywane, tzn. czy zamierzamy je pozyskać do własnej, wewnętrznej eksploatacji, czy też istnieje możliwość korzystania z BD w trybie bezpośrednim, za pośrednictwem wyspecjalizowanego serwisu online.

Można przyjąć jeden z dwóch wariantów działania:

- z góry ustalić tryb wykorzystania BD i pod tym kątem dokonywać wyboru potrzebnych BD,
- założyć, że tryb korzystania z BD uzależnimy od wyników wyboru przydatnych BD.

Oczywiście . każde rozwiązanie (wyłącznie za pośrednictwem serwisu online; wyłącznie własna eksploatacja BD; kilka najpotrzebniejszych baz - własna eksploatacja, a pozostałe przydatne - online) ma wady i zalety. Na przykład, użytkownik, korzystający z BD za pośrednictwem serwisu online, zazwyczaj ma do dyspozycji dużo więcej BD, a ponadto może z nich uzyskać najnowsze informacje na-

tychmiast po ich wprowadzeniu do bazy, ale każda minuta namysłu podczas wyszukiwania zwiększa koszt tego procesu.

W wypadku wyboru małej liczby BD (jednej, dwóch, co najwyżej kilku BD), które są stosunkowo rzadko (co pół roku, co kwartał) aktualizowane, a u nas będą bardzo często wykorzystywane (przeszukiwane), warto pozyskać te BD do własnej eksploatacji. W tej sytuacji trzeba wstępnie ustalić, na jakim nośniku (np. taśma magnetyczna, CD-ROM) wybrane BD mają nam być dostarczane.

Przypomnijmy, że eksploatacja BD wymaga spełnienia odpowiednich warunków technicznych (sprzęt, oprogramowanie). Warto zwrócić uwagę na fakt, iż w ciągu ostatniej dekady zaszły duże zmiany w dziedzinie nośników informacyjnych. Jeszcze stosunkowo niedawno instytucje, zainteresowane "wewnętrzną" eksploatacją BD, otrzymywały te bazy na taśmach magnetycznych. Obecnie coraz większą popularność zdobywa CD-ROM /21,32,36,62,79/.

Podobnie jak w wypadku każdego innego nośnika informacji, korzystanie z CD-ROM wymaga posiadania specjalnego urządzenia odczytującego. Przy zakupie BD trzeba sobie zapewnić oprogramowanie, umożliwiające jej właściwe wykorzystanie. Warto pamiętać, że informacje, zapisane przez jeden system obsługi BD, są czytelne tylko przez ten system i systemy z nim spokrewnione. Ponadto, należy tak dobrać oprogramowanie, by umożliwić właściwą realizację wszvstkich niezbędnych operacji, wykonywanych na BD. Można dodać, że w przypadku BD na CD-ROM oprogramowanie zazwyczaj jest dodawane do BD (na tym samym dysku CD-ROM lub oddzielnie, na dyskietce), należy więc ustalić, czy to oprogramowanie wystarczy do eksploatacji danej BD w ramach naszego systemu. Jeżeli nie - to należy zatroszczyć się o pozyskanie innego oprogramowania.

Decydując się na własne przetwarzanie BD, trzeba bardzo dokładnie sprecyzować zakres wykorzystania BD (RWI; SOI; częstotliwość wykorzystania; liczba egzemplarzy wydruku, przekazywanych jednemu użytkownikowi; charakterystyka użytkowników - czy tylko z jednej instytucji, czy z wielu, czy tylko z terenu jednego kraju), gdyż wiele danych tego typu jest niezbędnych do ustalenia warunków umowy z dysponentem BD, od nich też czasami zależy koszt prenumeraty. Trzeba pamiętać, że przestrzeganie zawartej umowy bywa kontrolowane przez dysponenta BD, a stwierdzenie naruszenia przyjętych zobowiązań pociąga za sobą karę.

Na zakończenie należy także zwrócić uwagę na fakt, że sprowadzenie większej liczby BD do wewnętrznej eksploatacji jest stosunkowo kosztowne i, jak na nasze warunki, dość złożone organizacyjnie. Warto więc przeprowadzić wstępną kalkulację opłacalności planowanego przedsięwzięcia (ta uwaga dotyczy także i korzystania z usług serwisu online).

Wszystkie problemy, związane z techniką obsługi BD, nie mają znaczenia, jeśli decydujemy się na skorzystanie z usług serwisu online. Słowo "serwis" znaczy obsługa, a więc wszelkie działania techniczne i organizacyjne, związane z przetwarzaniem BD, przejmuje serwis online. My musimy jedynie rozporządzać odpowiednim sprzętem, zapewniającym właściwe połączenie online z serwisem. Szczegółowe informacje na temat koniecznego wyposażenia technicznego i oprogramowania, służącego do komunikacji z serwisem, znaleźć można w publikacjach, wydawanych przez poszczególne serwisy (np. /15,35/). Jednak wysokie koszty łączności i nadal niezadowalający poziom usług w tym zakresie w naszym kraju mogą zniechęcać wielu potencjalnych użytkowników do korzystania online z BD. Warto także pamiętać, że serwis, pośrednicząc między nami a producentem BD,

pobiera za swe usługi opłaty, niekiedy dość wysokie. Jednak w sytuacji, gdy się okaże, że wybraliśmy stosunkowo dużo BD, a zwłaszcza - gdy te BD będą bardzo często aktualizowane (np. kilka razy dziennie), nie ma sensu sprowadzać tych BD do własnego przetwarzania i wówczas trzeba zdać się na pośrednictwo serwisu online. W tej sytuacji pozostaje jeszcze wybór odpowiedniego serwisu online. Warto pamiętać, że oferty poszczególnych serwisów różnią się znacznie: stosowanymi algorytmami przeszukiwania BD, możliwościami prezentowania wydruków, cenami, itd. Należy więc porównać warunki udostępniania interesującej nas BD we wszystkich serwisach, które obsługują rynek europejski i oferują tę bazę.

Problem wyboru (z kilku serwisów online, utrzymujących tę samą BD) jednego serwisu, z którego usług będziemy korzystać, wymaga szczególnej uwagi. Temu zagadnieniu poświęcona zostanie 3. część niniejszego artykułu.

Reasumując, w ramach wstępnego etapu prac, poprzedzającego działania poświęcone właściwemu wyborowi BD, trzeba określić przynajmniej wstępnie warunki informatyczne (rozważyć ewentualność zakupu sprzętu, np. czytnika CD-ROM i oprogramowania) i finansowe. Warto pamiętać, że w odniesieniu do coraz większej liczby BD użytkownik staje wobec konieczności wyboru: zakup BD na CD-ROM lub korzystanie z BD za pośrednictwem serwisu online. Dopiero szczegółowe przeanalizowanie obu wariantów pozwoli podjąć właściwą decyzję. Na przykład, A.S. Pollitt /63/ przytoczył szacunkowe koszty korzystania z bazy ABI/INFORM w obu wskazanych wypadkach, z których wynika, że przy 55-60 standardowych wyszukiwaniach (czas połączenia - ok. 10 minut, 5 rekordów wyświetlonych online, 50 rekordów otrzymanych offline) w serwisie DIALOG zdecydowanie korzystniejsza okazuje się roczna prenumerata ABI/INFORM na CD-ROM.

Tak więc, trzeba w pełni zdawać sobie sprawę, że praca nad wyborem BD nie kończy się w momencie przygotowania zestawu potrzebnych BD. Taki zestaw powinien zostać uzupełniony sugestiami na temat sposobu wykorzystania wybranych BD, przy czym przede wszystkim trzeba ustalić, który z dwóch podstawowych wariantów działania (najogólniej mówiąc: (1) płacisz raz (więcej) - pozyskujesz BD na określony czas i korzystasz, ile chcesz, (2) jednorazowo płacisz mniej, ale za każdym razem płacisz serwisowi online za każde wykorzystanie BD) będzie lepszy w naszej sytuacji.

2. WYBÓR BAZ DANYCH

2.1. Wybór źródeł informacji o bazach danych

Podstawową kwestią, którą należy rozstrzygnąć, przystępując do wybierania BD, stanowi dobranie odpowiednich materiałów źródłowych o BD. Wiadomości o BD czerpać można z następujących źródeł /7,28,29,65/:

- literatura o BD (artykuły z czasopism, monografie, materiały konferencyjne, publikacje reklamowe).
- katalogi BD,
- katalogi firm pośredniczących: serwisów online, serwisów CD-ROM,
- literatura fachowa z danej dziedziny tematycznej (często zamieszczane są omówienia BD i systemów informacyjnych z tej dziedziny),
- kontakty osobiste.

Niewątpliwie podstawowym źródłem informacji o BD są renomowane katalogi BD, oferujące pełny, a przede wszystkim aktualny zestaw informacji o BD a więc, w odniesieniu do poszczególnych BD: aktualne nazwy, adresy, zasady rozpowszechniania, charaktery-

styki ilościowe, informacje o zakresie tematycznym, itp. Ponadto, aktualne wydanie dobrego katalogu odzwierciedla sytuację na rynku BD, gdyż: zawiera także i najnowsze, niedawno wprowadzone BD; wskazuje BB, które są już niedostępne (w ogóle lub praktycznie).

Pierwszy katalog BD, rozpowszechnianych na nośniku maszynowym, opublikowano w USA w 1976 r., a już w 1985 r. liczba wydanych katalogów BD przekroczyła czterdzieści /63/. Kilka przykładów katalogów BD: /4,9,12,17,20,24,38,41,61,70,79/.

Niektóre z tych katalogów są już dostępne na nośniku maszynowym, a nawet - w trybie bezpośrednim. W 1987 r. np. udostępniano online trzy katalogi BD /33/:

- katalog /17/ wydawany przez Cuadra Associates,
- katalog DATABASE OF DATABASES, stanowiący skomputeryzowaną wersję katalogu /9/, opracowywanego pod kierownictwem prof. M.E. Williams,
- katalog /12/ wydawany przez Knowledge Industry Publications.

Pierwszy z wymienionych informatorów udostępniany był przez pięć serwisów online: DATA-STAR, Westlaw, Questel, DataArkiv i ORBIT; drugi - przez serwis DIALOG; trzeci - przez serwis BRS. Warto podkreślić, że katalogi (zarówno wydawane tradycyjnie, jak i udostępniane online) różnią się znacznie częstotliwością aktualizacji. W wersji online katalog Cuadra Associates jest aktualizowany co kwartał, katalog Knowledge Industry Publications - co miesiąc /33/, natomiast katalog DATABASE OF DATABASES aż do jesieni 1989 r. udostępniany był w wersji wprowadzonej w sierpniu 1985 r. Uzupełniająco można wspomnieć, że już w 1989 r. w Cuadra Associates rozważano możliwość udostępniania katalogu także i na CD-ROM (informacja o zrealizowaniu tego zamiaru jeszcze do nas nie dotarła).

W literaturze przedmiotu najczęściej wzmiankowane są pierwsze dwa z wymienionych amerykańskich katalogów BD (trzeci z nich zawiera opisy baz produkowanych w Ameryce Północnej). Tak więc, zdaniem wielu specjalistów, najbardziej godnymi polecenia informatorami o BD są:

- "Directory of Online Databases" /17/, opracowywany przez Cuadra Associates,
- "Computer-Readable Databases: a Directory and Data Sourcebook" /9/, przygotowywany pod kierunkiem prof. Marthy E. Williams.

Pierwszy z wymienionych katalogów BD cieszy się od dawna w świecie wielką i zasłużoną popularnością, jako "najbardziej pełne i wiarogodne źródło informacji o BD" /72/. Publikowany jest od jesieni 1979 r., cztery razy w roku (dwa razy - pełne wydania, dwa razy - uzupełnienia). W lipcu 1989 r. w katalogu Cuadra Associates przedstawiono: 4245 BD, 1870 producentów BD, 622 serwisy online i

Tabela 1

Liczba baz danych, producentów baz danych i serwisów online, uwzględnionych w poszczególnych wydaniach katalogu Cuadra Associates /17/

okres	liczba BD	liczba producentów BD	liczba serwisów online
1979/80	400	221	59
1980/81	600	340	93
1981/82	965	512	170
1982/83	1350	718	213
1983/84	1878	927	272
1984/85	2453	1189	362
1986	2901	1379	454
1987	3369	1568	528
1988	3699	1685	555
1989	4062	1813	600
1989(lipiec)	4245	1870	622

83 śluz (gateways). Dane o poszczególnych rocznikach tego katalogu przytoczono w tab. 1. Warto zwrócić uwagę na kryteria doboru BD, przyjęte przez Cuadra Associates: w katalogu uwzględniane są wyłącznie BD, powszechnie dostępne online, przy czym każda z nich musi być udostępniana przez serwis online, podłączony do jednej lub kilku międzynarodowych sieci telekomunikacyjnych i/lub do sieci, obejmującej swym zasięgiem jeden cały kraj lub kilka krajów (z tego powodu na próżno szukać w omawianym katalogu polskich BD). W wersji skomputeryzowanej omawiany katalog nosi nazwę CUADRA DIRECTORY OF ONLINE DATABASES. Od niedawna można korzystać online z tego katalogu we Francuskim Ośrodku Kształcenia i Informacji Kadr w Warszawie (Data-France, via serwis Questel). W serwisie Questel omawiana BD jest aktualizowana co pół roku. O wielkiej popularności katalogu Cuadra Associates świadczy następująca okoliczność: liczni autorzy przytaczają dane z tego katalogu bez powołania się na źródło, tzn. zakładają, że czytelnik wie, skąd pochodzi cytowana informacja.

Drugi z podanych katalogów jest dostępny w Polsce przez działający od kilku lat serwis BRIDOLIS (Instytut Brytyjski w Warszawie), umożliwiający korzystanie z serwisu online DIALOG, który udostępnił bazę skierowaną (informatorową) COMPUTER-READABLE DATABASES, stanowiącą maszynową wersję katalogu "Computer-Readable Databases". Katalog ten wydawany jest od 1976 r. Pierwsze wydanie zawierało opisy 301 BD, drugie, z 1979 r. - 528 BD, trzecie, z 1982 r. - 773 BD, czwarte, z 1985 r. - 2805 BD. Od października 1985 r. katalog ten dostępny jest także online w serwisie DIALOG, do końca 1989 r. - pod nazwą DATABASE OF DATABASES. W styczniu 1990 r. omawiany informator zawierał opisy ponad 4200 BD. Wadą przedstawianego informatora jest niezbyt częsta aktualizacja, dla porównania przypomnijmy, że omówiony wyżej katalog Cuadra Associates podlega aktualizacji co

kwartał (w serwisie Questel - co pół roku). We wrześniu 1989 r. wersję katalogu z 1985 r., oferowaną do tego czasu przez DIALOG, zastąpiono nową wersją, tak więc obecnie dane zawarte w katalogu dotyczą sytuacji na rynku BD z połowy 1989 r.

Podsumowując, w celu dokonania wstępnej selekcji BD pod kątem wyboru przydatnych BD warto posłużyć się jednym z dwóch omówionych katalogów BD, przy czym wybrać należy, oczywiście, możliwie najnowsze wydania. Lepiej jednak nie ograniczać się do jednego źródła, lecz móc równocześnie korzystać z kilku katalogów, gdyż opisy BD w różnych katalogach różnią się, niekiedy nawet znacznie (różne elementy opisu BD, różny stopień szczegółowości opisu zawartości BD). Ponadto, niekiedy w katalogach zdarzają się jednak drobne omyłki, dlatego bezpieczniej jest informację o każdej BD zawsze czerpać z co najmniej dwóch źródeł.

Oczywiście, przy ostatecznym wyborze BD pomocne są i inne katalogi, a także opinie o poszczególnych BD, przedstawione explicite lub implicite (np.: częstotliwość pozytywnych omówień danej BD, liczba serwisów online, udostępniających BD) w literaturze. Kilka przykładów omówień BD: /55/; bibliograficzne BD - /22,24,26,64/; niebibliograficzne BD - /74,84/; faktograficzne BD - /10,47/; numeryczne BD - /1,40,43,44,48,57,80/; pełnotekstowe BD - /31,81,82/. Warto także pamiętać o tematycznych katalogach BD, tzn. o informatorach, opisujących BD z jednej dziedziny tematycznej, takich jak np.: wydane w 1987 r. przez Cuadra/Elsevier wyciągi tematyczne z katalogów Cuadra Associates: "Online databases in the medical and life sciences", "Online databases in the securities and financial markets"; informatory o BD publikowane przez Aslib (The Association for Information Management) w serii "An Aslib Guide Online", np.: "Online business and company databases" /2/, "Building,

construction and architecture databases" /14/, "Law databases" /71/, "Online medical databases" /45/, "Management and marketing databases" /23/, "Online patent and trademarks databases" /46/; przygotowane przez J.L. Stauda zestawienie BD "Online Wirtschaftsdatenbanken" /77/ (BD dotyczące zagadnień gospodarczych); cztery artykuły D.E. Meyera i D. Ruizy /49,50,51,73/; opracowania /1,8,74,75, 78/.

Uzupełniająco trzeba zauważyć, że przy wyborze BD użytkownik z Zachodu ma do dyspozycji liczne informatory o BD, dostępne online (o czym już zresztą wspomniano wyżej). Ponadto, wiele serwisów online ułatwia proces wyboru BD, utrzymując online specjalne, pomocnicze BD (indeksy przedmiotowe), informujące o BD oferowanych przez dany serwis, np.: DIALINDEX w serwisie DIALOG, Database Index (DBI) w SDC, CROSS w BRS /30/. Z tych materiałów pomocniczych, bardzo przydatnych podczas wyboru BD, kilka jest dostępnych także i polskiemu użytkownikowi: skomputeryzowana, uaktualniona wersja katalogu światowych BD /9/, udostępniana w serwisie DIALOG (BRIO LIS); DIALINDEX, rejestr BD serwisu DIALOG; maszynowa wersja katalogu Cuadra Associates /17/, udostępniana w serwisie Questel (Data-France).

2.2. Kryteria przydatności baz danych

Jak wspomniano, w miarę stałego zwiększania się liczby BD coraz trudniejsze się staje dokonywanie wyboru potrzebnych BD. Jeszcze kilkanaście lat temu, gdy na całym świecie udostępniano wszystkiego dwieście czy trzysta BD, w większości - bibliograficznych, wybór odpowiednich BD nie nastęczał trudności. Obecnie natomiast istnieje już tak dużo BD i tak znacznie są one zróżnicowane, że użytkownik, zainteresowany określoną tematyką, staje

wobec (niekiedy nawet kłopotliwego) nadmiaru BD, gdyż może wybierać wśród kilku, kilkudziesięciu, a bywa, że nawet i wśród kilkuset BD. Tak więc, coraz częściej mówi się o "zakłopotaniu (konfuzji) użytkownika" (user confusion - określenie prof. M.E. Williams, cytowanej w /56/), wywołanym także i mnogością oraz zróżnicowaniem: pakietów programów do obsługi BD, serwisów online i serwisów dostarczających BD na CD-ROM.

Nieżej przedstawimy dziewięć zasadniczych elementów opisu BD, które będą rozpatrywane jako kryteria oceny przydatności i porównywania BD /67/.

Zakres tematyczny, przedmiotowy bazy danych

Oczywiście, najważniejsze kryterium wyboru BD stanowi kryterium tematyczne. Przede wszystkim należy więc starannie i precyzyjnie określić zakres tematyczny (przedmiotowy) interesującej nas dziedziny i zgodnie z tymi ustaleniami dokonywać przeglądu opisów BD.

Trzeba podkreślić, że nie jest możliwe dokonanie wyboru BD tylko na podstawie nazw BD, gdyż: po pierwsze, nazwy BD zazwyczaj nie mają dostatecznej wartości informacyjnej, np.: ASI, ZDE, AFEE, ALICE, AMEBEUS, ARGUS, RESEARCH, ARIANE, ASK MR. FREQ, BADICOM, CRONOS-INE, DOBIS, DOMA, ECO, FRIDAY MEMO, STOP, SPO, SPHINX, SCARABEE, SCORPIO, SCIPID, PASSPORT, THE MIDNIGHT TURTLE; po drugie, w nazwie BD mogą się pojawić słowa, nie mające zbyt wiele wspólnego z zawartością BD. Jeżeli np. przeszukamy mechanicznie tylko nazwy BD z katalogu COMPUTER-READABLE DATABASES pod kątem występowania słowa business, to otrzymamy także i bazę APPETITE FOR BUSINESS: NEW YDRK, zawierającą informacje o restauracjach

Nowego Jorku, być może bardzo przydatną biznesmenom z krajów rozwiniętych ekonomicznie, ale nam na razie raczej mało potrzebną.

Także kategorie, hasła przedmiotowe, deskryptory, itp., przyjęte w katalogach w celu przedstawienia zawartości BD, nie zawsze wystarczają do określenia przydatności BD. Na przykład, zakres przedmiotowy bazy ABI/INFORM (Abstracted Business Information/Inform) w różnych katalogach opisany jest następującymi hasłami:

katalog Cuadra Associates /16/:

Subject:

Administration and Management; Business and Industry.

katalog serwisu DIALOG /15/:

DIALINDEX Categories:

LABOR = LABOR; MANAGE = MANAGEMENT; TAXACCT = TAX AND ACCOUNTING.

katalog serwisu ESA-IRS /19/:

Main QUESTINDEX Categories:

Administration and management; Business and economics.

katalog serwisu ORBIT SEARCH SERVICE /11/:

Description:

Business, management, and industry including: accounting, advertising, banking, data processing, economics, employee benefits, and more.

baza COMPUTER-READABLE DATABASES:

Descriptors:

Management; Business and business administration; Accounting; Banking and finance; Economics; Electronic data processing; Real estate; Public administration; Advertising and marketing; Telecommunication.

monografia J.L. Halla: On-line information
retrieval sourcebook /25/:

Subject field:

Business, finance, management.

opracowanie J.L. Stauda: Online
Wirtschaftsdatenbanken /77/:

Descriptors:

Business, management.

Z przytoczonych hasłowych opisów bazy ABI/INFORM wynika, że hasła przyporządkowane tej BD w różnych katalogach różnią się dość znacznie, mimo wielu podobieństw. Jeżeli interesuje nas np. wyłącznie hasło public administration, to opisuje ono bazę ABI/INFORM tylko w jednym z siedmiu podanych wyżej przykładowo opisów.

Dopiero zapoznanie się z tekstowymi opisami zawartości BD w różnych katalogach (różnice w tych opisach, a zwłaszcza w szczególności potraktowania problemu, także bywają niekiedy znaczne), a przede wszystkim - z przykładowymi wydrukami rekordów BD pozwala zorientować się dość dobrze w rzeczywistej zawartości tematycznej rozpatrywanej BD.

W odniesieniu do bibliograficznych BD dużą pomoc w ustaleniu zakresu przedmiotowego BD stanowią wykazy tytułów czasopism, dokumentowanych na użytek danej BD. Warto zwrócić uwagę na informację, czy uwzględniane są wszystkie artykuły z interesującego nas czasopisma, czy też tylko niektóre.

Porównując różne bibliograficzne BD o zbliżonej tematyce, należał przeanalizować problem ewentualnego częściowego powtarzania się tych samych informacji (dokumentowania tych samych dokumentów) w BD, por. np. /64,83/.

Przy porównywaniu zakresu tematycznego różnych BD bardzo pomocne są opracowania, w których omówiono BD z poszczególnych dziedzin, np. /1,2,8,14,23,33,45,46,48,49,50,51,71,73,74,75,77,78,83/.

Typ bazy danych

Bardzo ważnym kryterium doboru BD jest kryterium typologiczne. Wybierając BD trzeba więc ustalić, jakiego typu BD są najbardziej przydatne w danej sytuacji. Przeglądając opisy BD, należy najwięcej uwagi poświęcić BD wybranego typu, nie pomijając jednak całkowicie bez innych typów.

Żałujemy, że mamy dokonać wyboru jednej z dwóch BD o tej samej tematyce, przy czym jedna z nich jest bazą numeryczną, druga - bibliograficzną. W tej sytuacji trzeba wiedzieć, które z informacji obu wskazanych typów są bardziej potrzebne naszym użytkownikom. Im większe i bardziej zróżnicowane jest to grono, tym bardziej przydatna (często wykorzystywana) będzie baza bibliograficzna (zwłaszcza w Polsce, wobec braku tradycji korzystania z BD innych typów). Ale i od tej reguły mogą być wyjątki, np. na pewno bardzo potrzebne licznym użytkownikom są numeryczne BD, zawierające wskaźniki ekonomiczno-statystyczne światowej gospodarki. Należy także pamiętać, że BD obu wskazanych typów do optymalnego ich wykorzystania wymagają oprogramowania (strategii wyszukiwania) o różnych własnościach, przystosowanego do specyfiki obsługi BD danego typu.

Określając typy potrzebnych BD, warto pamiętać, że pełnotekstowe BD udostępniają od razu pełne teksty dokumentów. Natomiast eksploatacja zachodnich bibliograficznych BD wymagać będzie rozwiązania, przynajmniej częściowego, problemu dostępu do zachodnich dokumentów pierwotnych, dokumentowanych w tych BD.

Dla przykładu można wspomnieć, że w dziedzinie biznesu i ekonomii za ważniejsze często uznaje się bazy źródłowe. Zresztą wielka liczba źródłowych BD dotyczy ekonomiki, finansów i marketingu i w tych dziedzinach najczęściej pojawiają się nowe BD. W związku z dużą dynamiką "wewnętrznej" informacji firmowej, bazy te wymagają częstej aktualizacji. Bazy danych zawierające informacje, dotyczące koniunktur ekonomicznych, poddawane są aktualizacji nawet i kilka razy dziennie. Analiza wykorzystywania BD przez użytkowników z przemysłu wykazała, że najbardziej potrzebne są BD, gromadzące wskaźniki statystyczne, informatory i streszczenia literatury. Bardzo cenione są też te pełnotekstowe BD, które zawierają nowości techniczno-ekonomiczne wraz z ich oceną, dokonaną przez ekspertów /74/.

Należy również wspomnieć o różnicach w typologiach BD - niekiedy ta sama nazwa oznacza u różnych autorów różne typy BD. Tak więc, każdorazowo należy przede wszystkim zapoznać się z typologiami BD, przyjętymi w rozpatrywanych opracowaniach.

Zakres terytorialny baz danych.

Wiele BD można określić jako lokalne, dotyczące jednego kraju, regionu, miasta, dzielnicy, np.: bazy ABI/INFORM i AGRICOLA mają zakres międzynarodowy; baza AFRICA NEWS dotyczy jednego kontynentu (Afryki); A-V ONLINE (dawniej znana jako NICEM) dotyczy USA; CRONOS - Hiszpanii; amerykańska baza ARTHUR D. LITTLE/ONLINE zawiera przede wszystkim informacje o USA, ale uzupełniająco także i informacje o innych krajach; ARIZONA BUSINESS GAZETTE dotyczy USA, głównie Arizony; AZURTEL - Francji, głównie rejonu Marsylii; LAS VEGAS HOTLINE - Las Vegas i Nevady; LOS ANGELES TIMES - USA, głównie Los Angeles i Kalifornii.

Zastanawiając się nad przydatnością lokalnych BD dla naszego systemu, trzeba rozstrzygnąć, które (którego kraju, regionu itp.) rozwiązania mogą być interesujące dla polskiego użytkownika, i te właśnie bazy uwzględnić, planując dalszy rozwój polskiego systemu.

Zakres chronologiczny

Przy analizowaniu różnych BD bardzo pomocna jest informacja o ich zakresie chronologicznym. Warto pamiętać, że rok, w którym zaczęto produkować daną BD, nie musi wyznaczać początku zakresu chronologicznego tej bazy - często wprowadzane są do BD także i informacje starsze. Oceniając BD pod kątem tego kryterium, musimy wiedzieć, jakie są zainteresowania naszych użytkowników: czy zależy im przede wszystkim na informacjach najnowszych (np. wyłącznie informacje bibliograficzne z bieżącego roku) czy też warto zapłacić więcej, ale pozyskać także i informacje z lat poprzednich.

Producent bazy danych

Warto zwrócić uwagę na informacje o producencie BD, wiedzieć, czy firma ta cieszy się uznaniem w środowisku fachowców, czy też jest to działająca od dawna firma, o której dotychczas nie słyszał nikt ze specjalistów naszej instytucji. Należy też wziąć pod uwagę lokalizację (kraj) producenta BD, np. stwierdzono /74/, że ważną cechą ekonomicznych BD jest ich specyficzna orientacja na określone kategorie użytkowników: przygotowywane we Francji dane o firmach francuskich są bardziej przydatne użytkownikom francuskim, niż np. specjalistom brytyjskim, zainteresowanym francuskimi firmami.

Język opisu informacji w bazie danych

W naszym kraju najpopularniejszym językiem zachodnim jest angielski. Tak więc, większości użytkowników naszego systemu najłatwiej będzie się posługiwać BD, opracowywanymi w tym języku. Z pozostałych języków - bardziej przydatne są BD, przygotowywane w którymś z europejskich języków kongresowych, niż np. baza prowadzona po fińsku czy po japońsku. Jednak kryterium językowe nie zawsze ma moc rozstrzygającą, jeżeli bowiem uznamy, że konkretna BD jest merytorycznie bardzo cenna, to nie można z niej od razu zrezygnować, tylko ze względu na język. Warto także pamiętać, że rozpatrywane kryterium ma dużo większe znaczenie w przypadku pełnotekstowych BD, niż np. w odniesieniu do statystycznej BD, gromadzącej szereg czasowy. Wówczas istnieje jedynie konieczność szybkiego tłumaczenia stosunkowo niewielkich fragmentów (tytuły tabel, opisy wierszy i kolumn, ewentualne komentarze tekstowe do poszczególnych tablic), a większość informacji w tej BD przedstawiona jest liczbowo.

Wyżej zwrócono uwagę na język naturalny (etniczny). Jednak problem języka opisu informacji w BD nie sprowadza się tylko do kwestii używanego w niej języka naturalnego. Analizując przydatność rozpatrywanej BD należy poznać sposób przedstawienia informacji w tej BD. W BD stosowane są różne języki informacyjne. Na przykład, w odniesieniu do bibliograficznej BD trzeba ustalić: jaki jest format rekordu oraz zawartość poszczególnych pól i podpól; jak opisano treść dokumentu (np.: język deskryptorowy, język słów kluczowych; język haseł przedmiotowych). Reasumując, podczas analizy i oceny BD należy uwzględnić wszystkie środki językowe, przyjęte w tej BD do opisu informacji. Można wspomnieć, że dane

o środkach językowych przytoczone są w opisach BD w omówionym wyżej katalogu /9/, przygotowanym pod kierunkiem prof. M.E. Williams.

Ograniczenia w korzystaniu z baz danych

Przy analizowaniu opisów BD trzeba także zwrócić uwagę na ewentualne ograniczenia, nakładane przez dysponenta BD na jej wykorzystanie. Nie wszystkie BD mogą być udostępniane wszystkim zgłaszającym się użytkownikom (por. np. /5,42,66/). Na rozpowszechnianie niektórych BD nałożone są ograniczenia, wynikające z konieczności ochrony pewnych informacji przed niepożądanym dostępem. I tak, np. ograniczenia takie dotyczyły 13 BD z ponad 220 BD oferowanych w Polsce przez serwis DIALOG w 1985 r., w tym bazy: DOE ENERGY, FOREIGN TRADERS INDEX, FEDERAL RESEARCH IN PROGRESS czy TRADE OPPORTUNITIES mogły być udostępniane wyłącznie użytkownikom z USA /68/. Inny przykład - w 1990 r. z bazy DOE ENERGY mogą korzystać jedynie użytkownicy z USA i jedenastu innych krajów /15/. Takie BD trzeba albo od razu wykluczyć, albo też porozumieć się z dysponentem BD (ograniczenia na udostępnianie BD mogą ulegać zmianom, złagodzeniom).

Ograniczenia w uzyskaniu dostępu do niektórych BD spowodowane są koniecznością ochrony określonych informacji (np. czasem w opisie BD figuruje zastrzeżenie, że baza ta może być udostępniana wyłącznie członkom określonego stowarzyszenia, związku itp.). Natomiast inne ograniczenia w wykorzystaniu BD wynikają z przyczyn handlowych, komercyjnych. Serwisy online i dystrybutorzy BD na nośnikach maszynowych prowadzą różne polityki rynkowe. I tak, jedni dysponenti BD wprowadzają wiele ograniczeń na wykorzystanie BD, np. serwis DIALOG w warunkach umów zastrzega, że użytkownik tego serwisu nie ma prawa kopiować uzyskanych wyników wyszukiwania na nośnik

maszynowy w celu ich przechowywania. Na nośnik maszynowy wolno te rezultaty skopiować wyłącznie tymczasowo, tj. na czas potrzebny do sporządzenia kopii na papierze. Dane uzyskane z serwisu DIALOG (niezależnie od ich formy) nie mogą być przekazywane, sprzedawane, ani w żaden inny sposób udostępniane komercyjnie, z wyjątkiem układu broker - klient czy biblioteka - czytelnik/użytkownik. Zgodę na odstępstwo od tych zasad trzeba każdorazowo uzyskać od producenta BD lub od serwisu DIALOG. W niektórych umowach z dystrybutorami BD bardzo precyzyjnie określone są ograniczenia, których musi przestrzegać użytkownik, np. czy BD może służyć do utworzenia własnej, lokalnej BD; czy jest możliwe RWI; czy można przekazywać wyniki wyszukiwania innym instytucjom.

Natomiast inni dystrybutorzy BD wprowadzają w umowach mało zastrzeżeń, np. zaznaczają jedynie, że nie wolno przekazywać innym użytkownikom BD na CD-ROM lub jej kopii (w całości czy fragmencie) na innym nośniku. Innymi słowy, w miarę upływu czasu ograniczenia możliwości wykorzystania BD na nośnikach maszynowych ulegają złagodzeniom. Często dystrybutor BD na CD-ROM zezwala na dowolne korzystanie z BD, oczywiście z uwzględnieniem zwyczajowych ograniczeń, takich jak przy każdej prenumeracie dzieła chronionego prawem autorskim.

Wbrew pozorom, na omawiany element charakterystyki BD warto zwracać baczną uwagę. Na przykład, jeżeli chcemy z dwóch BD wybrać jedną (gdyż sądzimy, że każda z nich z osobną zaspokoi nasze potrzeby informacyjne), to lepiej, oczywiście, zdecydować się na tę bazę, na której wykorzystania nałożone są mniejsze ograniczenia.

Wielkość bazy danych, częstotliwość aktualizacji, liczba każdorazowo dodawanych nowych rekordów

Informacje ilościowe o BD są niekiedy bardzo przydatne. Warto np. zdawać sobie sprawę, jak wielki (liczba rekordów) jest światowy zasób informacji, które mogłyby zainteresować użytkowników naszego systemu. Oprócz tego, w zależności od: wielkości BD, częstotliwości jej aktualizacji i liczby każdorazowo dodawanych rekordów planuje się parametry informatyczne naszego systemu. Niestety, informacje ilościowe nie zawsze są kompletne (brak danych) i aktualne. Oczywiście, w przypadku zainteresowania się konkretną BD, wszystkie informacje o niej można uzyskać od producenta. Warto zwrócić uwagę na częstotliwość aktualizacji BD - nie ma sensu sprowadzać do własnej eksploatacji BD, która jest aktualizowana kilka razy dziennie.

Przypomnijmy, że BD o wielkości do 100 000 rekordów są uznane za małe, od 100 000 do 1 mln - za średnie, ponad 1 mln - za wielkie /26/.

Cena bazy danych na nośniku maszynowym, cena na poszczególne usługi serwisu online

Warto zwracać baczność uwagę na ceny prenumeraty BD i na ceny poszczególnych usług w różnych serwisach online. Dla użytkowników z Zachodu ceny te są umiarkowane, dla nas natomiast - ciągle jeszcze wysokie (np. średnio za roczną prenumeratę BD na CD-ROM płaci się od 2 do 4 tysięcy USD). Nie należy jednak z góry mechanicznie zakładać, że zawsze wybierać będziemy najtańszą BD. Każdorazowo trzeba sprawdzić, jakie informacje dostarczane są za wskazaną cenę.

Zastanawiając się nad kosztami wykorzystania BD, warto także spróbować ustalić, czy na planowane przez nas pytania możemy uzyskać te same (lub prawie takie same) odpowiedzi z różnych BD. Jeżeli np. stwierdzimy, że nasze potrzeby informacyjne może zaspokoić zarówno baza ERIC, jak i baza SOCIAL SCISEARCH, to warto przy wyborze BD uwzględnić fakt, że godzina korzystania z pierwszej z wymienionych BD kosztuje obecnie w serwisie DIALOG 30 USD, a z drugiej - 120 USD.

Problemowi porównania kosztów korzystania z BD poświęcono wiele uwagi w literaturze przedmiotu, na przykład /34,37,39,40, 55,56,66/.

Wyżej wskazaliśmy kryteria oceny i porównywania BD. Uzupełniając przytoczymy kryteria wyboru BD, zaproponowane w /34/:

- zakres przedmiotowy BD,
- wszechstronność (comprehensiveness) BD,
- aktualność BD,
- częściowe pokrywanie się zakresów przedmiotowych różnych BD,
- elementy składowe rekordu,
- przyjęta metoda indeksowania dokumentów,
- dokumenty uwzględniane w BD,
- koszt BD.

Warto jednak zwrócić uwagę na następującą okoliczność: wprawdzie B. Houghton i J. Convey /34/ nie zaznaczają, że przedstawiony przez nich zestaw kryteriów odnosi się do bibliograficznych BD, ale z tekstu wyraźnie wynika, że Autorzy uwzględnili tylko bibliograficzne BD. Na przykład, w omówieniu kryterium zakres przedmiotowy BD stwierdzono, że konieczne jest przede wszystkim ustalenie, czy rozpatrywana BD jest: dziedzinowa (ukierunkowana na jedną dziedzinę), jak np. CA SEARCH; problemowa (ukierunko-

wana na jeden problem), jak np. TOXLINE czy ENVIROLINE; ukierunkowana na jedno zadanie, misję, jak np. NASA; wielod dziedzinowa, jak np. SCISEARCH czy NTIS.

Na zakończenie należy podkreślić, że każdy wybór (a więc także i wybór BD) ma charakter subiektywny. Najbardziej obiektywne kryterium przydatności bazy stanowi, niewątpliwie, zakres tematyczny, ale i przy posługiwaniu się tym kryterium pojawiają się niekiedy wątpliwości (może się okazać, że opisy zakresów tematycznych danej BD w jednym czy nawet w kilku katalogach nie wystarczają do ustalenia, czy ta BD zawiera informacje z interesującej nas bardzo wąskiej dziedziny). Do grupy kryteriów obiektywnych można także zaliczyć: typ, zakres terytorialny, zakres chronologiczny, częstotliwość aktualizacji i charakterystykę ilościową. Warto jednak pamiętać, że nawet rozstrzygnięcia, dokonane na podstawie tych kryteriów, nie zawsze są oczywiste. Zilustrujemy omawiany problem następującym przykładem.

Załóżmy, że musimy wybrać jedną z dwóch BD tego samego typu i o tej samej tematyce (hasła przedmiotowe w opisie zakresu tematycznego są jednakowe). W takiej sytuacji nie można zawsze mechanicznie przyjmować, że większa BD jest bardziej przydatna. Przede wszystkim trzeba ustalić, jakie były kryteria doboru materiału przy tworzeniu każdej z tych baz, i sprawdzić zakres chronologiczny obu baz. Jeżeli zasady tworzenia baz były te same, a większa z tych BD jest także bazą o większym zakresie chronologicznym, to bardziej sensowne wydaje się wybranie tej właśnie BD. Ale może ona także być droższa od bazy mniejszej, dlatego trzeba jeszcze ustalić, czy użytkowników naszego systemu interesują także i informacje starsze, dawniejsze, czy też są zainteresowani wyłącznie najnowszymi informacjami. Dodatkowo należy uwzględnić częstotliwość aktualizacji BD.

Jak wspomniano, rozpatrując dwie BD o różnych wielkościach, należałoby próbować określić jakość tych baz, a więc np. ustalić, na ile słusznie, naszym zdaniem, dokonywany jest wybór informacji, wprowadzanych do każdej z tych baz. Może się bowiem okazać, że w przypadku większej BD zastosowano mniej ostre (bardziej ogólne) kryteria selekcji. W takiej sytuacji trzeba zdecydować, co jest bardziej potrzebne użytkownikom naszego systemu: czy BD, zawierająca tylko informacje ściśle odpowiadające zadanej tematyce, czy też BD, dzięki której można także i szerzej rozpatrywać tematy, licząc się przy tym z możliwością uzyskania wielu informacji zbędnych lub mało przydatnych. Najtrudniejsza jest ocena jakości dwóch BD o tych samych parametrach. Aby tę kwestię uczciwie rozstrzygnąć, konieczne jest sprawdzenie obu baz dostateczną liczbą tych samych pytań, co wymaga równoczesnego korzystania z obu baz przez pewien okres. Uzupełniającą trzeba także rozważyć możliwości posiadanego lub projektowanego oprogramowania (jak i na ile szczegółowo można będzie przeszukiwać poszczególne rekordy BD), koszt eksploatacji obu BD, przewidywaną częstotliwość wykorzystania BD, możliwości finansowe naszej instytucji itp.

Przy wyborze jednej z dwóch BD o tej samej tematyce warto przejrzeć literaturę przedmiotu i poszukać informacji o obu BD. Trzeba przy tym pamiętać, że na ogół dużo częściej wzmiankowana jest w literaturze "stara" (tj. istniejąca od dawna), wielka, bibliograficzna BD, niż nowa, mała, bardzo specjalistyczna baza faktograficzna. Często przyjmuje się, że o jakości BD decyduje np. liczba serwisów online, udostępniających tę BD. Warto jednak pamiętać, że na program rozwoju serwisu mają wpływ różne, także i ekonomiczne, czynniki i dlatego nie zawsze obiektywnie doskonała merytoryczna ocena danej BD decyduje o włączeniu jej do grupy BD,

obsługiwanych przez dany serwis. Oczywiście, duża liczba serwisów online, oferujących określoną BD, świadczy o dużym zapotrzebowaniu na tę właśnie bazę.

2.3. Wieloetapowy proces wyboru baz danych

Pracę nad wyborem BD najlepiej przeprowadzać w kilku etapach (co najmniej - w dwóch), na każdym etapie uszczegółowiając wymagania w odniesieniu do niezbędnych BD. Tak więc, na początek trzeba przygotować wstępne zestawienie BD, tematycznie i typologicznie nam przydatnych, a następnie stopniowo dokonywać selekcji tego zestawienia, wieloaspektowo i szczegółowo analizując wybrane wstępnie BD, i próbując określić stopień ich przydatności w naszym systemie informacyjnym.

Wstępne przeglądanie zgromadzonych materiałów pod kątem wybrania potrzebnych BD warto przeprowadzać bardzo "szeroko", tzn. tak, by nie pominąć żadnej z użytecznych BD. W ten sposób otrzymany, oczywiście, wiele zbędnych BD (które można będzie potem, w następnych etapach pracy, szybko wyeliminować) i dużo BD obecnie nam nieprzydatnych lub mało przydatnych, ale będziemy mieli świadomość, że żadna ważna BD nie umknęła naszej uwadze. Ponadto, zyskamy pełny obraz sytuacji w dziedzinie BD dotyczących interesującej nas tematyki.

Przedstawiona zasada postępowania jest szczególnie istotna w sytuacji, gdy wykorzystujemy (na tym etapie wyboru BD) system skomputeryzowany. Pierwsze pytanie wyszukiwawcze warto wówczas sformułować dostatecznie ogólnie, tak by uzyskana odpowiedź raczej przeszyła nadmiarem, niż brakami. To pragmatyczne podejście jest niewątpliwie słuszne - łatwiej pominąć niektóre z baz, wyszukanych przez system, niż martwić się, że jakieś ważne dla nas BD nie

zostały uwzględnione w odpowiedzi systemu. Na przykład, w wyniku wstępnego wyszukiwania zachodnich BD, użytecznych dla polskiego systemu OZAGA, system DIALOG wygenerował zestawienie ok. 1000 BD. Z tego zestawienia wybrano początkowo ok. 500 BD, a następnie, kolejno: kilkadziesiąt, kilkanaście i ostatecznie kilka BD. Tak więc, w wyniku uzyskano kilka BD, uznanych wówczas (grudzień 1989 r.) za najbardziej potrzebne systemowi OZAGA, i zestawienie ok. 500 zachodnich BD, tematycznie zbliżonych do systemu OZAGA.

Na każdym etapie wybierania BD zwracać więc trzeba uwagę na istotne elementy charakterystyki BD (zob. 2.2.) i na wszystkie parametry naszego (istniejącego lub planowanego) systemu informacyjnego (informacyjno-wyszukiwawczego).

Uzupełniającą warto zauważyć, że analizując opisy BD pod kątem przydatności użytkownikom naszego systemu, trzeba dla każdej BD rozróżnić dwa przypadki. Po pierwsze, trzeba ustalić, czy ta baza jest potrzebna w naszym systemie, tzn. czy zawarte w niej informacje są niezbędne polskiemu użytkownikowi; po drugie, warto się zastanowić, czy ta baza, zawierająca np. informacje, dotyczące innych krajów, stosunkowo rzadko przydatne polskiemu specjalście, może być wykorzystana jako wzór do stworzenia podobnej polskiej BD, tzn. czy według schematu tej bazy nie warto budować jej polskiego odpowiednika.

Załóżmy, że uwzględniając wszystkie informacje o BD, dokonaliśmy wstępnego wyboru BD, tj. wybraliśmy grupę BD, przydatnych w naszym systemie. Teraz należy ustalić, czy wszystkie z wybranych BD są rzeczywiście potrzebne, a następnie zdecydować, czy można w wyselekcjonowanej grupie BD wyodrębnić podgrupę BD bardzo potrzebnych, niezbędnych.

Innymi słowami, na początku z bogatej oferty zachodnich producentów BD wybraliśmy grupę BD, które uznaliśmy za przydatne w naszym systemie informacyjnym. Następnie dokonaliśmy selekcji BD, wieloaspektowo i dokładnie analizując wybrane BD i uwzględniając znane nam wymagania użytkowników naszego systemu. W wyniku tej selekcji otrzymaliśmy grupę BD bardzo potrzebnych, niezbędnych, a pozostałe z wybranych na początku BD zaliczyliśmy do grupy BD mniej potrzebnych, albo - na razie - zbędnych.

Następnie należy przejrzeć literaturę na temat wybranych BD, zwracając uwagę na pokrywanie się zakresów BD (powtarzanie się tych samych informacji w różnych BD), tempo aktualizacji (jak szybko nowe informacje są wprowadzane do BD), analizy porównawcze i oceny wybranych BD.

Można też próbować dotrzeć do dotychczasowych użytkowników wybranych przez nas wstępnie BD i uwzględnić ich doświadczenia.

Warto zwrócić szczególną uwagę na nierozłączne BD, to jest na BD o częściowo pokrywających się zakresach. Trzeba specjalnie podkreślić, że ta sama informacja może być (i zazwyczaj bywa) różnie przedstawiona w różnych BD. Dotyczy to nawet bibliograficznych BD: ten sam dokument jest różnie opisany (różnice w wyborze elementów opisu, różny stopień szczegółowości potraktowania poszczególnych elementów opisu, różne pola przeznaczone do wyszukiwania i do sortowania, różne formaty wydruków). Zwłaszcza streszczenia dokumentów w różnych BD mogą się znacznie różnić między sobą, czasami więc może się nawet i okazać przydatne korzystanie np. z różnych nierozłącznych bibliograficznych BD /56/.

Przede wszystkim jednak celowe jest przeprowadzenie ankiety wśród użytkowników naszego systemu, tzn. warto przedstawić im proponowany zestaw BD i poprosić o określenie stopnia przydatności BD

(np. w skali: bardzo potrzebna, przydatna, mało przydatna, nieprzydatna).

Na tym etapie prac nad doбором BD dla naszego systemu można już także zacząć wstępne rozmowy (korespondencję) z dysponentami wytypowanych BD, w celu uściślenia niezbędnych nam informacji o BD i o ewentualnych warunkach eksploatacji BD w naszej instytucji.

Przypomnijmy, że od liczby i częstotliwości aktualizacji wybranych BD zależą rozwiązania organizacyjne:

- W sytuacji, gdy grupa bardzo potrzebnych BD jest stosunkowo duża, a wszystkie te BD będą u nas w miarę równomiernie wykorzystywane, lepiej korzystać z pośrednictwa serwisu online. Tak więc, należy przeprowadzić rozpoznanie ofert serwisów online, utrzymujących interesujące nas BD, i zacząć wstępne pertraktacje z wybranymi serwisami.
- W wypadku wyboru małej liczby BD, które nie podlegają bardzo częstej aktualizacji - warto je sprowadzić do własnej eksploatacji, tak więc na tym etapie prac trzeba nawiązać kontakt z dysponentami wybranych BD (nb. dysponent może, ale nie musi być równocześnie producentem BD), ustalić, na jakich nośnikach BD może być dostarczana, jakie są warunki uzyskania BD. Można też próbować oferować własny, polski wkład do BD, w zamian za bezpłatne otrzymywanie BD lub za obniżenie kosztów prenumeraty. Jeżeli jednak wybraliśmy mało BD, lecz są to często aktualizowane BD, to musimy z nich korzystać poprzez serwis online.
- Gdy zdecydowaliśmy się na wariant łączony, kombinowany (kilka BD do własnego przetwarzania, większość - do wykorzystania online), trzeba pamiętać, że jest to rozwiązanie optymalne, najbardziej ekonomicznie opłacalne, gwarantujące pełny dostęp do światowych zasobów informacji w naszej dziedzinie, ale - niestety - jak na nasze warunki, i kosztowne, i skomplikowane.

3. WYBÓR SERWISÓW ONLINE

Jeżeli zdecydowaliśmy się na skorzystanie z pośrednictwa serwisu online, a okazało się, że interesujące nas BD są dostępne w kilku serwisach, to musimy wybrać najbardziej nam odpowiadający serwis. Warto pamiętać, że oferty poszczególnych serwisów różnią się znacznie: stosowanymi algorytmami przeszukiwania BD, możliwościami prezentowania wydruków, cenami, itd. Należy więc porównać warunki udostępniania interesującej nas BD we wszystkich serwisach, które obsługują rynek europejski i oferują tę (te) BD.

Problem wyboru (z kilku serwisów online, utrzymujących tę samą BD) jednego serwisu, z którego usług będziemy korzystać, wymaga szczególnej uwagi. Panuje dość powszechne przekonanie, że wybór serwisu online jest trudniejszy w odniesieniu do źródłowych BD: oceniono, że dostęp do ok. 80-90% wszystkich informacji, zgromadzonych w odsyłających (tj. w bibliograficznych i skierowujących) BD zapewnia korzystanie z ok. 5 serwisów online, podczas gdy uzyskanie 80-90% informacji ze światowych źródłowych BD wymaga korzystania z pośrednictwa ok. 15 serwisów. Ta ocena, przytoczona w /56/, odnosi się do 1981 r. Obecnie, wobec zwiększenia się liczby BD, wymienione liczby (5 i 15) są już zbyt małe, ale na pewno nadal dużo więcej bibliograficznych BD oferowanych jest przez kilka serwisów, niż ma to miejsce w wypadku np. numerycznych BD: żaden serwis online, obsługujący niebibliograficzne BD, nie udostępnia bardzo wielu BD (w dziedzinie niebibliograficznych BD nie można skorzystać z usług takiego "supermarketu", jakim dla informacji bibliograficznych z różnych dziedzin tematycznych jest serwis DIALOG). Użytkownik wielu (kilku czy kilkunastu) niebibliograficznych (a zwłaszcza numerycznych) BD ma zazwyczaj do czynienia z różnymi: serwisami o bardzo wąskiej specjalizacji, systemami

i strategiami wyszukiwania, warunkami umowy na korzystanie z serwisu. Rynek niebibliograficznych BD jest bardzo specjalistyczny, a ceny na te BD są wysokie /30/. Reasumując, statystycznie użytkownik, zainteresowany jedną, konkretną, bibliograficzną BD, ma na ogół do wyboru więcej serwisów online, niż użytkownik, zamierzający skorzystać z niebibliograficznej BD. W opublikowanym stosunkowo niedawno (1990 r.) opracowaniu /13/ porównano liczbę BD powtarzających się w różnych serwisach, dane te przytoczono w tab. 2.

tabela 2

Liczba serwisów online udostępniających tę samą BD /13/:
w krajach Wspólnoty Europejskiej;
(uwzględniono 947 BD produkowanych w tych krajach)

Liczba BD	1	2	2	5	15	36	88	800
Liczba serwisów online	11	7	6	5	4	3	2	1

w USA:

Liczba BD	1	2	3	1	6	5	11	22	56	168	184	1604
Liczba serwisów online	13	11	10	9	8	7	6	5	4	3	2	1

Tak więc, z 947 BD krajów Wspólnoty Europejskiej, uwzględnionych w tabeli 2, aż 84% (800 BD) jest oferowanych tylko przez jeden serwis online, a jedynie 2,4% (23 BD) - przez więcej niż trzy serwisy. Dla USA zależności te przedstawiają się następująco: na 2063 uwzględnionych BD: 78% (1604 BD) udostępnianych jest wyłącznie przez jeden serwis, 5,2% (107 BD) - przez więcej niż trzy serwisy.

warto specjalnie uwypuklić następującą kwestię: z faktu, że ta sama BD jest udostępniana np. przez dwa serwisy online, bynajmniej nie wynika, że uzyskane w obu serwisach wyniki wyszukiwania będą identyczne. Różne serwisy stosują bowiem różne formaty wdruków, różne systemy przetwarzania BD, różnie też są kwalifikowane pola, dostępne do wyszukiwania i sortowania itp. Na przykład serwis I.P. SHARP specjalizujący się w obsłudze numerycznych BD, oferuje dużo bogatsze możliwości korzystania z bazy DISCLOSURE II, niż serwis DIALOG /56/.

Wyżej wspomniano o dość rozpowszechnionym mniemaniu, że łatwiej dokonać wyboru serwisu online w wypadku bibliograficznej BD. Z tym poglądem można, oczywiście, polemizować. Wydaje się bowiem, że (wobec wąskiej specjalizacji serwisów obsługujących numeryczne BD i większej liczby serwisów, utrzymujących te same bibliograficzne BD) użytkownik, zainteresowany np. jedną bibliograficzną BD, często może wybierać wśród większej liczby serwisów, a użytkownik, pragnący korzystać z jednej numerycznej BD, ma zazwyczaj do dyspozycji dużo mniej serwisów, ale te serwisy są dużo bardziej zróżnicowane. Tak więc, raczej trudno z góry ustalić, który z dwóch przedstawionych wyżej przykładowo wyborów jest trudniejszy.

Problemi wyboru serwisu online poświęcono wiele uwagi np. w /34,56,67,70/. Niżej przedstawiam zestaw pytań, które warto postawić podczas porównywania i wybierania serwisów online:

- Czy znamy wszystkie serwisy online, które oferują interesujące nas BD?
- Czy liczba kontraktów, zawieranych z serwisami online, ma dla nas istotne znaczenie? To znaczy czy możemy sobie pozwolić na korzystanie z usług kilku serwisów (np. dwie BD w jednym serwisie,

- jedna BD - w drugim serwisie, pięć BD - w trzecim, trzy BD - w czwartym) czy też musimy się zdecydować na jeden serwis (lub najwyżej - na dwa serwisy)?
- Jakie są geograficzne (ewentualnie inne) ograniczenia serwisu?
 - W jakiej walucie należy opłacać faktury?
 - Czy płaci się tylko za efektywne korzystanie z systemu, czy też serwis wymaga jeszcze jakiś (jakich?) opłat dodatkowych?
 - Czy serwis wymaga opłat z góry (awansem)? Jeżeli tak, to czy pierwsza wniesiona opłata stanowi zadatek na poczet przyszłych wyszukiwań, czy też jest to tylko opłata wstępna ("wpisowe") za korzystanie z usług serwisu?
- Czy jest możliwa miesięczna, zryczałtowana, opłata za usługi serwisu, niezależna od liczby przeprowadzonych wyszukiwań?
- Jak są skalkulowane koszty wyszukiwania? Czy płaci się za czas korzystania (cena za godzinę połączenia), czy za liczbę wyświetlonych lub wydrukowanych rekordów (czy pół rekordów), czy za liczbę trafnych odpowiedzi na pytanie?
 - Czy system pokazuje online koszt wyszukiwania? Jeżeli nie - to czy można go samemu łatwo obliczyć?
 - Czy nasze wyszukiwania, realizowane w danym serwisie, będą regularne, czy sporadyczne? A może - zamiast samemu prowadzić wyszukiwanie online - będzie się bardziej opłacało zlecić wyszukiwanie producentowi BD (jeśli istnieje taka możliwość), lub wyspecjalizowanemu pośrednikowi informacyjnemu? (Na Zachodzie istnieje wiele takich instytucji - firm pośredniczących).
 - Czy na podstawie materiałów otrzymywanych z serwisu można zakładać własne, "prywatne" BD? Warto zwrócić szczególną uwagę na ten problem. Można bowiem "oczekiwać, że zachodnie BD będą u nas wykorzystywane nie tylko jako bezpośrednie źródło informacji,

lucz także, a może nawet w przeważającej mierze, jako źródło zasilania polskich BD lub jako źródło wspomagające własne BD^{1/}

- Czy jest możliwe SOI?
- Czy można online zamawiać kopie dokumentów pierwotnych?
- Czy producent BD jest zobowiązany do udzielenia pomocy użytkownikowi serwisu? Czy też uzgodniono z producentem BD, że serwis zapewnia użytkownikowi współpracę w przeszukiwaniu BD?
- Czy producent BD lub serwis online organizują szkolenia użytkowników? W jakim zakresie? Za jaką opłatą?
- Jakie materiały pomocnicze udostępniane są użytkownikowi?
- Czy wprowadzono ograniczenia w udostępnianiu poszczególnych BD? Jakież?
- Czy serwis oferuje także i usługi "śluzowe" (gateway services), tzn. czy pośredniczy w dostępie do innych serwisów?
- Kiedy (w jakich godzinach i w jakich dniach tygodnia) można korzystać z serwisu?
- Czy są stosowane zniżki uzależnione od wybranej pory korzystania z serwisu?
- Czy zakres chronologiczny interesującej nas BD jest taki sam w różnych serwisach?
- Po jakim czasie w różnych serwisach nowa informacja jest wprowadzana do BD?
- Czy łatwy jest proces wyszukiwania? Czy język zapytań jest logiczny, łatwy do opanowania i wygodny w stosowaniu?

^{1/} Z recenzji dr Józefa Oleńskiego opracowania H. Popowskiej: Rozpoznanie zachodnich BD, tematycznie zbliżonych do systemu DZAGA, i określenie ich przydatności w tym systemie. Warszawa, grudzień 1989, maszynopis.

- Czy można prowadzić wyszukiwanie sterowane przez menu (tzw. "jadłospisowe")?
- Czy jest możliwe zapamiętanie strategii wyszukiwania i SDI? Porównaj koszt tych opcji z częstotliwością ich wykorzystania.
- Czy wszystkie (dla nas niezbędne) pola rekordu mogą służyć do wyszukiwania? do sortowania?
- Czy system przetwarzania oferuje możliwości operacji arytmetycznych na polach i rekordach (procesor numeryczny)? Jakie są w ogóle możliwości przetwarzania w systemie? Dla przykładu, porównaj możliwości korzystania z bazy DISCLOSURE II w serwisie CONTROL DATA czy BUSINESS INTERNATIONAL HISTORICAL DATA w serwisie I.P. SHARP z możliwościami korzystania z obu tych BD za pośrednictwem serwisu DIALOG.
- Jakie przewidziano możliwości drukowania (formaty wydruków) odpowiedzi? Czy użytkownik może sam definiować format wydruku, czy też musi wybrać jeden ze standardowych formatów, przyjętych w danym serwisie?
- Czy istnieje możliwość uzyskiwania wydruków offline? Jaki jest ich koszt? Jak szybko są dostarczane?
- Czy informacje, uzyskane w wyniku wyszukiwania, są dokładne i wiarygodne? Zdarza się bowiem, że wydruki zawierają błędy i omyłki, np. w wydruku offline z bazy DISCLOSURE II, o długości mniejszej od dwudziestu wierszy, znaleziono aż siedem błędów /56/; w wydruku offline z bazy DATABASE OF DATABASES, uzyskanym z BRIOLIS w listopadzie 1989 r., pomyłone zostały dane dotyczące trzech BD: IMSPACT, IEIDATA, INSTOOKAB (InstDokAB). Czy te same błędy występują przy wyszukiwaniu z tej BD we wszystkich serwisach? (por. /58/)

- Jak często jest przeprowadzana aktualizacja BD w poszczególnych serwisach? Czy zadeklarowana częstotliwość aktualizacji jest przestrzegana (realizowana w praktyce)?
- Jak wygląda wyświetlanie online i wydruk offline (kompozycja, przejrzystość, użyte środki graficzne)? Wszystkie te elementy mają duże znaczenie dla użytkownika, który skazany jest wyłącznie na korzystanie z możliwości edytorskich, oferowanych przez system (nie może wykorzystać dodatkowo innych edytorów tekstu).
- 1- Na ile różne (treściowo) są odpowiedzi uzyskane z tej samej BD w różnych serwisach? To znaczy - w jakim stopniu są redundantne, a w jakim - uzupełniają się?

Wyżej przedstawiono ponad trzydzieści pytań, które warto zadać przy wybieraniu serwisu (serwisów) online. Oczywiście, nie wszystkie ze wskazanych zagadnień są jednakowo ważne, nie wszystkie więc muszą zawsze być brane pod uwagę. Wybór kryteriów oceny serwisu (ustalenie cech, które są z naszego punktu widzenia najbardziej istotne) każdorazowo zależy od naszych potrzeb.

Na zakończenie należy podkreślić konieczność korzystania z aktualnych informacji o serwisach online. Serwisy często modyfikują oferty, dostosowując je do zmian w sytuacji rynkowej (np. w styczniu 1989 r. uległa całkowitej zmianie polityka cenowa serwisu ESA-IRS /39/). Dlatego przy porównywaniu warunków udostępniania interesujących nas BD w różnych serwisach należy rozporządzać możliwie najnowszymi informacjami o tych serwisach.

W podsumowaniu tego artykułu zwrócimy uwagę na sześć zagadnień.

Wybieranie BD i warunków korzystania z nich jest złożone, pracochłonne, wymaga czasu i pozyskania odpowiednich (szczegółowych, aktualnych) informacji o "podobszarach" rynku BD, odnoszą-

cych się do interesującej nas dziedziny tematycznej. Proces wyboru polega na porównywaniu poszczególnych ofert z wymaganiami, wynikającymi z potrzeb i możliwości grupy użytkowników, dla których dokonujemy tego wyboru. Raczej rzadko się zdarza, by np. dwie BD (czy dwa serwisy online) różniły się tylko jedną cechą (wtedy wybór nie byłby specjalnie trudny). Na ogół trzeba porównywać dwa zestawy cech znacznie zróżnicowanych. Osoba dokonująca wyboru musi więc wiedzieć, jakie parametry BD (czy serwisu) są najważniejsze z punktu widzenia przyszłych użytkowników.

*W tym opracowaniu nie raz i nie bez przyczyny powtórzyło się słowo aktualny. Wypada więc także i w zakończeniu podkreślić, że praca nad wyborem BD czy serwisu online musi być realizowana na podstawie aktualnej analizy rynku. W przeciwnym wypadku po prostu nie ma sensu.

*W literaturze przedmiotu rozróżniane są tzw. historyczne BD i BD udostępniane w trybie czasu rzeczywistego (real-time). Tak samo dzielone są serwisy online. Do pierwszej grupy zaliczono BD, które nie wymagają bardzo częstej aktualizacji, do drugiej - BD podlegające aktualizacji na bieżąco (informacje dotyczące giełdy, finansów, aukcji itp.). Wyżej wspomniano, że do własnego przetwarzania pozyskiwane są historyczne BD, natomiast z bardzo często aktualizowanych BD trzeba korzystać online. Praktyka wykazała jednak, że omawiane różnice muszą się stopniowo zacierać: coraz częściej użytkownik historycznych BD pragnie otrzymać także i najnowsze, dzisiejsze informacje, a klient serwisu, obsługującego BD działające w trybie czasu rzeczywistego, chce odwołać się do jakiś danych starszych, archiwalnych. Dystrybutorzy BD szybko dostosowują się do tych zmian, np. od pewnego czasu w wielu czasopismach firma H.W. WILSON reklamuje swe usługi w ramach serwisu WILSONDISC:

użytkownik, prenumerujący daną BD na CD-ROM, może bezpłatnie (tzn. płacąc tylko za transmisję danych czyli za czas połączenia online) korzystać online z tej BD, co daje mu możliwość szybkiego uzyskiwania najbardziej aktualnych informacji natychmiast po ich wprowadzeniu do BD.

*Bazy danych są towarem, więc podlegają prawom rynku. Innymi słowy, przy wybieraniu BD nie zawsze można ufać reklamowym zapewnieniom producentów BD. Także specjaliści od reklamy, zatrudnieni przez dany serwis online, starają się w możliwie jak najkorzystniej przedstawić warunki, proponowane użytkownikom tego serwisu. W literaturze przedmiotu natomiast pojawiają się zalecenia bardzo dokładnego zapoznawania się z ofertami serwisów. Na przykład, K.M. Nichol /56/ zwraca uwagę na tzw. koszty ukryte: przyszłemu użytkownikowi serwisu może się wydawać, że jakąś usługę dostanie darmo lub prawie darmo - w ofercie jest mała wzmianka o niewielkiej opłacie manipulacyjnej. Ale po kilku miesiącach okazuje się, że te małe kwoty naliczane są regularnie co pewien czas, niezależnie od częstotliwości korzystania z tej usługi, i po pewnym czasie stanowią już dla użytkownika wydatek dużo większy od spodziewanego (zakładanego).

*W tym artykule omówiono dwa zagadnienia, dotyczące BD: problem wyboru BD i problem wyboru serwisu online. Wielokrotnie wspomniano natomiast także i o możliwości pozyskania BD na CD-ROM. Tak więc, należałoby w przyszłości uzupełnić to opracowanie o problematykę wyboru serwisów BD na CD-ROM. Oczywiście, wiele uwag, przedstawionych w 3. części artykułu, można wykorzystać podczas wybierania firm - dystrybutorów BD na CD-ROM.

*Na zakończenie warto poświęcić kilka słów komentarza przytoczonej bibliografii. Nie jest to pełny wykaz literatury, dotyczącej

wszystkich poruszonych w tej sprawie problemów. Na ogół wybierano jedną, dwie prace, ilustrujące omawiane zagadnienie. Zresztą wyczerpująca bibliografia np. tematu BD na CD-ROM nie zmieściłaby się zapewne na wszystkich stronach, zajętych przez ten artykuł. Część wykrzystanych opracowań stanowią publikacje sprzed kilku lat, ale są to na ogół prace uznane za klasyczne w dziedzinie BD (wprawdzie zawarte w nich dane liczbowe dawno się zdezaktualizowały, ale wiele poglądów mimo upływu czasu nie straciło aktualności). Wypada także zaznaczyć, że w spisie literatury uwzględniono dwie pozycje /3,53/, które niestety nie zostały wykorzystane do przygotowania tego artykułu (nie udało mi się jeszcze do nich dotrzeć), ale obie wskazane prace są cytowane w literaturze przedmiotu, tytuły brzmią obiecująco, a monografii /3/ poświęcono pochlebne recenzje (np. w "Aslib Information" z września 1990 r.).

Literatura

1. ALLAN F.C., FERREL W.R. Numeric databases in science and technology: an overview. Database 1989 nr 3 s. 50-58.
2. BATER P., PARKINSON H. Online business and company databases. London: Aslib 1989. - (An Aslib Online Guide).
3. BAWDEN D. User-oriented evaluation of information systems and services. Aldershot: Gower 1990.
4. BAZY danych stran-členov MCNTI po tematike prioritetnyh napravlenij KP NTP: spravocnik. Moskva; MCNTI 1988.
5. BECKER J. Access limitations to US databases for users: a critique of a report to the Commission of the European Communities. Online Rev. 1989 nr 3 s.219-232.
6. BOYCE B.R. Online database search directory. JASIS=Journal of the American Society for Information Science 1986 nr 3 s. 170.

7. BYERLY G. Online searching: a dictionary and bibliographic guide. Littleton; Libraries Unlimited 1983.
8. CHADWICK T.B. International trade information: other business databases online. Database 1990 nr 4 s.26-31.
9. COMPUTER-READABLE databases: a directory and data sourcebook. M.E. Williams (ed.-in-chief). vol. 1-2. Amsterdam;North-Holland 1985.
vol.1. Science, technology, medicine.
vol.2. Business, law, humanities, social sciences.
10. CRONIN B. Databanks. Aslib Proc. 1981 nr 6 s. 243-250.
11. DATABASES at a glance. McLean, VA : ORBIT Search Service 1989.
12. DATABASE directory, 1984-1985. White Plains; Knowledge Industry Publications 1984.
13. DATABASE production. Information Market 1990 nr 62 s.8-9.
14. DEUNETTE J., PINSENT P. Building, construction and architecture databases. London;Aslib 1986. - (An Aslib Online Guide).
15. DIALOG database catalog January 1990. [Palo Alto]; DIALOG Information Services 1990.
16. DIRECTORY of online databases. [R.N. Cuadra (chief ed.)] New York;Cuadra/Elsevier 1988 vol.9 nr 1.
17. DIRECTORY of online databases. [R.N. Cuadra (chief ed.)] New York;Cuadra/Elsevier 1989 vol. 10 nr 3.
18. DIRECTORY of United Nations databases and information systems. 4th ed. New York;United Nations 1990.
19. ESA-IRS directory of databases and services. Frascati;ESA - Information Retrieval Service 1989.
20. EUSIDIC database guide. Oxford;Learned Information 1983.
21. FOX E.A. Optical disks and CD-ROM: publishing and access. W: ARIST=Annual Review of Information Science and Technology vol. 23, 1988 s.85-124.
22. GECHMAN M.C. Machine-readable bibliographic data bases. W: ARIST vol.7, 1972 s.323-378.

23. HAODON A. Management and marketing databases. London:Aslib 1987.
- (An Aslib Online Guide).
24. HALL J.L. Online bibliographic databases: a directory and sourcebook. 4th ed. London:Aslib 1986.
25. HALL J.L. On-line information retrieval sourcebook. London:Aslib 1977.
26. HALL J.L., BROWN M.J. Online bibliographic databases : a directory and sourcebook. 3rd ed. London:Aslib 1983.
27. HANSON T. A survey of European Communities databases. Aslib Proc. 1990 nr 6 s. 171-188.
28. HAWKINS O.T. Online information retrieval bibliography, 1964-1982. Medford: Learned Information 1983.
29. HAWKINS O.T. Online information retrieval reading list. JASIS 1986 nr 4 s.220-221.
30. HAWKINS O.T. Online information retrieval systems. W: ARIST vol.16, 1981 s.171-208.
31. HEARTY J.A., ROHRBAUGH V.K. Current state of full text primary information online with recommendations for the future. Online Rev. 1989 nr 2 s.135-140.
32. HEIMBORGER A. Guide to CO-ROM. Paris:UNESCO 1988.
33. HOUCK M.: A look at the databases directories online. Database 1987 nr 3 s.67, 69-75.
(Infor.EI=Informatika : Ekspress-Informacija 1988 nr 7).
34. HOUGHTON B., CONVEY J. Online information retrieval systems: an introductory manual to principles and practice. 2nd ed., rev. London:Clive Bingley 1984.
35. HOW to access ORBIT Search Service : international telecommunications procedures. McLean ;ORBIT Search Service 1989.
36. HUANG S.T. CD-ROM databases : unique resources for libraries. Access Faxon Summer 1989 vol.2 nr 1 s.6-9.
37. INTERNATIONAL comparative price guide to databases online. Online Rev. 1987 nr 4 s.255-266.
38. IOB directory of United Nations information systems. vol.I-II. Geneva:IOB 1980.

39. JACK R.F. The new ESA/IRS pricing scheme: a comparison with DIALOG. Online 1990 nr 1 s.35-39.
(Infor.EI 1990 nr 29 s.8-11).
40. JÄRVELIN K. An approach to query cost modelling in numeric databases. JASIS 1989 nr 4 s.236-245.
41. KATALOG faktograficeskich baz dannyh, sozdannyh v stranach-
-členach MCNTI. Moskva;MCNTI 1989.
42. LEONARD P.G., SPENDER P.A. Contractual and intellectual property protection of databases. Inf. Serv. and Use 1989 nr 1-2 s. 33-43.
(Infor. EI 1990 nr 21).
43. LUEDKE J.A. Numeric data-bases on-line. On-line Rev. 1977 nr 3 s. 207-215.
44. LUEDKE J.A., KOVACS G.J., FRIED J.B. Numeric data bases and systems. W: ARIST vol. 12, 1977 s.119-181.
45. LYON E. Online medical databases. London;Aslib 1989, - (An Aslib Online Guide).
46. MARCHANT P. Online patents and trademarks databases. London;Aslib 1989. - (An Aslib Online Guide).
47. MARTYN J. Factual databases. Aslib Proc. 1985 nr 5 s. 231-238.
48. MESCHEL S.V. Numeric databases in the sciences. Online Rev.1984 nr 1 s.77-103.
49. MEYER D.E., RUIZ D. End-user selection of databases: part I: science/technology/medicine. Database 1990 nr 3 s.21-29.
50. MEYER D.E., RUIZ D. End-user selection of databases : part II: business/law. Database 1990 nr 4 s.35-42.
51. MEYER E.O., RUIZ D. End-user selection of databases: part IV: people/news/general reference. Database 1990 nr 5 s.65-67.
52. MISCHO W.H., LEE J. End-user searching of bibliographic databases. W: ARIST vol.22, 1987 s.227-263.
53. MORRIS A., TSENG G., NEWHAM G. The selection of online databases and hosts : an expert system approach. W: Proc. 12th Int. Online Inf. Meeting, Oxford;Learned Information 1988 s.139-148.

54. NEUFELD M.L., CORNOG M. Database history: from dinosaurs to compact discs. JASIS 1986 nr 4 s.183-190.
55. NEUFELD M.L., CORNOG M. Secondary information systems and services. W: ARIST vol. 18, 1983 s.151-183.
56. NICHOL K.M.: Database proliferation: implications for librarians. W: Readings in technology. N. Vigiano (ed.). New York; Special Library Association 1984 s. 95-103.
57. NUMERIC databases. C.-c.Chen, P. Herson (eds.). Norwood; Ablex 1984.
58. O'NEILL E.T., VIZINE-GOETZ D. Quality control in online databases. W: ARIST vol. 23, 1988 s.125-156.
59. OJALA M. Views of end-user searching. JASIS 1986 nr 4 s.197-203.
60. OLMSTEAD M., LABRECHE S. DIALOG version 2/QUESTEL Plus: a comparison. Part III: accessing and treating the results. Online 1986 nr 3 s.68-72.
(Infor. EI 1987 nr 2).
61. ONLINE database search services directory. 1st ed. J. Schmittroth, Jr., D. Morris Maxfield (eds.). 2 parts. Detroit; Gale Research Company 1984.
62. PAISLEY W. Old technology, new decade: looking ahead to CD-ROM in the 1990s. Access Faxon Winter 1990 vol.2 nr 2 s.46-48.
63. POLLITT A.S. Information storage and retrieval systems: origin, development and applications. Chichester; Ellis Horwood 1989. - (Ellis Horwood Books in Information Technology).
64. POPOWSKA H. Bibliograficzne bazy danych dostępne w trybie bezpośrednim. Aktual. Probl. Inf. Dok. 1986 nr 3 s. 9-18.
65. POPOWSKA H. Literatura o bazach danych. Aktual. Probl. Inf. Dokl. 1988 nr 2 s.25-30.
66. POPOWSKA H. Warunki korzystania z baz danych. W: [69].
67. POPOWSKA H. Wybór baz danych: wskazówki metodyczne. W: [69].
68. POPOWSKA H. Zachodnie bazy danych w Polsce. Aktual. Probl. Inf. Dok. 1987 nr 1 s.11-22.
69. POPOWSKA H., STEFANIĄK B., WYSOCKI A. Możliwości wykorzystania

i finansowania wybranych zagranicznych źródeł informacji na nośnikach maszynowych. Warszawa; IINTE 1990, maszynopis.

70. PROVIDING access to United Nations databases : a guide for United Nations database producers. Geneva; ACCIS=Advisory Committee for the Co-Ordination of Information Systems 1988.
71. RAPER D. Law databases. London : Aslib 1988. - (An Aslib Online Guide).
72. REBROVA M.P., ENGELHARDT A.V. B "Spravočnike baz danih dostupnih v režime dialoga" (SSA). NII ser. 1 1988 nr 5 s.22-27, 30.
73. RUIZ D., MEYER D.E. End-user selection of databases: part III: social science/arts and humanities. Database 1990 nr 5 s.59-64.
74. SHIELDS M. Business information sources: the growing importance of non-bibliographic online databases. IATUL Quart. 1987 nr 1 s.44-50.
(Infor.EI 1988 nr 11).
75. SIECK S.K. Business information systems and databases. W: ARIST vol. 19, 1984 s.311-327.
76. SMITH D.H. Online government databases : into the maelstrom. Database 1988 nr 3 s.56-62.
77. STAUD J.L. Online Wirtschaftsdatenbanken. 2 Bände. Frankfurt am Main; Peter Lang 1987.
78. STERN D. Artificial intelligence databases : a survey and comparison. Database 1990 nr 4 s. 19-24.
79. SUBJECT catalog of CD-ROM titles : winter 1990. Access Faxon Winter 1990 vol. 2 nr 2 s.7-40.
80. SUOZZI P. By the numbers: an introduction to numeric databases. Database 1987 nr 1 s. 15-22.
(Infor. EI 1987 nr 34).
81. TENOPIR C. Full-text databases. W: ARIST vol. 19, 1984 s.215-246.
82. TENOPIR C., SHU M.E. Magazines in full text: uses and search strategies. Online Rev. 1989 nr 2 s.107-118.
83. WALKER G. Searching the humanities : subject overlap and search vocabulary. Database 1990 nr 5 s.37-46.

84. WÄNGER J., LANDAU R.N. Nonbibliographic on line data base services. JASIS 1980 nr 3 s.171-180.
85. WILLIAMS M.E. Online government databases: an analysis. Online Rev. 1986 nr 4 s.227-236.
86. ZETTELMAÑN A.-M. DatenbankenaktivitÄten in Polen. Nachr. Dok.= Nachrichten für Dokumentation 1987 nr 3 s.131-135.

27.02.1991 r.

THE SELECTION OF THE DATA BASES AND ONLINE SERVICES
(Some methodological advices)

S u m m a r y

There are proposed in the article some methodological advices which might be of use when selecting the data bases and online services. There are presented criterions for evaluating, comparing and selecting of the bases such as : the subject coverage of the data base, type of the data base, the regional extent of the data base, time limits of the data base, data base producer language used in the data base (natural language, information retrieval language), limitations in the use of the data base, size of the file, costs of searching. The author points out that the process of the data bases' selection does not end with the indication of the set of needed information files; such a list should be supplemented with the information about the most efficient way of use of the particular files. Two ways of using data bases are discussed: either through getting the file for processing on user's own, or through online services. There are pointed out the considerable

changes which have occurred recently in the field of the data carriers - more and more data bases have been available on CD-ROM. The instructions concerning the selection of the relevant online service were formulated.

Выбор баз данных и диалоговых сервисов
(Несколько методических указаний)

В статье предложены несколько методические указания пригодные при выборе баз данных и диалоговых сервисов. Представлены критерии оценки, сравнения и выбора баз данных: тематический охват базы данных, тип базы данных, территориальный охват, временный диапазон, производитель баз данных, язык описи информации в базе данных (естественный язык, информационный язык), ограничения в использовании базы данных, количественная характеристика базы данных, оплаты за пользование базой данных. Обращается внимание на то, что процесс выбора баз не кончается в момент приготовления списка нужных баз: такой список должен быть дополнен информацией о наиболее благоприятном способе использования отдельных баз. Охарактеризованы две возможности использования баз данных: путём приобретения базы для собственной переработки или её использование посредством диалогового сервиса. Обращается внимание на большие изменения в области машинных носителей данных: все больше баз делается доступными на CD-ROM. Представлены указания касающиеся выбора соответственного диалогового сервиса.

MATERIAŁY I PRZYCZYNKI

TERESA GAŁCZYŃSKA

Ośrodek Informacji Naukowej PAN

SYSTEM INFORMACJI O UKSZTAŁTOWANIU ŚRODOWISKA - SINUS

Funkcje systemu. Źródła informacji. Zarządzanie bazą danych. Przetwarzanie danych. Udostępnianie informacji w systemie. Możliwości wykorzystania.

Wśród systemów informacyjnych o środowisku przyrodniczym wyróżniają się zdecydowanie systemy informacji geograficznej. Przeznaczone do pracy z danymi przestrzennymi pozwalają na numeryczne modelowanie komponentów środowiska, ich inwentaryzowanie, integrowanie informacji pozyskiwanych z różnych źródeł, analizowanie i przetwarzanie informacji przestrzennej według różnych metod i sposobów, a także prezentowanie wybranych informacji w postaci kartograficznej. Jednym z takich systemów jest SINUS - System Informacji o Ukształtowaniu Środowiska. System ten opracowany został w ramach prac badawczych Instytutu Geodezji i Kartografii objętych podprogramem 12 Centralnego Programu Badań Podstawowych 04.10 "Ochrona i kształtowanie środowiska".

Źródła informacji

System SINUS jest zasilany z różnych źródeł informacji, które można pogrupować następująco:

"Zagadnienia Informacji Naukowej" 1991 nr 1(58)

- dane teledetekcyjne,
- istniejące mapy tematyczne,
- istniejące mapy topograficzne,
- wyniki pomiarów i obserwacji specjalistycznych np.: monitoring środowiska czy zagospodarowanie przestrzenne,
- dane pochodzące z istniejących systemów informacyjnych.

Niezwykle bogatym źródłem informacji dla systemu SINUS są dane teledetekcyjne. Tego rodzaju dane muszą podlegać konwersji. Polega to na przekształceniu komputerowym tzw. pikseli (elementarnych pól powierzchniowych, których obraz przesyłany jest przez satelity) na jednostki odniesień przestrzennych obowiązujące w systemie SINUS. Aby informacje charakteryzujące pod różnym względem obszar kraju gromadzić w sposób uporządkowany należy skonstruować specjalny układ. Taki układ został zaprojektowany i opracowany w oparciu o współrzędne geograficzne, jest to tak zwany układ odniesień przestrzennych.

Początek układu odniesień przestrzennych, przyjętego w systemie SINUS, przypada w punkcie przecięcia się 14 południka z 44 równoleżnikiem, co ma miejsce na wodach Zatoki Pomorskiej na północ od Świnoujścia. Tu rozpoczyna się układ składający się z pól czworokątnych zwanych blokami, zbliżonych kształtem do prostokąta o wymiarach w mierze łukowej 10' na 10', co odpowiada w mierze liniowej około 12 000 m x 18 000 m. Bloki te są zgrupowane w 36 pasach i 61 słupach. Jest ich więc w sumie 2196. Bloki dzielą się z kolei na pola pierwszego rzędu, oznaczone jako pola P_1 . W każdym bloku można wyróżnić 6 kolumn i 9 rzędów czyli 54 P_1 o wymiarach 2 km x 2 km. Dalszy podział pól stosuje sekwencje czwórki z założeniem, że pola wyższego rzędu dzielą się bez reszty na 4 pola niższego rzędu. A zatem każde pole P_1 dzieli

się na 4 pola P_2 o wymiarach 1000 m x 1000 m, pole P_2 dzieli się na 4 pola P_3 o wymiarach 500 m x 500 m, pole P_3 na 4 pola P_4 (250 m) itd. Ze względów technicznych wymiar najmniejszego pola ustalono na 31,25 m x 31,25 m. Przyjęta sekwencja wielkości pól sprawia, że układ odniesień przestrzennych można stosować zarówno dla całego kraju, jak też lokalnie dla niewielkich regionów przy zachowaniu pełnej wzajemnej kompatybilności.

Tak przyjęty układ pól odniesień przestrzennych umożliwi wykorzystanie informacji zgromadzonych również w innych systemach informacji przestrzennej i zasilania nimi bazy danych systemu SINUS.

W systemie SINUS możliwa jest digitalizacja map, realizowana przy użyciu urządzeń zwanych digimetrami. Służą one do przetwarzania informacji analogowej (graficznej) na postać cyfrową. Dane, powstałe w wyniku tego procesu mają charakter wektorowy, to znaczy obszary występowania różnych zjawisk są zapisywane jako ciągi par współrzędnych, wyznaczające granice poszczególnych jednostek powierzchniowych (wydzielań).

Jest to etap wstępny, a utworzone w ten sposób zbiory poddawane są przetwarzaniu komputerowemu, polegającemu na konwersji (przekształceniu) zapisu wektorowego na rastrowy. Metoda ta pozwala na definiowanie dowolnej wielkości rastra, tj. pola podstawowego, w zależności od potrzeb użytkownika.

Ukierunkowanie systemu na przetwarzanie informacji ujętej w formę rastrową oznacza, że operuje on geometrycznie standardowymi jednostkami powierzchniowymi o określonej wielkości, w których gromadzone są informacje o środowisku przyrodniczym. Jednakże przestrzenne zjawiska przyrodnicze mają zwykle charakter nieregularny, więc formę rastrową stworzono dla celów porządkowania zbioru danych.

Mając mapę tematyczną, której obraz ma być ujęty w formę rastrową, należy dokonać nałożenia wykreślonej siatki pól na materiał źródłowy, a następnie zapisać (na odpowiednich formularzach) odczytane wartości zjawisk w poszczególnych elementach powierzchniowych wyznaczonych przez tę siatkę. Następnie dane z formularzy wprowadza się z klawiatury do pamięci komputera.

Dane zbierane przez specjalistyczne stacje pomiarowe poddawane są procesowi interpolacji realizowanemu numerycznie za pomocą komputera. Zarówno miejsce położenia stacji jak i wyniki wartości (lub klasy) obserwowanych zjawisk odnoszone są do odpowiednich pól jednostkowych.

Przed wprowadzeniem do bazy systemu, dane pozyskiwane przy użyciu wyżej wymienionych metod są przedmiotem weryfikacji i korygowania, a także lokalizacji i poprawiania błędów. Dokonuje się przy tym podziału na pewne tematy jednostkowe, które tworzą zawartość poszczególnych tzw. warstw informacyjnych. Należy dodać, że moduł aktualizacji i korygowania posiada funkcje, dzięki którym można zasilić bazę unowocześnionymi informacjami z jednoczesną modyfikacją jej zawartości.

Zarządzanie bazą danych

Procedury zarządzania bazą danych stanowią narzędzie sprawnego manipulowania na danych zgromadzonych w zbiorach baz danych, których struktura odzwierciedla przyjęty model danych, utworzony z myślą o zasadniczym przeznaczeniu systemu. Pozwala to na dowolne definiowanie podstawowych parametrów bazy, takich jak zasięg terytorialny, zakres tematyczny, rozdzielczość przestrzenna (wielkość pól podstawowych) i klasyfikacyjna (szczegółowość wydzielenia klas zjawiska). Umożliwia również szybki dostęp do zawartych

w niej informacji oraz dokonywanie podstawowych operacji, w tym różnych konwersji.

Baza danych

W bazie danych systemu SINUS wykorzystuje się model danych zwanych w literaturze modelem rastrowym lub komórkowym, w którym komponenty środowiska przyrodniczego są opisywane przez ich cechy lub właściwości zwane atrybutami.

Rozkład przestrzenny atrybutu wyrażonego wartościami liczbowymi lub taksonomicznymi może być odwzorowany w dwuwymiarowej tablicy, której elementy są przyporządkowane określonym polom powierzchni, a wartości elementów tablicy są zgodne z wartościami atrybutów w opowiadających im polach. Każdy element bazy danych, podobnie jak każde pole powstałe w wyniku umownego podziału przestrzeni, jest wyróżniane przez przypisaną do niego identyfikator. Identyfikator pól w logiczny sposób wiąże się z numeracją stosowaną w przyjętym układzie odniesień przestrzennych /2/.

Przyjęto, że każda informacja tematyczna w bazie danych jest zorganizowana w pojedynczą warstwę informacyjną opisującą wybrany atrybut. Grupa atrybutów zorganizowana na jednym wybranym poziomie podziału przestrzeni (od P1 do P7) w zasięgu całej bazy, tworzy tzw. zbiór zasadniczy. W bazie danych może być jeden lub kilka zbiorów zasadniczych, którym celem oszczędności pamięciowej, towarzyszą tzw. zbiory uzupełniające. Są one listami identyfikatorów tych pól na których występuje określona kombinacja cech lub właściwości. Operowanie w bazie danych dużą liczbą obiektów przestrzennych stwarza potrzebę organizowania katalogów zbiorów uzupełniających. Określając kształt i lokalizację obiektu należy z kolei określić ich atrybuty przy pomocy tworzonych tablic zawierających te wartości.

Wybór sposobu przechowywania informacji w bazie danych systemu SINUS należy do użytkownika.

Baza danych systemu SINUS może być organizowana dla:

- dowolnego zasięgu terytorialnego w obszarze kraju (wybierany jest obszar prostokątny),
- dowolnego zakresu informacyjnego ustalonego przez użytkownika (wybierane są atrybuty i ustalana ich kolejność),
- wybranej przez użytkownika wielkości pól elementarnych dla poszczególnych zbiorów zasadniczych, zgodnie z przyjętym układem odniesień przestrzennych,
- wybranego przez użytkownika sposobu reprezentacji poszczególnych atrybutów w zbiorach zasadniczych.

Należy zauważyć, że poszczególne atrybuty mogą się odnosić do różnych aspektów środowiska zarejestrowanych w jednym czasie albo do jednego aspektu uchwyconego w różnych przekrojach czasowych. Sprzyja temu swobodą ustalania rodzaju atrybutów wyrażanych przez dane pierwotne, fizyczne (mianowane) statystyczne lub też dane wtórne taksonomiczne, sklasyfikowane /1/.

Działania na zbiorach danych systemu SINUS można podzielić na:

1. Operacje podstawowe włączone bezpośrednio do procedur zarządzania bazą danych:

- a) operacje organizacyjne,
- b) operacje wejścia/wyjścia,
- c) operacje arytmetyczne i logiczne.

2. Operacje specjalistyczne systemu informacyjnego.

Wśród operacji organizacyjnych należy wyróżnić:

- utworzenie nowego zbioru zasadniczego, utworzenie katalogu obiektów, utworzenie obiektu,

- ustawienie atrybutu na określoną wartość wstępną,
- utworzenie tabeli klas danego atrybutu.

Import czy eksport danych z/do świata zewnętrznego odbywa się dzięki operacjom wejścia/wyjścia do których należą m.in.:

- wprowadzenie danych zbioru zasadniczego bazy danych ze zbioru wejściowego oraz wyprowadzenie danych ze zbioru zasadniczego bazy danych do zbioru wyjściowego. Przewiduje się kilka formatów zbiorów wejściowych i wyjściowych, aby zapewnić możliwość wprowadzenia danych pozyskiwanych z różnych źródeł oraz wyprowadzania danych stosownie do różnych sposobów ich prezentacji i wykorzystania,

- wprowadzanie i wyprowadzanie obiektów w postaci zbiorów uzupełniających,

- wprowadzanie, aktualizowanie, sprawdzanie elementów bazy danych z klawiatury komputera,

- wyprowadzanie wybranych danych zgromadzonych w bazie na drukarkę lub monitor komputera.

Do operacji arytmetycznych i logicznych należą m.in.:

- klasyfikacja wybranego atrybutu, charakteryzowanego przez wartości pierwotne, według pomocniczej tabeli klasyfikacyjnej,

- reklasyfikacja wybranego atrybutu, charakteryzowanego przez wielkości taksonomiczne, według tabeli reklasyfikacyjnej,

- operacje arytmetyczne sumy, różnicy i iloczynu wykonane na dwóch dowolnych atrybutach zbioru zasadniczego,

- operacje przekształcenia atrybutu w obiekt według warunku logicznego (typu: $<$, $<=$, $< >$, $> =$, $>$), sprawdzoną na wszystkich polach bazy,

- operacje boolowskie: iloczynu logicznego, różnicy symetrycznej, sumy logicznej wykonywane na dwóch dowolnych obiektach,

- obliczanie statystycznych miar centralnych wybranego atrybutu, sporządzanie tabeli wyróżnień dla danego atrybutu,
- przeliczanie współrzędnych geograficznych na identyfikatory, pól układu odniesienia przestrzennego SINUS i odwrotnie.

Taka mnogość wymienionych operacji pozwala również na transfer danych pomiędzy dowolnymi niezależnymi bazami systemu SINUS, jeśli ich zasięgi terytorialne w jakiejś części pokrywają się. Możliwa jest także wymiana danych między bazą systemu SINUS a innymi systemami.

Przetwarzanie danych w systemie

Grupa procesów związana z przetwarzaniem danych została tak zaprojektowana, aby zawierała możliwie dużo modułów funkcjonalnych czy też pojedynczych funkcji o charakterze uniwersalnym, dooieranych przez użytkownika zgodnie z jego potrzebami.

Obecnie system wyposażony jest w następujące moduły funkcjonalne:

- złożone wyszukiwanie danych,
- operacje na jednostkach powierzchniowych,
- operacje pomiarowe,
- elementarne modelowanie,
- przetwarzanie rastrowe,
- generalizacja map (uproszczona),
- tworzenie informacji pochodnych.

Funkcje związane z przetwarzaniem danych podzielono na te, które ściśle służą procesom analiz przestrzennych oraz te stosowane do celów udostępniania danych.

Analizom przestrzennym służą następujące funkcje elementarne:

- przecinanie warstw informacyjnych,

- obliczanie gęstości obiektów liniowych w polach odniesień przestrzennych,
- dokonywanie reklasyfikacji jednostek powierzchniowych,
- obliczanie powierzchni.

Procesowi przecinania mogą być poddawane warstwy informacyjne o różnej rozdzielczości przestrzennej i różnym zasięgu terytorialnym, fragmentami wspólnym. Umożliwia to dokonanie selekcji i klasyfikacji otrzymanych par wartości zjawisk.

Określenie stopnia rozdrobnienia czy rozczłonkowania zjawisk elementarnych sprowadza się do operacji nałożenia podziału rastrowego na numeryczny zapis nieregularnych linii i obliczenia ich długości w poszczególnych polach układu odniesień.

Reklasyfikacja jednostek powierzchniowych polega na przypisaniu już zgromadzonym w bazie charakterystykom nowych charakterystyk niezbędnych do dalszych analiz.

Obliczanie powierzchni wiąże się z obliczaniem w różnych przekrojach przestrzennych i tematycznych zbiorach charakterystyk poszczególnych zjawisk.

System SINUS nie wyczerpuje wszystkich możliwych analiz przestrzennych. Istnieje jednak możliwość opracowania nowych funkcji analitycznych w przypadku, gdy takie potrzeby występują.

Udostępnianie informacji

Prezentacja analiz przestrzennych, jak też zawartość baz danych w systemie SINUS przedstawiane są najczęściej w formie map tematycznych. Do tego celu służy podsystem wspomagane komputerowo redagowania map. Użytkownik ma możliwość wyboru metod prezentacji kartograficznej, środków wyrazu graficznego, odwzorowania, skali, zasięgu terytorialnego czy formatu mapy, projektowania znaków graficznych oraz konstruowania legendy.

W systemie SINUS istnieją trzy formy prezentacji map: w postaci wyświetlonej na ekranie monitora graficznego, wydruku na drukarce mozaikowej oraz rysunku uzyskanego za pomocą plotera. Dostępne są również wydruki tabelaryczne oraz zbiory numeryczne. Wydruki wyświetlane na ekranie bądź utrwalone za pomocą drukarki wierszowej zawierają zestawienia zbiorcze lub wyniki obliczeń prowadzonych analiz przestrzennych czy statystycznych. Zbiory numeryczne może określać użytkownik. Zawierają one zwykle dane z bazy systemu SINUS lub wyniki analiz przestrzennych. Mogą one pełnić rolę pośrednika między systemami, a także służyć celom archiwizacji i ochrony zgromadzonych danych. Obie te funkcje (przygotowanie wydruków i tworzenie zbiorów numerycznych) realizowane są na wyjściu wielu omawianych już modułów, a w szczególności wchodzi w skład procedur zarządzania bazą danych.

Sprzęt komputerowy

Wyposażenie systemu w poszczególne zestawy sprzętowe będzie się różniło w zależności od przeznaczenia systemu i potrzeb użytkownika. Szacując potrzeby pamięciowe należy mieć na uwadze, że pamięć dyskowa powinna być trzykrotnie większa niż objętość zbiorów zasadniczych bazy. Przy dużej objętości tych zbiorów (> 10 MB) zaleca się stosowanie komputera AT lub Super AT, których czas wykonania operacji są 18 razy krótsze niż komputera XT. Przy zbiorach archiwizacyjnych małej objętości wystarczą dyskietki 1.2 MB lub nawet 360 KB. W przypadku większych zbiorów należy przewidzieć transfer lub archiwizację z użyciem pamięci kasetowej typu streamer /1/.

Wprowadzanie danych do systemu SINUS metodą dygitalizacji odbywa się przy pomocy digimetrów dostępnych na rynku w tym np. digimetru SummaSetch Professional firmy SummaGraphics.

Zestaw urządzeń wyjściowych zależnie od potrzeb użytkowych powinien zawierać: monitor kolorowy i kartę grafiki EGA oraz jeden z ploterów np. firmy Hewlett Packard lub firmy Rolland.

Monitor Mono z kartą Hercules oraz drukarka mozaikowa z opcją graficzną może wystarczyć w przypadku skromniejszych wymagań.

Oto przykładowy zestaw komputera właściwy dla przeciętnych zastosowań w systemie SINUS:

- Komputer PC-AT, 1 MB pamięci operacyjnej
- + Dyski - HDD 40 MB, 1 x FDD 1.2 MB, 1 x FDD 360 KB
- Monitor Mono lub Color
- Drukarka plotująca
- Digimetr lub Ploter.

Obsługa użytkownika

Spełnienie oczekiwań użytkownika dokonuje się przez dowolne projektowanie bazy danych oraz specjalne dobieranie zestawu stosowanych procedur. Przykładowy ciąg technologiczny składa się z następujących etapów:

1. Zbieranie materiałów źródłowych (mapy tematyczne, wyniki pomiarów itp.)
2. Przypisanie informacji do pól układu odniesień przestrzennych obowiązujących w systemie SINUS:
 - a) digitalizacja map,
 - b) weryfikacja wyników digitalizacji,
 - c) utworzenie wektorowej bazy danych,
 - d) korygowanie wektorowej bazy danych,
 - e) konwersja wektorowej bazy danych na zbiory o organizacji rastrowej.
3. Założenie rastrowej bazy danych w postaci kolejnych warstw informacyjnych, o różnych zakresach tematycznych i ewentualnie

różnym stopniu rozdzielczości przestrzennej (wielkości pola podstawowego).

4. Dokonanie numerycznego nałożenia warstw informacyjnych z utworzeniem nowej warstwy poddanej procesowi klasyfikacji.
5. Eventualne opracowanie programu realizującego funkcję analityczną zaprojektowaną przez użytkownika oraz wykonanie stosowanych obliczeń przy jego zastosowaniu.
6. Redagowanie mapy prezentującej wyniki przeprowadzanych analiz:
 - a) dobór zasięgu terytorialnego i skali mapy,
 - b) dobór odwzorowania,
 - c) dobór znaków kartograficznych,
 - d) zaprojektowanie legendy.
7. Wyświetlenie, wydrukowanie lub wykreślenie mapy przy użyciu urządzeń komputerowych oraz wykonanie uzupełniających wydruków tabelarycznych /2/.

Możliwości wykorzystania systemu

Zastosowanie systemu SINUS uwzględniające zasięg baz danych (oazy o zasięgu krajowym, regionalnym i lokalnym) może m.in. dotyczyć:

- inwentaryzacji przestrzennej zasobów naturalnych i przyrodniczych środowiska oraz działań antropogenicznych z możliwością selektywnego numerycznego i graficznego ich opracowania,

- gromadzenia i przetwarzania informacji o obiektach, zjawiskach i procesach przestrzennych w celu wspomagania procesów decyzyjnych dotyczących zagospodarowania przestrzennego na szczeblu krajowym i regionalnym,

- gromadzenia, przetwarzania i prezentacji informacji pozyskiwanych w sieci monitoringu środowiska dla śledzenia stopnia zagrożeń i zmian w obrębie środowiska naturalnego,

- tworzenia różnotematycznych baz danych i dokonywania analiz na potrzeby prac naukowych dotyczących przestrzennych aspektów badań przyrodniczych, a w szczególności kompleksowo ujmowanych problemów środowiska przyrodniczego.

System SINUS jest systemem uniwersalnym nadającym się do zastosowania w wielu zadaniach i różnych warunkach użytkowych. Zastosowanie techniki komputerowej pozwala na sprowadzanie różnego rodzaju informacji do postaci liczbowej i dokonywania najróżniejszych operacji na utworzonych zbiorach liczb. Umożliwia to uzyskanie informacji w czasie niemal rzeczywistym. Taka informacja powinna być podstawą poczynań decyzyjnych i interwencyjnych kształtujących środowisko przyrodnicze w naszym kraju.

Literatura

1. BARANOWSKI M. Prace badawcze nad Systemem Informacyjnym o Ukształtowaniu Środowiska. Materiały Instytutu Geodezji i Kartografii (maszyn.)
2. CIŁKOSZ A., BARANOWSKI M., JANKOWSKI R. Synteza Podprogramu "SINUS - System informacji o środowisku przyrodniczym".

10.11.1990 r.

INFORMATION SYSTEM ON CONFIGURATION OF ENVIRONMENT

Summary

The information system on configuration of environment was established at the Institute of Geodesy and Cartography. The information sources for the system make the teledetection data, available thematic and topographic maps, results of measurements and special observations (monitoring of the environment, spacial management) and the data from the other information systems.

The special setup of spacial references which enables to collect various data the well-ordered way was designed for the system. The raster (cellular) data model is used in the SINUS data base. The spacial analyses are presented most often in the form of thematic maps. The editing of maps is realized by means of the computer-aided subsystem, and at the same time the user has the possibility of choosing the method of the cartographic presentation, graphic tools, mapping, scale, the territorial scope, format of the map and so on.

Three forms of the presentation of maps are available in the discussed system: displaying on the computer monitor, printouts or draughts. The technical equipment of the system is also characterized as well as the possibilities of the further use of the system.

Р е з ю м е

Информационная система об рельефе естественной среды была разработана в Институте геодезии и картографии. Как источники информации используются данные теледетектирования, существующие тематические и топографические карты, результаты обмеров и наблюдения специалистов (наблюдение среды, благоустройство среды), а также данные из других информационных систем. Для целей системы была запрограммирована специальная система пространственных соотношений позволяющая собирать разнообразные данные упорядоченным путем.

В базе данных SINUS используется растровая (клеточная) модель данных. Пространственные анализы, а также содержание базы данных SINUS представлены чаще всего в форме тематических карт. Этой цели служит подсистема машинновоспроизводимого редактирования карт, при чём потребитель имеет возможность, между прочим, выбора метода картографической презентации, графических средств отображения, масштаба, территориального охвата, формата карт итп. В системе допускается три формы представления карт: дисплей на мониторе эвм, отпечатка и чертёж. Представлено также оборудование используемое в системе и возможности дальнейшего использования системы.

HASŁO PRZEDMIOTOWE A ADEKWATNY OPIS PRZEDMIOTOWY*

Adekwatność opracowania przedmiotowego. Użycie określników ogólnych i szczegółowych. Hasło przedmiotowe adekwatne a adekwatny opis przedmiotowy dokumentu.

Teoretycy i praktycy katalogu przedmiotowego omawiający metodykę tematowania i redakcji haseł przedmiotowych poświęcają tu i ówdzie po kilka zaledwie zdań problemowi adekwatności w opracowaniu rzeczowym^{1/}. Odrębnych prac poświęconych *expressis verbis* temu zagadnieniu do niedawna u nas nie było. Zapewne dlatego, że mieści się ono w ogólnej metodyce tematowania, jako zasadnicze wymaganie stawiane hasłu przedmiotowemu i charakterystyce treściowej dokumentu czyli jego opisowi przedmiotowemu. Wydaje się tu konieczne zacytowanie różnych wypowiedzi na temat adekwatności.

*Tekst referatu przygotowanego na Ogólnopolską konferencję nt. Opracowania rzeczowego, która odbędzie się 7-9 maja 1991 r. w Gdańsku.

^{1/}J.Ćwiekowa. Opracowanie przedmiotowe piśmiennictwa. Warszawa 1974 s.36, 111-112; łaż, Opracowanie tematyczne piśmiennictwa. Warszawa 1988 s.140, 142-143. A. Łysakowski, Katalog przedmiotowy. Podręcznik. Warszawa 1946 s.61.

J. Ćwiekowa w Opracowaniu przedmiotowym piśmiennictwa^{2/}, pisze: "W systemach przedmiotowych stosowanych w opracowaniu piśmiennictwa bardzo istotna jest adekwatność znaczeniowa tematów zawartych w jednostkach piśmiennictwa do odbicia tego znaczenia w ich nazwach hasłowych".

W drugim kompendium^{3/}, do najważniejszych postulatów, które należy uwzględnić w czasie opracowania przedmiotowego, zalicza: "Stosowanie szczegółowej analizy dzieła według zawartych w treści tematów; uzyskanie przybliżonego pokrywania się zakresów tematów dzieł z hasłami tematowymi katalogu przedmiotowego, przy czym adekwatność zakresów jest rzadko osiągalna ale chodzi o formułowanie tematów jak najbliższej "zrównanych" z zakresem opisywanego dzieła"^{4/}. używając określenia tematów zrównanych Autorka nawiązuje do założeń naszego klasyka katalogu przedmiotowego A. Łysakowskiego^{5/}.

J. Sadowska w Instrukcji tematowania i katalogu przedmiotowego zakłada m.in., że: "Każdy przedmiot dokumentu jest wyrażony w sposób adekwatny, tzn. nieuogólniony, aspekt przedmiotu może być uogólniony"^{6/}.

W syntetyczny sposób ujmuje problematykę opracowania rzeczowego T. Głowacka: "Bez względu na przyjętą metodę konstrukcji charakterystyk wszelkim typom katalogów rzeczowych stawia się takie same wymagania jak językom informacyjnym. Żąda się od nich

^{2/} J. Ćwiekowa Opracowanie przedmiotowe s. 36.

^{3/} J. Ćwiekowa Opracowanie tematyczne s. 140.

^{4/} ibidem s.142-143.

^{5/} Łysakowski, Podręcznik s. 61.

^{6/} J. Sadowska Koncepcja opracowania przedmiotowego w świetle Instrukcji tematowania i katalogu przedmiotowego. W: Biul.inform.BN 1988 nr 3-4 s.7.

siły semantycznej tj. zdolności dokładnego i całkowitego odtwarzania istoty informacji zawartej w tekście, jednoznaczności przyjętych sformułowań, sprawności gramatycznej, oddawania relacji zachodzących między pojęciami i wreszcie adekwatności i obiektywności charakterystyk wyszukiwawczych informacji zawartej w tekście"^{7/}.

W wypowiedziach tych trzech Auterek przewija się pojęcie adekwatności w nieco innych ujęciach. I. Głowacka podkreśla konieczność adekwatności całych charakterystyk wyszukiwawczych sporządzonych dla dokumentów. J. Ćwiekowa mówi o adekwatności znaczeniowej hasła przedmiotowego wobec tematów zawartych w treści opracowywanego dzieła, zaznaczając przy tym, że adekwatność taką trudno jest osiągnąć. Według J. Sadowskiej, nadanie adekwatnej nazwy dla przedmiotu dzieła oznacza nadanie mu nazwy nieuogólnionej, podczas gdy aspekt przedmiotu może być uogólniony, czyli według takiego ujęcia nieadekwatny. Zobaczmy to zalecenie na konkretnym przykładzie. Dla pracy pt. Budowa komórki roślinnej, hasło przedmiotowe adekwatne powinno składać się z tematu Komórka, z dwoma określnikami: - budowa-botanika. Jeżeli zaś według zalecenia aspekt przedmiotu uogólnimy, to hasło będzie miało postać albo Komórka-budowa, albo Komórka-botanika. W żadnym wypadku nie będzie to hasło w pełni adekwatne. Pod pierwszym znajdzie czytelnik prace ogólne o budowie wszelkiej komórki, pod drugim, wszystko o komórce roślinnej, nie tylko o jej budowie. Uogólnienie aspektu głównego przedmiotu dzieła spowoduje szum informacyjny, co będzie szczególnie niekorzystne przy większej ilości opracowań

^{7/}I. Głowacka Melioracje strukturalne w katalogach rzeczowych. W: Opracowanie rzeczowe zbiorów w dużych bibliotekach uniwersalnych. Warszawa 1989 s.217-231.

na dany temat. Zalecenie zawarte w Instrukcji tematowania, że każdy przedmiot dokumentu jest wyrażony w sposób adekwatny tzn. nieuoogólniony, podczas gdy aspekt przedmiotu może być uoogólniony, nie jest wystarczająco precyzyjne i jednoznaczne.

Nasuwa się tu jeszcze druga uwaga, dotycząca użycia określników tzw. ogólnych, w tym klasowych. W tym samym przykładzie, określnik - budowa, chociaż ogólny (Sadowska zalicza go do przedmiotowych^{8/}), dla jednej z właściwości głównego przedmiotu jest adekwatny. Drugi określnik, klasowy, -botanika, jest również adekwatny, ale dla innej z cech przedmiotu, wskazuje mianowicie na komórkę roślinną. Określniki te, z kategorii ogólnych, zastosowane pojedynczo przy adekwatnym temacie dały charakterystykę dokumentu uogólnioną, nieadekwatną, ale użyte łącznie uzupełniały się nazwajem i określając temat w sposób odpowiadający treści utworzyły adekwatną charakterystykę wyszukiwawczą. Użycie więc określników ogólnych nie oznacza niespełnienia zasady adekwatności. Zastosowanie bowiem w grupie określeniowej dwu określników ogólnych, a według Ćwikowej^{9/} "w stosunku do tej grupy określników stosuje się najczęściej metodę uogólniania wyrażanych przez nie pojęć", powinno oznaczać uogólnienie aspektu. Tymczasem w tym przypadku, aspekt przedmiotu jest jak najbardziej adekwatny do treści. Wygląda to jak paradoks, gdy nieadekwatne, ogólne określniki dają adekwatną charakterystykę treściową dokumentu. Ważny jeszcze przykład odwrotny, z określnikiem szczegółowym, a więc adekwatnym. Dla pracy: Katalog wystaw Muzeum Archeologicz-

^{8/} J. Sadowska, Z problemów katalogu przedmiotowego. Możliwości informacyjne określników. Rocznik BN 1984 nr XX s.211-233.

^{9/} Ćwikowa Opracowanie tematyczne s. 116.

nego w Warszawie, jeżeli zastosujemy przy temacie Warszawa tylko sam określnik szczegółowy -Muzeum Archeologiczne, to taka charakterystyka będzie nieadekwatna ponieważ określnik ten, mimo, że szczegółowy nie wyczerpał zakresu treściowego dokumentu i dla pełnego ujęcia tego zakresu potrzebne są jeszcze określniki -wystawy-katalog. Nie jest więc zasadne używanie zamiennie określeń adekwatny-nieuogólniony i nieadekwatny-uogólniony. Niezależnie bowiem od kategorii określników użytych (obok adekwatnego tematu) do charakterystyki treściowej dokumentu, adekwatność oznacza przede wszystkim wyodrębnienie najważniejszych aspektów przedmiotu i wyrażenie ich przy pomocy równoważnych im znaczeniowo określników. Rola określników zmienia się w zależności od zastosowania ich do scharakteryzowania takiej czy innej treści^{10/}. Jednak najważniejszym, znaczącym elementem hasła przedmiotowego jest grupa nominalna czyli temat, który powinien być szczególnie starannie sprecyzowany w stosunku do pojęcia przedmiotu. Z codziennej praktyki tematowania i korekty opracowania wiemy, że określenie głównego przedmiotu (najczęściej przedmiotów) wymaga nieraz dodatkowych poszukiwań w kompendiach, wyjaśnień merytorycznych, porównań, zwłaszcza, gdy opracowywać musimy dzieła z bardzo różnych dziedzin wiedzy. O stopniu trudności świadczyć może i to, że nawet przy opracowywaniu tematów modelowych dla słownika tematów, a więc z wyznaczonych przez siebie dziedzin, popełniamy czasami błędy, podając niewłaściwe albo niewystarczające relacje.

Może w tym miejscu należałoby, uwzględniając cytowane wcześniej wypowiedzi, spróbować określić co właściwie rozumiemy przez

^{10/}J. Sadowska Określniki w systemie języka haseł przedmiotowych. Aspekty semantyczno-syntaktyczne. Warszawa 1983 s.73-77.

hasło przedmiotowe adekwatne i adekwatny opis przedmiotowy dokumentu, bo te dwie rzeczy należy odróżnić.

Samo pojęcie adekwatności oznacza zrównanie, dokładne odpowiadanie czemuś. Adekwatność w opracowaniu rzeczowym powinna oznaczać dokładne odzwierciedlenie treści dzieła. Hasło przedmiotowe adekwatne, to takie hasło, którego zakres jest równoważny znaczeniowo zakresowi głównego tematu dzieła wraz z jego ujęciami. Inaczej można powiedzieć, że są to odpowiedniki słowne tematu dzieła wyrażone w formie hasła przedmiotowego przy czym przedmiot dzieła wyrażony jest przez tzw. temat hasła przedmiotowego, a ujęcie przedmiotu, czyli jego cechy, stany, wyrażamy za pomocą określników. Hasło przedmiotowe adekwatne jest więc dokładną informacją o przedmiocie i jego właściwościach, w której każda jednostka leksykalna odpowiada w sposób jednoznaczny lub bardzo przybliżony tematowi dzieła i jego aspektom.

Opis przedmiotowy adekwatny otrzymać możemy w kilku przypadkach:

1. Stosując adekwatne tematy w połączeniu z adekwatnymi określnikami
 - a/ mogą to być określniki szczegółowe, w tym jednostkowe lub specjalne dla specyficznych treści, np. dla pracy "Łączenie się łańcuchów polipeptydowych w cząsteczki białka", należałoby dać hasła Białko-biosynteza i Polipeptydy-biochemia.
 - b/ mogą też być przy adekwatnym temacie określniki ogólne, ale dla ogólnych treści, czyli też adekwatne, np. pracę "Fizyka powietrza" zatematujemy Powietrze-fizyka.
2. Stosując niezupełnie adekwatne hasła przedmiotowe, które w sumie dają adekwatny opis, np. dla pracy "Sceny batalistyczne w malarstwie polskim" damy hasło Malarstwo polskie-tematyka.

Nie będzie ono adekwatne, będzie miało uogólniony aspekt i ten aspekt powinien zostać wyrażony przez drugi temat - Wojna-ikonografia. Jeden temat dzieła, jeśli dzieło jest jednotematyczne, może dać się wyrazić przez jedno hasło adekwatne, ale często potrzebne jest więcej niż jedno hasło. Hasła przedmiotowe nadane przedmiotowi lub przedmiotom dzieła tworzą opis przedmiotowy nazywany też często charakterystyką wyszukiwawczą. Charakterystyka może być, podobnie jak hasło, adekwatna lub nie. Gdy w treści dzieła mowa jest o kilku przedmiotach, które możemy uznać za reprezentatywne, poszczególne hasła przedmiotowe opisujące te przedmioty powinny być adekwatne, aby w sumie otrzymać adekwatną charakterystykę dokumentu. Ale takiej idealnej sytuacji na pewno nie będzie, chociażby z powodu ograniczeń wynikających ze słownika tematów. W charakterystyce dokumentu mogą przeważać hasła adekwatne lub nieadekwatne, albo częściowo adekwatne, mogą być różne możliwości. W zależności od tego ile haseł w całym opisie przedmiotowym jest adekwatnych i w jakim stopniu poszczególne hasła są adekwatne wobec swoich przedmiotów, otrzymujemy charakterystyki o różnych stopniach adekwatności, czyli źródła informacji o dokumencie w sposób mniej lub bardziej precyzyjny mówiące o jego zawartości tematycznej i treściowej.

Mogą powstawać trudności w wyodrębnianiu głównego tematu dzieła, np. wtedy gdy nie orientujemy się w dziedzinie, do której należy opracowywana książka i gdy nie napotykamy w słowniku tematów hasła związanego z tą dziedziną. Wtedy musimy przyjąć jako temat taką nazwę lub nazwy, które uważamy za najbliższe zakresowi omawianych przedmiotów. Ale wtedy hasło przedmiotowe nie będzie adekwatne, może być tylko zbliżone do adekwatności. Czasami trudności wynikają z płynności pojęć, np. w naukach humanistycznych.

Tu często przedmiotem opracowania są nie realne przedmioty rzeczywistości lecz zjawiska, stany, pojęcia abstrakcyjne. Pojęcia te same w sobie mogą być złożone i niejednoznaczne i rozumienie ich przez różne osoby może być różne. W takich przypadkach więcej czasu i uwagi trzeba poświęcić na analizę dzieła i jego różnych kontekstów aby wyodrębnić główne tematy. I w takich przypadkach opracowuje się hasła przedmiotowe, które w sumie odtwarzają znaczenie przedmiotu danego dzieła i ukazują go dodatkowo w różnych aspektach. W takich przypadkach, kiedy brak jest ustalonej terminologii dla jakichś zjawisk, określenie tematu dzieła jest bardziej subiektywne, adekwatność hasła przedmiotowego może nie być osiągnięta, chociaż charakterystyka dokumentu może mieć nawet dużą wartość informacyjną.

Bardziej precyzyjną niż humanistyka terminologię posiadają nauki matematyczno-przyrodnicze, ale w tych dziedzinach napotykamy innego rodzaju trudności. Jest wiele bardzo wąskich specjalności czy zagadnień jak np.: Elektrodynamika kwantowa, Kombinatoryka czy Chemia jądrowa, gdzie część tekstu mogą stanowić wzory matematyczne i niezrozumiała dla bibliotekarza terminologia. W takich przypadkach, jako zło konieczne lepiej stosować temat uogólniony niż narażać się na ryzyko błędu. Teoretycznie w trudnych przypadkach powinno się korzystać z konsultacji specjalistów.

Specjalnie zagadnieniu adekwatności poświęcony został ostatnio artykuł J. Sadowskiej w serii *Predmetnyj poisk v tradicionnyh i netradicionnyh informacionno-poiskovyh sistemach* z 1989 r.^{11/}. Jest on ukierunkowany nieco na warunki radzieckie, ale istota zagadnienia powinna być taka sama niezależnie od systemu.

^{11/} J. Sadowska Adekwatnaja substka . *Uchoditno katalogo. W: Predmetnyj poisk v tradicionnyh i netradicionnyh informacionno-poiskovyh sistemach. Vyp.9 Leningrad 1989 s.50-60.*

Stwierdza w nim Autorka, że indeksowanie adekwatne, to nie tylko indeksowanie nieuogólnione, ale i zgodne z zakresem treści dokumentu. Przyczące też z metodyki opracowania przedmiotowego w ZSRR określenie hasła przedmiotowego adekwatnego - jest to hasło, którego treść wyraża zakres pojęcia najdokładniej odpowiadającego zakresowi pojęć o przedmiocie danego dzieła. Zastanawia się w związku z tym, czy hasło przedmiotowe przestanie być adekwatne, gdy do jego budowy użyjemy wyrażen bliskich lub szerszych zakresowo, skoro i tak wiadomo, że żaden dokument nie jest "czysty" treściowo. Nazywa to pytaniem o "granice adekwatności"^{12/}. Odpowiadając kolejno na te wątpliwości, wydaje się, że hasło przestanie być adekwatne w takim przypadku. Niejednorodność treściowa dokumentu nie powinna uzasadniać nieprecyzyjności w tematowaniu i powołując się na nią nie możemy przypisać adekwatności hasłu oddającemu treść dzieła w sposób przybliżony czy rozszerzający. Szczególnie jaskrawym przykładem nieadekwatności w przypadku zastosowania szerszych lub przybliżonych zakresowo terminów mogą być nauki przyrodnicze, gdzie np. w określeniu gatunku biologicznego czy związku chemicznego, jeżeli użyjemy nazw przybliżonych, to nie będzie to to samo. Co do "granic adekwatności", to wydaje się, że możemy się raczej zastanawiać czy spełniliśmy wymóg adekwatności z informacyjnego punktu widzenia, tzn. czy dostarczyliśmy czytelnikowi dostatecznie wyczerpującej informacji o zawartości tematycznej danego dzieła. Użycie znormalizowanych określników (a więc i ograniczenie ich liczby) zakłada potencjalne uogólnienie haseł przedmiotowych i w tym sensie nie można tu mówić o przekroczeniu "granicy adekwatności" lecz raczej o tym, że warunek adekwatności nie zawsze może być spełniony.

^{12/} ibidem s.51.

Kolejny rozpatrywany problem, to adekwatność hasła przedmiotowego, która oceniana jest zwykle w stosunku do dokumentu, a która powinna też być, zdaniem Autorki^{13/}, oceniana w stosunku do systemu, tzn. powinno się oceniać przydatność haseł adekwatnych w katalogu jako kluczy wyszukiwawczych.

I tu nasuwa się wątpliwość czy nie ma sprzeczności w takim podejściu do zagadnienia. Z jednej bowiem strony teoretycy katalogu przedmiotowego zalecają aby tematować adekwatnie, i w praktyce dążymy do tego, a z drugiej strony wyrażamy wątpliwości czy w wyniku takiego tematowania stworzone charakterystyki wyszukiwawcze będą przydatne jako klucze do wyszukiwania. Wątpliwości takich być chyba nie powinno przynajmniej w odniesieniu do bibliotek naukowych, ponieważ rzeczywistą wartość informacyjną może mieć tylko i wyłącznie ścisła informacja o piśmiennictwie naukowym. Wątpliwości te wiąże Autorka z rozbudową linearną hasła przedmiotowego, wydaje się jednak, że nie będą miały racji bytu, jeżeli w systemie, tj. w katalogu przedmiotowym będą zachowane konsekwentnie zasady przyjęte dla jego organizacji, jeżeli również będą przestrzegane zasady gramatyki języka haseł przedmiotowych, a także kiedy czytelnik będzie dobrze poinformowany o zasadach katalogu i będzie świadomie prowadził w nim poszukiwania. A to, że hasło będzie dłuższe, że trzeba przejść przez kolejne jego elementy przy wyszukiwaniu ręcznym nie powinno być takie istotne, jeżeli doprowadzić ma do odnalezienia potrzebnej szczegółowej informacji. Rozbudowane hasło adekwatne może mieć co najwyżej walor dydaktyczny. Zresztą rozbudowa hasła i kolejność poszczególnych elementów nie będą miały znaczenia w systemie zautomatyzowa-

13/ *ibidem* s. 53.

nym, gdzie każda jednostka leksykalna hasła zakodowana będzie oddzielnie i jej położenie w obrębie hasła nie będzie miało takiego wpływu na wyszukiwanie jak w tradycyjnym systemie. Inny polemiczny problem dotyczy korelacji. Zdaniem Autorki^{14/}, jeśli tworzymy hasło przedmiotowe z myślą o katalogu, to musimy pamiętać, że powinno ono być adekwatne, poprawnie zbudowane, ale także "skorelowane z innymi hasłami, powinno stanowić element systemu. Dlatego musi nieraz podlegać modyfikacjom, które umożliwią tworzenie kompleksów tematycznych"^{15/}.

W pierwszej chwili wydają się te poglądy słuszne, ale z drugiej strony, jeżeli tworzymy hasło przedmiotowe, obojętnie, adekwatne czy nieadekwatne, według reguł gramatycznych obowiązujących dla tego języka, to każde nowe hasło musi pasować do systemu, który przecież został stworzony dla funkcjonowania zdań tegoż języka. System był wtórny wobec haseł. Jeżeli dalej przestrzegamy relacji paradygmatycznych między tematami języka haseł przedmiotowych zawartych w słowniku tematów, to hasło przedmiotowe będzie też skorelowane z innymi hasłami przy pomocy przewidzianych powiązań odsyłaczowych, bo chyba tylko o takiej korelacji powinna być mowa przy alfabetycznym układzie haseł przedmiotowych. A dla takiej korelacji i "pasowania" haseł do systemu powinno wystarczyć tematowanie i redakcja katalogu zgodne z ustalonymi czy przyjętymi zasadami.

Jako przykład modyfikacji mającej dostosować hasło do systemu, podaje Autorka, uogólnianie i hierarchizację określników adekwatnych. W hasła - Język polski z określnikami -hydronimia, -nazwy

^{14/} ibidem s. 51.

^{15/} ibidem s. 53-55.

geograficzne, -toponimia (przykłady dostosowane do słownictwa ra-
dzieckiego) określniki te są bardzo bliskiej treści ale w katalogu
oddalone są od siebie alfabetycznie i mogą być jeszcze dodatkowo
rozdzielone innymi określnikami, które wejdą w szereg alfabetyczny.
Aby zapobiec nadmiernemu rozproszeniu określniki te można skupić
pou ¹ jednym bardziej ogólnym -onomastyka. Nie ma jednak żadnej
pewności czy tego rodzaju modyfikacja będzie wygodna dla wszystkich
użytkowników.

Za przykład hierarchizacji posłużyło hasło Biblioteki z
określnikami geograficznymi -Austria, -Kraków, -Polska, -Warszawa,
-Wiedeń. Modyfikacja polegałaby na tym, że w ciągu nazw geograficz-
nych miast i państw, wyłączono by nazwy państw, a pod nimi dopiero
ułożono by nazwy miast, uzyskując dwupoziomowy układ. Podobnie
można postąpić z określnikami chronologicznymi, grupując wieki
i lata, oraz z określnikami szczegółowymi będącymi nazwami instytu-
cji. Już z podanych przykładów wynika, że zabieg uogólniania
czy hierarchizacji określników adekwatnych na ograniczone zastosowa-
nie, dotyczy przeważnie tematów i określników geograficznych,
w dużo mniejszym stopniu ogólnych. Większość haseł przedmiotowych
takim modyfikacjom nie podlega i bardzo dobrze funkcjonuje w
systemie jako klucze wyszukiwawcze.

Do tej, uważa Autorka, że hasło przedmiotowe adekwatne wykorzy-
stywane w katalogu bez żadnych zmian, potwierdza zarzut stawiany
często katalogowi przedmiotowemu, że nie podaje informacji komplek-
sowej, że jednorodne informacje są rozproszone, że adekwatność
na poziomie tematów jest cenna, ale na poziomie określników rozbija
informacje i że nie jest konieczne wyrażanie drugoplanowych infor-
macji przy pomocy bardzo szczegółowych określników i bliskoznacz-
nych, które wpływają tylko na wzrost synonimii wyszukiwawczej^{16/}.

^{16/} ibidem s 54-55.

Zarzut taki nie wydaje się zasadny. Przyjęcie układu formalnego dla szeregowania haseł w katalogu przedmiotowym wynika z innych cech tego katalogu w porównaniu z katalogami w układzie logicznym. Alfabetyczne rozproszenie informacji w obrębie danego hasła tematowego nie jest specyficzną właściwością wyłącznie haseł z adekwatnymi określnikami. Wszystkie określniki z założenia różnicują i precyzują informacje odnoszące się do wyodrębnionego tematu, ale układ ich jest alfabetyczny. W uzasadnionych przypadkach jak przytoczone wcześniej można wprowadzić modyfikacje kompleksujące informacje, ale w sumie są one nieliczne i chyba nie da się ułożyć instrukcji, którą sugeruje Autorka, która podawałaby reguły tworzenia jednorodnych skupień tematycznych na poziomie określników. Byłaby to w pewnym sensie instrukcja "odprzedmiotowienia" katalogu przedmiotowego. Zdaniem Autorki^{17/}, hasło przedmiotowe adekwatne zakłada, że użytkownik podchodzi do informacji w sposób analityczny i że przychodząc do katalogu wie czego szuka. W zasadzie tak powinno być. A jeżeli tak nie jest, to trzeba będzie czytelników dokształcać i nauczyć ich jak mają korzystać z informacji zawartych w katalogu przedmiotowym.

Niektóre zagadnienia związane z adekwatnością porusza J. Sadowska w kolejnym artykule na temat - hasło przedmiotowe jako adnotacja i klucz wyszukiwawczy^{18/}. Jej zdaniem, hasło przedmiotowe, jak każda adnotacja powinno być adekwatne w stosunku do dokumentu. W tradycyjnym katalogu funkcjonuje ono jako całość niezależnie od tego z ilu elementów się składa. Wbrew pozorom jest to

^{17/} *ibidem* s.55.

^{18/} J. Sadowska Hasło przedmiotowe jako adnotacja i klucz wyszukiwawczy w systemach tradycyjnych i zautomatyzowanych. *Zagad.Infor.Nauk.* 1988, nr 2, s.73-93.

sytuacja niekorzystna, ponieważ tylko pierwsze elementy hasła spełniają funkcję informacyjną. Im bardziej adekwatne i rozbudowane hasło, tym trudniej jest się nim posługiwać w katalogu i tym trudniej jest szybko odnaleźć informację wyrażoną przez takie hasło. J. Ćwiekowa uważa, że przy temacie nie należy umieszczać więcej niż 4-5 określników^{19/}. A ponieważ nie wszystkie elementy hasła adekwatnego są jednakowo użyteczne z punktu widzenia potrzeb katalogu, często hasło przedmiotowe jest korygowane. Zależy to od oceny potrzeb informacyjnych lub intencji prowadzących katalog. Po pierwsze, nie każde hasło adekwatne musi być długie. Dalej zaś sama Autorka stwierdza, że hasło przedmiotowe funkcjonuje jako całość. Oznacza to, że tylko połączenie głównego tematu ze wszystkimi określnikami jako zdanie języka haseł przedmiotowych tworzy logiczną całość, która daje pewną informację o przedmiocie. I tylko to całe zdanie powinno być właściwym kryterium wyszukiwania informacji. Dlaczego więc tylko pierwsze elementy tej integralnej całości miałyby pełnić funkcję informacyjną? Jeżeli mamy np. hasło Serce-choroby-leczenie-podręcznik, to nie tylko pierwszy określnik jest użyteczny. Dla studenta medycyny równie ważna będzie informacja wyrażona w ostatnim określniku, że jest to podręcznik. Dalej Autorka uważa, że w zależności od potrzeb informacyjnych lub intencji prowadzących katalog hasło przedmiotowe powinno być korygowane. Co do intencji to na pewno są one dobre, jednak potrzeby informacyjne użytkowników rzadko są badane i są na pewno zróżnicowane.

W pewnym stopniu odpowiedzią na niektóre wątpliwości co do przydatności hasła adekwatnego w wyszukiwaniu informacji może być artykuł Grininy i Serebriannikowej dotyczący wykorzystania

^{19/} Ćwiekowa Opracowanie tematyczne s. 130.

hasła adekwatnego i hasła uogólnionego w warunkach ręcznego i zautomatyzowanego indeksowania^{20/}. Autorki powołując się na wyniki badań w ZSRR i za granicą, podają, że głównymi wymaganiami stawianymi przez czytelników wobec systemów są szczegółowość (detalność) i głębokość indeksowania. Uogólnione zaś hasła przedmiotowe powodują straty informacji i szum informacyjny. Szczegółowość określana jest jako dokładność odbicia przedmiotu dzieła w hasłach przedmiotowych. Wówczas mówi się o takiej samej pojemności pojęcia sformułowanego w postaci hasła i pojęcia odzwierciedlającego przedmiot dokumentu wyrażonego w języku naturalnym. Dla oznaczenia tej odpowiedniości używa się terminu adekwatne hasło przedmiotowe. Natomiast głębokość tematowania, to pełność odzwierciedlenia tematyki, którą osiąga się w charakterystyce wyszukiwawczej wówczas, gdy wyczerpane są w niej wszystkie tematy czy aspekty dokumentu z uwzględnieniem współzależności tych aspektów. Adekwatność jest pojęciem względnym z dwu powodów. Po pierwsze porównuje się tylko zakresy hasła i zakresy pojęcia wyrażającego przedmiot dokumentu, po drugie przedmiot może występować jako makroprzedmiot oddający treść dokumentu jako całości lub jako mikroprzedmioty wyizolowane drogą indeksowania analitycznego. I tak np. w Bibliotece AN poradnik bibliograficzny nt. ogólne zagadnienia sztuki, którego treść jako całość można wyrazić adekwatnie tematem Sztuka, otrzymał aż 20 haseł przedmiotowych. Prizment^{21/} zaś uważa, że aby szczegółowo oddać treść potrzeba od 1000 do 1500 haseł, co można zreali-

20/ P.F. Grinina, I. O. Serebriannikova Problemy ispolzovaniija adekwatnoj i obosčajuščej predmetnoj rubriki v uslovijach ručnoj i avtomatizirovannoj predmetizacii. W: Sovershenstvovanie sistemy katalogov naučnych bibliotek Leningrad 1988 s.86-97.

21/ cyt. za Autorkami j.w.

zować tylko przy pomocy indeksu przedmiotowego do książki. Natomiast wzrost liczby haseł do 5 czy 10, w stosunku do przeciętnej wynoszącej 1.5 hasła, nie zmienia w sposób istotny charakterystyki dokumentu. Autorki przyznają, że jest to krańcowy punkt widzenia.

Na ogół wszyscy zgadzają się z postulatem adekwatności opracowania przedmiotowego piśmiennictwa, chociaż pojawiają się i różne zastrzeżenia. Charakterystyczna pod tym względem jest wypowiedź E. Stępniakowej i J. Trzczińskiej we Wprowadzeniu do Słownika tematów języka haseł przedmiotowych DN^{22/}. Rozważmy ją po kolei. Autorki piszą: "W ujęciu modelowym hasło przedmiotowe winno wyrażać przedmiot dokumentu oraz istotne cechy, które przedmiot ten charakteryzują. Gdy sam temat ma znaczenie zbyt ogólne dodajemy określnik albo temat bardziej szczegółowy. Pewien stopień uogólnienia jest jednak niezbędny. Niemożliwe jest nie tylko pełne przedstawienie treści dokumentu (byłoby ono bowiem powieleniem go) ale i dokładne scharakteryzowanie jego głównego przedmiotu. Stopień uogólnienia zależy od ogólnych zasad jego, zasobu słownictwa danej jego odmiany, reguł jego uzupełniania, a także zasad metodyki tematowania".

Pierwsza wątpliwość, która się po tym fragmencie nasuwa, to czy pewien stopień uogólniania jest rzeczywiście niezbędny. Według polskich reguł tematowania przyjmuje się uogólniony aspekt przedmiotu na poziomie określników, dla samego zaś przedmiotu postuluje się wyszczególnianie. Wydaje się jednak, że i w przypadku określników nie może to być uogólnianie dla zasady, raczej powinna to być ostateczność. Każde uogólnienie jest w pewnym stopniu zniekształceniem rzeczywistego obrazu treściowego dzieła. Toteż

^{22/} J. Trzczińska, E. Stępniakowa Słownik języka haseł przedmiotowych Biblioteki Narodowej, Warszawa 1989 s. IV.

dla szczegółowych właściwości przedmiotu uogólniania nie powinno być stosowane.

W kolejnym twierdzeniu, że stopień uogólnienia zależy od zasad języka haseł przedmiotowych, zasobu jego słownictwa, od metod tematowania jest trochę racji, ale jednocześnie wszystkie te zasady nie mogą być niezmiennie. Poza tym zawsze się uważało, że to schematy klasyfikacyjne są sztywne i trudno je modyfikować, natomiast język haseł przedmiotowych odznacza się elastycznością, bowiem układ alfabetyczny nie krępuje wprowadzania nowych haseł. W dążeniu więc do adekwatności w tematowaniu piśmiennictwa naukowego, zasób słownictwa powinien być w uzasadnionych przypadkach wzbogacany o nowe szczegółowe tematy i określniki. Słownik tematów nie powinien ograniczać możliwości wyrażania nowych zjawisk, powinien być raczej traktowany jako spis tematów powszechnie spotykanych i określników typowych, które powiązane relacjami paradygmatycznymi ułatwiają naszą pracę, a nie ograniczają adekwatność i nowoczesność opracowania. Dlatego trudno zgodzić się z tym, żeby ze względów formalnych trzeba było wprowadzać uogólnienia. Jako argument za koniecznością uogólniania nie może chyba służyć zdanie o niemożliwości pełnego przedstawienia treści dokumentu i dokładnego scharakteryzowania głównego przedmiotu. Dotyczy ono opisu treści dokumentu w języku naturalnym, a w procesie tematowania nie chodzi o dosłowne odtwarzanie treści dokumentu aż do powielenia go i żadna metodyka tego nie zaleca. Nie chodzi też o rozważania teoretyczne nad treścią, z których miałyby wyniknąć dokładna charakterystyka głównego przedmiotu. W czasie tematowania działamy przede wszystkim na płaszczyźnie informacyjnej, nie interpretujemy przedmiotu lecz podajemy informacje o nim i jego cechach, jednak informację naukową zgodną z zawartością

treściową dzieła. I tak np. dla pracy - Oddziaływania cząstek elementarnych i jąder atomowych, powinniśmy podać informację w postaci haseł: Cząstki elementarne-fizyka, Jądro atomowe-fizyka i Reakcje jądrowe. Mimo że nie umiemy dokładnie scharakteryzować w języku naturalnym żadnego z wymienionych tematów i nie rozumiemy mechanizmów reakcji jądrowych, podajemy o tym wszystkim informację w języku haseł przedmiotowych, z wyszczególnieniem w temacie. Aspekt przedmiotu jest wyrażony uogólniającym określnikiem fizyka, ale nie dlatego, że takie uogólnienie jest niezbędne, tylko dlatego, że ten określnik jest optymalny dla treści tego dokumentu.

Dalszy ciąg wypowiedzi Autorek jest następujący: "Na ogół postuluje się wprowadzenie haseł adekwatnych, współmiernych (coextensive) wobec przedmiotu dokumentu, jest to jednak adekwatność względna: należy uwzględnić cechy przedmiotu uznane arbitralnie za istotne. Dopuszcza się również uogólniające traktowanie pewnych kategorii przedmiotów, np. niektórych grup osób lub obiektów geograficznych... Problem ten nie może być jednak rozstrzygany na płaszczyźnie czysto teoretycznej: o tym, jaki stopień szczegółowości haseł jest właściwy trzeba decydować w odniesieniu do każdego z osobna, uwzględniając wielkość opracowywanego zbioru piśmiennictwa, jego profil, zainteresowania i poziom przygotowania jego użytkowników^{23/}.

Co do niemożności rozstrzygnięcia problemu uogólniania na płaszczyźnie teoretycznej, to wydaje się, że problem ten powinien być jednak rozstrzygnięty i to w formie obowiązującej w ogólnych zarysach dla wszystkich bibliotek naukowych. Celem bowiem katalogo-

^{23/} ibidem

gów rzeczowych biblioteki naukowej nie jest zaszyfrowywanie informacji przez uogólnianie, lecz wszechstronne ukazanie tematyki piśmiennictwa w sposób adekwatny do treści i co się z tym wiąże w sposób obiektywny. Natomiast specjalne zapotrzebowania środowiska czytelniczego mogą sobie poszczególne biblioteki realizować w opisie wielokrotnym, formułując dodatkowe hasła tematyczne czy też uwzględniając specjalne aspekty przedmiotu w grubie określeniowej hasła przedmiotowego.

Nie wydaje się całkiem słuszne, aby stopień uogólnienia czy szczególności haseł uzależniać np. od wielkości czy profilu księgozbioru lub zainteresowań czy poziomu przygotowania użytkowników. To są argumenty, które mogą być w pewnym stopniu uzasadnione w bibliotekach wąsko specjalistycznych. Ogólnie jednak nie możemy dowolnie szacować poziomu użytkowników i według swoich wyobrażeń przeznaczać im informację mniej lub bardziej szczegółową. Wydaje się, że niezależnie od poziomu użytkowników informacje zawarte w katalogach bibliotek naukowych powinny być obiektywne i najwyższej jakości.

J. Cwiekowa w metodyce tematowania^{24/}, też co prawda zaznacza, że adekwatność zakresów tematów dzieł i haseł przedmiotowych jest rzadko osiągalna, ale chodzi nam o ich maksymalne przybliżenie. I właśnie, to też jest ważne. Opracowaniu przedmiotowemu powinno towarzyszyć dążenie do adekwatnego wyszczególnienia tematów dzieła, a nie przekonanie o niemożliwości scharakteryzowania nawet jego głównych przedmiotów.

Tylko zasada adekwatności może zapewnić podanie obiektywnej i w miarę możliwości dokładnej informacji o zawartości tematycz-

^{24/} Cwiekowa Opracowanie tematyczne s. 142-143.

nej piśmiennictwa. W przypadku bibliotek naukowych tylko taka informacja może mieć wartość zarówno dla studentów jak i naukowców, nawet, jeżeli jest to tylko zasada adekwatności w sensie informacyjnym, dotycząca najistotniejszych tematów dzieła, które jest w stanie określić przeciętnie wykształcony indeksator.

28.03.1991 r.

SUBJECT HEADING AND THE ADEQUATE SUBJECT CHARACTERISTICS

S u m m a r y

The problem of adequacy in the subject description of documents was treated as one of a minor importance in the theoretical works devoted to the methodology of formulating subject headings and to ascribing them to documents. This problem appears in works by J. Œwiekowa, J. Sadowska and T. Głowacka. Arguing their points, the author presents her own interpretation of the terms "adequate subject heading" and "adequacy of subject description" explicitly distinguishing one from the other. The adequate subject heading is understood as a heading, the scope of which is equivalent to the scope of the main theme of the document together with all its interpretations.

The ways of obtaining the adequate subject description are presented on some examples. When considering limits of adequacy, the author argues views of J. Sadowska on this problem and states: "only the principle of adequacy (in the subject description) can ensure the objectiveness and the possible detailness of information about the subject content of literature".

Р е з ю м е

Проблема адекватности в реальной каталоге являлась ещё до недавних пор в теоретических работах по методике тематизации и редактирования предметных заголовков поднималась мимоходом. Адекватность появляется в работах Я.Цвекоой, Я.Садовской и Т.Гловацкой. Полемизируя с их взглядами автор представляет собственную попытку понимания терминов "предметный заголовок" и "адекватное предметное описание" четко их разограничая. Предметный заголовок понимается как заголовок, которого объем равновеликий объему главной темы произведения вместе с её интерпретациями.

Рассуждая границы адекватности автор полемизирует со взглядами Я.Садовской по данной проблеме. Констатирует : " только принцип адекватности (при предметной обработке) может обеспечить объективную и по мере возможности точную информацию о тематическом содержанию литературы."

BARBARA SOSIŃSKA-KALATA

Instytut Bibliotekoznawstwa
i Informacji Naukowej UW

PROJEKT SYSTEMU NOTACYJNEGO DLA TEZAUZUSA INFORMACJI NAUKOWEJ

Charakterystyka struktury i metodyka opracowania

Cele i ogólne zasady budowy systemu notacyjnego dla tezaursów. Metoda opracowania systemu notacyjnego dla Tezaurusu Informacji Naukowej (charakterystyka struktury i metodyka opracowania).

CEL I OGÓLNE ZASADY BUDOWY SYSTEMU NOTACYJNEGO DLA TEZAUZUSA

Sprawne funkcjonowanie języka deskryptorowego w systemie informacyjno-wyszukiwawczym wymaga uporządkowania jego słownictwa według co najmniej dwóch uzupełniających się kryteriów: czytelnego dla użytkownika kryterium formalnego oraz trudniejszego w użytkowaniu, lecz znacznie pomocniejszego zarówno w indeksowaniu, jak i przy wyszukiwaniu informacji, kryterium semantycznego (logicznego).

W opracowanym dla systemu informacji bibliograficznej z zakresu informacji naukowej Tezaurusie Informacji Naukowej /1/ zastosowano wymagane zarówno przez obowiązującą normę polską /3/, jak i przez normę międzynarodową ISO /2/ formalne kryterium alfabetycznego porządkowania artykułów deskryptorowych, prowadzące

"Zagadnienia Informacji Naukowej" 1991 nr 1(58)

do utworzenia części alfabetyczno-hierararchicznej tezaurusa. Część ta ex definitione nie zawiera wszystkich informacji dotyczących znaczenia poszczególnych deskryptorów - nie zapewnia prezentacji ich pełnego kontekstu w polu semantycznym języka deskryptorowego, ani też nie umożliwia ustalenia formy deskryptora przyjętej dla oznaczenia pewnego zagadnienia (tematu czy przedmiotu dokumentu) w przypadku, gdy użytkownik nie zna terminów nazywających to zagadnienie i uwzględnionych w teaurusie jako deskryptory lub askryptory.

Zarówno ustalenie pełnego kontekstu deskryptorów w polu semantycznym, jak i ustalenie formy deskryptorów wyrażających określone znaczenie zapewnione są przez uporządkowanie słownictwa języka deskryptorowego według kryterium semantycznego, a więc wyposażenie tezaurusa w tzw. część semantyczną (systematyczną, fasetową, gniazdową, graficzną), w której deskryptory rozmieszczone są w pewnych kategoriach znaczeniowych lub grupach tematycznych. Warto zwrócić uwagę, że jeszcze do niedawna część semantyczna traktowana była jako fakultatywna część pomocnicza tezaurusa, w najnowszej natomiast normie ISO - zgodnie z coraz częstszą praktyką - wartość funkcjonalna tego typu wykazu słownictwa jest znacznie wyżej oceniana, a zalecenia budowy tezaursów umożliwiają przyjęcie jej jako głównej części tezaurusa.

Polska i międzynarodowa norma budowy tezaursów jednojęzycznych dopuszczają stosowanie różnych form części semantycznej - tradycyjne uporządkowanie hierararchiczne, systematyczne, działowe, fasetowe, różnego typu wykresy strzałkowe, tablice itp. W teaurusie Informacji Naukowej zastosowano formę obecnie najpopularniejszą - porządek deskryptorów według klasyfikacji fasetowej, w której na pierwszym szczeblu podziału wyodrębniono działy dziedzi-

nowe a wewnątrz nich materiał językowy poddano systematyce kategoryalnej stosując zróżnicowane kryteria podziału.

Korzystanie z jakiegokolwiek części tezaurusa, w której materiał leksykalny zorganizowany jest według pewnego kryterium semantycznego wymaga wyposażenia wszystkich uwzględnionych w tej części jednostek językowych w sztuczne kody określające miejsce każdej z nich w zaprojektowanej dla danego języka deskryptorowej strukturze paradygmatycznej. W tym celu konieczne jest opracowanie odpowiedniego dla struktury tezaurusa systemu notacyjnego, generującego kody o cechach zapewniających realizację następujących funkcji:

1. utrwalanie pionowego porządku wyrażen w części semantycznej tezaurusa,

2. zapewnienie więzi między częścią alfabetyczną lub alfabetyczno-hierarchiczną i częścią semantyczną,

3. ułatwienie identyfikacji pod semantycznych (lub tematycznych), do których należą poszczególne jednostki leksykalne i wzbogacenie informacji o treści jednostek leksykalnych zawartej w części alfabetycznej lub alfabetyczno-hierarchicznej,

4. umożliwienie automatycznego wyszukiwania (w tym też hierarchicznego).

Realizacja funkcji pierwszej wymaga, aby system notacyjny generował kody o określonej, powszechnie znanej wartości porządkowej oraz, aby wartość ta wzrastała zgodnie z następstwem jednostek leksykalnych (najczęściej deskryptorów) w porządku pionowym. Konieczne jest więc przyjęcie bazy notacyjnej (alfabetu systemu notacyjnego) w postaci zbioru znaków o powszechnie znanych zasadach porządkowania, np. cyfry arabskie lub litery alfabetu łacińskiego. Kody wyrażen znajdujących się na początku semantycznego wykazu

słownictwa powinny mieć wartość niższą aniżeli kody wyrazów występujących w wykazie w dalszej kolejności; deskryptor poprzedzający inny deskryptor w wykazie powinien mieć przypisany kod o wartości niższej od wartości kodu deskryptora po nim następującego. Tak więc symbole notacyjne powinny funkcjonować podobnie jak numery kolejnych deskryptorów i w najprostszej postaci systemu notacyjnego, notacji porządkowej, mogłyby mieć formę kolejnych liczb naturalnych, np.:

- 1 INFORMACJA NAUKOWA
 - 2 NAUKA O INFORMACJI
 - 3 INFORMACJA
 - 4 DEFINICJA INFORMACJI
 - 5 TYPY INFORMACJI
- (ze względu na zasięg rozpowszechniania)
- 6 INFORMACJA ADRESOWANA
 - 7 INFORMACJA POWSZECHNA
 - 8 INFORMACJA POUFNA
- (ze względu na zasięg chronologiczny)
- 9 INFORMACJA RETROSPEKTYWNA
 - 10 INFORMACJA BIEŻĄCA
 - 11 INFORMACJA PROSPEKTYWNA

Ten rodzaj notacji, choć spotykany w tezaurusach, nie jest zalecany, gdyż cechuje się bardzo ograniczoną produktywnością (małą zdolnością generowania kodów dla nowych deskryptorów wprowadzanych do tezaurusa w miarę jego aktualizacji) i nie zapewnia realizacji funkcji trzeciej i czwartej. Wygodniejszym i częściej stosowanym w tezaurusach rodzajem notacji jest notacja hierarchiczna - system znakowania o znacznie wyższej produktywności umożliwiający odwzorowanie w strukturze kodów semantycznej struktury deskryptorów,

co oznacza, iż kody te zapewniają także możliwość automatycznego wyszukiwania hierarchicznego, np.

- 1 INFORMACJA NAUKOWA
- 11 NAUKA O INFORMACJI
- 12 INFORMACJA
- 121 DEFINICJA INFORMACJI
- 122 TYPY INFORMACJI
- (za względu na zasięg rozpowszechniania)
- 1221 INFORMACJA ADRESOWANA
- 1222 INFORMACJA POWSZECHNA
- 1223 INFORMACJA POUFNA
- (ze względu na zasięg chronologiczny)
- 1224 INFORMACJA RETROSPEKTYWNA
- 1225 INFORMACJA BIEŻĄCA
- 1226 INFORMACJA PROSPEKTYWNA

Funkcja druga realizowana jest poprzez umieszczenie przy każdym wyrażeniu (zwykle deskryptorze) w części alfabetycznej lub alfabetyczno-hierarchicznej jego kodu traktowanego jako adres tego wyrażenia w części semantycznej. Aby funkcja ta realizowana była prawidłowo konieczne jest jedynie zapewnienie łatwej identyfikacji wartości porządkowej każdego kodu oraz spełnienie przez system notacyjny funkcji pierwszej.

Realizacja funkcji trzeciej i czwartej wiąże się ze zdolnością systemu znakowania do ujawniania w strukturze generowanych kodów relacji w jakich poszczególne deskryptory pozostają wobec innych deskryptorów. Szczególnie istotne jest wskazanie w strukturze kodu relacji hierarchicznych i relacji współrzędności, stanowi ono bowiem warunek zapewnienia możliwości automatycznego wyszukiwania hierarchicznego za pomocą kodów deskryptorów użytych w instrukcji wy-

szukiwawczej. Wprowadzenie do struktury kodu deskryptora informacji o tym, jakie deskryptory pozostają wobec niego w relacjach hierarchicznych i współrzędności można osiągnąć stosując pewną notację strukturalną, hierarchiczną lub komponentalną (składnikową).

W przypadku stosowania notacji hierarchicznej kody poszczególnych deskryptorów zawierają w swojej strukturze pełne kody deskryptorów wobec nich hierarchicznie nadrzędnych (tzw. deskryptorów szerszych), różniąc się od nich znakami wskazującymi miejsce danego deskryptora w wykazie deskryptorów współrzędnych. Ten sam kod deskryptora szerszego zostanie więc umieszczony w kodach wszystkich deskryptorów wobec niego podporządkowanych, spośród których deskryptory o kodach różniących się taką liczbą znaków, jaką zarezerwowano dla zróżnicowania elementów każdego poziomu hierarchii, są deskryptorami współrzędnymi. Warunkiem stosowania notacji hierarchicznej jest dokładne oznaczenie w części semantycznej lezaurusu relacji hierarchicznych oraz ukształtowanie porządku pionowego zgodnie z tą właśnie relacją.

W przypadku stosowania notacji komponentalnej kody poszczególnych deskryptorów zawierają znaki reprezentujące znaczenia częściowe składające się na znaczenie deskryptora. Deskryptory o znaczeniu pokrewnym, a więc takim, w skład którego wchodzi pewne wspólne znaczenia częściowe, otrzymują kody zawierające część znaków identycznych. Metoda ta zapewnia więc ujawnianie różnego rodzaju relacji semantycznych w strukturze kodu deskryptora, zarówno związków hierarchicznych, jak i silniejszych i słabszych związków asocjacyjnych. Zastosowanie notacji komponentalnej możliwe jest jednak wyłącznie wtedy, gdy dokładnie określona jest struktura semantyczna każdego deskryptora oraz gdy porządek pionowy części semantycznej zachowuje bezwzględną konsekwencję

w kolejności stosowania analogicznych kryteriów podziału w odniesieniu do różnych zagadnień cytowania analogicznych kategorii semantycznych w obrębie różnych faset zawierających deskryptory o znaczeniu pokrewnym.

METODA OPRACOWANIA SYSTEMU NOTACYJNEGO DLA TEZAURUSA INFORMACJI NAUKOWEJ

Opracowanie systemu notacyjnego dla Tezaurusu Informacji Naukowej, który pełniłby cztery wyżej omówione funkcje, poprzedziło przeprowadzenie grupy działań, których celem było ustalenie optymalnej formy notacji dla przyjętej w tym Tezaurusie metody organizacji słownictwa, jego charakteru i liczebności. Działania te to:

1. analiza struktury części semantycznej tezaurusu, rodzajów zastosowanych kryteriów podziału, zasad porządkowania faset i ustalenia związków hierarchicznych,

2. ocena liczebności zawartego w tezaurusie słownictwa i jego ilościowego rozkładu w poszczególnych działach, grupach tematycznych i szeregach deskryptorów współrzędnych (szeregach klasyfikacyjnych),

3. ocena stopnia szczegółowości zastosowanego w części semantycznej systemu klasyfikacyjnego deskryptorów,

4. ocena możliwości stosowania techniki rozbudowy równoległej kodów lub notacji komponentalnej,

5. ustalenie koniecznego poziomu produktywności systemu notacyjnego i metod jej zapewnienia.

Struktura części semantycznej Tezaurusu Informacji Naukowej

Jak wspomniano już wcześniej, Tezaurus Informacji Naukowej (TIN) opracowany dla Ośrodka Informacji Naukowej PAN i jego systemu informacji bibliograficznej z dziedziny nauki o informacji naukowej WIST, zawiera dwa podstawowe wykazy słownictwa - część alfabetyczno-hierarchiczną stanowiącą alfabetyczny wykaz artykułów deskryptorowych i artykułów askryptorowych oraz część systematyczną w formie klasyfikacji fasetowej deskryptorów pogrupowanych w 6 głównych działach o nierównomiernej liczebności /1/. Pięć pierwszych działów określono jako podstawowe dziedziny wchodzące w zakres tematyczny tezaurusu (informacja naukowa, bibliotekarstwo, bibliografia, czytelnictwo i informatyka), dział szósty natomiast obejmuje wykaz 15 dziedzin pokrewnych informacji naukowej wraz z nierównomiernie i raczej w małym stopniu uwzględnioną ich terminologią. Na początku każdego działu głównego umieszczono wykaz modyfikatorów, które można łączyć z deskryptorami należącymi do danego działu - w następnej redakcji tezaurusu zapowiedziano skumulowanie wszystkich modyfikatorów w jednym wykazie.

Porządek pionowy części systematycznej TIN odwzorowuje układ zastosowanej w nim klasyfikacji fasetowej. Wewnątrz poszczególnych szeregów klasyfikacyjnych umieszczone są deskryptory należące do różnych kategorii semantycznych, które w wielu przypadkach należy traktować raczej jako kategorie tematyczne czy zagadnieniowe, a nie jako kategorie semantyczne sensu stricto, np.

ŹRÓDŁA INFORMACJI

DOKUMENT

ODSTĘP DO ŹRÓDEŁ

WYKORZYSTYWANIE ŹRÓDEŁ

NOŚNIK INFORMACJI

INDEKS

RODZAJE INDEKSÓW

FUNKCJA INDEKSU

STRUKTURA INDEKSU

TWORZENIE INDEKSU

W fasetach zawierających nazwy szczegółowe rodzajów obiektów i procesów stosowane są zróżnicowane kryteria podziału, które niekiedy oznaczone są za pomocą umieszczonego w nawiasie i zapisanego kursywą kwalifikatora, a niekiedy sygnalizuje się je jedynie poprzez zwiększenie przerwy między wykazami deskryptorów należących do różnych faset. Tego rodzaju fasety poprzedza zwykle tzw. deskryptor podziałowy, np. RODZAJE INDEKSÓW, TYPY TEMATÓW, TYPY SIW, RODZAJE BIBLIOGRAFII.

W dziale Informatyka, w fascie obejmującej języki programowania uwzględniono obok deskryptorów w postaci nazw gatunkowych także pewną grupę nazw własnych, czyli identyfikatorów. Identyfikatory te zgromadzono w wyróżnionej liście alfabetycznej poprzedzonej kwalifikatorem "Nazwy własne języków programowania". W innych fasetach identyfikatory podporządkowane są bezpośrednio deskryptorom o postaci nazwy gatunkowej obiektów przez nazwy te wskazywanych i najczęściej również uporządkowane alfabetycznie.

Podstawową relacją semantyczną organizującą słownictwo w części systematycznej jest relacja hierarchiczna, która przyjmuje postać najczęściej związku generycznego, ale niekiedy także relacji hierarchii tematycznej (związku kojarzeniowego, któremu w strukturze systemu leksykalnego języka deskryptorowego nadano własności porządku logicznego), np.

relacja generyczna:

klasyfikacja specjalna > klasyfikacja branżowa
klasyfikacja specjalna > klasyfikacja dziedzinaowa
klasyfikacja specjalna > klasyfikacja patentowa
temat indywidualny > temat osobowy
temat indywidualny > temat geograficzny
temat indywidualny > temat chronologiczny
informacja (typy informacji) > informacja adresowana
informacja (typy informacji) > informacja powszechna
informacja (typy informacji) > informacja poufna

relacja hierarchii tematycznej:

indeks > tworzenie indeksu
indeks > struktura indeksu
źródła informacji > nośnik informacji
źródła informacji > dostęp do źródeł
teoria JIW > definicja JIW
teoria JIW > funkcje JIW

Obie te relacje hierarchiczne mogą być wykorzystywane w automatycznym wyszukiwaniu hierarchicznym, jednakże obecność relacji hierarchii tematycznej przemawia przeciwko tworzeniu dla IIN notacji komponentalnej.

Liczebność słownictwa i jego rozkład ilościowy w poszczególnych działach i szeregach klasyfikacyjnych

Tezaurus Informacji Naukowej w wersji obecnej obejmuje 1404 deskryptory szczegółowe wraz z identyfikatorami włączonymi do części systematycznej, 6 deskryptorów stanowiących nazwy działów głównych oraz 6 modyfikatorów powtarzających się częściowo w trzech działach. Rozkład ilościowy deskryptorów w działach głównych przedstawia się następująco:

"Informacja naukowa"	658
"Bibliotekarstwo"	305
"Bibliografia"	90
"Czytelnictwo"	53
"Informatyka"	231
"Dziedziny pokrewne"	67

W TIN występuje 20 nazw dziedzinowych: 5 stanowi nazwy działów głównych, 15 to deskryptory szczegółowe będące nazwami dziedzin pokrewnych zawartych w dziale szóstym o quasi-dziedzinowej nazwie "Dziedziny pokrewne". Nazwy dziedzinowe 5 działów głównych oraz nazwa "Dziedziny pokrewne" występują dwukrotnie - jako deskryptory dziedzinowe i jako deskryptory szczegółowe.

Utworzone w systematycznej części Tezaurusu szeregi klasyfikacyjne mają bardzo zróżnicowaną liczebność. Występują tu zarówno szeregi szcążkowe, jedno- lub dwuelementowe, jak i szeregi rozwinięte, kilkunasto-, a nawet kilkudziesięcioelementowe, na ogół podzielone na sektory obejmujące kilka lub kilkanaście deskryptorów należących do poszczególnych kategorii lub faset. Na przykład szereg podporządkowany deskryptorowi INFORMACJA za pośrednictwem deskryptora podziałowego TYPY INFORMACJI obejmuje 14 deskryptorów szczegółowych, a szereg podporządkowany deskryptorowi BIBLIOTEKA za pośrednictwem deskryptora podziałowego TYPY BIBLIOTEK zawiera 42 deskryptory. Najwyższa liczba sektorów wydzielanych wewnątrz szeregów klasyfikacyjnych nie przekracza 10 (np. w szeregu podporządkowanym deskryptorowi BIBLIOTEKA).

Szczegółowość klasyfikacji zastosowanej w Tezaurusie

Szczegółowość słownictwa zawartego w Tezaurusie i równocześnie szczegółowość organizującej je klasyfikacji fasetowej jest znacznie zróżnicowana w poszczególnych działach głównych tej klasyfikacji.

Najbardziej szczegółowe są działy "Informacja naukowa" i "Informatyka". W dziale "Informacja naukowa" wyróżniono maksimum 9 szczebli podziału, np. na 9 szczebli szczegółowości umieszczone są deskryptory SŁOWNIK INFORMACJI NAUKOWEJ, SŁOWNIK INFORMATYKI, ENCYKLOPEDIA BIBLIOTEKOZNAWSTWA. W dziale "Informatyka" najbardziej szczegółowe podziały sięgają 8 szczebli, np. deskryptory CD ROM i WORM umieszczone zostały na ósmym poziomie hierarchii. W działach "Bibliotekarstwo", "Bibliografia", "Czytelnictwo" i "Dziedziny pokrewne" (Językoznawstwo) najbardziej szczegółowy podział sięga do 4 szczebla hierarchii.

Ocena możliwości stosowania notacji komponentalnej i techniki rozbudowy równoległej kodów

Przyjęty w części systematycznej pionowy porządek deskryptorów praktycznie uniemożliwia stosowanie techniki rozbudowy równoległej kodów deskryptorów i wyposażenie TIN w notację komponentalną, gdyż nie zachowano w nim konsekwentnie stałej kolejności cytowania tych samych kategorii semantycznych w obrębie poszczególnych działów i faset, np.

TYPY SŁOWNIKÓW JIW

SŁOWNIK HASEŁ PRZEDMIOTOWYCH

TABLICE KLASYFIKACYJNE

TEZAUROS

WYKAZ SŁÓW KLUCZOWYCH

STRUKTURA SŁOWNIKA JIW

STRUKTURA TEZAUUSA

STRUKTURA TABLIC KLASYFIKACYJNYCH

STRUKTURA SŁOWNIKA HASEŁ PRZEDMIOTOWYCH

Aby usprawnić proces indeksowania i zapewnić związek między przekładową częścią alfabetyczną lub alfabetyczno-hierarchiczną tezauryusa i częścią o organizacji semantycznej, notacja przede wszystkim musi zapewnić utrwalenie porządku deskryptorów przyjętego w części semantycznej. Konstrukcją notacji należy więc podporządkować warunkom narzuconym przez ten właśnie porządek. W sytuacji, kiedy analogiczne kryteria podziału, czy kategorie semantyczne wymieniane są w różnej kolejności w różnych fasetach, nie jest możliwe równoczesne zapewnienie wskazania w kodach deskryptorów należących do danej kategorii ich miejsca w wykazie elementów danej fasety oraz występowania wśród cech składających się na treść deskryptora tej właśnie kategorii. W niektórych przypadkach, kiedy zachowana jest identyczna kolejność kryteriów podziału możliwe jest wykorzystanie do budowy kodów techniki rozbudowy równoległej, jakkolwiek przypadki takie w IIN są nieliczne, co uniemożliwia stosowanie tej techniki na większą skalę.

Konieczny poziom produktywności i metody jej zapewnienia

Zapewnienie produktywności systemu notacyjnego jest warunkiem jego trwałości. Prawidłowo skonstruowany system notacyjny musi gwarantować możliwość generowania nowych kodów dla deskryptorów wprowadzanych do tezauryusa w miarę jego aktualizacji, przy czym kody te muszą spełniać wszystkie warunki stawiane symbolom notacyjnym utworzonym dla pierwotnego zbioru leksykalnego. Tak więc kody te muszą prawidłowo lokalizować nowe deskryptory w semantycznej strukturze języka uwidocznionej w części systematycznej tezauryusa. Równocześnie kody nowych deskryptorów powinny być symbolami tego samego typu notacji, do którego należą kody deskryptorów wcześniej wprowadzonych do tezauryusa.

Zapewnienie tak rozumianej produktywności systemu notacyjnego jest bodaj najtrudniejszym problemem projektowania notacji i często powoduje znaczne skomplikowanie systemu notacyjnego, struktury generowanych przez niego kodów oraz ich wydłużenie, co w konsekwencji utrudnia posługiwanie się symbolami notacyjnymi i zwiększa prawdopodobieństwo popełnienia błędów przy ich zapisie. Z tego powodu określa się pewną skalę preferencji cech symboli notacyjnych, na początku której usytuowana jest zdolność prawidłowej lokalizacji deskryptora i optymalna sprawność użytkowa, czyli łatwość posługiwania się kodami oraz minimalizacja prawdopodobieństwa błędnego zapisu lub odczytania. Jeśli zachowanie innych cech notacji nie wpływa negatywnie na te dwie podstawowe własności, zaleca się zapewnienie zdolności ujawniania relacji hierarchicznych w strukturze kodów, a w dalszej kolejności innych relacji semantycznych oraz zapewnienie mnemoniki naturalnej kodów lub ich części (podobieństwa kodów lub ich fragmentów do postaci wyrażen naturalnych, których znaczenie reprezentują te kody).

Produktywność systemu notacyjnego należy rozpatrywać w dwóch podstawowych jej aspektach: tzw. gościnności systemu w szeregu klasyfikacyjnym oraz gościnności systemu w łańcuchu klasyfikacyjnym. Gościnność w łańcuchu klasyfikacyjnym oznacza zdolność dołączania do każdego z istniejących łańcuchów składających się na pewną klasyfikację nowych klas (w przypadku notacji dla tezauryusa - nowych deskryptorów szczegółowych), dla których system będzie w stanie wygenerować odpowiednie kody. W przypadku notacji hierarchicznej, którą po przeprowadzeniu analizy struktury części systematycznej TIN uznano za optymalny rodzaj systemu znakowania dla tego tezauryusa, gościnność w łańcuchu w sensie jego ekstrapolacji (wydłużania) jest praktycznie nieograniczona, choć warto zdawać

sobie sprawę z tego, iż jej realizacja zapewniona jest przez wydłużanie kodów, co w przypadku bazy notacyjnej o małej liczebności może znacznie utrudnić posługiwanie się kodami i zwiększyć prawdopodobieństwo ich błędnego zapisu lub odczytania.

Gościnność systemu notacyjnego w szeregu klasyfikacyjnym oznacza zdolność wprowadzania nowych klas (deskryptorów szczegółowych) do istniejących już szeregów klasyfikacyjnych, która w przypadku notacji hierarchicznej ograniczona jest liczbą znaków przewidzianych do oznaczenia elementów poszczególnych szeregów i mocą alfabetu (liczebnością bazy notacyjnej). Wykorzystanie metod ekstrapolacji szeregu opracowanych przez S.R. Ranganathana /4,5,6/ umożliwia przełamanie tego ograniczenia, w przypadku jednak tezauryasa zawierającego niespełna 1500 deskryptorów wydaje się, że metody te nadmiernie skomplikowałyby postać kodów, a w praktyce wykorzystywane byłyby w małym stopniu. Stąd też za najwłaściwsze uznano zapewnienie relatywnej produktywności systemu notacyjnego za pomocą tzw. rezerwy znaków. Za przyjęciem takiego rozwiązania przemawia także konieczność stosowania liczego magazynu znaków dla już istniejących szeregów klasyfikacyjnych - stosując jako bazę notacyjną dziesięcioelementowy zbiór cyfr arabskich przy liczebności szeregów przekraczającej 10 jednostek należy przyjąć wariant notacji hierarchicznej w postaci notacji setnej, która rezerwuje dla oznaczenia elementów każdego szeregu 100 znaków dwucyfrowych. W większości przypadków rezerwa znaków dla poszczególnych szeregów będzie więc obejmować 50 - 80 znaków, co automatycznie gwarantuje dość wysoką produktywność notacji i równocześnie pozwala efektywniej wykorzystać zbiór znaków przeznaczony dla szeregu (por. fragment Oznaczenia faset o zróżnicowanych kryteriach podziału).

W przypadku szeregów klasyfikacyjnych uporządkowanych alfabetycznie (np. wykazów identyfikatorów) technika rezerwy znaków dla poszczególnych szeregów nie zapewnia istotnej tu interpolacji szeregu, czyli wprowadzania nowych deskryptorów (identyfikatorów) wewnątrz istniejącego już szeregu na miejsce właściwe dla porządku alfabetycznego. W odniesieniu do tego rodzaju sytuacji przyjęto więc dodatkową zasadę, iż wartość porządkowa kodu zapewniająca właściwą lokalizację deskryptora czy identyfikatora w części systematycznej powinna być uznana za własność ważniejszą aniżeli zdolność ujawniania struktury semantycznej deskryptorów w strukturze formalnej ich kodów. Zasada ta pozwala w razie konieczności tworzyć kody notacji tysięcznej i dziesięciotysięcznej, jeśli kody generowane przez notację setną przyjętą dla całego słownictwa nie zapewniają odpowiedniej interpolacji szeregów uporządkowanych alfabetycznie, np.

E04.21	ABEL
E04.22	ALGOL
E04.23	APL
E04.24	BASIC
E04.25	COBOL
E04.26	FORTH
E04.261	FORTRAN
E04.27	JĘZYK C
E04.271	LISP
E04.28	LOGO
E04.29	PASCAL
E04.30	PL/1
E04.31	PROLOG
E04.32	SAKO

rezerwa znaków dla szeregu: E04.33 - E04.99

CHARAKTERYSTYKA SYSTEMU NOTACYJNEGO DLA TEZAUROSA INFORMACJI NAUKOWEJ

Na podstawie analizy struktury części systematycznej IIN, oceny wielkości jego zbioru leksykalnego oraz jego ilościowego rozkładu w poszczególnych częściach zastosowanej w tym Tezaurusie klasyfikacji fasetowej uznano, że optymalną formą projektowanej notacji jest hierarchiczny system notacji setnej korzystającej z mieszanej bazy notacyjnej, tj. zbioru wielkich liter alfabetu łacińskiego oraz zbioru cyfr arabskich. Zaprojektowana dla IIN notacja jest więc notacją alfanumeryczną, hierarchiczną, a w zakresie zasad tworzenia ciągów cyfrowych - notacją setną.

Forma kodu deskryptora

Kody deskryptorów szczegółowych mają postać ciągów literowo-cyfrowych, w których na pierwszym miejscu znajduje się wielka litera alfabetu łacińskiego, natomiast następne pozycje zajmują cyfry arabskie delimitowane za pomocą kropki po sekwencjach dwucyfrowych. Długość kodu jest wprost proporcjonalna do stopnia szczególności deskryptora, czyli szczebli hierarchii, na którym został on umieszczony w strukturze klasyfikacji fasetowej. Liczba pełniących funkcję delimitacyjną kropek umieszczonych po sekwencjach dwucyfrowych kodu wskazuje równocześnie stopień szczególności deskryptora w odniesieniu do szczebli hierarchii od drugiego do bezpośrednio nadrzędnego dla danego deskryptora - kropki nie umieszczono po symbolach literowych pierwszego poziomu hierarchii oraz na końcu kodu, czyli po dwucyfrowych oznaczeniach ostatniego poziomu hierarchii, np.

A02.02.44	INFORMACJA SPECJALISTYCZNA (3 szczebel hierarchii w dziale A - Informacja naukowa)
A07.01.31.01.22	INDEKS CYTOWAŃ (5 szczebel hierarchii w dziale A - Informacja naukowa)

Oznaczenia deskryptorów dziedzinowych

Ze względów praktycznych pożądanym jest, aby w trakcie indeksowania można było w sposób możliwie najprostszy zapewnić informowanie indeksatora o tym, do jakiej dziedziny należy deskryptor, którego ma on zamiar użyć w tworzonej przez siebie charakterystyce lub instrukcji wyszukiwawczej. Informacja taka ma niekiedy zasadnicze znaczenie dla właściwego interpretowania znaczenia poszczególnych deskryptorów, a rzadko można ją odnaleźć w artykułach deskryptorowych części alfabetyczno-hierarchicznej tezauryusa. Dlatego też w zaprojektowanej dla TIN notacji wszystkie deskryptory reprezentujące nazwy dziedzinowe otrzymały kody literowe wyraźnie wyodrębniające się w strukturze kodów szczegółowych. Dzięki zasadzie notacji hierarchicznej kody te znajdują się na początku każdego kodu deskryptora szczegółowego, lokalizując go tym samym w określonym polu dziedzinowym.

W TIN nazwy dziedzinowe należą do dwóch kategorii deskryptorów - kategorii deskryptorów dziedzinowych sensu stricto, które są nazwami głównych działów Tezauryusa oraz kategorii deskryptorów szczegółowych będących nazwami dziedzin pokrewnych informacji naukowej zawartymi w dziale VI "Dziedziny pokrewne". Deskryptorów pierwszej kategorii jest 5 (nie należy do niej jedynie nazwa działu szóstego), natomiast deskryptorów kategorii drugiej jest 15. Łączna liczba tego typu wyrażeń wskazuje więc, że najprostszą metodą wyróżnienia ich treści jest właśnie nadanie im oznaczeń literowych (dysponujemy tu 26 jednoelementowymi znakami), które

zawsze występować będą na początku kodu każdego deskryptora w TIN. Aby nie wydłużać kodów deskryptorów należących do działu VI "Dziedziny pokrewne" i zapewnić równocześnie jednolitą formę kodów wszystkich deskryptorów dziedzinowych, działowi temu nadano symbol podwójny złożony z dwóch dużych liter przedzielonych kreską ukośną; litera pierwsza jest oznaczeniem pierwszego w porządku pionowym TIN deskryptora dziedzinowego należącego do tego działu, litera druga natomiast jest oznaczeniem ostatniego zawartego w tym dziale deskryptora dziedzinowego. Poniżej podano wszystkie przyjęte w systemie notacyjnym TIN oznaczenia deskryptorów dziedzinowych:

- A INFORMACJA NAUKOWA (dział I)
- B BIBLIOTEKARSTWO (dział II)
- C BIBLIOGRAFIA (dział III)
- D CZYTELNICZTWO (dział IV)
- E INFORMATYKA (dział V)
- F/W DZIEDZINY POKREWNE (dział VI)
- H DZIEDZINY POKREWNE INFORMACJI NAUKOWEJ
- I ARCHIWISTYKA
- J BIBLIOFILSTWO
- K BIBLIOLOGIA
- L EDYTORSTWO
- M JĘZYKOZNAWSTWO
- N KSIĘGARSTWO
- P LOGIKA
- Q MUZEALNICTWO
- R NAUKOZNAWSTWO
- S NORMALIZACJA
- T POLIGRAFIA

U REPROGRAFIA
V STATYSTYKA
W TEORIA KOMUNIKACJI

Oznaczenia deskryptorów dziedzinowych dysponują następującą rezerwą znaków: F/G, X/Z; zgodnie z powszechnie stosowaną w odniesieniu do notacji alfanumerycznej praktyką, wyłączono z użycia literę 0 ze względu na jej podobieństwo do cyfry 0.

Kompresja

Zgodnie z zasadami notacji setnej połączonej ze stosowaną na jednym tylko szczeblu podziału notacją alfabetyczną oraz regułą delimitacji sekwencji dwucyfrowych za pomocą kropki, najdłuższy kod deskryptora należącego do TIN składałby się z 30 znaków - 1 litery, 20 cyfr i 9 kropek, np. A06.01.01.31.03.11.02.02.10.02 (słownik informatyki). Ponieważ prawdopodobieństwo uszczegółowienia deskryptorów znajdujących się na 9 szczeblu podziału w strukturze tego tezausa (obecnie są to takie deskryptory jak SŁOWNIK INFORMACJI NAUKOWEJ, ENCYKLOPEDIA BIBLIOTEKARSKA, CD ROM, WORM) wydaje się znikome, a istniejąca tu rezerwa znaków jest bardzo duża, od 9 szczebla hierarchii zaproponowano stosowanie kompresji, dzięki czemu maksymalna długość kodu nie powinna przekraczać 28 znaków i dla wszystkich deskryptorów szczegółowych znajdujących się na 8 i dalszych stopniach podziału w hierarchicznej strukturze słownictwa TIN powinna być jednakowa. W obecnej wersji Tezausa, dzięki korzystaniu z setnego wariantu hierarchicznej notacji numerycznej, kompresja 8 i 9 szczebla podziału nie powoduje utraty możliwości identyfikacji relacji hierarchicznych, gdyż dla szczebla 8 przyjęto oznaczenia ze zbioru znaków dwucyfrowych (10, 20; 30, 40, 50, 60, 70, 80, 90), a dla szczebla 9 - oznaczenia ze zbioru

ru znaków dwucyfrowych (11...19, 21...29,.....,91...99), przy czym kody deskryptorów sześciu 9 cyfrę przedostatnią mają taką samą jak kod deskryptora dla nich bezpośrednio nadrzędnego, np.

A06.01.01.01.31.03.11.02.02.10	SŁOWNIK TERMINOLOGICZNY
A06.01.01.01.31.03.11.02.02.11	SŁOWNIK INFORMACJI NAUKOWEJ
A06.01.01.01.31.03.11.02.02.12	SŁOWNIK INFORMATYKI

Oznaczenia faset o zróżnicowanych kryteriach podziału

Dla oznaczenia elementów faset, w których deskryptory uporządkowane są według więcej niż jednego kryterium, przyjęto zasadę podziału zarezerwowanych dla oznaczenia poszczególnych szeregów klasyfikacyjnych 99 znaków na dziewięcioelementowe sektory, z których na ogół każdy służy do znakowania grupy deskryptorów podporządkowanych jednemu kryterium. W nielicznych przypadkach, tam gdzie grupa deskryptorów jest liczniejsza niż 9 jednostek, lub też gdzie uznano, że istnieje duże prawdopodobieństwo szybkiego pojawienia się znacznej liczby nowych deskryptorów należących do tej grupy, zarezerwowano po dwa lub trzy sektory znaków. Poniżej podano przykład oznaczenia rozbudowanej fasety:

A07.01.31	INDEKS
A07.01.31.01	RODZAJE INDEKSÓW (ze względu na układ) - sektor 11 - 19
A07.01.31.01.11	INDEKS FORMALNY
A07.01.31.01.11.01	INDEKS ALFABETYCZNY
A07.01.31.01.11.02	INDEKS AUTORSKI
A07.01.31.01.11.03	INDEKS CHRONOLOGICZNY
A07.01.31.01.11.04	INDEKS TYTUŁOWY
A07.01.31.01.12	INDEKS RZECZOWY
A07.01.31.01.12.01	INDEKS KLASOWY

A07.01.31.01.12.02.	INDEKS PRZEDMIOTOWY
A07.01.31.01.12.03	INDEKS SŁÓW KLUCZOWYCH
A07.01.31.01.12.04	INDEKS SYSTEMATYCZNY
A07.01.31.01.13	INDEKS KRZYŻOWY (ze względu na charakter haseł) - sektor 21 - 29
A07.01.31.01.21	INDEKS CYTOWAŃ
A07.01.31.01.22	INDEKS NAZW
A07.01.31.01.22.01	INDEKS NAZW GEOGRAFICZNYCH
A07.01.31.01.22.02	INDEKS NAZW INSTYTUCJI
A07.01.31.01.22.03	INDEKS NAZW OSOBOWYCH (ze względu na rodzaj zbioru, do którego sporządzony jest indeks) - sektor 31 - 39
A07.01.31.01.31	INDEKS DO CZASOPISM
A07.01.31.01.32	INDEKS DO KATALOGU
A07.01.31.01.33	INDEKS DO KSIĄŻKI
A07.01.31.01.34	INDEKS DO TABLIC (KLASYFIKACYJNYCH)
A07.01.31.01.35	INDEKS DO WYDAWNICTW INFORMACYJNYCH
A07.01.31.01.35.01	INDEKS DO BIBLIOGRAFII (ze względu na technikę sporządzania) - sektor 41 - 49
A07.01.31.01.41	INDEKS ŁAŃCUCHOWY
A07.01.31.01.42	INDEKS PERMUTACYJNY
A07.01.31.01.42.01	INDEKS KWIC
A07.01.31.01.42.02	INDEKS KWOC
A07.01.31.01.42.03	INDEKS WADEX
A07.01.31.01.43	INDEKS ROTACYJNY (ze względu na środki służące do sporządzania i wykorzystywania indeksu) - sektor 51 - 59
A07.01.31.01.51	INDEKS AUTOMATYCZNY
A07.01.31.01.52	INDEKS TRADYCYJNY
A07.01.31.01.53	INDEKS KUMULACYJNY

Oznaczenia modyfikatorów

W obecnej wersji Tezaurusu Informacji Naukowej modyfikatory umieszczone zostały na początku trzech działów głównych: "Informacja naukowa", "Bibliotekarstwo" i "Informatyka". W przyszłości przewidziana jest łączna lista wszystkich modyfikatorów, które mają być wyrażeniami o nieograniczonej formalnie łączliwości, tj. można je będzie łączyć w syntagmy z dowolnymi deskryptorami, niezależnie od tego w jakim dziale są te deskryptory umieszczone. Ze względu na odmienne funkcje zarówno semantyczne, jak i wyszukiwawcze tej kategorii wyrażań języka deskryptorowego, powinno się zróżnicować także ich kody w stosunku do formy kodów deskryptorów. Dzięki temu kod syntagmy deskryptor - modyfikator wyraźnie zostanie odróżniony od kodów szczegółowych deskryptorów i nie będzie komplikował automatycznego rozpoznawania deskryptorów powiązanych relacjami hierarchicznymi. Obecnie dla oznakowania modyfikatorów wykorzystano małe litery alfabetu łacińskiego - liczebność tego zbioru znaków zapewnia znaczną produktywność podsystemu notacyjnego przeznaczonego dla modyfikatorów, przy czym w razie dużej rozbudowy samego systemu modyfikatorów i uporządkowania go za pomocą relacji hierarchicznych, będzie można opracować niezależny system notacji hierarchicznej alfabetycznej dla tej kategorii jednostek leksykalnych IIN.

Określenie wartości porządkowej elementów bazy notacyjnej IIN. Zapis kodów w części systematycznej i w części alfabetyczno-hierarchicznej Tezaurusu

Opracowany system notacyjny utrwała pionowy porządek deskryptorów przyjęty w części systematycznej Tezaurusu, toteż kody w tej części umieszczono w lewej kolumnie tak, aby poprzedzały odpowia-

dające im Deskryptory. Zgodnie z porządkiem pionowym jako pierwsze w każdym dziale cytowane są modyfikatory, które otrzymały kody o najwyższej wartości porządkowej (małe litery alfabetu łacińskiego poprzedzają więc w przyjętym systemie wartości znaków bazy notacyjnej IIN zarówno litery wielkie, jak i cyfry arabskie). Po wykazie modyfikatorów w części systematycznej następuje wykaz deskryptorów uporządkowanych w działach, z których każdy otwiera deskryptor dziedziczny o kodzie w postaci wielkiej litery alfabetu łacińskiego - litery te mają więc wartość porządkową niższą od wartości małej litery z, ale wyższą od wartości porządkowej cyfry 0. Kod każdego deskryptora szczegółowego jest ciągiem znaków alfanumerycznych, których pierwszym elementem jest wielka litera łacińska, a następnymi cyfry arabskie rozdzielane co dwa znaki kropką. Cyfry arabskie mają więc wartości porządkowe niższe od wartości litery Z, ale wyższe od wartości porządkowej przypisanej kropce. Ostatni wreszcie element bazy notacyjnej systemu znakowania opracowanego dla IIN, kreska ukośna wykorzystana do utworzenia znaku podwójnego dla kodu działu VI, ma przypisaną najwyższą wartość porządkową w całym systemie notacyjnym, co zapewnia możliwość tworzenia dowolnych znaków podwójnych, które w porządku pionowym poprzedzać będą wykaz kodów zawierających się w przedziale objętym przez ten znak podwójny. Tak więc przyjęto następujący porządek znaków w bazie notacyjnej IIN: / a/z A/Z 0/9 .

W części alfabetyczno-hierarchicznej tezauryza kody deskryptorów pełnią przede wszystkim funkcję adresu wskazującego miejsce deskryptora w części systematycznej, wzbogacając jednocześnie informacje o treści poszczególnych deskryptorów zawarte w artykule deskryptorowym lub askryptorowym, toteż uznano, że najwygodniej jest umieścić je bezpośrednio po każdym deskryptorze. Kody umiesz-

czono więc zarówno w artykułach deskryptorowych przy wszystkich ich elementach będących deskryptorami, jak i przy deskryptorach w artykułach askryptorowych oraz przy deskryptorach cytowanych zarówno w części alfabetyczno-hierarchicznej, jak i w części systematycznej w tekstach instrukcji dotyczących indeksowania.

Literatura

1. CHMIELEWSKA-GORCZYCA E. Tezaurus Informacji Naukowej. Zagad. Inf. Naukowej 1989 nr 2(56) s.51-96
2. ISO 2788-1986(E) Documentation - Guidelines for the establishment and development of monolingual thesauri. International Organization for Standardization 1986
3. PN-81/N-09018 Tezaurus jednojęzyczny dla polskich systemów informacyjnych. Warszawa: Wydaw. Normalizacyjne 1981
4. RANGANATHAN S.R. Prolegomena to Library Classification. 3rd ed. London: Asia Publishing House 1967
5. SOSIŃSKA-KALATA B. Relacje między planem treści i planem wyrażania w językach informacyjno-wyszukiwawczych. Warszawa: CİNTE 1989
6. UNGURIAN O. Teoria i praktyka klasyfikacji fasetowej S.R. Ranganathana. Warszawa: CİNTE 1975
7. TEZAURUS Informacji Naukowej. Oprac. E. Chmielewska-Gorczyca. Warszawa: OIN PAN 1990 (maszynopis).

14.03.1991 r.

PROJECT OF THE NOTATIONAL SYSTEM FOR THE THESAURUS OF INFORMATION
SCIENCE

Characteristics of the structure and methods of establishment

Summary

There are presented aims and general rules of working out of the notational systems for thesauri. Methods of elaboration of the notation for the Thesaurus of Information Science (TIN - Tezaurus Informacji Naukowej) to be used in the information retrieval system on information science (WIST - Wyszukiwawczo-Informacyjny System Tezaurusowy) are discussed in detail including: analysis of the structure of the systematic part, evaluation of the quantity of terms in the TIN and of their distribution in the adopted classification system, evaluation of the detailness of the used classification, evaluation of possibilities of use in TIN of the componental notation and of the techniques of the parallel extension of codes, determining of the optimal level of the productivity of the notational system together with the methods of ensuring such productivity.

The designed notational system for the Thesaurus of Information Science is described - it makes the hierarchical system of decimal notation in which mixed notational basis was used. The form of descriptor's and of modifier's codes in the proposed TIN's notational system is presented as well as marking of the field descriptors, rules of compression of hierarchy, marking of facettes based upon the differentiated criterions of partition, ordinal values of the elements of notational basis for the TIN's system of marking and rules of recording codes in this thesaurus.

Проект нотационной системы для Тезауруса научной информации.
Характеристика структуры и методика разработки.

Резюме

Представлены цели и общие принципы создания нотационных систем для тезаурусов. Подробно обсужден метод разработки нотации для Тезауруса научной информации (TIN - *Thesaurus Informacii Naukovej*) информационно-поисковой системы в области научной информации (*Wyszukiwaczo-Informacyjny System Thesaurusowy - WIS*). Проведен анализ его систематической части, оценка количества лексики содержимой в Тезаурусе и её дистрибуции в отдельных частях принятой классификационной системы, оценка детальности использованной классификации, оценка возможности использования в TIN компонентальной нотации и техник параллельного развертывания кодов, определения оптимального уровня продуктивности нотационной системы и методов её обеспечения.

Охарактеризована проектируемая нотационная система для Тезауруса научной информации - она является иерархической системой сотовой нотации использующей смешанную нотационную базу. Обсуждена форма кодов дескрипторов и модификаторов в предлагаемой нотационной системе, обозначения отраслевых дескрипторов, принципы использования компрессии иерархии, обозначения фасет обоснованных на разных критериях деления, порядковые стоимости элементов нотационной базы системы обозначений для TIN и принципы записи кодов в Тезаурусе.

RECENZJE I OMÓWIENIA

HIPERTEKSTEM O HIPERTEKSCIE

"Hypertext Hands-On!" jest jedną z niewielu pozycji książkowych dotyczących nowej techniki organizowania informacji - hipertekstu. Jak wiele innych nowatorskich dziedzin hipertekst był do tychczas przedmiotem przede wszystkim doniesień naukowych, raportów, referatów konferencyjnych i artykułów w czasopismach naukowych. Książka "Hypertext Hands-On!" stanowi ich podsumowanie. Zgodnie z podtytułem - "An Introduction to a New Way of Organizing and Accessing Information" może służyć jako wprowadzenie do zagadnień hipertekstu. Szczególnie atrakcyjne jest wyposażenie poszczególnych egzemplarzy w dyskietki z tekstem książki zapisanym przy pomocy oprogramowania Hyperties.

Książka składa się z części wstępnej, dziewięciu rozdziałów i pięciu aneksów. Rozdziały dzielą się z kolei na drobne, jedno lub dwustronicowe podrozdziały zwane tu artykułami, z których każdy napisany jest tak, by mogły być czytane w dowolnej kolejności.

"Hypertext Hands-On!" zaczyna się, nietypowo jak na publikację naukowo-techniczną, pięknym filozoficznym wierszem Roberta Frosta pt. "The Road Not Taken". W przedmowie autorzy - Ben Shneiderman i Greg Kearsley przedstawiają pokrótce założenia, jakimi kierowali się przy opracowywaniu niniejszej książki, a we wstępie omawiają

* SHNEIDERMAN B., KEARSLEY G.: Hypertext Hands-On! An Introduction to a New Way of Organizing and Accessing Information. Reading Mass.: Addison - Wesley 1989, 187 s.

jej treść. Pierwszy rozdział poświęcony jest podstawowym pojęciom i ideom hipertekstu. Opisana jest struktura baz hipertekstowych, zasady ich budowy i eksploatacji. W drugim rozdziale wymienione są dotychczasowe dziedziny zastosowań. Są wśród nich wydawnictwa komputerowe (encyklopedie, słowniki, katalogi), wspomaganie zespołowej twórczości naukowej, inżynierskiej czy literackiej oraz wszelkiego rodzaju materiały instruktażowe. Rozwinięcie tego tematu znajduje się w rozdziałach: szóstym i dziewiątym poświęconych perspektywom dalszych zastosowań, przy wykorzystaniu innych narzędzi informatycznych takich jak techniki sztucznej inteligencji, hipermedia etc.

Problemy występujące przy projektowaniu hipertekstowych baz danych, zarówno specyficzne dla technik hipertekstu jak i ogólne wymagania jakościowe stawiane wszystkim użytkowym produktom informatycznym, są przedmiotem ponad trzydziestu krótkich artykułów pogrupowanych w trzy duże rozdziały. Są to "System Design Issues", "Implementation Issues" oraz "Authoring". Dwa pierwsze tzn. problemy projektowania i implementacji systemów nie wymagają dodatkowych wyjaśnień, natomiast "Authoring" oznacza proces tworzenia dokumentów hipertekstowych.

W rozdziale szóstym opisane są w porządku chronologicznym najważniejsze systemy hipertekstowe, a w rozdziale siódmym krótkie biografie ich twórców. Przegląd systemów rozpoczyna opis automatu Memex, którego koncepcję opracował w połowie lat pięćdziesiątych Vannevar Bush, uważany za twórcę idei hipertekstu. Kolejne pionierskie przedsięwzięcia to projekt XANADU Teda Nelsona oraz AUGMENT Douglasa Engelbarta. Następnie przedstawiono współczesne rozwiązania, zarówno eksperymentalne, stanowiące przedmiot badań naukowych, jak i produkty komercyjne. Dla wersji komercyjnych opracowanych na mikrokomputery IBM PC podano adresy dystrybutorów tych pakietów.

Jeden z artykułów poświęcony został oprogramowaniu Hyperties. W związku z tym przedstawiono sylwetkę Bena Shneidermana, autora niniejszej książki, gdyż był on szefem projektu Hyperties na Uniwersytecie w Maryland, USA.

"Hypertext Hands-On!" zawiera bibliografię, opis wersji komputerowej tekstu książki, raport z procesu tworzenia obu wersji - drukowanej i komputerowej, słownik definicyjny i indeks przedmiotowy.

Większość twórców prezentowanych systemów hipertekstowych to autorzy prac wymienionych w załączonej bibliografii oraz uczestnicy konferencji "Hypertext 87" - pierwszej dużej imprezy naukowej poświęconej hipertekstowi. Książka zawiera również wykaz materiałów z tej konferencji.

Zgodnie z zainteresowaniami Bena Shneidermana, m.in. autora publikacji "Strategies for Effective Human-Computer Interaction", szefa Human-Computer Interaction Laboratory Uniwersytetu w Maryland, w omawianej publikacji położony jest duży nacisk na aspekt użytkowy techniki hipertekstu, szeroko omawiany jest tzw. "czynnik ludzki" w systemach informatycznych. Wymienione zostały organizacje naukowe oraz wydawnictwa podejmujące tę problematykę.

Założeniem autorów było odwzorowanie struktury bazy hipertekstowej w układzie książki "Hypertext Hands-On!". Ponieważ wersja komputerowa oparta jest na oprogramowaniu Hyperties, przyjęte zasady organizacji tekstu książki odpowiadają ideowo rozwiązaniom tego oprogramowania.

Podstawową jednostką bazy hipertekstowej jest węzeł (node), w Hyperties zwany artykułem. Każdy artykuł posiada tytuł, stanowiący odsyłacz do tego artykułu, identyfikujący go. Nazwa tytułu musi być więc unikalna na poziomie całej bazy. W wersji książkowej,

oprócz tytułu właściwego, wyodrębnionego poprzez zapisanie wersalikami jest jeszcze podtytuł. Zawartość treściowa podtytułów pełni rolę wyjaśniającą w stosunku do tytułów właściwych, przy okazji będąc w wielu wypadkach dowodem rozwiniętego poczucia humoru autorów.

Jakkolwiek przyjęte oprogramowanie nie narzuca ograniczeń co do długości artykułów, gdyż w wersji maszynowej zastosowano stronowanie, autorzy starali się zmieścić na jednej lub dwu stronach formatu książki, toteż poszczególne podrozdziały (artykuły) są opracowane w sposób skondensowany, niestety niekiedy kosztem jasności i szczegółowości zawartych w nich informacji. Zresztą celem "Hypertext Hands-On!" nie jest wyczerpujące, analityczne omówienie hipertekstu, lecz wprowadzenie do tej tematyki, a częste odwołania do bogatej bibliografii pozwalają zainteresowanemu czytelnikowi na zgłębienie interesujących go zagadnień. W tekstach artykułów znajdują się terminy specjalnie wyodrębnione poprzez wytłuszczenie druku. Umieszczone przy nich w nawiasach liczby oznaczają numery stron książki. Aby przeczytać inny artykuł, poświęcony danemu problemowi, zaopatrzony tytułem o brzmieniu takim samym jak oznaczony termin, należy otworzyć książkę na stronie podanej w nawiasach. W ten oto sposób realizowane jest w formie tradycyjnego wydawnictwa połączenie sieciowe w bazie hipertekstowej.

Połączenia te (links) wraz z węzłami stanowią zasadnicze elementy struktury hipertekstu. Opisane powyżej rozwiązanie wydaje się banalne i być może nie usprawnia posługiwania się książką w sposób jakościowo różny niż tradycyjny spis treści i dobry indeks przedmiotowy, ale należy pamiętać, że podstawowym celem była tu prezentacja zasady budowy i funkcjonowania baz hipertekstowych przy pomocy tradycyjnego medium - książki. Dlatego więc to, co w formie książko-

wej może być uznane jedynie za rodzaj tricku, ewentualnie drobnego usprawnienia, w przypadku pełnotekstowych baz komputerowych prowadzi do wręcz rewolucyjnych zmian w tworzeniu i wykorzystywaniu systemów informacyjnych, zważywszy, że hipertekstowej strukturze towarzyszy wiele wyrafinowanych narzędzi programowych, służących do tworzenia jak i eksploatacji poszczególnych baz.

Dodatkowo tytuły artykułów tworzą słownik definicyjny, na który składa się ponad 120 terminów, zamieszczony na końcu książki (Hyper Glossary). Także w słowniku przy terminach umieszczono odwołania (w postaci numerów stron) do odpowiadających im artykułów. Zgromadzenie definicji na końcu książki jest odstępstwem od stosowanego w systemie Hyperties ulokowania definicji w zakończeniu poszczególnych artykułów, jak to zostało zrealizowane w wersji komputerowej "Hypertext Hands-On". Dzięki takiemu rozwiązaniu użytkownik w niektórych przypadkach mógłby się zadowolić objaśnieniem wyodrębnionych w artykule terminów bez potrzeby przywoływania nowych artykułów, ich właśnie dotyczących, czyli bez wykorzystania połączenia (link) z nowym węzłem.

Struktura, jaką umożliwia Hyperties, jest niehierarchiczna a kolejność artykułów przypadkowa. Wyjątek stanowi tzw. "root document", który w Hyperties nazwano artykułem wstępnym (Introductory article). W wersji książkowej rolę tę pełni początkowy artykuł zatytułowany po prostu "Introduction", stanowiący krótkie omówienie problematyki poruszanej w książce.

Ze względu na wymogi tekstu publikowanego należało przyjąć jakąś logiczną kolejność poszczególnych artykułów. Dodatkowo, dla wygody przyszłego czytelnika pogrupowano je w dziewięć tradycyjnych rozdziałów. Aby nie łamać hipertekstowej konwencji każdy z rozdziałów zaopatrzony jest w rodzaj lokalnego artykułu wstępnego,

co można interpretować jako propozycję dla czytelnika zapoznawania się z materiałem książki w kolejności zalecanej przez autorów. Analogicznie w wersji komputerowej twórcy baz hipertekstowych spośród wielu możliwych "ścieżek" (paths) sugerują optymalne metody poruszania się po bazie, preferowane "trasy" (tours), zatem taki a nie inny układ artykułów w książce pogrupowanych w rozdziały, których spis znajduje się także w "Introduction", jest po prostu preferowaną przez autorów "trasą" pomocną przy korzystaniu z maszynowej wersji "Hypertext Hands-On!".

Pełny spis treści, - "Contents" - w którym wymienione są wszystkie tytuły artykułów. poprzedzający artykuł wstępny ("Introduction") ma odpowiednik w postaci TABLE OF CONTENTS, natomiast indeks przedmiotowy zamieszczony na końcu książki przeznaczony jest tylko dla wersji tradycyjnej, gdyż INDEX komputerowy tworzą tytuły artykułów w porządku alfabetycznym. Natomiast funkcje indeksu przedmiotowego, znacznie poszerzone, realizowane są w bazie hipertekstowej za pomocą procedury SEARCH.

W książce zamieszczony jest materiał ilustracyjny w postaci wykresów oraz przykładowych wydruków ekranowych przy użyciu różnych systemów hipertekstowych. Ograniczenia mikrokomputera typu IBM PC, na którym zaimplementowano oprogramowanie Hyperties w wersji demonstracyjnej, uniemożliwiły włączenie ich do bazy hipertekstowej, natomiast znalazły się tam dodatkowo przykłady zastosowań hipertekstu, także z wykorzystaniem grafiki komputerowej. Zademonstrowano fragmenty przewodnika turystycznego, katalogu budowlanego, instrukcji podatkowej, a nawet próby pisania powieści i żartów językowych.

Jak wspomniano wyżej, wersja maszynowa "Hypertext Hands-On!" opracowana jest na mikrokomputer kompatybilny z IBM/PC. Podstawowe wymagania sprzętowe to pamięć 384K, dwie stacje dysków elastycznych

5 1/4 i jeden dysk twardy. Monitor może być monochromatyczny lub kolorowy (CGA, EGA lub VGA), ale na monitorze monochromatycznym niektóre funkcje wersji demonstracyjnej hipertekstowej bazy "Hypertext Hands-On!" będą pominięte. Można nawet posługiwać się mikrokomputerem o pamięci 256K, lecz wówczas dostępna będzie tylko procedura przeglądania bazy bez możliwości wyszukiwania, gdyż wyłączona będzie funkcja SEARCH. Instalacja systemu jest bardzo prosta. Obszerna instrukcja zamieszczona w książce omawia również sytuacje awaryjne.

System Hyperties, jak wspomniano wyżej, został opracowany na Uniwersytecie w Maryland, a szefem zespołu projektowego był Ben Shneiderman. System powstał w 1983 roku. Początkowo nosił nazwę TIES - The Interactive Encyclopedia System i w wersji eksperymentalnej zainstalowany był na mikrokomputerach SUN. Wersja komercyjna na IBM/PC rozpowszechniana jest przez Cognetics Corporation. Notabene formularz zamówienia na pełne oprogramowanie w cenie ok. 350 dol. dołączony jest do dwóch dyskietek demonstracyjnych.

Pełne oprogramowanie składa się z dwu podstawowych modułów. Dyskietki zawierają tylko jeden z nich (zwany "browser"), który służy do przeglądania gotowej bazy bez możliwości zmiany jej zawartości. Drugi z modułów podstawowych ("authoring system") przeznaczony jest do tworzenia bazy. W książce zamieszczono opis obu modułów. Funkcjonowanie przeglądania bazy i wyszukiwania informacji można przetestować dzięki udostępnionym dyskietkom, natomiast w zapoznaniu się z problemami kreowania własnej bazy hipertekstowej pomocne są refleksje autorów zawarte w epilogu ("Epilog. The Making of Hypertext Hands-On!"). Shneiderman i Kearsley przedstawiają tam swoje doświadczenia przy tworzeniu książki i jej wersji komputerowej. Mimo wielu trudności, jak np. braku dobrych procedur

modyfikacji zawartości bazy w oprogramowaniu Hyperties, uważają oni pracę nad "Hypertext Hands-On!" za niezwykle satysfakcjonującą i inspirującą. Zechęcają też przvszłych czytelników do tworzenia własnych hiperbaz, poczynając od najprostszych, z wykorzystaniem różnych dostępnych systemów hipertekstowych. Sami z kolei obiecują, że w ciągu najbliższych lat opracują nową edycję z uwzględnieniem dodatkowych możliwości, m.in. podziału ekranu monitora na tzw. okna czy wpisywania uwag użytkownika (annotations), których obecnie komercyjna realizacja oprogramowania Hyperties na IBM/PC jest pozbawiona.

W "Epilogu" zamieszczono również fragmenty opinii studentów z Uniwersytetu w Maryland na temat książki i odpowiadającej jej bazy danych. Szkoda, że nie udało się zamieścić ich w formie komentarzy (annotations) do poszczególnych artykułów. Większość przytoczonych ocen jest pozytywna, chociaż nie brakuje uwag krytycznych. Generalnie należy podkreślić, że było to bardzo pożyteczne przedsięwzięcie, a pewne niedociągnięcia nie powinny wpływać na ostateczną ocenę pozytywną. Charakterystyczne, że znacznie większą aprobatą cieszy się wersja książkowa. Istotnie jest napisana w sposób przystępny. W skondensowanej formie zapoznaje czytelnika z dotychczasowym rozwojem technik hipertekstowych i, co ważniejsze, wyjaśnia podstawowe zasady hipertekstu. Może być wykorzystywana jako dobry podręcznik dla adeptów tej nowej dziedziny. Z kolei pozbawiona jest informacji natury technicznej, umożliwiających zapoznanie się z warstwą programową hipertekstu. Szczególnie daje się to odczuć w rozdziale poświęconym poszczególnym systemom, których opisy są zbyt lakoniczne. Książka jest nieco chaotyczna. Nie tylko poprzez pewną przypadkowość kolejności artykułów, co w końcu jest dopuszczalne, skoro ma ona odwzorowywać strukturę bazy hipertekstowej, ale

i przez ich redundancję. Wynika to prawdopodobnie z podziału pracy nad tekstem przez dwu autorów, o czym sami wspominają w podrozdziale (artykule) epilogu "Collaboration at a distance".

Zaletą książki jest rozbudzenie zainteresowania nową techniką informacyjną. Entuzjazm autorów udziela się czytelnikowi i niestety w kontekście nadmiernie rozbudzonych oczekiwań zaprezentowany przykład realizacji komputerowej musi rozczarować. Nawet świadomość, że nie mamy doczynienia z pełnym oprogramowaniem, a tylko spreparowaną do celów demonstracyjnych częścią użytkową, nie usprawiedliwia wszystkich niedoskonałości. W ogóle nasuwa się przypuszczenie, że znacznie więcej satysfakcji ze stosowania nowych, wyrafinowanych rozwiązań mają twórcy hipertekstowych baz danych niż ich późniejsi użytkownicy.

Korzystanie z programu "browser" nie następuje problemów dzięki stosowaniu kursora do wyboru poszczególnych opcji. Jedyne instrukcja SEARCH wymaga napisania wyrażenia wyszukiwawczego. "Browser" pozwala na przeglądanie bazy poprzez wybór połączeń (links). Połączenia można wybierać z indeksu, ze spisu treści lub z tekstów artykułów. SEARCH ma cztery warianty - wyszukiwanie pojedynczego terminu, dwóch terminów przy użyciu operatorów AND oraz OR i dwóch terminów oddzielonych spacją. Procedura HISTORY służy do wyświetlania tytułów artykułów przeglądanych od początku sesji wyszukiwawczej. Wbudowany program pomocniczy uruchamiany instrukcją CTRL-H, do którego można odwołać się w razie wątpliwości, jest bogato rozbudowany. Jest to więc produkt łatwy w użytkowaniu i dobrze ilustrujący najprostsze funkcje bazy hipertekstowej, ale nic ponad to. Można mówić o niewątpliwym postępie w porównaniu z typowymi systemami baz danych czy edytorów tekstu, lecz oczekiwania idą dalej, tym bardziej, że znane są realizacje z uwzględnieniem elementów

szucznej inteligencji (np. KnowledgePro - hipertekstowy system ekspertowy), hipermediów oraz pracy nad tworzeniem i eksploatacją hiperbaz rozproszonych w sieciach komputerowych.

Rosnąca liczba eksperymentalnych i komercyjnych realizacji z wykorzystaniem techniki hipertekstu świadczy, że jej obecność na rynku informacyjnym w krajach rozwiniętych stała się faktem. Wszystkie liczące się konferencje, sympozja itp. spotkania przedstawicieli nauki i przemysłu informacyjnego w ciągu ostatnich kilku lat uwzględniają tę tematykę. Podczas kongresu "The First European Congress for the Education of Information for Europe: Information Specialists for Europe" (Hannover 17-19 X 1989) wyrażono potrzebę włączenia zagadnień hipertekstu do rutynowych programów szkolenia specjalistów informacji. Także w Polsce niektóre placówki zajmują się tą problematyką, m.in. Politechnika Wrocławska, IBIN UW oraz OIN PAN. Bariery jest trudny dostęp do zachodnich pakietów programowych oraz do najnowszych publikacji. Można mieć nadzieję, że "Hypertext Hands-On!" dostępny w Bibliotece OIN PAN w Warszawie wraz z dyskietkami demonstracyjnymi przynajmniej częściowo wypełni tę lukę.

Krystyna Siwek

26.03.1991 r.

KRONIKA

DZIAŁALNOŚĆ SEMINARYJNA OIN PAN

w 1990 r.

Rok 1990 był piątym kolejnym rokiem organizowania przez OIN PAN seminariów naukowych. Ich tematyka była dość zróżnicowana: spośród przygotowanych 9 spotkań seminaryjnych, cztery dotyczyły problematyki języków informacyjno-wyszukiawczych, cztery - zautomatyzowanych systemów informacyjnych i jedno - naukoznawstwa.

W grupie omawiającej języki informacyjno-wyszukiawcze znalazły się m.in. 2 wykłady doc.dr hab. Eugeniusza Ścibora: "Języki informacyjne w bazach danych"^y i "Polska Klasyfikacja Tematyczna" oraz "Tezaurus zagadnień wspólnych". W pierwszym z nich Referent przedstawił i scharakteryzował różnego rodzaju języki informacyjne, zwane niekiedy środkami językowymi. Należą do nich m.in. słowniki, indeksy, słowa kluczowe, deskryptory, wykazy słów kluczowych. Różne języki informacyjne mają różne zastosowanie w bazach danych zależnie od ich rodzaju, przy czym można wyodrębnić dwa podziały: na bazy danych pełnotekstowe i niepełnotekstowe oraz bazy źródłowe (faktograficzne) i dosyłaające (dokumentacyjne). Autor przedstawił wnioski wynikające z analizy katalogu światowych baz danych, odnoszące się do stosowania języków informacyjnych i występujących najczęściej prawidłowości w zakresie doboru określonych języków do pewnych typów baz.

^y zob.: L.A. Bielicka, E. Ścibor: Środki językowe stosowane w bazach danych. ZIN 1990 nr 2(57) s. 109-145.

W drugim wykładzie doc.dr hab. E. Ścibor scharakteryzował dwa znane języki informacyjne i sposoby ich wykorzystywania, przedstawiając równocześnie historię ich powstawania na tle przemian organizacyjnych informacji naukowej w Polsce.

W referacie mgr Marka Nahotki na temat "Klasyfikacje fasetowe" przedstawiono przykłady i charakterystykę tych klasyfikacji, poświęcając równocześnie sporo miejsca ich genezie, a także ich twórcom takim jak Ranganathan i Vickery.

Referat dr Barbary Sosińskiej-Kalaty - "Systemy notacyjne w językach informacyjnych"^{2/} - stanowił próbę typologii systemów notacyjnych według różnych kryteriów: rodzaju bazy notacyjnej, rodzaju generowanego słownictwa, zdolności ujawniania struktury semantycznej wyrażen języka informacyjno-wyszukiwawczego (JIW). Omawiając poszczególne rodzaje notacji Autorka zwróciła szczególną uwagę na zasady, jakimi należy się kierować przy wyborze notacji do projektowanego JIW.

Referat mgr inż. Jacką Pańnika (grupa zautomatyzowane systemy informacyjne) dotyczył charakterystyki i zastosowań pakietu Mikro CDS/ISIS. Zmienna długość pól i wynikająca z niej zmienna długość rekordu sprawia, że pakiet ten ma głównie zastosowanie w systemach bibliotecznych i bibliograficznych. Autor przedstawił szczegółowo możliwości wykorzystania tego pakietu.

Dokum.dypl. mgr inż. Krystyna Siwek przedstawiła "Wykorzystanie doświadczeń budowy systemu informacji z dziedziny informacji naukowej przy tworzeniu systemów dziedzinowych". System informacyjny z dziedziny informacji naukowej^{3/}, którego budowa została

^{2/}Zob. B.Sosińska-Kalata: Systemy notacyjne w językach informacyjno-wyszukiwawczych. Próba typologii. ZIN 1990 nr 2(57) s.21-44.

^{3/}Zob.K.Siwek: System informacji o informacji. ZIN 1989 nr 2(55) s. 29-50.

zapoczątkowana w 1989 r., składa się z 4 modułów funkcjonalnych: Tezaurusu Informacji Naukowej^{4/}, bibliograficznej bazy danych opisów artykułów z czasopism polskich z dziedziny informacji naukowej i dziedzin pokrewnych, zautomatyzowanego katalogu czasopism biblioteki OIN PAN oraz biblioteki IINTE oraz baz faktograficznych tworzących system informacji skierowującej. Tezaurus Informacji Naukowej składa się ze wstępu, części systematycznej i alfabetycznej. Zawiera około 1600 artykułów deskryptorowych (370 stron wydruku z bazy komputerowej założonej przy pomocy adytora Chiwriter). Tezaurus powstał równolegle z bibliograficzną bazą WIST (Wyszukiwawczo-Informacyjny System Tezaurusowy) zawierającą obecnie ponad 1800 abstraktów i był wykorzystany do indeksowania tej bazy. Zautomatyzowany katalog jest zaczątkiem centralnego dziedzinowego katalogu zbiorów z zakresu informacji naukowej. Bazy faktograficzne są w stadium eksperymentalnym.

Mgr inż. Andrzej Zienkiewicz z Centrum Informatycznego Uniwersytetu Warszawskiego (CIUW) omówił powstanie i zasady działania Europejskiej Sieci Akademickiej i Badawczej - European Academic and Research Network - EARN, oraz podłączenie Polski do tej Sieci. Węzeł krajowy EARN w Polsce znajduje się na terenie CIUW, gdzie również odbywały się zajęcia seminarium, co dało uczestnikom możliwość bezpośredniej obserwacji funkcjonowania tej Sieci w Polsce. Węzeł krajowy jest podłączony z węzłem w Danii. Spośród licznych rodzajów usług^{5/} dostępnych dla użytkowników EARN (m.in. dostępu

^{4/} Zob. E. Chmielewska-Gorczyca: Tezaurus Informacji Naukowej. ZIN 1989 nr 2(55) s.51-96 oraz D. Ohnsorge: Projekt techniczny systemu informacyjno-wyszukiwawczego z zakresu informacji naukowej. tamże s. 97-118.

^{5/} porównaj: B. Rykaczewska-Wirogórska: Sieć komputerowa EARN w Polsce. Aktualne Problemy Informacji i Dokumentacji 1990 nr 4 s. 5-8.

do baz danych i bibliotek poszczególnych członków systemu? użytkownicy polscy mogą korzystać obecnie wyłącznie z poczty elektronicznej i transferu zbiorów

EARN jest siecią komputerową dostępną dla wszystkich uczelni i niekomercyjnych instytucji badawczych z terenu Europy, Afryki i Środkowego Wschodu. Jej głównym zadaniem jest wymiana informacji. Przy użyciu sieci EARN nie wolno przysyłać informacji o charakterze komercyjnym, politycznym, tajnym i religijnym. Transmitowane dane nie podlegają specjalnej ochronie. W roku 1990 koszty zorganizowania Węzła Krajowego EARN w Polsce są pokrywane ze specjalnej dotacji Komitetu d/s Nauki i Postępu Technicznego przy Radzie Ministrów. Zasady finansowania na przyszłość będą zależały od sformułowań ustawy o organizacji i finansowaniu nauki w Polsce.

W wykładzie prof.dr Jerzego Ruszkiewicza omówiono kierunki zmian w zarządzaniu i finansowaniu nauki w Polsce na tle ewolucji instytucji i zasad finansowania nauki w przeszłości. W niegalekiej przyszłości będzie powołana krajowa Rada Nauki pochodząca częściowo z urzędu, a częściowo z wyboru. Będzie ona dysponowała całością pieniędzy przeznaczonych na granty. Sekretariat Rady będzie sprawował Urząd Rady Nauki. Część funduszu finansowania nauki ma pochodzić z budżetu państwa. Finansowanie będzie podmiotowe i będą z niego korzystały przede wszystkim wyższe uczelnie i instytuty badań podstawowych. Krajowa Rada Nauki, powołana w przyszłości na podstawie ustawy, dysponująca funduszami na finansowanie badań, będzie się dzieliła na 2 komisje: Komisję Badań Podstawowych i Komisję Badań Stosowanych. W 1990 r. będzie jeszcze istniał Centralny Fundusz Finansowania Nauki. Z tego Funduszu będą finansowane tylko badania podstawowe oraz instytuty związane z infrastrukturą kraju. takie jak medyczne, farmaceutyczne, gospodarki żywnościowej, obsługi nauki.

28.03.1991 r.

A. Pietrzak

**OPRACOWANIE PRZEDMIOTOWE DOKUMENTÓW
ORAZ NAUCZANIE OPRACOWANIA RZECZOWEGO
W SZKOLACH WYŻSZYCH**

Ogólnopolska konferencja, Gdańsk, 7-9 maja 1991 r.

W dniach od 7 do 9 maja 1991 r. odbyła się w Gdańsku ogólnopolska konferencja nt. "Opracowanie przedmiotowe dokumentów oraz nauczanie opracowania rzeczowego w szkołach wyższych". zorganizowana przez Komisję ds. opracowania rzeczowego przy Zarządzie Głównym Stowarzyszenia Bibliotekarzy Polskich (SBP), tenże Zarząd SBP, Bibliotekę Gdańską PAN oraz Gdański Oddział SBP.

Było to trzecie spotkanie z cyklu zainicjowanego konferencją nt. opracowania rzeczowego zbiorów w dużych bibliotekach uniwersalnych (Jarocin, 22-24 maja 1986 r.)^{1/}. Drugim była zorganizowana przez Bibliotekę Główną Uniwersytetu Mikołaja Kopernika w maju 1988 r. konferencja nt. opracowania przedmiotowego w bibliotekach uniwersalnych (Toruń, 16-17 maja 1988 r.)^{2/}

Omawiana konferencja zgromadziła około 80. uczestników, przedstawicieli bibliotek naukowych, dużych bibliotek publicznych oraz placówek zajmujących się pomaturalnym i akademickim kształceniem

1/ Opracowanie rzeczowe zbiorów w dużych bibliotekach uniwersalnych. Ogólnopolska konferencja, Jarocin, 22-24 maja 1986 r. Zagadnienia Informacji Naukowej 1986 nr 2 s.157-161.

2/ Ogólnopolska konferencja na temat opracowania przedmiotowego w bibliotekach uniwersalnych. Toruń 16-17 maja 1988 r. Zagadnienia Informacji Naukowej 1988 nr 1 s.163-165.

bibliotekarzy i pracowników inte. Obrady miały miejsce w pomieszczeniach Biblioteka Gdańskiej PAN i były znakomicie organizacyjnie przygotowane przez pracowników tej Biblioteki. Wygłoszono 12 referatów. W ostatnim dniu odbyły się dwa pokazy: poprzedzona prelekcją Jadwigi Wołodźko-Sarosiek prezentacja katalogu przedmiotowego Biblioteki Gdańskiej PAN i słownika haseł przedmiotowych tego języka oraz pokaz najcenniejszych rękopisów, inkunabułów i starych druków znajdujących się w zbiorach Biblioteki.

Tematyka referatów była różnorodna, ale dały się wyróżnić dwie grupy tematyczne odpowiadające dwóm członom tytułu konferencji. Jedna dotyczyła zagadnień nauczania opracowania rzeczowego w szkołach wyższych. Druga była poświęcona wybranym problemom opracowania przedmiotowego zbiorów w bibliotekach.

W grupie dotyczącej dydaktyki ulokowały się następujące referaty:

1. Stan nauczania przedmiotu "Opracowanie rzeczowe zbiorów bibliotecznych" na wyższych studiach bibliotekoznawczych w Polsce - Zdzisław Gębołyś, Jolanta Paw (Zakład Bibliotekoznawstwa Uniwersytetu Śląskiego).

2. Metodyka nauczania przedmiotu "Źródła informacyjne" w zakresie opracowania rzeczowego zbiorów w Zakładzie Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu Jagiellońskiego - Alina Misiowa (Biblioteka Jagiellońska).

3. Praktyczne nauczanie opracowania rzeczowego zbiorów w Uniwersytecie Jagiellońskim - Zofia Steczowicz-Sajderowa (Biblioteka Jagiellońska).

4. Praktyka bibliotekarska studentów po II roku studiów. Doświadczenia Biblioteki Głównej Uniwersytetu Mikołaja Kopernika w Toruniu - Anna Tomczyk-Churska (Biblioteka Główna UMK).

Problematyki opracowania przedmiotowego dotyczyły następujące referaty:

1. Katalogi rzeczowe w heurystyce informacyjnej - Katarzyna Materska (Instytut Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu Warszawskiego).

2. Opis wielokrotny w języku haseł przedmiotowych - Ewa Stępnia-kowa (Biblioteka Narodowa).

3. Hasło przedmiotowe a adekwatny opis przedmiotowy - Maria Błaziak (Biblioteka Narodowa)^{3/}.

4. Niektóre reguły "Instrukcji tematowania i katalogu przed-miotowego" w ocenie użytkowników katalogów - Jadwiga Sadowska (Biblioteka Narodowa).

5. Język haseł przedmiotowych w katalogach online. (Z doświad-czeń francuskich, kanadyjskich i amerykańskich) - Teresa Głowacka (Biblioteka Uniwersytetu Warszawskiego).

6. Katalog przedmiotowy dla księgozbioru biologiczno-geografi-cznego w Bibliotece Głównej Uniwersytetu Gdańskiego - Grażyna Jaśkowiak (Biblioteka Główna UG).

7. Tematy jednostkowe w praktyce opracowania przedmiotowego Biblioteki Narodowej. Wybrane problemy - Janina Trzecińska (Bibliote-ka Narodowa).

8. Sprawność informacyjna katalogów rzeczowych w opinii użytkow-ników (komunikat z badań) - Teresa Turowska (Biblioteka Narodowa).

Problemy przedstawione w referatach znalazły szeroki odzew w dyskusjach. Niezwykle żywa i, ze względu na dobór jej uczestników (konferencja ta była jednym z nielicznych spotkań praktyków opra-

^{3/} Tekst referatu publikujemy w bieżącym numerze "Zagadnień ... " s. 155-175.

cowania rzeczowego i dydaktyków tego przedmiotu), twórcza i nieru-
lynowa była dyskusja dotycząca problemów dydaktyki opracowania
rzeczowego. Starano się znaleźć odpowiedzi między innymi na następu-
jące pytania: czego środowisko bibliotekarskie oczekuje od absolwen-
tów szkół wyższych; w jaki sposób środowisko może formułować
swoje wymagania i sugestie wobec placówek zajmujących się kształce-
niem; jakie cele dydaktyczne stawiają sobie uczelnie ucząc opraco-
wania rzeczowego; czy i jak uczelnie realizują cel kształcenia
specjalistów, którzy twórczo pracowaliby nad teorią i praktyką
już istniejących i nowych języków informacyjno-wyszukiwawczych?
Stosunkowo wiele dyskutowano na temat studenckich praktyk zawodo-
wych. Zebrani zaakceptowali wnioski z dyskusji dotyczące form
kształcenia, organizacji i niektórych treści programów opracowania
rzeczowego, wśród których najważniejszymi wydają się być następu-
jące:

1. Istnieje potrzeba poddyplomowych form kształcenia i doskona-
lenia zawodowego w zakresie bibliotekoznawstwa i informacji nauko-
wej. Zwrócono jednak uwagę na konieczność większej specjalizacji
i zróżnicowania programów nauczania, na przykład bądź poprzez
wprowadzenie formuły frekwencyjnej zmiany profilu (specjalizacji)
programu w danym ośrodku kształcenia bądź poprzez ustalenie spe-
cjalizacji zakresowej różnych ośrodków.

2. Ogromne jest zapotrzebowanie środowiska bibliotekarskiego
na nowe organizacyjne formy kształcenia i doskonalenia zawodowego,
zwłaszcza specjalistycznego. Uczestnicy dyskusji wskazali na
takie możliwe rozwiązania jak na przykład: rozwinięcie i uelasty-
cznienie instytucji wolnego słuchacza; umożliwienie udziału w
wybranych zajęciach dydaktycznych osobom spoza społeczności stu-
denckiej i środowisk wyższych uczelni; organizowanie przez wyższe
uczelnie kursów doskonalenia zawodowego dla bibliotekarzy i pracow-

ników inte. Ostatnia z propozycji prawdopodobnie wynika z podkreślanej przez zgromadzonych potrzeby wypełnienia luki w obszarze doskonalenia zawodowego, jaka wytworzyła się po rozwiązaniu CINTÉ.

3. Zwrócono uwagę na potrzebę zmiany charakteru egzaminu dla kandydatów na stanowiska bibliotekarzy dyplomowanych, idącą w kierunku zwiększenia jego specjalistycznego charakteru, tzn. skorelowania go ze specjalizacją i specyfiką warsztatu pracy osoby składającej egzamin. Zakwestionowano zasadność warunku złożenia egzaminów państwowych z dwóch języków obcych. Sformułowano wniosek pod adresem Zarządu Głównego Stowarzyszenia Bibliotekarzy Polskich o zajęcie przez Stowarzyszenie oficjalnego stanowiska w sprawie statusu bibliotekarzy dyplomowanych w świetle nowej "Ustawy o szkolnictwie wyższym" i taryfikatorów płacowych.

4. Niezbędne jest zintensyfikowanie kształcenia akademickiego w zakresie teorii i praktyki języków informacyjno-wyszukiwawczych. W związku z tym zebrani zwrócili się z apelem do szkół wyższych, które zatrudniają kadrę dydaktyczną o odpowiednim profilu specjalizacji o rozważenie możliwości uruchomienia i/lub rozwinięcia tej specjalizacji zarówno na poziomie studiów magisterskich, podyplomowych, jak i innych form kształcenia.

5. Sformułowano następujące postulaty co do programu studiów bibliotekoznawczych:

a) przesunięcie uczenia zagadnień opracowania rzeczowego z drugiego roku na wyższe lata studiów i co za tym idzie zmianę czasu odbywania odpowiedniej praktyki zawodowej;

b) zwrócenie uwagi na potrzebę szerszego uwzględnienia w programach studiów lektoratów języków obcych, przedmiotów rozwijających wiedzę ogólną oraz umiejętności samodzielnego rozwiązywania problemów;

c) wzbogacenie programów nauczania opracowania rzeczowego o problematykę katalogów bibliotecznych widzianych z trzech równorzędnych punktów widzenia - teorii, metodyki tworzenia oraz stanu ich istnienia i funkcjonowania; z wnioskiem tym w pewnym stopniu łączy się kolejny postulat.

d) szerszego włączenia praktyków opracowania rzeczowego do procesów dydaktycznych, na przykład poprzez delegowanie realizacji odpowiednich zajęć dydaktycznych osobom zajmującym się prowadzeniem realnie istniejących katalogów rzeczowych w bibliotekach.

6. Uznano, iż praktyki zawodowe są niezwykle ważnym elementem procesu dydaktycznego i w związku z tym postulowano co następuje:

a) zrewidowanie programów praktyki. Zdaniem zgromadzonych wskazane byłoby poprzedzenie tych działań zapoznaniem się z opiniami i sugestiami placówek realizujących praktyki tak, aby z jednej strony realizacja praktyk mogła odzwierciedlać specyfikę konkretnej biblioteki, z drugiej zaś - pozwalała instytucjom przyjmującym praktykantów z II roku studiów na skoncentrowanie się na problemach opracowania rzeczowego;

b) powrót do tradycji wizytowania praktyk przez przedstawicieli szkół;

c) rozszerzenie formuły praktyk zawodowych w zakresie opracowania rzeczowego o praktyki realizowane w trakcie roku akademickiego lub rozważenie możliwości wprowadzenia dodatkowych praktyk w ramach zajęć fakultatywnych realizowanych w różnych okresach toku studiów;

d) objęcie studentów zaocznych takim samym obowiązkiem praktyk, jak studentów studiów stacjonarnych.

Drugi dzień obrad był poświęcony wybranym problemom opracowania przedmiotowego zbiorów. Referaty dotyczyły zarówno zagadnień

warsztatowych, metodycznych, jak też związanych z przyszłością opracowania przedmiotowego w perspektywie automatyzacji. Dwa z nich, choć dotyczyły bezpośrednio problemów warsztatowych, były próbą przybliżenia się do rozwiązania problemów ogólniejszej natury ("Opis wielokrotny w języku haseł przedmiotowych" i "Hasło przedmiotowe a adekwatny opis przedmiotowy. Niektóre zagadnienia"). Problemy przedstawione we wszystkich referatach były przedmiotem interesującej dyskusji. Warto w tym miejscu odnotować, iż problemy warsztatowe nie przysłoniły uczestnikom dyskusji celu głównego tworzenia katalogów, jakim jest danie użytkownikom (czytelnikom, ale także bibliotekarzom) sprawnego i "przyjaznego" narzędzia selekcjonowania informacji.

W dyskusji zwracano uwagę na potrzebę kontaktów nieformalnych w środowisku bibliotekarskim, wymiany doświadczeń i współpracy. Sformułowano wnioski, wśród których znalazły się następujące.

1. Zebrani uznali, iż istnieje pilna potrzeba ogłoszenia polskich norm dla nazw geograficznych i nazw ciał zbiorowych.

2. Konieczne jest podjęcie przez Bibliotekę Narodową prac nad polskimi kartotekami autorytatywnymi zgodnie z zaleceniem IFLA oraz kartoteką wzorcową haseł przedmiotowych.

3. Widzi się potrzebę podjęcia prac nad porządkiem określników występujących przy tematach różnych kategorii. W związku z tym zostały sformułowane pewne wnioski co do zakresu i form działania Komisji ds. opracowania rzeczowego.

Z inicjatywy przedstawicieli Biblioteki Głównej Uniwersytetu Mikołaja Kopernika przyjęto wniosek, iż kolejna konferencja, której zakres roboczo został określony jako "kartoteki wzorcowe haseł przedmiotowych" odbędzie się za dwa lata w Toruniu.

Jadwiga Woźniak

17.05.1991 r.

**BAZY DANYCH I ICH WYKORZYSTANIE W SYSTEMIE
MONITORINGU I OCHRONY ŚRODOWISKA**

Seminarium naukowe

Warszawa, 28 maja 1991 r.

W dniu 28 maja 1991 roku staraniem Ośrodka Informacji Naukowej Polskiej Akademii Nauk oraz Instytutu Geodezji i Kartografii odbyło się w Pałacu Staszica w Warszawie seminarium naukowe na temat baz danych i możliwości ich wykorzystania w systemie monitoringu i ochrony środowiska.

Celem Seminarium było przedstawienie Systemu Informacji o Ukształtowaniu Środowiska - SINUS i doświadczeń uzyskanych przy jego budowie.

W problematykę współczesnej informacji naukowej i roli Ośrodka, jako placówki współodpowiedzialnej za tworzenie i eksploatację systemów informacyjnych z zakresu nauk ścisłych i przyrodniczych, wprowadził dyrektor OIN PAN doc.dr hab. A. Gromek.

Następnie zabrał głos prof.dr hab. Andrzej Ciołkosz z Instytutu Geodezji i Kartografii. W swoim referacie - "System Informacji o Ukształtowaniu Środowiska - SINUS" przedstawił główne założenia i cele, jakie postawiono przy jego budowie. Podkreślił przy tym, że informacje, charakteryzujące pod różnym względem obszar kraju, wymagają skonstruowania układu pozwalającego gromadzić je w sposób uporządkowany. Taki układ zwany układem odniesień przestrzennych zaprojektowano specjalnie dla potrzeb systemu^y.

^y bliższe informacje nt. układu odniesień przestrzennych oraz systemu SINUS zob. bieżący numer "Zagadnień..." s.139-153.

Układ odniesień przestrzennych nawiązuje do struktury informacji pozyskiwanych za pomocą teledetekcji satelitarnej, która stanowi jedno z ważniejszych źródeł zasilania systemu. System SINUS jest systemem narzędziowym, co oznacza, że posiada on wiele opcji przetwarzania danych dobieranych zgodnie z wymogami zastosowania. Przewidziano również możliwości uzupełnienia systemu oprogramowaniem specjalistycznym. Na uwagę zasługuje fakt, że SINUS został potraktowany przez United Nations Environment Programme - UNEP jako początek systemu informacyjnego o środowisku w Polsce. W związku z tym Ministerstwo Ochrony Środowiska, Zasobów Naturalnych i Leśnictwa powierzyło zorganizowanie i prowadzenie Narodowego Centrum Informacji o Środowisku Instytutowi Geodezji i Kartografii.

Dr A. Walanus z Politechniki Śląskiej omówił program "Czwartorzędowa palinologiczna baza danych". Założeniem jest, aby program ten stworzył podstawy dla pierwszego etapu działania Paleoeologicznego Regionalnego Banku Danych, który stałby się w pełni kompatybilną częścią składową Europejskiej Palinologicznej Bazy Danych z siedzibą w Marsylii.

Prace nad opracowaniem regulaminu wprowadzania i użytkowania danych podjęto na podstawie projektu regulaminu Europejskiego Banku Danych zaproponowanego na posiedzeniu Rady Konsultacyjnej Banku w 1990 roku. Program ten pozwala wprowadzać i przetwarzać wyniki analizy pyłkowej jednostek taksonomicznych występujących w osadach organogenicznych przy pomocy komputera i przedstawiać wyniki tej analizy w formie graficznej. Dotychczas opracowano listy taksonomiczne roślin występujących w formie pyłku w osadach późnego glacjału i holocenu oraz rozpoczęto wprowadzanie danych z poszczególnych stanowisk.

Omawiając "Bazę danych jeziora Gościąg" dr A. Walański wskazał na wyjątkowość tego obiektu badawczego ze względu na zachowanie w niezminionej formie jego struktury sedymentacyjnej. Na podstawie ilości i grubości warstw osadu dennego (20 m głębokości wody i 16 m grubości osadu) można się doliczyć okresu rzędu 12 tysięcy lat. Z kolei mnogość cech tego osadu zmusza do zainteresowania się korelacjami, jakie występują między poszczególnymi danymi. Baza ta jest programowana w języku Pascal, a otrzymywane w dowolnej skali wykresy pozwalają zaobserwować roczną periodyzację zjawisk.

"Wykorzystanie programu CLUSTER do badań taksonomicznych i fitosocjologicznych" przedstawił mgr M. Verey z Instytutu Botaniki im. W. Szafera PAN w Krakowie. Wykładowca omówił sposób tworzenia biologicznej bazy danych za pomocą klastryzacji różnych obiektów w wielowymiarowej przestrzeni. Obiekty te podlegają opisowi za pomocą kilku lub całego ciągu cyfr, co w konsekwencji pozwala na graficzne generowanie programu. Powstające w ten sposób dendrogramy informują, jak się mają do siebie poszczególne gatunki o określonej liczbie cech, umieszczone w wielowymiarowej przestrzeni. Program ten pozwala analizować około 400 gatunków roślin jednocześnie.

W czasie przerwy prezentowano bazy danych Ośrodka Informacji Naukowej PAN.

Dużym zainteresowaniem cieszyła się baza CAB-Abstracts wydana przez Commonwealth Agriculture Bureau. Baza ta zawiera abstrakty z ponad 10 000 czasopism oraz monografii, materiałów konferencyjnych, raportów technicznych i patentów wydawanych w 100 krajach dotyczących 24 dyscyplin nauk rolniczych.

Równie ciekawy dla uczestników Seminarium był pokaz znanego Indeksu Cytowań Naukowych (Science Citation Index) na dyskach optycznych.

Baza "Birds of America" - "Ptaki Ameryki" opracowana przez Johna Jamesa Audubona składa się z 500 kolorowych litografii ptaków, których oglądanie na monitorze można połączyć z wykorzystaniem indeksu tekstów opisujących oraz nagraniem ich głosu. Wykorzystanie tej bazy deklarowali ornitolodzy oraz dydaktycy Wydziałów Leśnictwa.

Po przerwie jako pierwszy zabrał głos dr J. Maroński z Instytutu Podstaw Informatyki PAN. Charakteryzowany "System Wczesnego Ostrzegania o Skażeniu Promieniotwórczym" składa się z detektorów (wykrywaczy) oraz systemu komputerowego z odpowiednim oprogramowaniem. Dane z poszczególnych detektorów zbierane są w określonych odstępach czasu za pomocą łącz telefonicznych bądź wspomnianej sieci komputerowej. W sytuacji przekraczania progu alarmowego w dowolnym detektorze podejmowana jest akcja alarmowa w miejscu jego zainstalowania oraz akcja sygnalizacyjna wywołana przez komputer umieszczony w centrum sterowania systemem. Dodatkowo system generuje dane, które mogą być umieszczane w dowolnej bazie faktycznej.

W ubiegłym roku na bazie wspomnianych opracowań zostały zainstalowane tego typu systemy w Syrii (w oparciu o sieć komputerową) oraz trzy systemy resortowe w kraju. Obecnie instalowany jest analogiczny system w Jordanii i Iranie.

Należy nadmienić, że system ten można stosować w monitorowaniu różnych skażeń środowiska naturalnego.

Kolejnym referatem była "Faunistyczna baza danych" autorstwa dr W. Jędrzykowskiego z Instytutu Zoologii PAN.

Zaprojektowana przez Centrum Dokumentacji Faunistycznej tego Instytutu faunistyczna baza danych spełnia następujące funkcje:

- umożliwia wprowadzenie danych niepublikowanych, stanowiących uzupełnienie do wydawanych już katalogów fauny,
- zawiera informacje o danych niepublikowanych, do których istnieją materiały w postaci zbiorów,
- zawiera informacje o rozmieszczeniu gatunków w Polsce i umożliwia wydruk map z siatką UTM lub geograficzną,
- umożliwia szybkie wyszukiwanie informacji i ich wymianę z podobnie tworzonymi bazami w Polsce i za granicą.

W chwili obecnej baza zawiera informacje dotyczące trzech następujących grup zwierząt: Pseudoscorpiones, Opiliones i Aranei. Informacja ta dotyczy 100 gatunków, a na podobny opis w Muzeum Kolekcji Instytutu Zoologii PAN czeka około 5 milionów egzemplarzy.

Program aplikacyjny został napisany w języku dBASE i przewiduje się jego kompilację po planowanej rozbudowie bazy.

"Ornitologiczne bazy danych w świetle badań międzynarodowych"

- to temat referowany przez doc. M. Gromadzkiego ze Stacji Ornitologicznej PAN w Gdańsku. Rozpoczynając swoje wystąpienie prelegent podkreślił, że oprócz całego systemu pomiarów różnych niepożądanym związków w środowisku, ptaki są dla nas również aparaturą monitoringową. Cały system obserwacji ornitologicznych prowadzony jest w oparciu o obowiązujące sposoby unifikacji i standaryzacji metod badawczych, a także odpowiednie gromadzenie i przetwarzanie danych. Opracowaniem metodyk, głównie z zakresu obrączkowania, zajmuje się Europejska Unia Obrączkowania Ptaków (EURING) z siedzibą w Heteren w Holandii. W Polsce badaniami ornitologicznymi zajmuje się Stacja Ornitologiczna Instytutu Ekologii PAN w Gdańsku tworząc w tym zakresie dwa systemy informacyjne. Pierwszym z nich, jest Centrala Obrączkowania Ptaków (członek EURING) prowadząca ewidencję obrączkowania na terenie całej Polski. Drugi system sta-

nowi PAO - Polski Atlas Ornitologiczny, zawierający informację o występowaniu ptaków lęgowych w Polsce. Polski Atlas Ornitologiczny jest prowadzony zgodnie z zasadami przyjętymi przez Komitet Europejskiego Atlasu Ornitologicznego, by zebrane dane mogły posłużyć do opracowania Atlasu Ptaków Lęgowych Polski, a także Atlasu Ptaków Lęgowych Europy. Systemy te są oprogramowane w języku dBASE III plus. Zastosowano w nich układ odniesień przestrzennych.

Wszystkie referaty były bogato ilustrowane pokazami na monitorach komputerowych, foliogramami i przezroczami.

W seminarium wzięło udział 85 przedstawicieli reprezentujących ośrodki informacyjne, placówki naukowe oraz instytucje centralne i terenowe. Stworzyło okazję do wymiany poglądów na temat zakresu dotychczasowych poczynąń i osiągnięć w budowie baz informacyjnych. Jednocześnie wykazało konieczność przyjęcia jednolitego układu pól odniesień przestrzennych, który umożliwi wykorzystanie informacji zgromadzonych w innych systemach, co obniży koszty, przyspieszy opracowanie, a także pozwoli na uniknięcie błędów i przekłamań przy wprowadzaniu danych.

Duża frekwencja, interesujące referaty, przebieg dyskusji i końcowe wystąpienie prof. A. Ciołkosza oraz doc. A. Gromka podsumowujące przebieg Seminarium pozwalają mieć nadzieję, że Seminarium było nie tylko źródłem wartości poznawczych na temat systemu informacyjnego służącego ochronie środowiska naturalnego, lecz przyczyni się do jego dalszego rozwoju i wykorzystania.

I. Gałczyńska

10.06.1991

SPIS TREŚCI

1. E. CHMIELEWSKA-GORCZYCA: Język wyszukiwawczy a potrzeby informacyjne użytkowników	3
2. B. SOSIŃSKA-KALATA: Terminologiczna baza danych z zakresu teorii języków informacyjno-wyszukiwawczych	41
3. D. DHNSORGE: Projekt techniczny terminologicznej bazy danych z zakresu teorii języków informacyjno-wyszukiwawczych	73
4. H. POPOWSKA: Wybór baz danych i serwisów online (kilka wskazówek metodycznych)	91

Materiały i przyczynki

1. T. GAŁCZYŃSKA: System Informacji o Ukształtowaniu Środowiska - SINUS	139
2. M. BŁAZIAK: Hasło przedmiotowe a adekwatny opis przedmiotowy	155
3. B. SOSIŃSKA-KALATA: Projekt systemu notacyjnego dla Tezaurusu Informacji Naukowej. Charakterystyka struktury i metodyka opracowania	177

Recenzje i omówienia

1. Hipertekstem o hipertekście - K. Siwek	205
---	-----

Kronika

1. Działalność seminaryjna OIN PAN w 1990 r. - A. Pietrzak ..	215
2. Opracowanie przedmiotowe dokumentów oraz nauczanie opracowania rzeczowego w szkołach wyższych. Ogólnopolska Konferencja (Gdańsk 7-9 maja 1991) - J. Woźniak	219
3. Bazy danych i ich wykorzystanie w systemie monitoringu i ochrony środowiska. Seminarium naukowe (Warszawa 28 maja 1991) - T. Gałczyńska	226

CONTENTS

1. E. Chmielewska-Gorczyca: Information retrieval language and information users' needs	3
2. B. Sosińska-Kalata: The terminological data base on the theory of information retrieval languages	41
3. D. Ohnsorge: The technical project of the terminological data base on the theory of information retrieval languages	73
4. H. Popowska: The selection of the data bases and online services (Some methodological advices)	91

Materials and Contributions

1. T. Gałczyńska Information System on Configuration of Environment - SINUS	139
2. M. Błaziak: Subject heading and the adequate subject characteristics	155
3. B. Sosińska-Kalata: Project of the notational system for the Thesaurus of Information Science. Characteristics of the structure and methods of establishment	177

Reviews and Surveys

1. On hypertext in terms of hypertext - K. Siwek	205
--	-----

Chronicles	219
------------------	-----

СОДЕРЖАНИЕ

1. Э. Хмелевска - Горчица Информационно - поисковый язык
э информационные потребности потребителей 3
2. В. Сосянска - Калята Терминологическая база данных
в области теории информационно - поисковых языков 41
3. Д. Ожисорге Технический проект терминологической
базы данных в области теории информационно - поисковых
языков 73
4. И. Поповска Выбор баз данных и диалоговых сервисов
(Несколько методических указаний) 91

М а т е р и а л ы и п р и м е ч а н и я

1. Т. Галчиньска Информационная система об рельефе среды 139
2. М. Влазяк Предметный заголовок а адекватное предмет-
ное описание 155
3. В. Сосянска - Калята Проект нотационной системы для
Тезауруса научной информации. Характеристика структуры
и методика разработки..... 177

Р е ц е н з и и о б з о р ы

1. О гипертексте на гипертексте - К. Сивек 205
- Х р о н и к а 215

Wskazówki dla Autorów

Redakcja „Zagadnień Informatyki Naukowej” uprzejmie prosi Autorów o przestrzeganie następujących zasad przy nadsyłaniu materiałów.

Artykuł nie powinien przekraczać 30 stron maszynopisu formatu A4 wraz z przypisami i ewentualnymi tablicami (podwójny odstęp między wierszami, ok. 30 wierszy na stronie, margines 3,5 cm lewy, 1 cm prawy).

Maszynopis artykułu należy dostarczyć w 2 egzemplarzach. Ilustracje (tablice, wykresy) powinny być umieszczone na osobnych ponumerowanych stronach z zaznaczeniem ich miejsca w tekście.

Przypisy i bibliografię należy umieszczać na końcu tekstu i na osobnych stronach. Opis bibliograficzny powinien składać się z: nazwiska i imienia autora, tytułu publikacji, miejsca wydania, nazwy wydawnictwa i roku wydania oraz stron. W wypadku powoływania się na artykuł opis powinien zawierać: nazwisko i imię autora, tytuł artykułu, nazwę czasopisma, rok, tom (numer), strony lub stronę zawierającą cytowany fragment.

Każdy artykuł powinien być zaopatrzony w streszczenie autorские ok. $\frac{1}{2}$ strony maszynopisu.

Autorzy proszeni są o podawanie do wiadomości redakcji (oprócz imienia i nazwiska) również tytułu naukowego, nazwy i adresu instytucji, w której pracują, prywatnego adresu zamieszkania, numeru telefonu i ewentualnie, jeśli honorarium autorskie ma być przekazane na konto, numer konta PKO.

O przyjęciu do druku decyduje Rada Redakcyjna.

Maszynopisów artykułów nie zamówionych nie przyjętych do druku redakcja nie zwraca.

