

POLSKA AKADEMIA NAUK

OŚRODEK INFORMACJI NAUKOWEJ

PL ISSN 0324-8194

ZAGADNIENIA INFORMACJI NAUKOWEJ

1990

OSSOLINEUM

Nr 1 (56)

POLSKA AKADEMIA NAUK

OŚRODEK INFORMACJI NAUKOWEJ

ZAGADNIENIA
INFORMACJI
NAUKOWEJ

1990

Nr 1(56)

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, ul. Nowy Świat 72 (Pałac Staszica)**

PL ISSN 0324-8194

Maszynopis niniejszego numeru przekazano do Wydawnictwa 14.05.1990

Wykonano ze składu pisma dostarczonego przez OIN PAN

Zakład Narodowy im. Ossolińskich-Wydawnictwo. Wrocław 1990.

Objętość: art. wyd. 10,27; art. druk. 14,63.

Zakład Graficzny Politechniki Wrocławskiej

Zam. 901/90

BARBARA WERESZCZYŃSKA-CISŁO

Instytut Biotechnologii
Przemysłu Rolno-Spożywczego

DEFINIOWANIE JĘZYKA DESKRYPTOROWEGO

Leksyka języka deskryptorowego: jednostki autosyntaktyczne i jednostki synsyntaktyczne. Zdanie i tekst języka deskryptorowego, Podjęzyki języka deskryptorowego: język charakterystyk wyszukiwawczych dokumentów (JCHW) a język wyszukiwania informacji (JWI). Leksyka i reguły gramatyczne JCHW oraz JWI. Dwa znaczenia operatora różnicy zbiorów. Charakterystyka formalna eksperymentalnego języka deskryptorowego o symbolu RK.

Wśród języków informacyjno-wyszukiwawczych (JIW) o standaryzowanym słownictwie wyróżnia się języki klasyfikacyjne (hierarchiczne) oraz języki deskryptorowe. Rozłączność tego podziału budzi wątpliwości, gdyż trudno jest określić cechę dystyngtywną JIW uzasadniającą takie rozróżnienie, ale związana z tym podziałem terminologia jest powszechnie przyjęta i stosowana. Tym, co w praktyce wyraźnie dzieli te języki, są pewne cechy strukturalne słowników i reguły ich opracowywania oraz sposoby wyszukiwania dokumentów za pomocą tych słowników. W wypadku języków deskryptorowych do takich cech wyróżniających należą:

- niektóre reguły indeksowania za pomocą tezaurusów, na przykład zasada dowolnej kolejności zestawiania terminów w charak-

terystyce wyszukiwawczej dokumentu (CHWD) i instrukcji wyszukiwawczej (IW), wynikająca z równorzędnego traktowania tych terminów, zwana zasadą indeksowania współrzędnego;

- możliwość stosowania negacji jako operatora umieszczanego w instrukcjach wyszukiwawczych, które jest niemożliwe lub trudne do zrealizowania w wypadku języków klasyfikacyjnych.

Języki klasyfikacyjne są na ogół szczegółowo i precyzyjnie definiowane przez podanie ich charakterystyki kategorialnej i leksykalnej oraz reguł gramatycznych. Natomiast definicje języków deskryptorowych, występujące w opisach wykorzystywania tezaurusów, są najczęściej nieprecyzyjnie i niepełne, co może mieć znaczący wpływ na prawidłowość i jednolitość ich stosowania. Często opis deskryptorowego JIW sprowadza się do omówienia form deskryptorów, znaków delimitujących i ewentualnie podania zasady dowolnej kolejności deskryptorów w CHWD lub IW. W wypadku deskryptorowego JIW wykorzystującego jednostki leksyki pomocniczej, takie jak modyfikatory lub relatory (wskaźniki więzi, roli, wagi itp.), podawane są także, zwykle na przykładach, reguły ich łączenia z deskryptorami. Tymczasem pełny opis deskryptorowego JIW powinien uwzględniać zarówno szczegółową charakterystykę kategorii leksyki, jak i pełny zestaw reguł gramatycznych oraz tekstotwórczych danego języka, z wyróżnieniem specyfiki podjęzyka charakterystyk wyszukiwawczych dokumentów (lub informacji faktograficznych) oraz podjęzyka wyszukiwawczego.

Wiąże się to z jednej strony z potrzebą takiego przedstawienia struktury i wartości semantycznej leksyki JIW oraz jej dopuszczalnych połączeń syntaktycznych, aby użytkownik JIW mógł zawsze jednoznacznie stwierdzić, czy dane wyrażenie, zdanie lub tekst są poprawnie zbudowane, czyli zgodne z zasadami tego języka.

Z drugiej zaś strony jest to związane z potrzebą przedstawienia użytkownikowi JIW wszystkich zróżnicowanych dopuszczalnych sposobów wykorzystywania języka w procesach tworzenia CHWD lub IW oraz wyszukiwania informacji. Bowiem zarówno brak możliwości oceny prawidłowości zbudowanej CHWD lub IW oraz stosowanych zasad strategii wyszukiwawczej, jak i niewykorzystywanie pewnych działań (opcji) dopuszczalnych na gruncie danego JIW mogą powodować niewłaściwe wyniki wyszukiwania informacji w postaci znacznego szumu informacyjnego lub niskiej kompletności wyszukiwawczych zbiorów dokumentów (albo zbiorów informacji faktograficznych).

Ogólne zagadnienia teoretyczne struktury języków informacyjno-wyszukiwawczych zostały szczegółowo omówione w pracy: "Struktura języka informacyjno-wyszukiwawczego" /2/. Omówienie to zostało zilustrowane licznymi przykładami z JIW różnych typów, między innymi języków deskryptorowych i dlatego może być pomocne przy definiowaniu budowanego języka deskryptorowego. Przedstawiono tam m.in. problemy poziomów JIW, a w tym poziomów semantycznie pełnych jednostek języka, którymi są słowa (leksyka JIW) oraz zdania i teksty wyrażone w danym języku; problem autosyntaktycznych (samodzielnych) i synsyntaktycznych (niesamodzielnych) jednostek leksykalnych JIW; problem konkatenacji (składania) jednostek niższego poziomu w jednostki poziomu wyższego z wykorzystaniem reguł składniowych zdaniotwórczych i tekstotwórczych, a także zagadnienie wykładników przynależności do zdania oraz łączników zdaniowych.

Leksykę deskryptorowego JIW stanowi zbiór elementarnych jednostek leksykalnych (EJL) obejmujący obligatoryjnie słownictwo podstawowe, czyli deskryptory i fakultatywnie słownictwo

pomocnicze (modyfikatory, relatory, atrybutory itd.). Deskryptory są zawsze wyrażeniami autosyntaktycznymi, natomiast leksykę pomocniczą stanowią wyrażenia, które nie mogą występować samodzielnie, czyli wyrażenia synsyntaktyczne. Z deskryptorów buduje się zdania danego JIW, a w przypadku uwzględnienia w leksyce języka pomocniczych EJL również złożone wyrażenia nazwowe (złożone z deskryptora i pomocniczej jednostki leksykalnej), które są autosyntaktyczne i tworzą zdania JIW na równi z deskryptorami /3/. Na przykład złożone wyrażenia nazwowe:

a) 1 SOK JABŁKOWY,

b) EKSPORT.= prognozy

gdzie:

a) to wyrażenie złożone z deskryptora SOK JABŁKOWY. i wskaźnika roli "surowiec"

b) to wyrażenie złożone z deskryptora EKSPORT. i modyfikatora "prognozy"

Zdaniem deskryptorowego JIW jest ciąg zestawionych w dowolnej kolejności wyrażeń autosyntaktycznych tego języka, charakteryzujących treść dokumentu lub pytania informacyjnego jednotreściowego albo jedną z treści dokumentu lub pytania informacyjnego wielotreściowego. Inaczej mówiąc, jest to zbiór samodzielnych wyrażeń JIW tworzących spójną logicznie całość dotyczącą jednego zagadnienia (omówienia przedmiotu głównego dokumentu w określonym aspekcie) przedstawionego w dokumencie lub pytaniu informacyjnym. Zbiór ten może zawierać również deskryptory określające cechy formalne dokumentu. Zdaniem języka deskryptorowego może być:

- ciąg deskryptorów,

np. SOK JABŁKOWY ZAGĘSZCZONY. EKSPORT. POLSKA. J. ANG.

- ciąg złożonych wyrażeń nazwowych,

np. 1 SOK JABŁKOWY ZAGĘSZCZONY. EKSPORT.= prognozy

- ciąg deskryptorów i złożonych wyrażeń nazwowych,
np. 1 SOK JABŁKOWY ZAGĘSZCZONY. SKŁAD CHEMICZNY. POLSKA.

EKSPORT.= prognozy

Szczególnym przypadkiem jest zdanie języka deskryptorowego składające się z jednego deskryptora albo z jednego złożonego wyrażenia nazwowego (zbiór jednoelementowy), na przykład:

a) CHROMATOGRAF CIECZOWY.

b) 1 JABŁKA GOLDEN DELICIOUS.

Tekstem deskryptorowego JIW jest spójna syntagmatyczna struktura składająca się ze zdań tego języka, charakteryzująca dokument wielotreściowy albo wielotreściowe pytanie informacyjne. Szczególnym rodzajem tekstu deskryptorowego JIW jest tekst jednozdaniowy, który - co wynika z charakterystyki zdania - w pewnych przypadkach może się składać tylko z jednego wyrażenia autosyntaktycznego, tzn. deskryptora albo złożonego wyrażenia nazwowego.

W systemie informacyjno-wyszukiwawczym (SIW) posługującym się językiem deskryptorowym charakterystyka wyszukiwawcza dokumentu oraz instrukcja wyszukiwawcza (wraz z jej transformacją według zasad stosowanej strategii wyszukiwawczej) są zawsze tekstami danego języka deskryptorowego, a zarazem jedynymi możliwymi formami realizacji tego języka.

Ze względu na to, że wyszukiwanie dokumentów (lub informacji faktograficznych) realizowane jest zawsze z wykorzystaniem zdania języka deskryptorowego i jego elementów, język taki powinien umożliwiać wyróżnianie poszczególnych zdań w tekstach wielozdaniowych. Stosowane są różne sposoby rozwiązywania tego problemu. Jednym z nich jest wykorzystywanie w roli łączników zdaniowych wskaźników więzi w postaci liter, cyfr lub innych znaków graficznych. Taki znak umieszczony obok odpowiedniego wyrażenia w CHWD lub IW stano-

wi wykładnik relacji przynależności tego wyrażenia do określonego zdania, czyli wykładnik jednoargumentowej relacji syntagmatycznej treściowej /3/. Na przykład w poniższej charakterystyce wyszukiwawczej dokumentu:

a OWOCE JAGODOWE. ^b WINO.b EKSPORT.= prognozy a SKŁAD CHEMICZNY.
b RFN.

gdzie: zdanie dotyczące składu chemicznego wina z owoców jagodowych składa się z następujących deskryptorów, które wyżej poprzedzono wskaźnikiem więzi a

OWOCE JAGODOWE. SKŁAD CHEMICZNY. WINO. RFN.

a zdanie dotyczące prognoz nt. eksportu wina z RFN. składa się z wyrażen, które poprzedzono wskaźnikiem więzi b
WINO. EKSPORT.= prognozy RFN.

Inny sposób to wykorzystywanie jako łączników symboli logicznych - w roli łącznika zdaniowego symbolu koniunkcji, a w roli łącznika międzyzdaniowego (rozdzielającego dwa zdania) symbolu alternatywy. Na przykład:

OWOCE JAGODOWE. ^ SKŁAD CHEMICZNY. ^ WINO. V WINO. ^ EKSPORT.-
prognozy ^ RFN.

Istnieją również deskryptorowe JIW, w których niedopuszczalne jest tworzenie tekstów wielozdaniowych. W ich wypadku w trakcie indeksowania wielotreściowego dokumentu lub pytania informacyjnego tworzy się oddzielne CHWD lub IW dla każdej treści.

Maksymalną-dopuszczalną liczbę autosyntaktycznych wyrażen (na ogół tylko deskryptorów) w CHWD lub IW określają zasady budowy tych tekstów w danym deskryptorowym JIW, umieszczane zazwyczaj w instrukcji indeksowania dokumentów i pytań informacyjnych dołączonej do tezaurusa. Wyprowadzenie ograniczenia liczby wyrażen tworzących tekst w deskryptorowym JIW (ograniczenie szerokości indeksowania),

podyktowaną, względami ekonomicznymi, powoduje niejednokrotnie pomijanie w trakcie indeksowania wyrażen charakteryzujących drugoplanowe zagadnienia przedstawione w dokumentach. Brak tych wyrażen uniemożliwia często wyszukanie pożądaných informacji w odpowiedzi na pytania bardzo szczegółowe lub dotyczące zagadnień z pogranicza danej dziedziny. Na przykład jeśli w dokumencie na temat analizy chromatograficznej składników smakowo-zapachowych produktów żywnościowych występują, oprócz nazw różnych metod chromatograficznych, nazwy stosowanych kolumn chromatograficznych, które nie zostaną odzwierciedlone w CHWD dokumentu, wówczas dokument ten nie zostanie prawdopodobnie wyszukany w odpowiedzi na pytanie o typy kolumn chromatograficznych wykorzystywanych w analizie żywności.

Przekład charakterystyki słownej (zbioru słów kluczowych) dokumentu albo pytania informacyjnego na język informacyjno-wyszukiawczy daje w rezultacie odpowiednio: charakterystykę wyszukiawczą dokumentu albo instrukcję wyszukiawczą. Następnie instrukcja wyszukiawcza podlega przetworzeniu według zasad stosowanej w danym SIW strategii wyszukiawczej w celu sformułowania alternatywy dopuszczalnych odpowiedzi na dane pytanie /4/. Porównywanie przetworzonej IW z poszczególnymi CHWD przechowywanymi w bazie (lub bazach) danych SIW prowadzi do wydania zbioru dokumentów stanowiących odpowiedzi na określone pytanie informacyjne skierowane do systemu.

Język wyszukiwania informacji (JWI), obejmujący język instrukcji wyszukiawczych i język strategii wyszukiawczej nie jest jednak taki sam jak język charakterystyk wyszukiawczych dokumentów (JCHW). Są to różne podjęzyki JIW, choć mają wiele cech wspólnych /1/. W wypadku języka deskryptorowego wspólnymi

dla obu tych podjęzyków są zbiór leksyki podstawowej, czyli deskryptorów i ewentualnie część zbioru leksyki pomocniczej JIW (relatory w postaci wskaźników roli, wskaźników więzi, wykładniki kategorii itp.). Leksyka pomocnicza JWI może obejmować również elementy nie występujące w JCHW, do których należą operatory logiczne (operatory algebry Boole'a) "i", "lub", "oprócz" odpowiadające działaniem na zbiorach dokumentów - tworzenie iloczynów, sum lub różnic zbiorów z wykorzystaniem wyrażeń auto-syntaktycznych JIW traktowanych jako nazwy zbiorów dokumentów - dokonywanym w trakcie realizowania strategii wyszukiwawczej określonej na podstawie instrukcji wyszukiwawczej. Są one dość powszechnie nazywane "spójnikami logicznymi" (koniunkcja, alternatywa, negacja) chociaż ich interpretacja na gruncie SIW dotyczy rachunku zbiorów, a nie rachunku zdań tworzących instrukcje wyszukiwawcze. Operatory te oznaczane (odpowiednio): znakiem \cdot albo \wedge , znakiem $+$ albo \cup , znakiem \neg albo \bar{A} stanowią symbole rozkazów systemu wyszukiwawczego. Ich stosowanie wpływa korzystnie zarówno na trafność wyszukiwania, jak i na kompletność wyszukanych zbiorów informacji. Tak na przykład następujące wykorzystanie operatora \wedge

WINO \wedge EKSPORT.= prognozy \wedge RFN.

spowoduje wyszukanie tylko tych dokumentów, których CHWD zawierają te trzy samodzielne wyrażenia języka deskryptorowego, czyli dokumentów należących do zbioru będącego iloczynem zbiorów dokumentów wyznaczonych przez poszczególne z tych wyrażeń. Natomiast następujące wykorzystanie operatora \cup

WINO OWOCOWE. \wedge EKSPORT. \wedge RFN. \cup WINO GRONOWE. \wedge EKSPORT. \wedge RFN.

spowoduje wyszukanie i połączenie dwóch podzbiorów dokumentów re-

lewantnych do pytania o dokumenty na temat eksportu wina owocowego i eksportu wina gronowego z RFN. Powstały zbiór dokumentów może zawierać dokumenty dotyczące eksportu obu rodzajów win (suma zbiorów interpretowana analogicznie do alternatywy nierozłącznej). Jeśli dokumenty takie są niepożądane dla użytkownika, w niektórych SIW stosuje się wówczas wyszukiwanie z wykorzystaniem innego symbolu sumy zbiorów (sumy zbiorów interpretowanej analogicznie do alternatywy rozłącznej). Stosowanie operatora różnicy zbiorów stanowi bardziej złożony problem. Jest on wykorzystywany zarówno do wyłączenia z wyszukanego zbioru informacji dokumentów o określonych cechach formalnych (język dokumentu, typ dokumentu itd.), jak i do wyłączenia z wyszukanego zbioru dokumentów o określonych cechach zawartości treściowej. Na przykład zapis:

WINO OWOCOWE.∩ SKŁAD CHEMICZNY.∩ Ā J PORT.

spowoduje usunięcie z wyszukanego zbioru dokumentów na temat składu chemicznego win owocowych tych, które są napisane w języku portugalskim, a zapis:

WINO OWOCOWE.∩ SKŁAD CHEMICZNY.∩ Ā WINO JABŁKOWE.

spowoduje usunięcie z wyszukanego zbioru tych dokumentów, w których CHWD znajduje się deskryptor WINO JABŁKOWE. Wyszukiwanie prowadzone zgodnie z drugim z tych zapisów wydaje się nieprawidłowe, ponieważ może spowodować straty informacyjne (obniżenie kompletności wyszukanego zbioru dokumentów), wynikające z wyłączenia z wyszukanego zbioru tych dokumentów, które dotyczą zarówno wina jabłkowego, jak i innych win owocowych. To samo odnosi się do poniższego podawanego w literaturze przedmiotu przykładu

IW
ZDERZENIA.∩ SAMOLOTY.∩ POWIETRZE.∩ ∞ PTAKI.

ponieważ zgodne z nią wyszukiwanie informacji doprowadzi do wyłączenia z wyszukanego zbioru także tych dokumentów, które dotyczą zarówno zderzeń samolotów z ptakami, jak i zderzeń samolotów. Wydaje się więc, że stanowiący nazwę cechy treściowej dokumentu deskryptor (albo złożone wyrażenie nazwowe) wyznaczający zbiór nie chcianych dokumentów powinien być poprzedzany w IW operatorem (np. symbolem \complement) wskazującym, iż nie można go wykorzystywać do wyszukiwania informacji, innym niż operator różnicy zbiorów (np. \bar{A}), który powoduje wykluczenie z wyszukanego zbioru wszystkich dokumentów reprezentowanych przez ten deskryptor /4/. Pierwszy z nich będzie więc operatorem jednoargumentowym (odpowiednikiem funktora negacji w rachunku zdań), a drugi operatorem dwuargumentowym, zgodnie z jego intencją w pracy E. Chmielewskiej-Gorczyicy /1/. Wykorzystywanie w obu przedstawionych wyżej wypadkach tego samego operatora wymaga sformułowania odpowiednich zastrzeżeń definiujących dwie różne role tego symbolu /3/.

Do elementów leksyki pomocniczej deskryptorowego JIW występujących tylko w JWI mogą należeć również pewne relatory umożliwiające precyzowanie IW (wskaźniki wagi, relator oznaczający deskryptor główny, itp.). Strategia wyszukiwawcza dostosowana do spełniania wymagań użytkowników danego SIW może przybierać zróżnicowane formy. W wypadku wyszukiwania informacji w oparciu o IW sformułowane w deskryptorowym JIW polega ona m.in. na dokonywaniu zamian leksykalnych, czyli zastępowaniu w IW deskryptora stanowiącego przekład słowa kluczowego pytania informacyjnego odpowiednim deskryptorem ważnym, szerszym lub kojarzeniowym wybranym z tezaurusa. Dokonywanie zamian danego deskryptora na deskryptory węższe lub szersze z wykorzystaniem tezaurusa o budowie hierarchicznej nie przedstawia na ogół trudności i od-

dywa się zazwyczaj w sposób zautomatyzowany (z wyjątkiem najprostszych systemów, w których dopisuje się je w instrukcji wyszukiwawczej ręcznie). Natomiast dokonywanie zamian deskryptora na odpowiednie dla danego pytania informacyjnego deskryptory (lub deskryptor) kojarzeniowe wymaga wybrania ich z artykułu deskryptorowego, umieszczenia w IW i takiego oznaczenia, aby było wiadomo, który z deskryptorów występujących już w instrukcji mają zastępować. Jeśli strategia wyszukiwawcza przewiduje zastępowanie deskryptorami kojarzeniowymi tylko deskryptora głównego, wówczas można je poprzedzać jedynie wskaźnikiem relacji kojarzeniowej (np. symbolem KD zapożyczonym z tezaury), a deskryptor główny odpowiednim wskaźnikiem relacji syntagmatycznej (np. relatorem GD). Mają one również charakter rozkazów systemu. Relatory tego typu oraz relatory arytmetyczne, wykorzystywane głównie w SIW udostępniających dane faktograficzne, mogą również stanowić jedynie elementy leksyki pomocniczej JWI. Jak z tego wynika, w deskryptorowym JIW słownictwo podjęzyka CHWD i słownictwo podjęzyka wyszukiwania informacji jest częściowo różne, choć leksyki te pokrywają się w dużym stopniu.

Zestawy reguł gramatycznych obu podjęzyków deskryptorowego JIW są także częściowo różne. Zarówno w JCHW jak i JWI występuje tekstotwórcza reguła indeksowania współrzędnego, ale w JWI występują dodatkowo reguły zapisu strategii wyszukiwawczej. Należą do nich reguły budowania zapisów iloczynów, sum i różnic zbiorów dokumentów oraz ewentualnie reguły tworzenia złożonych wyrażeń nazwowych z wykorzystaniem elementów leksyki pomocniczej występującej jedynie w słownictwie JWI. Zasady wykorzystywania jednostek leksyki pomocniczej należącej do słownictwa obu podjęzyków też nie zawsze są jednakowe w tych językach. Tak na przykład wskaźniki

roli stosowane obligatoryjnie w JCHW mogą być wykorzystywane w JWI obligatoryjnie albo fakultatywnie, tzn. jedynie w przypadku, gdy wyszukiwanie za pomocą samych deskryptorów doprowadza do otrzymania dużej liczby dokumentów nierelevantnych do pytania informacyjnego.

Tak więc JCHW i JWI są to różne podjęzyki deskryptorowego JIW, które należy definiować oddzielnie. Wydaje się, że najlepszą formą definiowania JIW jest jego charakterystyka formalna, ze względu na jasność i zwięzłość takiego opisu. Poniżej przedstawiono dla przykładu charakterystykę formalną prostego eksperymentalnego języka deskryptorowego oznaczonego symbolem RK, zbudowanego dla jednej z gałęzi technologii żywności /3/. Opis ten zawiera charakterystykę JCHW i charakterystykę podjęzyka instrukcji wyszukiwawczych oraz charakterystykę podjęzyka strategii wyszukiwawczej, przedstawioną oddzielnie w postaci formalnego opisu stosowanej w badaniach metody wyszukiwania dokumentów.

Charakterystyka języka RK

Język RK obejmuje dwa podjęzyki: język, w którym są budowane charakterystyki wyszukiwawcze dokumentów (RK1) oraz język, w którym są budowane instrukcje wyszukiwawcze (RK2).

1. CHARAKTERYSTYKA JĘZYKA RK1

1.1. Kategorie gramatyczne języka RK1

- a) deskryptory - wyrażenia pierwotne języka RK1 równokształtne z pewnymi nazwami języka naturalnego (języka polskiego);
- b) symbole kategorii tematycznych tekstów dokumentów o postaci cyfr rzymskich.

1.2. Słownik języka RK1.

1.2.1. Słownik podstawowy: deskryptory - skończony zbiór wyrazów prostych języka RK;

1.2.2. Słownik pomocniczy: symbole kategorii tematycznych: I - symbol kategorii "wytwarzanie", II - symbol kategorii "jakość", oraz III - symbol kategorii "ekonomika".

1.3. Definicja zdania języka RK1:

Zdanie języka RK1 jest to charakteryzujący dokument jednotreściowy ciąg zestawionych w dowolnej kolejności deskryptorów^{1/}, oznaczony symbolem albo symbolami kategorii tematycznych.

1.4. Definicja tekstu w języku RK1

a/ tekst prosty w języku RK1 jest to pojedyncze zdanie tego języka;

b) tekst złożony w języku RK1 jest to, charakteryzujący dokument wielotreściowy, ciąg n-elementowy ($n \geq 2$) zdań języka RK1.

1.5. Definicja charakterystyki wyszukiwawczej dokumentu w języku RK1:

Charakterystyka wyszukiwawcza dokumentu w języku RK1 jest to dowolny tekst w języku RK1.

2. CHARAKTERYSTYKA JĘZYKA RK2

2.1. Język RK2 to język, który zawiera w sobie również wyrażenia metajęzykowe^{2/}.

^{1/} W języku RK1 stosowane są obligatoryjnie deskryptory stanowiące nazwy języków dokumentów, a więc najkrótsza CHWD zawiera dwa deskryptory - deskryptor określający język dokumentu i deskryptor określający treść dokumentu.

^{2/} Mając na uwadze zasadę niemieszania języków A. Tarskiego należałoby traktować charakteryzowany język RK2 jako metajęzyk RK2, w którym występowałyby przekłady wyrazów języka RK2, ale dla potrzeb JIW odróżnianie tych poziomów za pomocą specjalnych środków przekładu nie jest konieczne.

2.2. Kategorie gramatyczne języka RK2:

- a) deskrytory - wyrażenia pierwotne języka RK2 równokształtne z pewnymi nazwami języka naturalnego (języka polskiego);
- b) symbole kategorii tematycznych tekstów dokumentów o postaci cyfr rzymskich;
- c) relatory literowe stanowiące symbole pewnych relacji w jakich pozostają wyrażenia pierwotne języka RK2.

2.3. Słownik języka RK2.

2.3.1. Słownik podstawowy: deskrytory - skończony zbiór wyrazów prostych języka RK.

2.3.2. Słownik pomocniczy.

2.3.2.1. Symbole kategorii tematycznych: I - symbol kategorii "wytworzenie", II - symbol kategorii "jakość" oraz III symbol kategorii "ekonomika".

2.3.2.2. Relatory:

- a) GD - symbol specyfikujący deskryptor główny oraz KD - symbol specyfikujący pomocniczy deskryptor kojarzeniowy;
- b) operator stanowiący wykładnik dwóch relacji syntagmatycznych specyfikujących - relacji specyfikującej deskryptor, który nie może występować w charakterystykach wyszukiwawczych wyszukanych dokumentów (deskryptor będący nazwą cechy formalnej dokumentu) oraz relacji specyfikującej deskryptor, który nie może być wykorzystany do wyszukiwania dokumentów (deskryptor będący nazwą cechy treściowej dokumentu).

2.4. Definicja złożonego wyrażenia nazwowego języka RK2:

- a) zestawienie deskryptora i relatora GD takie, że relator GD poprzedza deskryptor jest złożonym wyrażeniem nazwowym;

- b) zestawienie deskryptora i relatora KD takie, że relator KD poprzedza deskryptor jest złożonym wyrażeniem nazwowym;
- c) zestawienie deskryptora i relatora \cup takie, że operator \cup poprzedza deskryptor jest złożonym wyrażeniem nazwowym;
- d) żadne inne zestawienie wyrażań nie jest złożonym wyrażeniem nazwowym języka RK2.

2.5 Definicja zdania w języku RK2:

- a) zdaniem w języku RK2 jest deskryptor charakteryzujący treść pytania informacyjnego oraz symbol albo symbole kategorii tematycznych pytania;
- b) zdaniem w języku RK2 jest, charakteryzujący pytanie informacyjne jednotreściowe, ciąg zestawionych w dowolnej kolejności deskryptora lub deskryptorów i jednego lub więcej złożonych wyrażań nazwowych, z których jedno i tylko jedno zawiera relator GD, oznaczony symbolem albo symbolami kategorii tematycznych pytania;
- c) zdaniem w języku RK2 jest, charakteryzujący pytanie informacyjne jednotreściowe, ciąg zestawionych w dowolnej kolejności złożonych wyrażań nazwowych, z których jedno i tylko jedno zawiera relator GD, oznaczony symbolem albo symbolami kategorii tematycznych pytania;
- d) żaden inny ciąg wyrażań nie jest zdaniem języka RK2.

2.6 Definicja tekstu w języku RK2:

- a) tekst prosty w języku RK2 jest to pojedyncze zdanie tego języka;
- b) tekst złożony w języku RK2 jest to, charakteryzujący dokument wielotreściowy, ciąg n-elementowy ($n \geq 2$) zdań języka RK2.

2.7. Definicja instrukcji wyszukiwawczej w języku RK2:

instrukcja wyszukiwawcza w języku RK2 jest to dowolny tekst w języku RK2.

3. CHARAKTERYSTYKA METODY WYSZUKIWANIA DOKUMENTÓW ZA POMOCĄ JĘZYKA RK.

Wyszukiwanie informacji w zbiorze dokumentów eksperymentalnego SIW za pomocą języka RK obejmuje:

- a) sporządzenie charakterystyk wyszukiwawczych dokumentów w języku RK1 z wykorzystaniem instrukcji indeksowania w tym języku i tezaurusa języka RK;
- b) sporządzenie instrukcji wyszukiwawczej w języku RK2 z wykorzystaniem instrukcji indeksowania w tym języku i tezaurusa języka RK;
- c) stosowanie przyjętej strategii wyszukiwania dokumentów, opartej na wykorzystywaniu:
 - zbioru CHWD w języku RK1
 - IW w języku RK2
 - tezaurusa języka RK
 - leksyki pomocniczej języka strategii wyszukiwawczej w postaci operatorów logicznych \cap , \cup , \sim (stanowiących symbole odpowiednio: iloczynu zbiorów, sumy zbiorów, różnicy zbiorów) oraz indeksów o postaci cyfr arabskich, którymi oznacza się w razie potrzeby deskryptory podstawowe inne niż GD i ich opisane deskryptory kojarzeniowe
 - kartoteki inwersyjnej symboli kategorii tematycznych dokumentów
 - kartoteki inwersyjnej deskryptorów (zwanej też zbiorem odwrotnym deskryptorów).

3.1. Strategia wyszukiwawcza

3.1.1. Wyszukanie podzbioru dokumentów o określonej kategorii lub określonych kategoriach tematycznych:

- a) gdy w IW występuje jeden symbol kategorii tematycznej - wyszukanie podzbioru zawierającego dokumenty, w których CHWD występuje dany symbol;
- b) gdy w IW występuje więcej niż jeden symbol kategorii tematycznej - wyszukanie podzbioru zawierającego dokumenty, w których CHWD występuje co najmniej jeden z tych symboli.

3.1.2. Wyszukanie w zbiorze uzyskanym wg p. 3.1.1. podzbioru dokumentów o określonej w IW zawartości informacyjnej (dokumentów o określonej treści i ewentualnie o określonych cechach formalnych).

3.1.2.1. Wyszukanie podzbioru dokumentów reprezentowanych przez jeden deskryptor podstawowy w IW:

- a) gdy w IW występuje tylko jeden deskryptor (np. WIND GRONOWE.) - wyszukanie podzbioru dokumentów zaindeksowanych tym deskrytorem, a następnie, o ile jest to możliwe^{3/}, wyszukanie podzbiorów dokumentów zaindeksowanych jego kolejnymi deskryptorami węższymi;
- b) gdy w IW występuje tylko jeden deskryptor podstawowy oraz jego deskryptor albo deskryptory kojarzeniowe dopisane - wyszukanie podzbiorów dokumentów określonych w p. a) oraz podzbioru lub podzbiorów dokumentów zaindek-

^{3/} W wypadku występowania w IW deskryptora nie posiadającego w artykule deskryptorowym w teaurusie odpowiednich deskryptorów węższych, nie ma możliwości wykorzystywania takich deskryptorów do wyszukiwania hierarchicznego.

sowanych poszczególnymi dopisanymi deskryptorami kojarzeniowymi, a następnie, o ile jest to możliwe, wyszukanie podzbiorów dokumentów zaindeksowanych deskryptorami węższymi w stosunku do deskryptorów kojarzeniowych;

- c) utworzenie zbioru dokumentów A stanowiącego sumę podzbiorów wyszukanych zgodnie z p. a) lub b) wg zapisu z wykorzystaniem operatora \cup (np. WINO GRONOWE. \cup WINO TOKAJSKIE. \cup WINO BURGUNDZKIE.).

3.1.2.2. Wyszukanie podzbioru dokumentów reprezentowanych przez więcej niż jeden deskryptor podstawowy w IW - wyszukanie podzbioru stanowiącego sumę właściwych dla danej IW iloczynów zbiorów dokumentów, reprezentowanych przez poszczególne deskryptory podstawowe.

- (1) Przyjęte zasady poprawnego zapisu iloczynu zbiorów dokumentów:

a) zapis iloczynu zbiorów dokumentów składa się z dwóch lub więcej deskryptorów uszeregowanych w dowolnej kolejności i operatora \cap umieszczonego między sąsiadującymi deskryptorami (np. WINO. \cap SKŁAD CHEMICZNY. \cap EKSPORT.);

b) zapis iloczynu zbiorów dokumentów zawiera:

- albo dany deskryptor podstawowy,
- albo deskryptor węższy w stosunku do danego deskryptora podstawowego,
- albo dopisany deskryptor kojarzeniowy danego deskryptora podstawowego, poprzedzony w IW wskaźnikiem relacji KD, ewentualnie zaopatrzonym w odpowiedni indeks (np. KD₁ STERYLIZACJA. odsyła do

- deskryptora podstawowego STERYLIZATOR.₁, który nie jest deskryptorem głównym).
- albo deskryptor węższy w stosunku do dopisanego deskryptora kojarzeniowego;
- c) jeden i tylko jeden z deskryptorów występujących w zapisie iloczynu zbiorów dokumentów jest:
- albo deskryptorem głównym, czyli deskryptorem podstawowym stanowiącym nazwę przedmiotu głównego pytania informacyjnego, poprzedzonym w IW wskaźnikiem relacji GD,
 - albo deskryptorem węższym w stosunku do danego deskryptora głównego,
 - albo dopisanym deskryptorem kojarzeniowym danego deskryptora głównego, poprzedzonym w IW wskaźnikiem relacji KD
 - albo deskryptorem węższym w stosunku do dopisanego deskryptora kojarzeniowego danego deskryptora głównego;
- d) deskryptor stanowiący nazwę cechy treściowej dokumentu poprzedzony w IW operatorem \cup albo deskryptor węższy w stosunku do tego deskryptora nie może wystąpić w zapisie iloczynu zbiorów dokumentów (nie może być wykorzystany do wyszukiwania iloczynów zbiorów dokumentów).
- (2) Ustalenie i zapisanie iloczynów zbiorów dokumentów tworzonych przy użyciu deskryptorów podstawowych występujących w IW (np. PIMARYCYNA. \cap WIND. \cap KONSERWANT. oraz PIMARYCYNA. \cap WIND.).
- (3) Ustalenie i zapisanie iloczynów zbiorów dokumentów

tworzonych przy użyciu deskryptorów podstawowych oraz dopisanych w IW deskryptorów kojarzeniowych (np. PIMARYCYNA. \cap WIND. \cap TOKSYCZNOŚĆ. oraz PIMARYCYNA. \cap WIND. \cap KONSERWOWANIE CHEMICZNE.).

- (4) Ustalenie i zapisanie iloczynów zbiorów dokumentów przy użyciu wszystkich, wybranych z tezaury, deskryptorów węższych w stosunku do poszczególnych deskryptorów występujących w zapisach iloczynów zbiorów dokumentów utworzonych wg p. (2) i (3).
- (5) Ustalenie i zapisanie sumy podzbiorów dokumentów określonych w p. (2), (3) i (4) z wykorzystaniem operatora \cup (np. PIMARYCYNA. \cap WIND. \cup TOKSYCZNOŚĆ. PIMARYCYNA. \cap WIND. \cap KONSERWOWANIE CHEMICZNE.).
- (6) Wyszukanie dokumentów za pomocą kolejnych zapisów iloczynów zbiorów dokumentów - wszystkie deskryptory występujące w zapisie iloczynu muszą występować w CHWD dokumentu wyszukanego w oparciu o ten zapis.
- (7) Utworzenie zbioru dokumentów A, stanowiącego sumę iloczynów zbiorów wyszukanych wg p. (6), zgodnie z zapisem wg p. (5).

3.1.2.3. Wyszukanie, o ile konieczność tego wynika z zapisu IW, żadanego podzbioru zbioru dokumentów A (utworzonego wg p. 3.1.2.1. lub p.3.1.2.2.) metodą tworzenia różnicy lub kolejnych różnic zbiorów dokumentów wg zapisów z wykorzystaniem operatora - i literowych symboli podzbiorów:

- a) gdy w IW występuje tylko jeden poprzedzony operatorem \cap deskryptor będący nazwą cechy formalnej dokumentu wyszukiwanie prowadzi się wg następującego zapisu różnicy zbiorów

$$A - B = C$$

gdzie: B to podzbiór zbioru A, obejmujący dokumenty, których CHWD zawierają deskryptor poprzedzony operatorem \subset w IW (np. J PORT.).

C to podzbiór zbioru A, obejmujący dokumenty, których CHWD nie zawierają danego deskryptora (w ramach zbioru A stanowiącego uniwersum podzbiór C jest poszukiwanym zbiorem dokumentów, a podzbiór B jest jego dopełnieniem);

- b) gdy w IW występuje więcej niż jeden poprzedzony operatorem \subset deskryptor stanowiący nazwę cechy formalnej dokumentu (np. J. PORT. J. POL. ART PRZEGL.) wyszukiwana
- nie prowadzi się kolejno wg następujących zapisów różnic zbiorów

$$A - B = C$$

$$C - D = E$$

$$E - F = G$$

itd.

aż do uzyskania ostatniego podzbioru (np. zbioru G) obejmującego tylko te dokumenty, których CHWD nie zawierają żadnego z W.W. deskryptorów.

3.2. Sprawdzenie prawidłowości rezultatów wyszukiwania przez prowadzącego wyszukiwanie i uznanie podzbioru dokumentów określonego w wyniku postępowania wg p. 3.1.1. i 3.1.2. za zbiór dokumentów stanowiący odpowiedź SIW na dane pytanie informacyjne.

Wydaje się, że przedstawiony powyżej opis formalny w sposób dostatecznie jasny i dokładny definiuje eksperymentalny język RK. Sporządzanie charakterystyki tego typu w wypadku każdego budowane-

go języka deskryptorowego ułatwiłoby pracę indeksatorów, -a także wpłynęłoby prawdopodobnie korzystnie na jednolitość i efektywność wykorzystywania tego języka w wyszukiwaniu informacji.

Literatura

1. CHMIELEWSKA-GORCZYCA E. Struktura instrukcji wyszukiwawczej. Zagadnienia Informacji Naukowej 1984 nr 1(44) s.49-63
2. CHMIELEWSKA-GORCZYCA E. Struktura języka informacyjno-wyszukiwawczego. Zagadnienia Informacji Naukowej 1982 nr 2(41) s.41-64.
3. WERESZCZYŃSKA-CISŁO B. Przydatność specyfikacji relacji kojarzeniowych w procesie wyszukiwania informacji z zakresu wybranej gałęzi technologii żywności. Rozprawa doktorska. Warszawa: Uniwersytet Warszawski 1984, 431 s.
4. WERESZCZYŃSKA-CISŁO B., OGÓRKIEWICZ W. Pytania informacyjne w ujęciu teoretycznym. Zagadnienia Informacji Naukowej 1986 nr 1(48) s. 41-59.

15.01.1990 r.

DEFINING OF THE DESCRIPTOR LANGUAGE

Summary

There are discussed: the vocabulary of the descriptor language comprising its autosyntactic and syn-syntactic units, the sentence and the text of this language. The sublanguages of the descriptor language are distinguished: the language of the retrieval characteristics of documents and the information retrieval language. The

differences between them and their common features are shown. Vocabularies and grammatical rules of both languages are compared. It is proposed to define them separately. The author suggests to define the information retrieval language through its formal characteristics which should ensure the clarity and the consistence of the description. The author uses - as an example - the formal characteristics of the simple experimental descriptor language, marked by symbol RK, built for one of the branches of the food technology. The characteristics of the sub-language of the retrieval instructions (search statements) and of the sublanguage of the search strategies are presented separately in the form of the formal description of the searching method used within the research.

ОПРЕДЕЛЕНИЕ ДЕСКРИПТОРНОГО ЯЗЫКА

Резюме

Обсуждается лексика дескрипторного языка: автосинтактические и син-синтактические единицы, а также предложение и текст дескрипторного языка. Выделены под-языки дескрипторного языка: язык поисковых характеристик документов и язык поиска информации. Указаны различия между ними и их совместные черты. Сравнены лексика и грамматические правила обеих языков. Предлагается определить язык поиска информации (информационно-поисковый язык) с помощью его формальной характеристики в виду ясности и краткости описания. Авторка использует - в качестве примера - формальную характеристику простого, экспериментального дескрипторного языка, обозначенного символом RK, построенного для одной из отраслей пищевой технологии. Характеристика его под-языка, поисковых предписаний и под-языка поисковых стратегий представлена отдельно в виде формального описания метода поиска документов использованного в исследовании.

**UDOSTĘPNIANIE WIEDZY W DIALOGOWYM SYSTEMIE
INFORMACYJNO-WYSZUKIWAWCZYM (ON-LINE)**

Wyszukiwanie informacji bez transformacji oraz z transformacją semantyczną. Zasady funkcjonowania dialogowych systemów informacji on-line. Kryteria relewancji w systemie on-line i sposoby ich prezentacji. Wiedza o języku informacyjno-wyszukiwawczym w systemie i jej uwarunkowanie instrumentami informacyjnymi. Problemy komunikacji użytkownika z systemem.

Gromadzenie i strukturalizacja wiedzy systemu informacyjno-wyszukiwawczego według odwzorowanych w języku informacyjnym kryteriów relewancji należą do tych procesów informacyjnych, które poprzedzają udostępnianie informacji przechowywanej w systemie drogą jej transformacji, a więc przetwarzania. Termin udostępnianie wiedzy systemu został wprowadzony dopiero w pracach związanych z rozwojem systemów inteligentnych, w których wskazano dwa sposoby używania zbiorów informacyjnych systemu:

- a) wyszukanie informacji relewantnej, a więc jej wyselekcjonowanie,
- b) wygenerowanie informacji relewantnej w procesach wnioskowania na podstawie wiedzy o dokumentach oraz o obiektach pozadokumentacyjnych (tzw. wiedzy deklaratywnej) i reguł jej transformacji prawdziwościowej (tzw. wiedzy proceduralnej) /2/.

Na tej podstawie możemy wskazać systemy, w których:

1) udostępnianie informacji drogą jej selekcjonowania określone jest jedynie przez relację podobieństwa charakterystyki wyszukiwawczej dokumentu (lub obiektu pozadokumentacyjnego) do instrukcji wyszukiwawczej, a więc systemy udostępniające wiedzę bez transformacji,

2) udostępnianie informacji polega na dokonywaniu przez system transformacji semantycznej wyrażen ze zbioru systemu będącej równocześnie transformacją prawdziwościową, a więc taką, że z wyrażen wyjściowych (przesłanek) uznawanych w systemie za prawdziwe otrzymuje się inne wyrażenia (wnioski) prawdziwe w stopniu takim samym, jaki przysługuje przesłankom (wnioskowanie logiczne) lub mniejszym (wnioskowanie uprawdopodobniające), a więc systemy udostępniające wiedzę po jej transformacji.

W literaturze z zakresu języków i systemów informacyjnych termin transformacja stosowany jest często w znaczeniu bliższym selekcjonowaniu informacji - na przykład streszczanie tekstów dokumentów pierwotnych interpretowane jest jako transformacja tekstu połączona ze zmianą języka naturalnego na inny język naturalny, w wyniku której powstaje inny tekst krótszy, zawierający wyrażenia będące wykładnikami relewancji /13/. Transformacje wyrażen językowych dokonywane w procesie przekładu zapytania informacyjnego na instrukcję wyszukiwawczą i formułowania strategii wyszukiwawczej również są transformacjami semantycznymi i prawdziwościami wyznaczanymi przez funkcję presupozycji /1, 19/. Sprowadzenie różnic w sposobach udostępniania wiedzy przez system do dokonywania (lub nie) transformacji prawdziwościowej informacji w procesach wnioskowania byłoby jednak dużym uproszczeniem. Należy zauważyć, że chodzi tu właściwie o moment, w którym

transformacja jest dokonywana. W systemach tradycyjnych, w których ocena relewancji i pertynencji informacji wyselekcjonowanej przez system odbywa się na podstawie końcowych wyników wyszukiwania, proces prawdziwościowej transformacji informacji poprzedza realizację strategii wyszukiwawczej. Dążenie do zapewnienia maksymalnej kompletności i relewancji wyników wyszukiwania powoduje, że cała wiedza o rzeczywistości pozajęzykowej jest wbudowana w język informacyjno-wyszukiwawczy łącznie z wiedzą językową. W systemach dialogowych (on-line), w których użytkownik komunikuje się na zasadzie "akcja - reakcja" zależnie od oceny pertynencji wyszukanej informacji, transformacja odbywa się w trakcie sesji wyszukiwawczej i polega na zmianie zbioru wyrażen składowych instrukcji wyszukiwawczej odwzorowujących kryteria relewancji. Wpływa to na zakres wykładników relewancji wbudowanych w język informacyjno-wyszukiwawczy. W systemach inteligentnych (faktograficznych) transformacja semantyczna i prawdziwościowa informacji polegająca na wykorzystaniu jego wyrażen składowych jako przesłanek w procesie wnioskowania prowadzącego do wygenerowania odpowiedzi relewantnej realizowana jest przez system, formułujący strategię wyszukiwawczą na podstawie własnej interpretacji zapytania. System określa więc samodzielnie algorytm realizacji funkcji wyszukiwawczej na podstawie wiedzy proceduralnej oraz wiedzy deklaratywnej o pozajęzykowych obiektach, ich cechach oraz sytuacjach, w których są użyte.

Funkcjonowanie systemów dialogowych omawiamy przede wszystkim z punktu widzenia wpływu sposobu realizacji funkcji wyszukiwawczej na prezentację wykładników cech relewantnych. Należy przy tym podkreślić, że powstanie systemów dialogowych było możliwe dzięki rozwojowi technicznych instrumentów informacyjnych:

- powstaniu komputerów z pamięciami dyskowymi umożliwiającymi szybkie wyszukiwanie informacji i działających w trybie wielodostępny,
- rozwojowi sieci telekomunikacyjnych umożliwiającym szybki przekaz informacji między komputerami przy wykorzystaniu kanałów łączności satelitarnej, linii telefonicznych itp.,
- opracowaniu oprogramowania dostosowanego do organizacji informacji na nowych nośnikach (w pamięciach dyskowych) oraz do przetwarzania informacji w trybie interakcyjnym na podstawie procedur interrogacyjnych.

Czynnikiem pośrednio wpływającym na powstanie systemów dialogowych był ilościowy rozwój zbiorów informacji na nośnikach maszynowych (baz danych). Pierwsze wydanie katalogu opracowywanego od lat siedemdziesiątych na Uniwersytecie w Illinois (USA), uznawanego za jedną z najbardziej reprezentatywnych "baz danych o bazach danych" zawierało charakterystykę około trzystu skomputeryzowanych zbiorów informacji (według danych opublikowanych w 1976 roku); Wydanie czwarte z 1985 roku podaje informacje o około trzech tysiącach baz /5/.

STRATEGIA WYSZUKIWAWCZA W SYSTEMIE DIALOGOWYM

Przyjęcie zasady indywidualnego szybkiego komunikowania się użytkownika z systemem dysponującym zbiorami informacji na dyskach zawierających od kilkudziesięciu do kilkuset tysięcy rekordów z dowolnej dziedziny wiedzy wiązało się ze zmianą jego struktury funkcjonalnej. Zmiana ta polegała na rozróżnieniu ról producenta informacji (data base producer, supplier) oraz dystrybutora informacji nazywanego także operatorem systemu (vendor, operator). Zmianę tę zapoczątkowały w latach sześćdzie-

siątych w USA firmy Lockheed oraz Systems Development Corporation (SDC), które rozpoczęły udostępnianie informacji w systemach on-line bez ingerowania w zasady udostępniania zbiorów tradycyjnie opracowywanych przez producentów baz danych w postaci wydawnictw abstraktowych, katalogów bibliotecznych i indeksów przedmiotowych, a więc przez biblioteki naukowe, służby abstraktowe, instytucje naukowe i in.

Prace nad budową dwu najstarszych i największych pod względem zbiorów informacji oraz zasięgu działania systemów on-line Lockheed-DIALOG oraz SDC-ORBIT rozpoczęto w latach sześćdziesiątych. W roku 1963 firma Lockheed utworzyła Pracownię Informacji Naukowej (Information Sciences Laboratory) prowadzącą badania podstawowe nad automatyzacją przetwarzania informacji. W roku 1966 wygrała konkurs na budowę systemu informacji bibliograficznej on-line dla National Aeronautics and Space Administration (NASA), podejmując opracowanie oprogramowania programów wyszukiwawczych nazwanego RECON (REmote CONsole), za pomocą którego pierwotnie wyszukiwano informację w zbiorach "International Aerospace Abstracts", "Scientific and Technical Aerospace Reports" tworzonych w NASA. Programy RECON zaczęto następnie wykorzystywać w US Atomic Energy Commission, the US Office of Education oraz National Technical Information Service (NTIS, ESA). Od roku 1972 firma Lockheed stosuje zmodyfikowaną wersję RECON we własnym systemie on-line, udostępnianym użytkownikom na zasadach komercyjnych pod nazwą DIALOG. System zaczął funkcjonować z trzema bazami danych; liczba ta powiększała się systematycznie. W roku 1975 system posiadał 24 bazy, w roku 1978 - ponad 70, w roku 1980 - ponad sto, w połowie lat osiemdziesiątych - około 500.

Budowę systemu (programów wyszukiwawczych) SDC Orbit^{1/} podjęto w roku 1963 na zlecenie US Department of Defence Advanced Research Projects Agency (ARPA) dysponującego wówczas zbiorem informacji zawierającym ponad 200 tys. rekordów bibliograficznych. W roku 1968 firma SDC rozpoczęła prace na zlecenie amerykańskiej National Library of Medicine nad oprogramowaniem do interakcyjnego wyszukiwania informacji w zbiorach wydawnictwa abstraktowego Index Medicus. Programy te wraz z językiem wyszukiwawczym ELHILL (Lister Hill National Centre for Biomedical Communication) są stosowane od 1970 roku w systemie MEDLINE. Od 1973 roku SDC uruchomiła własny system SDC ORBIT funkcjonujący na zasadach komercyjnych, posiadający w momencie uruchomienia trzy bazy danych, w 1980 - około 60.

Kolejne systemy on-line o zasięgu międzynarodowym powstawały w drugiej połowie lat siedemdziesiątych; wśród nich m.in.:

- Bibliographic Retrieval Service (BRS) w State University of New York w 1976 r. funkcjonujący ze zmodyfikowaną wersją oprogramowania STAIRS,
- US New York Times Information Bank,
- Infoline w zakresie nauk biomedycznych (Wielka Brytania) - od 1979 r.,
- BLAISE (British Library Automated Information Service) - od 1977 r.,
- AUSINET w Australii,
- EURONET-DIANE (Europejska Wspólnota Gospodarcza) - od 1980 r.

Do zadań dystrybutorów baz danych w systemie on-line (instytucji i organizacji komercyjnych, dużych bibliotek naukowych o zasięgu międzynarodowym typu National Library of Medicine, National

^{1/} ORBIT - Online Retrieval of Bibliographic Information and Text

Agricultural Library, Library of Congress, Ohio College Library Centre. British Library, międzynarodowych organizacji naukowych typu European Space Agency) należało:

- dokonywanie konwersji indywidualnych formatów danych przyjętych w zbiorach informacji poszczególnych producentów na format systemu, a więc transformacja struktur tekstów charakterystyk wyszukiwawczych dokumentów (ChWD) w planie wyrażania na strukturę obowiązującą w systemie,

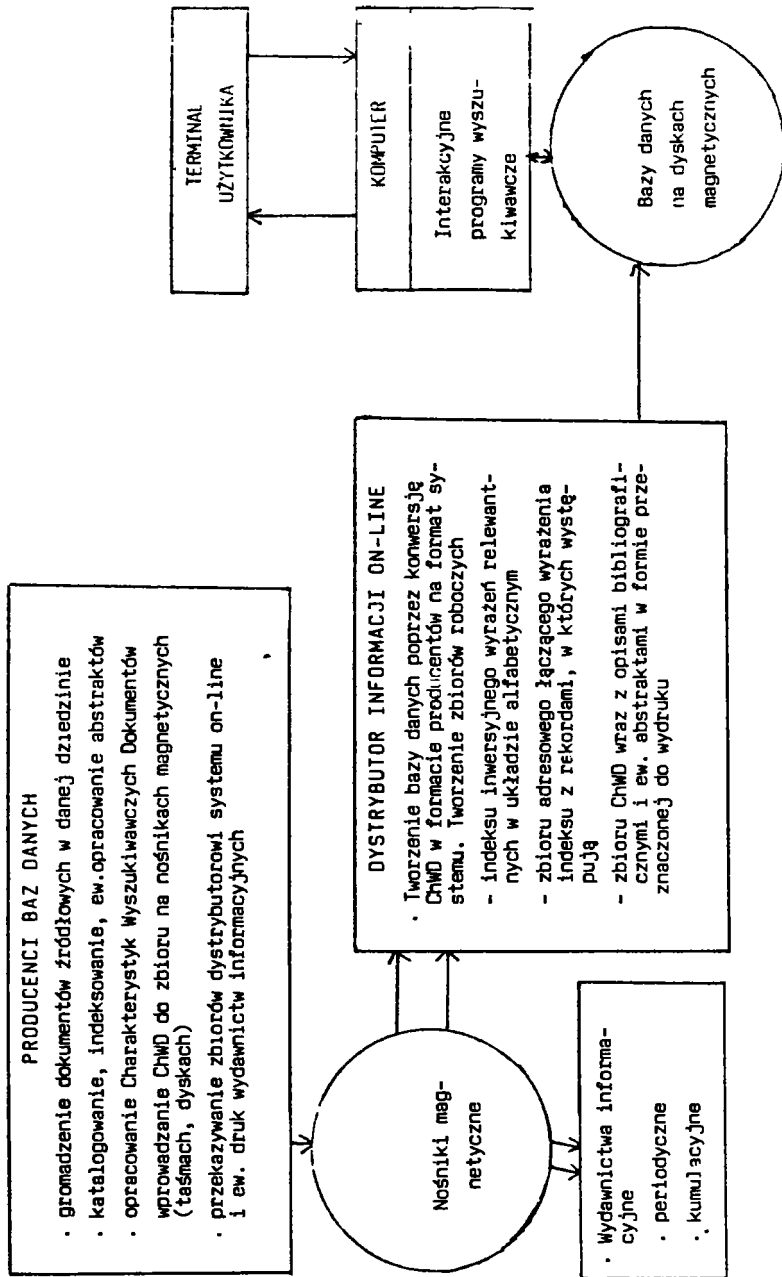
- tworzenie i aktualizacja bazy danych systemu zorganizowanej w sposób ułatwiający selekcję informacji według różnych kluczy wyszukiwawczych, a więc według wyrażen pełniących funkcję metatekstową, wskazujących wyrażenia relewantne w procesie wyszukiwania,

- opracowanie języka wyspecjalizowanego w funkcji wyszukiwawczej umożliwiającego użytkownikowi prowadzenie dialogu z systemem kurtuazyjnie reagującym na zapytania,

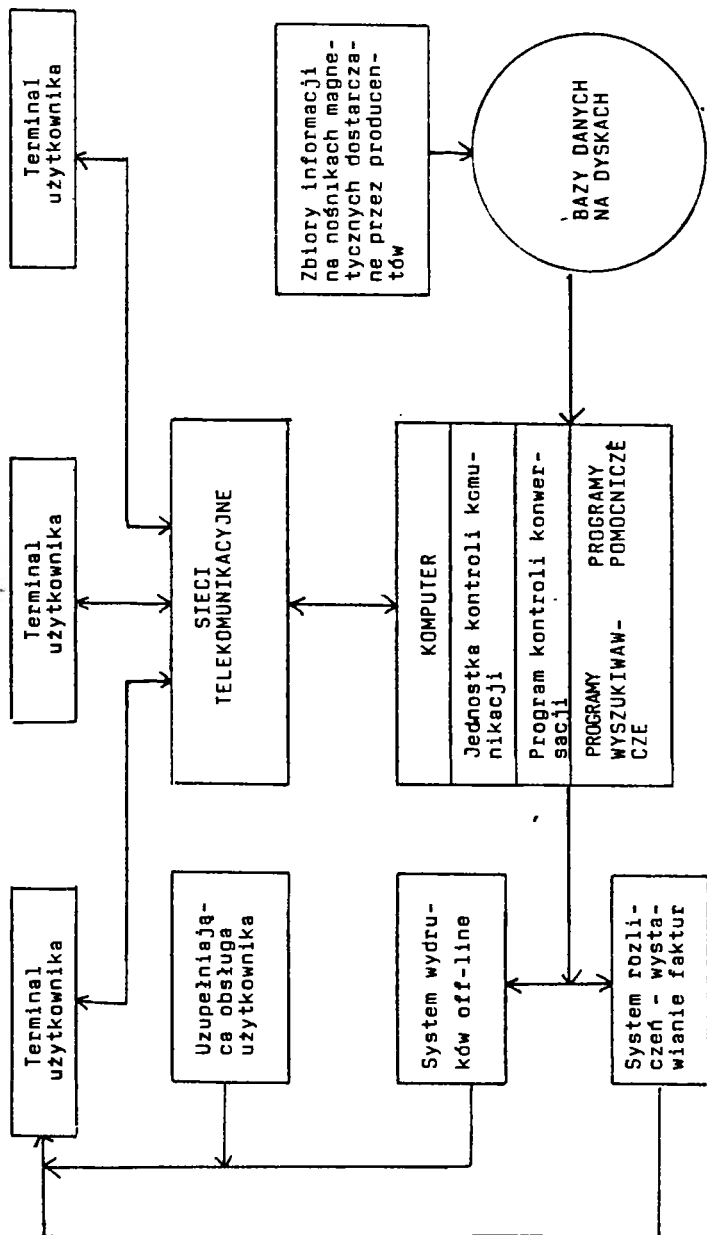
- dostarczanie użytkownikowi informacji ze zbiorów systemu opracowanej w trybie wsadowym (off-line) i ewentualnie kopii dokumentów źródłowych. Zasady podziału funkcji w systemie on-line oraz układ instrumentów informacyjnych tego systemu ilustrują schematy 1 i 2 /12/.

PREZENTACJA KRYTERIÓW RELEWANCJI W SYSTEMIE ON-LINE

Podstawą selekcjonowania informacji w systemie jest, jak zaznaczono w załączonym schemacie, indeks wyrażen relewantnych (indeks podstawowy - basic index) nazywanych terminami wyszukiwawczymi (search terms) uporządkowanymi alfabetycznie, w stosunku do którego zbiór adresowy oraz zbiór ChWD w systemie dokumentacyjnym pełnią funkcję pomocniczą związaną z technicznymi (informatycznymi) aspektami przetwarzania informacji. Świadczy to o przyjęciu



Schemat 1 FUNKCJE UCZESTNIKÓW SYSTEMU INFORMACYJNO-WYSZUKIWAWCZEGO ON-LINE



Schemat 2. INSTRUMENTY INFORMACYJNE W SYSTEMIE ON-LINE

innych niż w systemach tradycyjnych założeń określających minimalny warunek relewancji systemowej - każde wyrażenie umieszczone w indeksie inwersyjnym, niezależnie od jego miejsca w strukturze paradygmatycznej JIW, może być relewantne. Ewentualną rozbieżność między systemową oceną relewancji i oceną użytkownika (pertynencji) rekompensuje możliwość zmiany strategii wyszukiwawczej w trakcie sesji wyszukiwawczej dzięki szybkiej reakcji systemu na przekazywane przez użytkownika komendy sformułowane w języku wyszukiwawczym systemu (języku kwerend - query language).

Każdy z omawianych przykładowo systemów posiada własny schemat dialogu z użytkownikiem, określony w tzw. protokołach konwersacji, niezależnych od języków informacyjnych stosowanych w funkcji metainformacyjnej przez producentów baz danych. Protokoły te i związane z nimi języki kwerend, oparte najczęściej na alternatywnej notacji paranaturalnej i sztucznej, wskazują rodzaje operacji wykonywanych przez komputer w celu wyszukania informacji. Stanowią zasadniczą część dokumentacji instruktażowej każdego systemu on-line. Różniąc się w szczegółach pod względem repertuaru operacji i ich nazw, co stwarza kłopoty użytkownikom korzystającym równoległe z różnych systemów, są zbieżne pod względem etapów realizacji procesu wyszukiwania,

W dokumentacji instruktażowej poszczególnych systemów można wyróżnić następujące grupy operacji niezbędnych do przeprowadzenia sesji wyszukiwawczej:

1. Operacje telekomunikacyjne (podejmowania i kończenia dialogu z systemem - "Logon and Logoff"), do których należy:
 - a) uruchomienie sprzętu: terminala, modemu, drukarki,
 - b) uzyskanie połączenia z siecią telekomunikacyjną^{2/}, na przv-

^{2/} Systemy amerykańskie udostępniane są m.in. za pośrednictwem dwu sieci TELENET i TYMNET. Istnieją ponadto sieci regionalne, np. SCANNET, EURONET AUSINET, międzybiblioteczne - OCLC /15/

kład poprzez linię telefoniczną, po przekazaniu identyfikatora terminala, następnie hasła (password) otwierającego dostęp do sieci oraz kolejnego hasła (password) otwierającego dostęp do systemu on-line,

c) reagowanie na informacje przekazywane z sieci lub systemu zakłóceń w łączności lub niesprawności sprzętu dystrybutora,

d) ortograficzną korektę tekstów pisanych na ekranie przez użytkownika,

e) uzyskanie informacji o kosztach wyszukiwania.

2. Operacje selekcjonowania informacji za pomocą języka kwerend, wśród których można wyróżnić następujące grupy procedur interrogacyjnych umożliwiających:

a) wybór baz danych (zbioru informacyjnego), w których ma być wyszukiwana informacja,

b) przegląd indeksów baz danych zawierających wyrażenia uznane za wykładniki relewancji w systemie,

c) formułowanie strategii wyszukiwawczej - wprowadzanie wyrażen składających się na instrukcję wyszukiwawczą,

d) modyfikację instrukcji wyszukiwawczej (search statements), np. przez ograniczenie zbioru ChWD,

e) przegląd instrukcji wyszukiwawczych stosowanych w trakcie sesji wraz z wynikami wyszukiwania uzyskanymi na poszczególnych etapach,

f) wydruk (w trybie on-line lub of-line) wyselekcjonowanej informacji,

g) uzyskanie informacji pomocniczej o procedurach interrogacyjnych dopuszczalnych w ramach stosowanego języka wyszukiwawczego (języka kwerend),

h) uzyskanie informacji o bieżących zmianach w zakresie organizacji zbiorów informacji w systemie oraz ich zasobach.

Spośród wymienionych dwu grup operacji pozostających w dyspozycji użytkownika prowadzącego dialog z systemem, przedmiotem naszego zainteresowania są operacje grupy drugiej służące selekcjonowaniu informacji. Procedury telekomunikacyjne i inne związane z technicznymi i czysto pragmatycznymi aspektami funkcjonowania systemu stanowią w naszych rozważaniach zagadnienia marginalne.

Realizacja operacji selekcjonowania informacji możliwa jest - jak wspomnieliśmy - dzięki językowi kwerend, którego wyspecjalizowaną funkcją jest funkcja impresywna sterowania operacjami komputera /16/. Jest on funkcjonalnie komplementarny wobec języków informacyjnych, stosowanych w poszczególnych bazach danych w funkcji metainformacyjnej. Wyszukanie informacji w systemie on-line nie jest możliwe wyłącznie za pomocą języka służącego do opracowania CHMD lub charakterystyk obiektów pozajęzykowych w systemach faktograficznych. Funkcji tej nie można także zrealizować wyłącznie przy użyciu samego języka kwerend. Tak więc nowy rodzaj instrumentu informacyjnego w systemie (interakcyjne programy wyszukiwawcze) powodują wyodrębnienie w JIW podzbiorów jego wyrażeń wyspecjalizowanych w funkcjach: metainformacyjnej, wyszukiwawczej i impresywnej. Z realizacją funkcji wyszukiwawczej i impresywnej wiąże się stosowanie operatorów logicznych oraz operatorów tekstowych, które omawiamy w dalszej części artykułu. Interesującą cechą niektórych systemów on-line jest możliwość zmiany języka kwerend przy wyszukiwaniu w tym samym zbiorze, udostępnionym za pośrednictwem różnych sieci teleinformatycznych. Na przykład w systemie ESA-IRS (European Space Agency-Information Retrieval System) możliwa jest zmiana języka 'QUEST podobnego do Lockheed-DIALOG na język SESAME (EUROLANGUAGE), jeden z języków sieci EURONET przez podanie odpowiedniej komendy "SET SESAME ON" lub EUROLANGUAGE ON".

Użytkownik systemu on-line, dysponujący często dostępem równocześnie do kilku systemów teleinformatycznych (o zasięgu światowym lub przynajmniej regionalnym) staje nie tylko przed problemem metod selekcji informacji w zbiorach liczących miliony rekordów, lecz także przed problemem wyboru bazy danych, a więc jednego z podzbiorów systemu. Pojawia się pytanie "gdzie wyszukiwać informację", poprzedzające problem "jak wyszukiwać". W ten sposób w systemie on-line zostaje wskazane nowe kryterium relewancji, nie nazywane explicite w systemach tradycyjnych, mianowicie informacji o zasobach informacyjnych systemu, czyli o bazach danych bibliograficznych i faktograficznych^{3/}. Należy zauważyć, że ten istotny dla użytkownika rodzaj relewancji, oznacza wyróżnienie kolejnego poziomu metainformacji, gdyż jest to:

- informacja o metainformacyjnych bazach danych dokumentacyjnych (bibliograficznych), a więc ~~meta-metainformacja~~, a także
- informacja o informacyjnych bazach danych faktograficznych, a więc metainformacja.

Prezentacji kryteriów relewancji odwzorowujących rodzaj informacji w podzbiorach systemu służą komendy przeznaczone do:
- wykazania zasobów informacyjnych systemu (pkt."h", "g"),
- wyboru jednej lub więcej baz danych, w których ma być wyszukiwana informacja. (por. tab. 1).

Ten rodzaj relewancji odwzorowany jest także w indeksie przedmiotowym wspomnianego katalogu baz danych /5/. Scharakteryzowanie zakresu tematycznego baz danych jest istotne dla użytkownika ze względu na merytoryczne i ekonomiczne, ponieważ do zjawisk

^{3/} W systemach amerykańskich operuje się terminami "baza danych bibliograficznych" (bibliographic data base) i "baza danych niebibliograficznych" (non-bibliographic data base). Ten dychotomiczny podział zbiorów informacyjnych systemu pozwala (pragmatycznie) rozwiązać definicyjne dylematy związane z terminem "fakt" (faktograficzny).

Tab. 1. SŁOWNIKI JĘZYKÓW WYSZUKIWAWCZYCH NIEKTÓRYCH SYSTEMÓW ON-LINE
(nazwy podstawowych operacji selekcjonowania informacji)

Rodzaj operacji	DIALOG		SDC. ORBIT	BRS	DIANE	ESA-IRS
	Nazwa komendy	Skrót				
Uzyskiwanie (meta-) metainformacji o zbiorach systemu	?FILESn ?FILESUM		FILES?		INFO	?FILES
Uzyskiwanie metainformacji o JM	?EXPLAIN		EXPLAINitem(EXP,?)		HELP	?QUEST ?SESAME ?FIELDS
Uzyskiwanie metainformacji o strukturze CHWD (polach wyszukiwawczych)	?FIELDn					
Uzyskiwanie informacji o bieżących zmianach w systemie	?NEWS ?WHATNEW		NEWS	NEWS	NEWS	?NEWS
Uzyskiwanie informacji o aktualizacji zbiorów w bieżącym tygodniu i poprzednim	?UPDATE					?UPDATE
Wybór bazy danych do wyszukiwania	BEGIN	Bn I	FILE	CHANGE C	BASE	BEGINn
Przegląd indeksu wyrażen relewantnych	EXPAND	E "	NEIGHBOR	ROOT*	DISPLAY	EXPAND E
Polecenie wyszukania CHWD/CHWD według wskazanej instrukcji wyszukiwawczej	SELECT SUPERSELECT	S #	FIND	SEARCH S	FIND	SELECT S
Modyfikacja (ograniczenie) zakresu przeszukiwanego zbioru	LIMIT	L	FIND	LIMIT L	LIMIT	LIMIT I
Przegląd wyników stosowanej strategii wyszukiwawczej	DISPLAY SETS	DS a	HISTORY	DISPLAY ALL	BACK	DISPLAYSETS DS
Wydruk wyszukiwanej informacji: - on-line - off-line	TYPE PRINT	T % PR	PRINT PRINT OFFLINE - - PRT OF	PRINT PRINTOFF	SHOW PRINT	TYPE PRINT

powszechnych należy częściowe pokrywanie się zasobów informacyjnych poszczególnych baz tworzonych przy wykorzystaniu różnych JIW. Oprogramowanie większości systemów on-line umożliwia obecnie przeprowadzanie wyszukiwania równoległe w kilku bazach danych na podstawie tej samej instrukcji wyszukiwawczej, co z jednej strony decyduje o obniżeniu kosztów wyszukiwania, z drugiej zaś pozwala porównać rezultaty wyszukiwania w bazach o pokrewnym zakresie tematycznym, np. COMPENDEX (Computerized Engineering Index), ISMEC (Information Services for Mechanical Engineering), INSPEC zawierają wspólne pole tematyczne "fizyka". Analogicznie w systemach udostępniających informację katalogową, np. w systemie SOCRATES, Stanford University USA, możliwe jest wyszukiwanie informacji równocześnie w katalogach wydawnictw zwartych, ciągłych, filmów, map, archiwów, raportów z badań lub w jednym z nich, wskazanym, przez użytkownika /14/.

WIEDZA O ZBIORACH INFORMACYJNYCH SYSTEMU

Meta-metainformacyjny opis zasobów informacyjnych systemu dostarczany użytkownikowi w bardzo obszernej, zalgorytmizowanej dokumentacji instruktażowej^{4/}, w układzie podporządkowanym operacjom realizowanym za pomocą instrukcji języka wyszukiwawczego; stanowi jeden ze sposobów "humanizacji" procesu komunikowania się użytkownika z systemem. Służy temu wymieniona w tabeli komenda "NEWS", występująca w repertuarze instrukcji, jak też instrukcje umożliwiające uzyskiwanie informacji o zasobach informacyjnych

^{4/} Dokumentacja instruktażowa systemu Lockheed-DIALOG wydana w 1979r. składała się z czterech kilkusetstronicowych tomów, z których pierwszy zawierał opis procedur telekomunikacyjnych oraz języka wyszukiwawczego, natomiast trzy pozostałe - poświęcone były charakterystyce poszczególnych baz danych.

systemu, w postaci HELP, FILES, INFO, INSTRUCT oraz EXPLAIN związana z prezentacją wiedzy o JIW.

Wykorzystanie tych komend może dotyczyć różnych poziomów organizacji informacji w zbiorach. Na przykład w systemie DIALOG instrukcja:

:?EXPLAIN EXPLAIN umożliwia przekazanie przez system informacji o zasobie słownika języka kwerend

:?FILES powoduje wydruk listy wszystkich baz danych dostępnych w momencie wyszukiwania dla użytkownika legitymującego się określonym hasłem (Zbiory poufne "confidential files" są udostępniane nie wszystkim użytkownikom).

Możliwe jest ograniczenie wykazu baz danych przez dołączenie numeru bazy, np. FILES 90 spowoduje wydruk listy baz, których numer jest wyższy niż 90;

?FILE n - powoduje przekazanie syntetycznego opisu bazy danych, np.

?FILE 47 : MAGAZINE INDEX PROVIDES COVERAGE OF OVER 350 POPULAR AMERICAN MAGAZINES TO SUPPLY INFORMATION ON A BROAD RANGE OF TOPICS INCLUDING CURRENT AFFAIRS, CONSUMER PRODUCT EVALUATION, SPORTS AND RECREATION, BUSSINESS, ENVIRONMENT, SCIENCE AND TECHNOLOGY. THE FILE CONTAINS RECORDS FROM 1977 TO THE PRESENT. APPROXIMATELY 5000 NEW RECORDS ARE ADDED MONTHLY, RETROSPECTIVE COVERAGE FOR 1976 IS CURRENTLY IN PROCESS..AS OF SEPTEMBER 1978, THE FILE CONTAINED OVER 100.000 RECORDS. COST OF SEARCHING IS \$ 45/CONNECT HOUR AND \$ 10/FULL RECORD PRINTED OF LINE. SEE ALSO ?FIELD 47

?FIELD - służy prezentacji tych wszystkich cech dokumentu lub obiektu pozajęzykowego, których cechy są odwzorowywane w charakterystyce wyszukiwawczej, powodując wydruk wszystkich pól wyszukiwaw-

czych rekordu danej bazy. W wykazie pól wyróżnia się odpowiadające cechom formalnym dokumentu (tzw. prefix-coded fields) oraz pola tekstowe (full-text fields), np.

?FIELD 75

FIELD 75: MANAGEMENT CONTENTS

PREFIX-CODED FIELDS (EXPAND OR SELECT)

FIELD NAME	EXAMPLE
AUTHOR	AU-ZUCKER,S.
DESCRIPTOR CODE	DC=0933
PUBLICATION DATE	DT=76/08
JOURNAL CODE	JC=NAB
JOURNAL NAME	JN=NATIONS BUSSINESS
PUBLICATION YEAR	PY=1976
UPDATE	UD=9999

FULL-TEXT FIELDS (SELECT ONLY)

ABSTRACT	ADVERTISING (F)MODEL/AB - (F) operator tekstowy
DESCRIPTOR	MANAGEMENT STYLE
TITLE	MEDIA(F)RESPONSE/TI

?FIELD 106

FIELD 106: TRADE OPPORTUNITIES WEEKLY (baza faktograficzna)

PREFIX-CODED FIELDS	EXAMPLE
BUSSINESS CODE	BC- A
BUSSINESS NAME	BN=AGENT
COUNTRY CODE	CC=241
COUNTRY NAME	CN=BRAZIL
DATE OF INFORMATION	DI=790618
PRODUCT AMOUNT	PA=0050
PRODUCT CODE	PC=991
TYPE OF OPPORTUNITY CODE	TC=204

TYPE OF OPPORTUNITY TO=0IRECT SALE

UPDATE UD=9999

FULL-TEXT FIELDS

ABSTRACT DESIGN PLANNING/AB

COMPANY PORT(W)SAID(F)HOUSING/CO

W przytoczonych przykładach pól wyszukiwawczych baz danych dokumentacyjnej i faktograficznej równocześnie złączono uwagę o możliwości zastosowania wyrażeń języka kwerend związanych z selekcjonowaniem informacji. Każdy z systemów on-line posiada wykaz zbiorczy symboli pól wyszukiwawczych, załączany w dokumentacji instruktażowej. Jego zawartość uzależniona jest od liczby i rodzajów baz danych. Znacznie bogatszy jest repertuar pól wyszukiwawczych systemu DIALOG, w którym udostępnia się kilkadziesiąt baz danych faktograficznych, głównie zawierających informację handlową i statystyczną, np. bazy zaliczane do grup BUSINESS COMPANIES, PREDICAST - MARKET RESEARCH, BUSINESS STATISTICS funkcjonujące komplementarnie z bazami danych bibliograficznych, BUSINESS BIBLIOGRAPHIC, LAW GOVERNMENT, PUBLIC AFFAIRS i in. Symbole stosowane w funkcji etykiet pól wyszukiwawczych, pełniące rolę wykładników kryteriów relewancji dla obiektów danej bazy mają często przyporządkowanych kilka znaczeń. Wieloznaczność ta jest automatycznie eliminowana przez wybór zbioru informacji do wyszukiwania. Podany przykładowo sposób prezentacji wiedzy o zasobach informacyjnych całego systemu oraz poszczególnych baz danych wraz ze strukturą rekordów bazy, a więc strukturą tekstów spójnych języka informacyjnego, może stanowić podstawę jednej z dwu metod wyszukiwania informacji, tzw. metody jadłospisowej (menu driven), zalecanej dla użytkowników początkujących w korzystaniu z systemów on-line. Posługując się podanym przez system wykazem pól wyszukiwawczych, nazywanych opcjami wyszukiwawczymi,

użytkownik dokonuje wyboru pól odpowiadających odpowiednim cechom formalnym lub treściowym obiektu i wypełnia je wyrażeniami, które powinny zawierać charakterystyki wyszukiwawcze. Formułowanie instrukcji wyszukiwawczej odbywa się wówczas metodą indeksowania ankietowego, szczególnie często wykorzystywanego w bibliotecznych katalogach on-line, obsługujących wielu użytkowników, na przykład studentów bibliotek akademickich we wspomnianym systemie SOCRATES oraz w National Library of Medicine (USA). Ten sposób prezentacji wiedzy o cechach relewantnych zbioru informacyjnego decyduje o tym, że system jest "przyjazny użytkownikowi" (user cordial, user friendly), a ponadto uwalnia użytkownika od obowiązku znajomości często uciążliwej, bo zróżnicowanej w poszczególnych zbiorach, interpunkcji. Przykład wyszukiwania w katalogu autorskim w National Library of Medicine:

Nazwisko	Imię	Imię
<input type="text" value="FREUD"/>	<input type="text" value="SIGMUND"/>	<input type="text"/>

Informacja systemu:

"..Proszę nacisnąć klawisz ENTER po wypełnieniu poszczególnych podpól

"..Proszę nacisnąć klawisz TAB w celu przeniesienia się do kolejnego podpola.."

"..Pytanie zostało przyjęte. Proszę czekać..."

W systemie SOCRATES "menu" katalogów sieci bibliotecznej zawiera wykaz następujących pól wyszukiwawczych:

A	AUTHOR, personal name	np. J.D. Salinger
O	ORGANIZATION, as author	Smithsonian
TP	TITLE, Initial phrase	Portrai of the Artist as
T	TITLE, any words	Portrait Artist
S	SUBJECT, topical words	Genetic Engineering
SN	SUBJECT, personal name	Martin Luther King

Po wyborze opcji system prosi użytkownika o podanie wyrazów relewantnych, tzw. wartości wyszukiwawczych (search values), na podstawie których selekcjonuje informację po wprowadzeniu komendy "FIND".

Zaletą wykorzystania cech strukturalnych zbioru informacji w wyszukiwaniu jądłospisowym jest łatwość w posługiwaniu się systemem. Oznacza to równocześnie znaczne ograniczenie zasad indywidualizacji komunikowania się z systemem oraz schematyzację w formułowaniu strategii wyszukiwawczej, której sprecyzowanie i umiejętność modyfikacji wymaga pogłębionej wiedzy o języku systemu.

WIEDZA O JĘZYKU

Selekcjonowanie informacji według sformułowanej przez użytkownika strategii wyszukiwawczej odbywa się po wskazaniu, za pomocą komendy oznaczającej rozpoczęcie wyszukiwania (np. BEGIN, CHANGE, BASE):

- jednej bazy danych: np. ?BEGIN lub B3 (nr bazy INSPEC w systemie DIALOG)

- kilku baz danych a) przez wskazanie ich numerów :?SELECT FILE

lub SF 2,3,4,8,32,34,94, gdzie

FILE2: Chemical Abstract Search 67-71

FILE3: " " " 72-76

FILE4: " " " 77-80

FILE8 COMPENDEX 70-80

FILE32 METADEX 66-80

FILE34 SCISEARCH 78-80

FILE94 SCISEARCH 74-77

* b) przez polecenie wyboru baz według kryteriów przedmiotowych

?SF ENERGY, ENVIRONMENT

Odpowiedź systemu:

FILE6 NTIS 64-80

FILE8 COMPENDEX 70-80/OCT

FILE12 INSPEC 69-77/OCT

FILE13 INSPEC 78-80/OCT

FILE28 OCEANIC ABS 64-80/OCT

FILE29 MET/GEOASTRO ABS 70-79/DEC

FILE40 ENVIROLINE 71-80/AUG

FILE41 POLLUTION ABS 70-80/SEP

FILE44 AQUATIC SCI ABS 78-80/JUL

FILE45 APTIC 66-78/OCT

FILE68 ENVIRONMENTAL BIBLIOGRAPHY 74-80/OCT

FILE69 ENERGYLINE 71-80/AUG

Podstawą realizacji procesu wyszukiwania jest - jak wspominaliśmy - tzw. "indeks podstawowy" (basic index) każdej z baz danych, który stanowi słownik wyrazów relewantnych w systemie, odwzorowujących cechy obiektów według omówionych wyżej opcji wyszukiwawczych. Do cech tych należą:

- cechy formalne dokumentów
- cechy treściowe dokumentów
- cechy obiektów pozajęzykowych.

Repertuar cech formalnych dokumentów jest bardzo obszerny i wykorzystywany w różnym zakresie w poszczególnych bazach. Uogólniając można stwierdzić, że nie występuje w nim jedynie cecha paginacji jako opcja wyszukiwawcza.

Cechy treściowe dokumentów w bazach danych bibliograficznych najczęściej odwzorowywane są w słowniku bazy za pomocą wyrazów należących do języka naturalnego oraz języka informacyjno-wyszukiwawczego. Na podstawie przytoczonych wykazów pól wyszukiwawczych sys-

temów DIALOG i SDC-ORBIT możemy wymienić następujące JIW, najczęściej stosowane w funkcji metainformacyjnej:

- język słów kluczowych,
- język haseł przedmiotowych,
- język deskryptorowy,
- specjalistyczna kategoryzacja,
- klasyfikacje patentowe (narodowe i międzynarodowa);

Należy przy tym zastrzec, że na skutek zasady nieingerowania przez dystrybutora informacji w metody charakterystyki treściowej dokumentów, obserwuje się dużą dowolność i nieprecyzyjność w terminologii stosowanej w charakterystykach struktury rekordów w systemie i zawartości ich pól wyszukiwawczych. Częstość zjawiskiem jest stosowanie terminów "deskryptor", "termin", "hasło", "fraza", "pojęcie" (concept). Termin "identyfikator" definiowany w tradycji europejskich służb informacyjnych jako deskryptor równokształtny z nazwą własną, w systemach amerykańskich oznacza słowo kluczowe ustalone w ramach swobodnego indeksowania dokumentów. Dokładna charakterystyka JIW w bazach danych on-line wymagałaby dostępu do szczegółowej dokumentacji instruktażowej każdej z nich i zapoznania się z ich kontrolowanymi słownikami, co nie jest możliwe. Słowniki nazywane tezaurusami są niekiedy kontrolowanymi wykazami słów kluczowych z zaznaczoną relacją odpowiedniości wyszukiwawczej. Dla użytkownika nie posiadającego doświadczenia w korzystaniu z systemów on-line myśląc może okazać się informacja o stosowaniu "kodów" w charakterystykach wyszukiwawczych. W językach o notacji paranaturalnej, także deskryptorowych, stosuje się często równoległą notację sztuczną, np. cyfrową DC-0508 (b.d.MANAGEMENT CONTENT), która pozwala oszczędzać czas i koszty wyszukiwania.

Kolejną interesującą cechą w systemach on-line związaną z zakresem wykorzystania tradycyjnych JIW jest minimalny udział klasyfikacji w charakteryzowaniu treści dokumentów. W niektórych amerykańskich bazach stosowana jest Klasyfikacja Dziesiętna Deweya (np. w Government Printing Office - Monthly Catalogue). Ponadto w bazach BIOSIS z zakresu nauk przyrodniczych/biologicznych BIOSIS PREVIEWS równoległe z językiem deskryptorowym wykorzystuje się język klasyfikacyjny - tzw. "biosystematic code". Międzynarodowa Klasyfikacja Patentowa oraz klasyfikacje narodowe, znajdują zastosowanie w tych bazach danych, które zawierają charakterystyki opisów patentowych. (INPADOC - o zasięgu międzynarodowym, amerykańskich - CLAIMS/CHEM 50-70, CLAIMS/US PAT 71-77, CLAIMS/US PAT 78-, CLAIMS/WEEKLY).

PREZENTACJA LEKSYKI JIW

Prezentacja cech obiektów pozadokumentacyjnych w faktograficznych bazach oparta jest na kontrolowanych słownikach nazw obiektów usystematyzowanych w ramach pragmatycznie interpretowanych kategorii, nazywanych niekiedy "zdarzeniami", (events), bądź parametrów ilościowych zwłaszcza w dominujących w zbiorach DIALOG-u bazach informacji rynkowej, przemysłowej czy też w zakresie finansowania i sprawozdawczości z prac naukowo-badawczych i rozwojowych. Słowniki te konstruowane przy równoległym zastosowaniu notacji paranaturalnej i sztucznej (product, name, event name, phase name - product code, event code, phase code) są udostępniane użytkownikom uprawnionym do korzystania z baz. Na przykład obiekty bazy "EIS INDUSTRIAL PLANTS" (Economic Information Systems - US industrial activity), charakteryzowane są za pomocą następujących cech:

BC = Branch city (siedziba filii)
BN = Branch name (nazwa filii)
BS = Branch state (stan, w którym znajduje się filia)
BZ = Branch zip code (skrótowy kod siedziby filii)
CC = Company code (kod firmy)
CN = Company number (numer filii)
DE = Descriptor (deskryptor)
EX = Employee size (wielkość zatrudnienia)
HC = Headquarters city (miasto, siedziba centrali)
HN = Headquarters name (nazwa centrali)
HS = Headquarters state (stan, w którym znajduje się centrala)
HX = Headquarters code (kod centrali)
HZ = Headquarters zip code (skrótowy kod centrali)
PC = Product code (kod produktu)
PN = Product name (nazwa produktu)
SD = Sales (in million \$)
SF = Special feature (specjalne cechy)
SM = Share of market (udział w rynku)

Tak zwane nie-bibliograficzne bazy danych łączą faktycznie informację o dokumentach z informacją o innych obiektach, co odzwierciedlone jest w zestawie kategorii, w ramach których umieszczane są nazwy (i kody) obiektów i ich cech. Na przykład w bazie PTS FEDERAL INDEX (Predicasts FI) przyjęto następującą kategoryzację obiektów:

- 00 Annotation and Status Codes (Federal Index) - (Adnotacja i kod statusu w Federal Index)
- 10 Organization and institutions (Organizacje i instytucje)
- 20 Management procedures (Procedury zarządzania)
- 30 Products and processes (Produkty i procesy)
- 40 Resources and resource use (Zasoby i ich wykorzystanie)

- 50 People (Ludzie)
- 60 Market information (Informacja rynkowa)
- 70 Unit costs and prices (Jednostki kosztów i cen)
- 80 Financial data (Dane finansowe)
- 90 Government and Society (Rząd i społeczeństwo)

Drugim źródłem wyrażeń indeksu bazy danych, oprócz słowników JIW, są te pola charakterystyki wyszukiwawczej dokumentu lub obiektu pozadokumentacyjnego, które określane są jako pola pełnotekstowe (full-text fields), zawierające informację w języku naturalnym, a więc: TYTUŁ, ABSTRAKT, ADNOTACJA, a także w niektórych bazach fakto-graficznych nazwy własne instytucji lub przedsiębiorstw. Selekcja wyrażeń relewantnych z tych pól dokonywana jest automatycznie przez zastosowanie tzw. stop listy, a więc drogą eliminacji tych wyrażeń, które nie mogą pełnić funkcji wyszukiwawczej, tradycyjnie uznawa-nych za "asemantyczne" - spójników, przedimków, poimków, przyimków. Są one umieszczane w indeksie bazy w postaci unitermów. Odróż-nienie wyrażeń JIW od unitermów (słów kluczowych) selekcyjonowanych automatycznie umożliwia sposób ich prezentacji w indeksie i niepo-dzielność deskryptorów równokształtnych z wielowyrzowymi wyrażenia-mi języka naturalnego, np.: /12/:

E1 NUCLEAR	17564
E2 NUCLEAR/DE	5333
E3 NUCLEAR/ID	7288
E4 NUCLEAR FISSION/DE	4792
E5 NUCLEAR POWER/DE	4298
E6 NUCLEAR REACTOR	6736

Zapoznanie się z indeksem bazy jako słownikiem wyrażeń rele-wantnych, charakteryzujących przedmioty treści dokumentów lub objek-ty pozadokumentacyjne umożliwia komenda: EXPAND lub E (DIALOG),

NEIGHBOR lub NBR (SDC), ROOT (BRS) lub DISPLAY (DIANE). System automatycznie przypisuje pozycjom ukazywanego na ekranie indeksu numery, które można wykorzystać w instrukcji wyszukiwawczej zamiast pełnych wyrażeń, umieszczając ponadto po ostatniej pozycji pytanie o kontynuację, np. w systemie DIALOG 2 - "more". Potwierdzenie przez naciśnięcie klawisza "carriage return" powoduje dalsze wyświetlanie indeksu bez potrzeby powtarzania komendy.

W niektórych bazach danych zasób leksykalny JIW udostępniany jest w postaci odrębnego podzbioru - słownika lub tezaury on-line (dialist). Informację o możliwości korzystania ze słownika JIW on-line podaje się w charakterystyce bazy danych, w ERIC, PSYCHOLOGICAL ABSTRACTS GEOARCHIVE, US POLITICAL SCIENCES ABS, BIOSIS, MESH (Medical Subject Headings) i in.

Szczególnie rozbudowaną informacją słownikową odznaczają się bazy informacji chemicznej (CA -Chemical Abstracts) liczące kilka milionów rekordów. Dla baz tych (CA, CA Condensates, CASIA, CA PATents) zbudowano słownik konkordacyjny, funkcjonujący jako niezależna baza danych CHEMNAME.

Nazwy substancji chemicznych są reprezentowane za pomocą wyrażeń języka naturalnego oraz symboli chemicznych, dzięki czemu możliwe jest wyszukiwanie informacji według cech strukturalnych związków oraz według kryteriów przedmiotowych (subject search). Częsta aktualizacja zbiorów informacji, rozbudowa terminologii chemicznej oraz złożoność zależności między substancjami (strukturalna i aplikacyjna) wpłynęły na szczegółowość specyfikacji relacji w słowniku; Odwzorowano w nim następujące cechy wyrażeń i ich desygnatów:

- nazwa w języku deskryptorowym (IDE), obejmującym nazwy substancji zawarte w "CA Index Names", np. ACID, AZULENE, BENZ, CARBOXYLIC, DERIV, DI, GIBBAN

- wykaz elementów składowych (element count) oparty na formule molekularnej, umożliwiający dostęp do informacji o wszystkich atomach formuły /EC/

- słowa kluczowe (ID) (identifiers) wyselekcjonowane z pola synonimy (SY)

- nazwa nadrzędna (Heading Parent HP=) w CA Index Names, np. HP=AZULEN

- podstawniki (Substituent) (SB=), np. SB=METHYLEN

- modyfikacje nazwy (Name Modification, NM=) w CA Index Names

- ?name match (CN=), umożliwiające wyszukanie dokumentów stanowiących pierwotne publikacje o danej substancji z pominięciem późniejszych pochodnych

- synonimy (SY=) nazw umieszczonych w CA Index Names, np. SY=GIBBERELLIN

- formuła molekularna (Molecular Formula, MF=), np. MF=C19H24O4, według systemu Hilla

- Stereochemical descriptor (ST=) odwzorowujący układ przestrzenny substancji,

- numer grupy (Group number) w schemacie klasyfikacyjnym CA, np. GN=A1

- termin indeksu periodycznego (Periodic index term, PI=).

Prezentacja leksyki języków w postaci indeksów poszczególnych baz danych i ewentualnie słowników (tezaurusów słowników języka haseł przedmiotowych) on-line przy częstym pokrywaniu się zakresów tematycznych zbiorów stwarza użytkownikom systemów trudności spowodowane faktycznym ograniczeniem opisu znaczenia jednostek leksykalnych do planu wyrażania. Stanowi to także problem dla służb informacyjnych, opracowujących zbiory informacji, lecz równocześnie współdziałających w zakresie wymiany informacji w ramach

sieci biblioteczno-informacyjnych. Specjaliści zajmujący się globalnymi problemami rozwoju systemów on-line coraz częściej piszą o swoistej "dżungli językowej" w bazach danych, z których korzystanie staje się coraz trudniejsze i coraz bardziej kosztowne. Jako propozycję rozwiązania problemów semantycznej porównywalności i przekładalności języków poszczególnych baz danych przedstawia się koncepcję "lingwistycznych (terminologicznych) banków danych wbudowanych w architekturę systemu jako samodzielny moduł (podsystem). Zakłada się, że moduł taki spełniałby pewne funkcje:

a) wewnętrzny, w tym:

- kontrola i korekta zapytań niepoprawnych leksykalnie lub ortograficznie,
- integracja językowa zbiorów systemu,
- pomoc terminologiczna w formułowaniu strategii wyszukiwawczej.

b) zewnętrzne (międzysystemowe), w tym:

- ułatwianie dostępu do baz danych w ramach lokalnych sieci teleinformatycznych,
- słuzenie jako podstawa terminologiczna w budowie JIW nowych baz danych i aktualizacji istniejących.

Koncepcję budowy zintegrowanego leksykonu baz danych (authority file) skupiającego i charakteryzującego semantycznie słownictwo języków deskryptorowych, języków słów kluczowych, języków haseł przedmiotowych zaczęła realizować w latach siedemdziesiątych Biblioteka Kongresu USA współpracująca z Government Printing Office USA, bibliotekami uniwersyteckimi (Northwestern, Texas, Wisconsin), National Library of Medicine dysponująca indeksem haseł przedmiotowych Medical Subject Headings (MESH) oraz National Agricultural Library /18/.

Innym przykładem prac nad scalaniem zasobów leksykalnych słowników JIW baz danych udostępnianych w systemach on-line jest Vocabulary Switching System (VSS), tworzony w Battelle Columbus Laboratories, USA. Początkowo zasoby jego obejmowały kilka słowników JIW, m.in.: tezaurus TEST, tezaurus NASA, tezaurus INSPEC, indeksy przedmiotowe, obecnie ich liczba sięga dwudziestu o łącznych zasobach około 500 tys. jednostek leksykalnych, tworzących terminologiczny bank danych. Zbiory tego leksykonu zgrupowano w czterech modułach tematyczno-dziedzinowych /4/:

1) nauk technicznych i ścisłych, zawierający wyrażenia indeksów przedmiotowych Chemical Abstracts, deskryptory Tezaurusu Departamentu Energetyki USA, Tezaurusu TEST, INSPEC, NASA oraz trójjęzycznego Tezaurusu Żelaza i Stali,

2) nauk biomedycznych, zawierający wyrażenia indeksu przedmiotowego bazy BIOSIS PREVIEWS, MESH oraz część indeksu CA,

3) nauk społecznych, zawierający m.in. deskryptory tezaurusów ERIC i Psychological Abstracts,

4) businessu, administracji i zarządzania, zawierający leksykę tezaurusów stosowanych w bazach "ABSTRACTS BUSINESS inc. INFORM. MANAGEMENT CONTENT oraz ECONOMIC ABSTRACTS INTERNATIONAL.

Zasoby leksykalne integrowanych słowników prezentowane są w układzie alfabetycznym (term file) z uwzględnieniem wykazu relacji paradygmatycznych każdego wyrażenia, a więc wykładników relacji generycznych, skojarzeniowych, odsyłaczy związanych z międzyjęzykową relacją ekwiwalencji wyszukiwawczej, typu: "użyj - użyj zamiast", "zob.-zob.też" oraz precyzujących znaczenie "scope notes".

Funkcję indeksów pomocniczych ułatwiających wyszukanie wyrazów (terminów) w zbiorach leksykonu pełnią słowniki inwersyjne opracowe-

wane przy wykorzystaniu operacji permutacji, przetaczania i obcinania:

- indeks unitermów (word file) grupujący wyrazy mogące pełnić samodzielnie funkcję wyszukiwawczą,

- indeks rdzeni wyrazowych (stem file), tworzony za pomocą algorytmów obcinania końcówek fleksyjnych wyrazów i części morfemów słowotwórczych w elementach składowych terminów,

- indeks rdzeni terminów (stem phrase file), powstały analogicznie jak indeks wyrazów (unitermów).

Wykazaniu podobieństwa znaczeniowego terminów należących do leksyki różnych JIW, a tym samym wskazaniu ich synonimii wyszukiwawczej, służy tzw. "zbiór pojęć" (concept file), w którym odwzorowane są także relacje paradygmatyczne, organizujące pola semantyczne poszczególnych języków.

Podobnie zintegrowane są leksykony, pełniące funkcję makrotezaurusów dla systemów specjalistycznych, np. system SCORPIO (Subject Oriented Retriever for Processing Information On-line) stosowany przy wyszukiwaniu informacji do obsługi informacyjnej członków Kongresu USA /9/. Ich powstanie ilustruje interesującą tendencję rozbudowy w systemach on-line wiedzy o strukturze paradygmatycznej JIW oraz o relacjach międzyjęzykowej ekwiwalencji wyszukiwawczej przy ograniczaniu zakresu wiedzy o rzeczywistości charakterystycznej dla języków systemów tradycyjnych.

REGUŁY GRAMATYKI

Charakterystyczną cechą języka wyszukiwawczego w systemach on-line jest - jak wspominaliśmy - wzbogacenie jego funkcji związanych z instrumentami przetwarzania informacji o funkcję impresywną, tradycyjnie "odmawianą" językom sztucznym. Realizacja procesu

selekcjonowania informacji w zbiorach systemu po:

- uzyskaniu meta-metainformacji o ich zasobach,
- wyborze bazy/baz danych,
- zapoznaniu się z indeksem wyrażeń relewantnych reprezentujących cechy formalne i/lub treściowe dokumentów bądź pozadokumentacyjnych obiektów,

odbywa się za pomocą komendy o znaczeniu "wyselekcjonuj informację o obiektach" według wskazanych cech. Komenda ta ma postać (najczęściej) SELECT (DIALOG, ESA-IRS), FIND (SDC ORBIT, DIANE), SEARCH (BRS). Z jej użyciem wiąże się specyfikacja reguł pozycyjnej gramatyki języka.

Wyrażenie wskazane przez użytkownika jako wykładnik relewancji musi zawsze wystąpić po pełnej lub skrótowej nazwie operacji.

Najprostsza formuła zdaniowa komendy SELECT (w syst. DIALOG) ma postać: :?SELECT lub ?S WYRAŻENIE jednowyrazowe - ?S CANCER

Odpowiedź: nr wyszukanego podzbioru zawierającego
wskazane wyrażenie, liczba zawierających je ChW, wyrażenie

3655 CANCER

Ten sposób formułowania instrukcji interpretowany jest przez system jako polecenie wyszukania poprzez indeks bazy ChW, które zawierają w dowolnym polu tekstowym wskazane wyrażenie. Specyfikacji cech obiektów systemu odwzorowanych w przewidzianym dla danej bazy zestawie pól wyszukiwawczych służą posiadające stałe miejsce w formule zdaniowej języka kwerend.

W systemie DIALOG etykieta (symbol) pola, które powinno zawierać wskazane wyrażenie umieszczana jest prepozycyjnie lub postpozycyjnie w stosunku do tego wyrażenia, zależnie od przyjętego ogólnego kryterium selekcji (formalnych cech dokumentu, cech obiektu pozadokumentacyjnego lub treściowych cech dokumentu).

Komenda sformułowana w celu wyselekcjonowania informacji o dokumentach według cech formalnych powinna odpowiadać wzorcowi:

?S symbol pola tzw. prefix ze znakiem równości, wykładnik relewancji

?S AU= JONES C^{5/}

2 \$22 AU=JONES C

lub (przykład opcji wyszukiwawczych dla pól oznaczanych w zdaniu prepozycyjnie z bazy LANGUAGE AND LANGUAGE BEHAVIOR ABSTRACTS):

?S AU=FAISS, KLAUS

?S CA=CHOMSKY (CA= Cited author)

?S CO=LECI-4 (CO= Coden, dla wydawnictw ciągłych)

?S DT=REVIEW (DT= Document type)

?S JO=LINGUISTICS (JO= Journal)

?S LA=GER (LA= Language)

?S PD=1975 (PD= Publication Date)

?S SC=110 (SC= Section code, symbol wykazu działów
lingwistyki)

?S UD=76 (UO= Update, aktualizacja)

Gdy liczba pozycji związanych z niedokładnie określoną cechą jest duża system przekazuje informację, np.:

?S AU=SAWYER

3 1 000 TERMS; RESPECIFY

W układzie prepozycyjnym wskazywane są również cechy obiektów pozadokumentacyjnych w systemach faktograficznych lub dokumentacyjno-faktograficznych, np. w bazie FOUNDATION GRANTS INDEX (USA):

?S AM=00020 (AM= Grant amount, \$ thousands)

?S AG=00020 . (AG= Greater than or equal)

^{5/} Ogromnym praktycznym problemem np. przy wyszukiwaniu ChWD według nazwisk autorów dokumentów, jest interpunkcja, np. kropka lub przecinek umieszczane po inicjałach imienia/imion. Obserwuje się tendencję do eliminacji interpunkcji i zastępowanie jej spacjami.

- ?S Z(F)REYNOLD/FN - /FN nazwa fundacji
- ?S ART(W)CULTURE(W)CAROLINA/RN - /RN - recipient name
- ?S GOVERNMENTAL/RT - /RT Recipient type, według przyjętej kategorii, np. SCHOOLS, HOSPITALS

2) LLBA

- ?S PHONOLOGY /AB
- ?S BOCHUM /CS - /CS Corporate Source, źródło korporatywne
- ?S GENERATIVE PHONOLOGY/DE
- ?S BOOK(W)REVIEW/ID - /ID - słowa kluczowe, amer. "identifier"
- ?S SYLLABE(W)STRUCTURE/TI - /TI - title

W wielu systemach pozycja etykiety pola wyszukiwawczego zależy od rodzaju cech relewantnych określanych w zdaniu instrukcji wyszukiwawczej nie jest rozróżniana. Dotyczy to np. systemów SDC ORBIT, BLAISE, BRS, w których symbol pola wyszukiwawczego jest zawsze umieszczany postpozycyjnie, przy zastosowaniu różnych znaków delimitacji, np.:

SDC ORBIT - baza INSPEC	BRS - baza INSPEC
FIND SEISMOTECTONIC /TI	SEARCH SEISMOTECTONIC.TI
" NUCLEAR POWER STATIONS/IT	-
" WEBER,C/AU	SEARCH (WEBER ADJ C).AU
	ADJ - operator tekstu
" GERMAN/LA	SEARCH GE.LG

W bazach BRS przy wyszukiwaniu pełnotekstowym, nie jest wymagana kwalifikacja pól wyszukiwawczych za pomocą etykiet pól dołączanych do wyrażeń jako sufiksy. W systemie DIALOG możliwość taką zapewnia wersja komendy SELECT - SUPERSELECT: (SS).

Stosowanie podstawowej komendy umożliwiającej selekcjonowanie informacji ze wskazaniem wyrażenia określającego kryterium relewancji dla użytkownika związane jest z istotnym ograniczeniem w spe-

cyfikacji cech obiektów bazy danych. Ograniczenie to polega na wyszukiwaniu informacji według jednej cechy relewantnej, na przykład według cech treści dokumentów lub ich własności formalnych bez specyfikacji relacji zachodzących między tymi cechami w tekstach ChW.

We wszystkich systemach on-line podstawowym środkiem językowym służącym budowie złożonych zdań instrukcji wyszukiwawczej i zawężania zakresu podzbiorów wyselekcjonowanych w wyniku działania komendy SELECT, są operatory algebry Boole'a wyrażające działania na zbiorach (iloczyn, suma, różnica i ewentualne uzupełnienie zbioru AND, OR, NOT. Wykładnikiem operacji na zbiorach jest (najczęściej) komenda COMBINE (w skrócie C), która może być stosowana łącznie z numerami wyszukanych podzbiorów lub z wyrażeniami JIW. łączna liczba zbiorów poddawanych operacjom logicznym nie może przekraczać 29. Na przykład:

#SCIENCE CITATION INDEX - DIALOG/:

1) ?S CANCER

1 3655 CANCER

?S CARCINOMA

2 2597 CARCINOMA

?S LUNG

3 2190

?C (1 OR 2) AND (3 OR 4)

5 422(1 OR 2) AND (3 OR 4)

2) ?S PAPER(W)MILL OR PAPERMILL

1 177 PAPER(W)MILL OR PAPERMILL

?S AU= SAWYER S

2 498 SAWYER S

?C1AND2

3 0 1AND2

?S AU=JONES C

4 422 AU=JONES C

?CIAND4

5 0 IAND4

?SS AU=SAWYER S AND (PAPERMILL/TI OR WATER/TI)

6 498 AU=SAWYER S

7 2755 PAPERMILL/TI

8 21505 WATER/TI

9 7 6AND (7 OR 7)

Po uzyskaniu odpowiedzi systemu (podzbioru 9) użytkownik zwykle zapoznaje się z charakterystykami wyszukiwawczymi, przekazując komendę ich wydruku na ekranie (całego zbioru lub wybranych pozycji) w wybranym formacie:

?T9/2/1-3, gdzie , T - wydruk na ekranie (wydruk na papierze wymaga komendy P - print, sformułowanej wg tych samych zasad), /2/ nr formatu^{6/}, 1-3 numery pozycji ChWD w wyszukanym podzbiorze.

Modyfikacji zakresu wyszukanego podzbioru, przez ograniczenie liczby ChWD wskazanych przez system, służy także komenda LIMIT (L); najczęściej przydatna do ograniczenia wyszukiwania informacji o dokumentach według ich cech formalnych (chronologicznych, językowych, lokalizacyjnych itp.), np.

^{6/} Formaty prezentacji informacji wyszukanej w systemie są zróżnicowane w poszczególnych bazach. W systemie DIALOG użytkownik dysponuje ośmioma opcjami : Format 1 - DIALOG - numer akcesyjny, Format 2 - pełny rekord (ChWD) z wyłączeniem abstraktu, Format 3 - Cytacja bibliograficzna, Format 4 - Tytuł dokumentu i abstrakt, Format 5 - Pełny tekst ChWD, Format 6 - Numer akcesyjny DIALOG oraz tytuł dokumentu, Format 7 - Cytacja bibliograficzna i Abstrakt, Format 8 - Tytuł i ChWD w języku informacyjnym. Sortowaniu informacji według cech innych niż przewidziano służy komenda SORT i instrukcja zbudowana według formuły ?SORT nr podzbioru, nr ChWD, kryterium selekcji, np.: SORT8/1-48/AU TI; SORT10/1-54/SD,DT; a następnie P/11/5/AM,D.

?S 5/00550133-0669809 - wskazanie podzbioru według numerów systemowych ChWD

?L 2/ART - ograniczenie wyszukiwania tylko do ChWD artykułów w Science Citation Index

?L21/NAR - ograniczenie wyszukiwania tylko do ChWD nie będących artykułami (SCI)

?L9/ENG - ograniczenie wyszukiwania tylko do publikacji angielskojęzycznych

?L10/FOR - ograniczenie wyszukiwania tylko do publikacji obcojęzycznych

?L5/AVAIL - ograniczenie wyszukiwania do dokumentów ogólnie dostępnych (lub niedostępnych)

Możliwości zawężania zakresu podzbiorów są zróżnicowane w poszczególnych bazach. Najbardziej uniwersalnym kryterium jest numer akcesyjny ChWD w systemie.

Funkcję ograniczania zakresu przeszukiwanego zbioru według cech chronologicznych dokumentu może spełniać także komenda SELECT wraz ze znakiem ":" (dwukropek), np.:

?S PY=1976:PY=1978

Symbol ten jest szczególnie przydatny w wyszukiwaniu w zbiorach, w których JIW są klasyfikacje w celu wskazania jako wykładników relewancji symboli należących do jednego łańcucha klasyfikacyjnego, np.

?S CC=64072:CC 078.

Sesja wyszukiwawcza w systemie przeznaczona na wyselekcjonowanie informacji na jedno zapytanie użytkownika polega zazwyczaj na realizacji strategii złożonej z instrukcji wyszukiwawczych o zmieniających wykładnikach relewancji zależnie od wyników etapowych wyszukiwania.

Przyczyną niepowodzeń, niezależnie od niewłaściwego doboru wyrażen w instrukcji wyszukiwawczej, bywa rozbieżność ortograficzna wyrażen instrukcji i zbioru informacyjnego systemu lub ich zmienne rdzenie. Źródłem poważnych trudności dla użytkowników europejskich są różnice w ortografii brytyjskiej i amerykańskiej, np. CENTER - CENTRE, THEATER - THEATRE, łączna lub rozdzielna pisownia wyrażen złożonych (PAPERMILL - PAPER MILL). Środkiem systemowo włączanym do reguł gramatyki języka wyszukiwawczego jest operacja maskowania (obcinania - ang. truncation) polegająca na neutralizacji co najmniej jednej cechy dystynktywnej wyrażenia składowego instrukcji wyszukiwawczej, najczęściej litery. Wykładnikiem tej operacji jest symbol "?". W systemie DIALOG możliwości wyszukania wyrażen maskowanych ograniczają się do jednostek indeksowych (indeksu podstawowego lub indeksów innych pól wyszukiwawczych). Jest stosowana w zdaniach rozpoczynających się komendą EXPANDSELECT lub SUPERSELECT, na przykład:

?E ON?LINE - neutralizacja znaku "-"

?E TRANSFORMER? - neutralizacja końcówki liczby mnogiej lub innych morfemów, dzięki czemu możliwe jest wyszukanie znacznej liczby wyrażen indeksu należących zarówno do tezauryusza, jak i słów kluczowych.

- 1) ?E TRANSFORME?
 - E1 TRANSFORMED
 - E2 TRANSFORMEDTO
 - E3 TRANSFORMEE
 - E4 TRANSFORMEER
 - E5 TRANSFORMEES
 - E6 TRANSFORMER
 - E7 TRANSFORMER BUBLING

E8 TRANSFORMER CAPACITANCE

MODELS

E9 TRANSFORMER CHARGING

E10 TRANSFORMER COLIS

E11 TRANSFORMER CORE

E12 TRANSFORMER FEEDBACK

E13 TRANSFORMER INDUCTANCE MODELS

E14 TRANSFORMER LIFE

E15 TRANSFORMER LOADING POLICIES

E16 TRANSFORMER OIL

E17 TRANSFORMER SHEET

E18 TRANSFORMER SHEETS

E19 TRANSFORMER STEEL

E20 TRANSFORMER SUBSTANTIIONS

E21 TRANSFORMER WINDINGS

E22 TRANSFORMER-RECTIFIERS

E23 TRANSFORMER-TERMINATED LINES

E24 TRANSFORMER 'S

E25 TRANSFORMERETTES

E26 TRANSFORMERLESS

E27 TRANSFORMERS

E28 TRANSFORMERS,PULSE

E29 TRANSFORMERSAND

E30 TRANSFORMES

E31 TRANSFORMESS

E32 TRANSFORMHOLOGRAMS

E33 TRANSFORMIERENDE

2) ?S AU=SAWYER?

1 498 AU=SAWYER

- 3) ?S JN=SCIENTIFIC AMER?
 2 615 JN=SCIENTIFIC AMER?
- 4) ?S PY=197?
 3 40252389 PY=197?
- 5) ?S AGGRESS?/TI,DE
 4 780 AGGRESS?/TI,DE

Liczba neutralizowanych znaków jest niespecyfikowana. System wyszukuje wszelkie wyrażenia złożone ze znaków następujących po symbolu operacji.

W rozwiniętej wersji operacja maskowania może być realizowana następującymi sposobami:

a) ze specyfikacją maksymalnej liczby znaków neutralizowanych: na końcu wyrażenia:

?S SELECT HORSE??
 ?S THEAT???

w wyniku czego system wyszukuje wyrażenia o określonej długości;

b) przez wbudowanie symbolu maskowania między znaki składowe wyrażenia (embedded):

?S WOM?N
 ?S ORGANI?ATION/TI,DE

c) przez kombinację sposobu a) oraz b):

?S WORKM?n?

OPERATORY TEKSTOWE

Rozszerzenie opcji wyszukiwawczych w systemie on-line do w zasadzie tych wszystkich cech dystynktywnych obiektów, które są wyróżnione jako pola wyszukiwawcze reprezentowane przez wyrażenia odpowiednich indeksów bazy danych, powoduje - jak wskazywaliśmy - rozbudowę wiedzy o planie wyrażania języków informacyjnych o notacji

paranaturalnej dzięki transformacjom wyrazów złożonych. Wiedza o rzeczywistości pozajęzykowej jest ograniczana do tych wykładników, które są tradycyjnie stosowane w słownikach języków deskrypcyjnych (kategorii semantycznych, wykładników relacji mereologicznych i skojarzeniowych). Na poziomie syntagmatycznym nie stosuje się w praktyce wskaźników roli jako wykładników sytuacyjnych związków między desygnatami wyrazów. Możemy stwierdzić, że obowiązek znajomości relacji między wyrazami języka naturalnego i JIW pełniących funkcję kluczy wyszukiwawczych i obiektami rzeczywistości pozajęzykowej zostaje przeniesiony na użytkownika systemu (pracownika służb informacyjnych lub bezpośrednio zainteresowanego specjalistę). W porównaniu z systemami tradycyjnymi zmienia się także sposób, w jaki system pomaga użytkownikowi wskazać związki zachodzące w tekstach ChW między wyrazami składowymi. Sposób ten jest pochodną interakcyjnego trybu wyszukiwania i możliwości zmiany operacji dokonywanych na zbiorach bazy danych. W literaturze przedmiotu często jest określany nieprecyzyjnym terminem **wyszukiwanie pełnotekstowe** (full-text searching).

Wyszukiwaniem tekstowym nazywa się różne operacje selekcyjonowania informacji w systemie on-line, najczęściej interpretujące jąko:

a) wyszukiwanie w systemie, którego zbiór tworzą teksty dokumentów pierwotnych, a więc wyszukiwanie odbywa się w "pełnych tekstach" dokumentów. Do systemów takich funkcjonujących w skali pełnej eksploatacji należy komercyjny system informacji legislacyjnej LEXIS w USA /8/,

b) wyszukiwanie za pomocą operatorów tekstowych w pełnym tekście, na przykład w streszczeniu, charakterystyce wyszukiwawczej dokumentu,

c) wyszukiwanie w dowolnych wskazanych ciągach wyrażen charakterystyki wyszukiwanej dokumentu.

W omawianych przykładowo systemach on-line dostępnych w skali międzynarodowej nie stosuje się wyszukiwania pełnotekstowego w znaczeniu pierwszym (a), ponieważ zbiory informacyjne tych systemów tworzą bądź ChWD bądź ChW obiektów pozajęzykowych (zbiory faktograficzne). Jako cechę dystynktywną wyszukiwania pełnotekstowego w znaczeniu (b) i (c) podaje się możliwość stosowania operatorów tekstowych, interpretowanych jako wyrażenia języka wyszukiwanego wyspecjalizowane w funkcji metatekstowej do określania warunków, jakie powinny spełniać wyrażenia wskazane w instrukcji wyszukiwanej, aby mogły być uznane za relewantne. Do warunków tych należą:

- odległość wyrażen (pojedynczych wyrazów) w tekście, określana za pomocą tzw. "operatorów sąsiedztwa/bliskości" (proximity operators),

- rodzaj kontekstu, w którym powinno wystąpić wyrażenie w charakterystyce wyszukiwanej dokumentu.

Możliwość łącznego stosowania wykładników operacji maskowania ze wskaźnikami położenia wyrażenia w tekście jest różna w poszczególnych systemach. Należy podkreślić, że w odróżnieniu od wykładników relacji semantycznych w zdaniu JIW w systemach tradycyjnych operatory tekstowe systemów on-line są jedynie wykładnikami relewancji technicznej, a więc określającej zgodność wyrażen ChWD z wyrażeniami instrukcji wyszukiwanej w planie wyrażania.

Do wskazania kryteriów relewancji przez specyfikację odległości wyrażen w tekście służą najczęściej symbole interpretowane jako "przyległy, sąsiadujący" (adjacent):

(W) - zawsze umieszczany w nawiasie, pełniącym funkcję delimitacyjną (stosowany np. w SYSTEMIE DIALOG,ESA-IRS),

(ADJ) - również umieszczany w nawiasie, stosowany w systemie BRS.

(NBR) - w systemie SDO ORBIT, BLAISE

Umieszczenie tego operatora między dwoma wyrażeniami oznacza, że wyrażenie stojące po nim powinno następować bezpośrednio po wyrażeniu stojącym przed nim, np.:

?S SOLAR(W)ENERGY = SOLAR ENERGY

?S DEEP(ADJ)SPEEN = DEEP SPEEN

?S NUCLEAR(ADJ)POWER(ADJ) (PLANTS' OR STATIONS)

W celu określenia dopuszczalnej liczby wyrażeń dzielących w tekście ChWD wyrażenia instrukcji można rozbudować operator przez dodanie przed nim wskaźnika łączbowego "n", np. (1W),(3W). Precyzowanie liczby wyrażeń rozdzielających wyrażenia instrukcji jest istotne ze względu na możliwość wystąpienia między wyrażeniami relewantnymi wyrażen zakazanych tzw. "stop words", automatycznie pomijanych przez system przy tworzeniu indeksów pól tekstowych na podstawie wspomnianej "stop listy". Na przykład w celu wyszukania tytułu powieści "Przeminęło z wiatrem" (Gone with the wind) w języku angielskim konieczne jest użycie operatora w postaci: ?S GONE(2W)WIND.

Określenie rodzaju kontekstu, w którym powinno wystąpić wyrażenie bez względu na jego odległość od innych wyrażeń i ich kolejność umożliwiając operatory wskazujące jednostki strukturalne tekstu spójnego, a więc rekordu (ChWD), pola (field), podpola, (subfield), cytacji (citation), a w niej dowolnego pola/podpola. Stanowią one faktycznie ukryte wskaźniki wagi, wykorzystujące cechę współwystępowania wyrażeń w kontekście jako wykładnik związku semantycznego. Rozbudowany system operatorów kontekstowych opraco-

wano dla systemu DIALOG, stosowany jest łącznie z operatorami maskowania oraz operatorami logicznymi również umieszczanymi w nawiasach:

- (F) - wyrażenia powinny wystąpić w tym samym polu: SOLAR(F)ENERGY
- (C) - wyrażenia powinny wystąpić w tej samej cytacji, w dowolnym polu: SOLAR(F)ENERGY
- (L) - stosowany tylko w tych bazach, w którym JIW jest język (link) deskryptorowy, umożliwiający wyszukiwanie hierarchiczne, a więc tych ChWD, które zawierają w polu opisu deskryptorowego wyrażenia pozostające w relacji paradygmatycznej z wyrażeniami (deskryptorami) instrukcji wyszukiwawczej SOLAR(L)ENERGY ; ANTENNA?(L)TELEVISION
- (S) - stosowany w tych bazach, w których w strukturze rekordu wyróżnia się podpola wyszukiwawcze; wskazane wyrażenia instrukcji powinny wystąpić w podpolach w dowolnej kolejności SOLAR(S)ENERGY

Użycie operatorów kontekstowych jest niekiedy funkcjonalnie równoważne użyciu komendy SUPERSELECT, oznaczającej polecenie wyszukania ChWD według dowolnych pól, w połączeniu z operatorami logicznymi, np. ?DIALOG,Guide to searching):

?S ARTHRITIS(C)ACUPUNCTURE

1 41 ARTHRITIS(C)ACUPUNCUTRE

?SS ARTHRITIS AND ACUPUNCTURE

19555 ARTHRITIS

875 ACUPUNCTURE

2 41 ARTHRITIS AND ACUPUNCTURE

Zastosowanie operatorów odległości i kontekstowych nie jest ograniczone do pól tekstowych, jakkolwiek jest w nich najczęstsze. Okazuje się pomocne w wyszukiwaniu według cech formalnych dokumen-

tów w indeksach cytowań, na przykład:

?S CS=JOHNS?(C)CS=HOPKINS - CS=Cited source

?S CA=JOHNS?(C)CS=HOPKINS

Kombinacje przedstawionych operatorów dają różne możliwości precyzowania instrukcji wyszukiwawczej, które przytaczamy w porządku wzrastającej precyzji /12/:

?S ANTENNA?(C)MOBILE

?S ANTENNA?(C)LA=FRENCH

?S ANTENNA?(F)MOBILE

?S ANTENNA?(F)CS=WESTERN

?S ANTENNA?(F)MOBILE/OE

?S ANTENNA?(F)MOBILE/OE,/TI

?S MOBILE(nW)ANTENNA?

?S CS=LIBRARY(nW)CS=CONGRESS

?S MOBILE(W)ANTENNA?

?S CS=NEW(W)CS=JERSEY

?S MOBILE(W)ANTENNA?(F)MICROWAVE

?S CS=NATIONAL(W)CS=LIBRARY(F)CS=MEDICINE

?S ANTENNA(L)TELEVISION.

W literaturze przedmiotu od lat prezentowane są różne oceny wyników wyszukiwania pełnotekstowego oraz kontrolowanego w JIW ograniczonego do opisów deskryptorowych lub haseł przedmiotowych. Ocena efektywności różnych metod formułowania strategii stanowi bardzo obszerne zagadnienie wykraczające poza ramy naszych rozważań, wymagałaby ponadto dysponowania dostępem do któregoś ze wspomnianych systemów przez dłuższy czas. Opinie użytkowników profesjonalnych i nieprofesjonalnych prezentowane na łamach czasopism wyspecjalizowanych w problematyce systemów on-line (Database, 'On-line Review') formułowane są zwykle na podstawie doświadczeń, związanych z określoną

bazą danych lub grupą baz pokrewnych tematycznie, zebranych przy obsłudze określonej grupy użytkowników. Te spory o relewancję w systemie on-line kończą się zwykle wnioskiem o uzależnieniu sposobu formułowania strategii od potrzeb użytkowników, celu wyszukiwania, środków finansowych oraz kosztów wyszukiwania w poszczególnych bazach. Są de facto sporami o ocenę relewancji systemowej i pertynencji, przy różnych strategiach wyszukiwawczych. Z eksperymentalnych badań (case studies) nad wynikami wyszukiwania informacji w zakresie ochrony środowiska w USA (US ENVIRONMENTAL PROTECTION AGENCY, OHIO) wynika, że wbrew potocznym opiniom o efektywności wyszukiwania pełnotekstowego, lepszą trafność osiąga się posługując się kontrolowanymi słownikami JIW i instrukcjami wyszukiwawczymi zawierającymi od dwu do czterech wyrażen, reprezentujących terminy nadrzędne w systemie paradygmatycznym. Wyszukiwanie pełnotekstowe oceniano jako bardziej efektywne przy szerokich profilach tematycznych określanych w celu zebrania np. kompletnej bibliografii /3/.

WIEDZA O RZECZYWISTOŚCI W SYSTEMIE ON-LINE

Możliwość dokonywania transformacji tekstów charakterystyk wyszukiwawczych dokumentów w planie wyrażania i stosowania różnych kluczy wyszukiwawczych w systemie on-line dzięki oprogramowaniu powoduje - jak wskazywaliśmy - ograniczanie wiedzy językowej przy równoczesnej rozbudowie wiedzy meta-metainformacyjnej (informacji o bazach danych) oraz wiedzy o języku komunikowania się z systemem. Charakterystyczne jest ograniczenie wykorzystania w funkcji metainformacyjnej i wyszukiwawczej tradycyjnych klasyfikacji bibliotecznych. W nielicznych bazach danych katalogowych bibliotek amerykańskich stosowana jest Klasyfikacja Dziesiąta

Deweya. Wyjątek stanowią klasyfikacje patentowe pełniące de facto funkcje kodów semantycznych w odniesieniu do nowo powstałych obiektów rzeczywistości pozajęzykowej, a także klasyfikacje (systematyczne) obiektów w naukach przyrodniczych, np. BIOSIS PREVIEWS.

Spośród sposobów odwzorowania wiedzy pozajęzykowej należy wymienić przede wszystkim te, które są właściwe językom deskryptorowym na ich osi paradygmatycznej:

- kategoryzację obiektów pozajęzykowych, zróżnicowaną w tezę i zarysach poszczególnych baz danych,

- wykładniki relacji mereologicznych oraz skojarzeniowych.

Kategoryzacja obiektów jako zabieg organizujący pole semantyczne języka jest często stosowana w faktograficznych bazach danych, zwłaszcza zawierających informacje handlową, polityczną, których słowniki obejmują nazwy tysięcy obiektów (np. PREDICATS). Przyporządkowanie obiektom odpowiednich atrybutów i spełnienie warunku rozłączności ich zbiorów wymaga narzucenia JIW pewnego "szkieletu organizacyjnego" w postaci kategoryzacji. Potwierdza się przy tej okazji uniwersalność języków deskryptorowych jako języków systemów dokumentacyjnych i faktograficznych. Należy przy tym zauważyć umowność podziału systemów (baz danych) na faktograficzne i dokumentacyjne, wynikającą m.in. z tego, że zarówno charakterystyki wyszukiwawcze obiektów w bazach faktograficznych, jak i ich słowniki budowane są na podstawie analizy dokumentów: prasy codziennej (NEWSSEARCH)-DIALOG, wydawnictw statystycznych, biuletynów agend rządowych i międzynarodowych (PREDICATS FEDERAL INDEX - PTS, PTS FEDERAL INDEX WEEKLY, PTS Funk and Scott Indexes, PTS INTERNATIONAL STATISTICAL ABSTRACTS, PTS PROMT, PTS US ANNUAL TIME SERIES, PTS US STATISTICAL ABSTRACTS, PTS WEEKLY i in.).

PROBLEMY KOMUNIKACJI UŻYTKOWNIKA Z SYSTEMEM

Budowa sieci teleinformatycznych umożliwiającą szybkie przesyłanie informacji między systemem i użytkownikiem oraz opracowanie interakcyjnych programów wyszukiwawczych dokonujących transformacji zbiorów informacji w sposób rozszerzający repertuar kluczy wyszukiwawczych w systemie, a więc zbiór wykładników relewancji, wpłynęła na zmianę proporcji w zakresie wiedzy wbudowanej w JIW, powiększając zakres wiedzy o języku. Oznacza to równocześnie, że użytkownik korzystając z systemu rozstrzyga o relewancji wyselekcjonowanej informacji na podstawie własnej wiedzy specjalistycznej. Sposób reprezentacji wiedzy w systemie, głównie w postaci zbiorów terminów indeksów uporządkowanych alfabetycznie - w odróżnieniu od kompendium wiedzy, jakie stanowią tradycyjne klasyfikacje typu UKD, stanowi źródło zarzutów kierowanych pod adresem systemów on-line, iż ich zbiory stanowią zbiory nie powiązanych logicznie elementów (rekordów), które nie odzwierciedlają sposobu myślenia człowieka i nie są przystosowane do zindywidualizowanego z nich korzystania, stanowią "zbiory informacji lecz nie są zbiorami wiedzy" /11/.

Jako drugą zasadniczą wadę systemów on-line funkcjonujących w latach siedemdziesiątych wymienia się sztywność języka wyszukiwawczego (języka kwerend - query language) przystosowanego do protokołów konwersacji w poszczególnych systemach i dającego de facto niewielkie możliwości indywidualnego określania potrzeb informacyjnych w procesie konwersacji z systemem, a więc szczegółowej specyfikacji kryteriów relewancji. Wiąże się to z postulatem budowy odrębnych modułów komunikacji (interface), stanowiących element podsystemu lingwistycznego wraz ze zintegrowanymi leksykonami JIW (Vocabulary switching systems). Prace w tym zakresie prowadzone w MIT oraz jednostkach badawczych systemów komercyjnych

(Lockheed) - np. nad językiem CITE (Current Information Transfer in English), IIDA (Individualized Instruction for Data Access) dla systemu DIALOG /6/ Jako pożądane cechy języka komunikacji z systemem i odpowiedniego oprogramowania wymienia się:

- możliwość wykorzystania dokumentów w języku naturalnym do automatycznej budowy baz danych faktograficznych,

-- możliwość formułowania instrukcji wyszukiwawczych w języku zbliżonym do naturalnego i "rozumienia" tych instrukcji przez system, a także generowania odpowiedzi.

Budowa tzw. psychologicznych baz danych z dostępem w języku naturalnym przetwarzających "wiedzę" przeciwstawianą informacji stanowi odrębne zagadnienie, posiadające obszerną literaturę. Z ich powstaniem wiąże się nowa interpretacja relewancji uwzględniająca dynamizm procesów komunikacji oraz indywidualizację reprezentacji wiedzy.

Literatura

1. AJDUKIEWICZ K. Zdania pytańne. /W:/ Język i poznanie. Warszawa: Państwowe Wydawnictwo Naukowe 1985, s. 278-286
2. BARR A., FEGENBAUM E.A. The handbook of artificial intelligence. vol.1 Heuris Tech Press 1981, vol.2 Heuris Tech Press 1982
3. CALKINS M.L. Free text or controlled vocabulary? A case history step by step analysis, plus other aspects of search strategy. Database, June 1980 s.53-67
4. COLOMBO D.S., NIEHOFF R.T. Improved access to scientific and technical information through automated vocabulary switching. Final report Battele Columbus Labs. 1977
5. COMPUTER readable databases. A directory and data sourcebook. vol.1 Science, Technology, Medicine. vol.2 Business, Law, Humanities, Social Sciences. Ed. M.Williams, North Holland, Amsterdam: New York, Oxford 1985

6. GOLDSTEIN Ch. M. FORD W. The user-cordia³ interface. On-line Review 1978 vol.2 nr 3, s.270-278
7. GRABOWSKA M. Amerykański zautomatyzowany system informacji prawniczej oraz informacji prasowej - LEXIS/NEXIS. Zagadnienia Informacji Naukowej 1986 nr 2(49) s.89-102
8. GRABOWSKA M. Zautomatyzowane katalogi centralne w bibliotekach USA. Zagadnienia Informacji Naukowej 1987 nr 2(51) s.103-120
9. GREGORY N. The US Congress on-line users as policy makers. On-line Review 1979 vol.3 nr 4
10. GUIDE to DIALOG searching. Palo Alto, Lockheed Information Systems. 1979 vol. I-III
11. HILLMAN D.J. Model for the on-line management of knowledge transfer On-line Review 1977 vol.1 nr 1 s.23-30
12. HOOVER R.E. The library and information manager's guide to on-line services. New York: Knowledge Industry Publication Inc. 1980
13. MARCISZEWSKI W. Metody analizy tekstu naukowego. Warszawa: Państwowe Wydawnictwo Naukowe 1977
14. A REFERENCE guide to SOCRATES. The Online Library Catalog of Stanford University. 1985
15. SOSIŃSKA B. Międzynarodowe Centrum Koordynacyjne Sieci Systemów Informacji Bibliotecznej - nowa funkcja OCLC. Zagadnienia Informacji Naukowej 1983 nr 2(43) s.99-114
16. SOSIŃSKA B. Relacje między planem treści i planem wyrażania w języku informacyjno-wyszukiwawczym. Praca doktorska wykonana pod kierunkiem doc.dr hab. Bożenny Bojar. Warszawa, Uniwersytet Warszawski 1987 Wydział Neofilologii
17. SOSIŃSKA B. Reprezentacja wiedzy w systemach informacji dokumentacyjnej. Zagadnienia Informacji Naukowej 1985 nr 1(46) s.19-36
18. VENEZIANO V. Library automation: data processing and processing for data. Annual Review of Information Science and Technology. 1980 vol. 15 s.110-145
19. WERESZCZYŃSKA-CISŁO B., OGÓRKIEWICZ W. Pytania informacyjne w ujęciu teoretycznym. Zagadnienia Informacji Naukowej 1986 nr 1(48) s.41-60

5.02.1990

ACCESS TO THE KNOWLEDGE IN THE ON-LINE INFORMATION
RETRIEVAL SYSTEM

Summary

The article is devoted to the presentation of the methods of the access to the knowledge in the on-line information retrieval systems. Two types of methods of making the knowledge of the system available are distinguished: retrieval of information without its semantic transformation and with its transformation. The recent is done in expert (intelligent systems).

The principles of functioning of on-line information systems are discussed as well as the relevance criterions in these systems together with the ways of their presentation.

The scope of the knowledge about the information retrieval language is characterized with the special regard to the influence of the information tools (computer and its software) on the language. Some problems of communication of the user with the system are shown.

ДОСТУП К ЗНАНИЯМ В ДИАЛОГОВОЙ ИНФОРМАЦИОННО - ПОИСКОВОЙ
СИСТЕМЕ (ON-LINE)

Резюме

Статья посвящена представлению методов доступа к знаниям в диалоговых информационно-поисковых системах (on-line). Выделены два вида методов сделания доступными знаний системы: поиск без семантической трансформации информации и поиск с её трансформацией. Последняя проводится в системах - эксперт (интеллектуальных системах). Обсуждаются принципы функционирования диалоговых систем, а также

критерия релевантности в этих системах вместе со способами их представления.

Охарактеризован объём знаний об информационно-поисковой системе, особенно учитывая влияние информационных инструментов (эти с её программным обеспечением) на эту систему. Указаны некоторые проблемы коммуникации потребителя с системой.

EWA CHMIELEWSKA-GORCZYCA

Instytut Bibliotekoznawstwa
i Informacji Naukowej UW

**INSTRUKCJA INDEKSOWANIA
DO "TEZAUROSA INFORMACJI NAUKOWEJ"**

Zakres i struktura instrukcji. Metodyka indeksowania (postać charakterystyki wyszukiwawczej, szczególność indeksowania, koordynacja deskryptorów, wieloznaczność charakterystyk wyszukiwawczych, deskryptory dziedzinowe, proces indeksowania). Zalecenia ogólne (deskryptory geograficzne, modyfikatory). Zalecenia szczególne. Forma instrukcji.

Niezbędnymi elementami strukturalnymi każdego języka są słownik i gramatyka. W językach informacyjno-wyszukiwawczych (JIW) słownik reprezentowany jest przez wykaz elementarnych jednostek leksykalnych (EJL) w postaci tablic klasyfikacyjnych, słowników tematów i określników, tezaursów itp. wraz z zaznaczeniem relacji zachodzących między poszczególnymi EJL (relacji paradygmatycznych). Natomiast reguły gramatyczne przedstawiane są najczęściej w postaci instrukcji indeksowania, określającej postać wyrażen złożonych (charakterystyk wyszukiwawczych dokumentów - ChWD) oraz zależności między jednostkami leksykalnymi w tekście (relacje syntagmatyczne).

"Zagadnienia Informacji Naukowej" 1990 nr 1(56)

Informacja o relacjach między jednostkami leksykalnymi JIW zawarta w słowniku może się różnić w zależności od typu JIW, a w obrębie jednego typu w zależności od danego opracowania (wystarczy porównać proste tezauryusy alfabetyczne oraz rozbudowane tezauryusy alfabetyczno-fasetowe z bogatą siecią powiązań między deskryptorami oraz różnego rodzaju wykazami pomocniczymi). Jeszcze większe różnice zauważyć można w istniejących JIW w zakresie stopnia szczegółowości i formalizacji opracowania reguł gramatycznych. Istnieje szereg języków, których twórcy ograniczyli się do podania tylko wykazu EJL (z ewentualnym zaznaczeniem relacji paradygmatycznych zachodzących między nimi), a z całkowitym pominięciem ich gramatyki. W innych JIW wspomniane są tylko ogólne zasady gramatyczne (najczęściej nie w postaci oddzielnego opracowania, a jako fragmenty wstępu lub przedmowy do słownika). W nielicznych tylko istniejących JIW reguły gramatyczne potraktowane są na równi z leksyką języka i przedstawione explicite w sposób sformalizowany i szczegółowy (np. UKD i BSO^{1/}). Zauważyć można, iż wyczerpujące przedstawienie reguł gramatycznych charakterystyczne jest dla języków o słownictwie sztucznym (notacji sztucznej). W JIW czerpiących słownictwo z języka naturalnego (o notacji paranaturalnej) autorzy jakby zapominają, że tworząc język sztuczny, który tylko pozornie przypomina język naturalny (odwołuje się do reguł języka naturalnego) w zasadzie wprowadzają własne reguły semantyczne i syntaktyczne. Reguły te powinny być przedstawione w sposób w miarę możliwości jednoznaczny, szczegółowy i sformalizowany w celu zapewnienia spójnego (konsekwentnego) indeksowania oraz wyeliminowania (lub przynajmniej zmniejszenia) przy wyszukiwaniu ciszy i szumu informacyjnego.

^{1/}BSO - Broad System of Ordering

Przy tworzeniu systemu Wyszukiwawczo-Informacyjnego Systemu Iezaurusowego (WIST)^{2/} postanowiono rozpocząć opracowywanie instrukcji indeksowania w trakcie procesu tworzenia słownika - tezaurusa (tj. budować reguły syntagmatyczne jednocześnie z regułami paradygmatycznymi języka) i następnie przedstawić ją w formie oddzielnego opracowania, modyfikowanego, rozbudowywanego i ulepszanego w miarę rozwoju systemu oraz poszerzania jego zakresu i zasięgów.

ZAKRES I STRUKTURA INSTRUKCJI

Instrukcja zawiera reguły indeksowania przy wykorzystaniu słownictwa zgromadzonego w "Tezaurusie Informacji Naukowej" (IIN)^{3/}, tworząc wraz z nim język deskryptorowy dla powstającego systemu WIST. System ten indeksuje artykuły z czasopism polskich z zakresu szeroko rozumianej dziedziny informacji naukowej i pod takim kątem została opracowana instrukcja. Poszerzenie zakresu lub zasięgów systemu, np. o inne typy dokumentów (prace magisterskie, wydawnictwa zwarte, itp.) lub o artykuły także z czasopism obcych wymagać będzie rozbudowy i odpowiedniej modyfikacji instrukcji.

Instrukcja zawiera podstawowe uwagi dotyczące metodyki indeksowania, zalecenia ogólne, odnoszące się do całej dziedziny (obowiązujące we wszystkich działach, dla wszystkich deskryptorów IIN), oraz zalecenia szczegółowe, których zakres stosowania jest ograniczony (dotyczące niektórych tylko działów IIN) i zgodnie z tym schematem została podzielona. Rozdział pierwszy stanowi

^{2/} Zob. D. Ohnsorge: Projekt techniczny systemu informacyjno-wyszukiwawczego z zakresu informacji naukowej. ZIN 1989 nr 2(55) s.119-154

^{3/} Zob. też E. Chmielewska-Gorczyca: Tezaurus Informacji Naukowej. ZIN 1989 nr 2(55) s.51-96

metodyka indeksowania w systemie, rozdział drugi - zalecenia ogólne, rozdział trzeci - zalecenia szczegółowe, podzielone na szereg podrozdziałów uzupełnianych sukcesywnie w miarę prac przy tworzeniu zbioru wyszukiwawczego i przyjmowaniu przy tych pracach nowych ustaleń (reguł). Rozdział czwarty zawiera uwagi na temat formy instrukcji i możliwych jej realizacji. Ostatni rozdział omawiać będzie warsztat pracy indeksatora, możliwości opracowania własnych narzędzi pomocniczych przydatnych przy indeksowaniu, np. kartoteki identyfikatorów. Do instrukcji dodano kilka aneksów, np. artykuły podawane jako przykłady w rozdziale dotyczącym metodyki indeksowania, propozycje wykazu identyfikatorów, wykaz słowników, encyklopedii i innych wydawnictw informacyjnych przydatnych przy indeksowaniu oraz przykłady gotowych opisów (wzorcowy zbiór wyszukiwawczy).

Przyjęta struktura instrukcji (szczególnie podział na zalecenia ogólne i szczegółowe, a w ich obrębie na działy i deskryptory znajdujące się w tych działach wymagające oddzielnego omówienia) pozwala stosunkowo szybko odnaleźć odpowiednie instrukcje w ich licznych zbiorze przy konkretnym zapotrzebowaniu. Z zaleceniami ogólnymi powinien zapoznać się każdy indeksator, ze szczegółowymi tylko ten, który indeksuje artykuły danego działu. Skorzystanie z zaleceń innego działu jest zawsze możliwe, a łatwe dotarcie do odpowiednich zaleceń gwarantowane jest poprzez układ tych zaleceń zgodny w pewnym stopniu z układem deskryptorów w tezaursie. Idealnym rozwiązaniem byłoby umieszczenie każdej szczegółowej instrukcji w samym tezaursie przy deskryptorze, którego może dotyczyć (przy niektórych deskryptorach byłoby ich kilka), o czym będzie mowa w rozdziale poświęconym formie instrukcji.

Niektóre z zaleceń dotyczą więcej niż jednej grupy zagadnień wyodrębnionych w instrukcji, stąd czasami powtórzenia, a czasami odesłania do innego działu instrukcji.

Poniżej przedstawione zostaną obszernie fragmenty instrukcji ilustrujące jej budowę, charakter i zastosowanie.

METODYKA INDEKSOWANIA

Postać charakterystyki wyszukiwawczej

Wszystkie deskryptory w charakterystyce wyszukiwawczej dokumentu (ChWD) wpisujemy małymi literami, identyfikatory jako nazwy własne rozpoczynają się od dużych liter, przy czym identyfikatory będące tytułami dokumentów omawianych w artykule lub tytułem konferencji, podajemy w cudzysłowie. Służy to rozgraniczeniu znaczeń terminów równokształtnych (mających tę samą postać), a reprezentujących różne tematy, np. biblioteka narodowa jako typ biblioteki, Biblioteka Narodowa jako nazwa własna instytucji, lub ew. "Biblioteka Narodowa" jako wydawnictwo (informatoryjny) o bibliotece.

Deskryptory dziedzinowe (nazwy działów), np. informacja naukowa, bibliotekarstwo, informatyka, też są deskryptorami wykorzystywanymi przy indeksowaniu; wpisujemy je małymi literami, choć w TIN zapisane są literami dużymi (dla różnicowania ich w stosunku do pozostałych deskryptorów).

Nie ma formalnych ograniczeń wynikających z reguł indeksowania co do liczby deskryptorów składających się na jedną ChWD, jedyne ograniczenie wynika z przyjętego w systemie WIST formatu opisu, w którym pole deskryptorów nie może przekraczać 300 znaków. W praktyce sprowadza się to do większej liczby deskryptorów przy krótszych terminach, lub mniejszej przy dłuższych. Średnia długość

deskryptorów w różnych działach TIN jest różna - najmniejsza w działach "Informatyka" i "Oziedziny pokrewne", jako że przyjęto tam najwyższy stopień uogólnienia - największa w podstawowej części tezaurusa. Orientacyjnie średnia długość deskryptora w TIN nie przekracza 20 znaków, co oznacza możliwość umieszczania do 15 deskryptorów w jednym opisie. W praktyce liczba ta przy indeksowaniu artykułów nie przekracza dziesięciu.

Długość deskryptorów nie jest teoretycznie limitowana, ale pakiet wyszukiwawczy wykorzystywany w systemie narzuca pewne ograniczenia; terminy wyszukiwawcze są obcinane do 30 znaków. Dlatego przy indeksowaniu manualnym wprowadzamy pełne nazwy deskryptorów (w przypadku niektórych identyfikatorów, np. tytułów konferencji mogą być one długie), w realizacji maszynowej (przy wprowadzaniu danych na nośnik maszynowy) następuje ich obcięcie do 30 znaków (wliczając w to spację), co jednak i tak zapewnia wystarczającą identyfikację terminów wyszukiwawczych.

Kolejność deskryptorów w ChWD nie ma znaczenia.

Szczegółowość indeksowania

Nie wszystkie zagadnienia indeksowane są na jednakowym poziomie szczegółowości. Głębokość indeksowania limitowana jest już częściowo samym tezauresem i różną szczegółowością terminów w różnych działach TIN (niejednolitą rozbudową poszczególnych działów). Dalsze zróżnicowanie poziomu głębokości indeksowania osiągnąć można na etapie samego procesu indeksowania poprzez zastosowanie różnych reguł. I tak dla wielu zagadnień (dziedzin), nawet z tej samej grupy deskryptorów, zalecane są różne przepisy, w zależności od stopnia pokrewieństwa (związku) tych zagadnień z dziedziną informacji naukowej. Na przykład ze wszystkich urządzeń peryferyjnych w dziale "Informatyka" tylko drukarki potraktowane są

z większą szczegółowością (zarówno w tezaurusie, jak i w regułach indeksowania). W praktyce oznacza to, że podział tych zagadnień w tezaurusie jest bardziej rozbudowany (tu: wyszczególniono typy drukarek, np. drukarka mozaikowa, termiczna, laserowa, itp.), a zasady indeksowania narzucają konieczność bardziej precyzyjnego przydzielania deskryptorów i w większej ich liczbie (tu: artykuł o drukarkach indeksujemy trzema deskryptorami określającymi: typ drukarki, nazwę firmy oraz nazwę modelu, podczas gdy dla artykułów omawiających pozostałe urządzenia zalecany jest tylko jeden deskryptor określający jego typ, np. monitor, modem, ploter). Rozróżnienie to podyktowane jest potrzebami użytkowników systemu, którzy w większym stopniu będą się interesować informacjami na temat drukarek (z racji ich znacznie częstszego wykorzystywania w bibliotekach i ośrodkach informacji), niż innym sprzętem peryferyjnym nie znajdującym zastosowania w działalności informacyjnej. Informacje te mogą być przez nich np. poszukiwane przed podjęciem decyzji o kupnie lub w celu poznania parametrów i zasad eksploatacji drukarek już przez nich posiadanych. Podobnie rozwiązano problemy architektury komputera (jego poszczególnych elementów), np. indeksowanie artykułów o różnych typach procesorów tylko jednym deskryptorem "procesor", bez wyszczególniania typu lub nazwy handlowej może wydawać się specjalistom innych dziedzin zbyt uogólnieniem ("płytkością" systemu). W systemie WISi można jednak przypuszczać, że temat ten nie będzie wzbudzać większego zainteresowania u przewidywanych użytkowników i dlatego uznano, że nie wymaga szczegółowego przedstawienia w tezaurusie i w charakterystykach wyszukiwawczych tworzonego zbioru. Fakt, że opisy artykułów wzbogacone są o abstrakty usprawiedliwia częściowo tak płytkie indeksowanie, gdyż przy ewentualnych pytaniach

na ten temat (z pewnością nielicznych) użytkownik otrzymując nawet bardzo dużą liczbę wyszukanych opisów ma zawsze możliwość jej zmniejszenia (dokonania dodatkowej selekcji) na podstawie informacji uzyskanych z abstraktów lub nawet z samych tytułów dokumentów (w większości artykułów typ i nazwa handlowa omawianych urządzeń wyszczególnione są w tytule).

Koordinacja deskryptorów

Język deskryptorowy, szczególnie z tezauresem o strukturze fasetowej, jest typowym językiem stosującym postkoordinację. TIN zawiera jednak wiele deskryptorów prekoordinowanych (złożonych), co umożliwia wyrażenie niektórych tematów indeksowanych artykułów za pomocą jednej jednostki leksykalnej (jednego deskryptora). Podyktowane to było wieloma czynnikami, między innymi próbą wyeliminowania szumu informacyjnego spowodowanego "fałszywą koordynacją", wygodą użytkownika, krótkością ChWD, itp. Dla wyrażenia wielu tematów potrzebna jest jednak koordynacja deskryptorów, np. artykuł na temat dostępu do zagranicznej literatury naukowej i fachowej zaindeksujemy:

dostęp do źródeł + literatura zagraniczna + literatura naukowa +
+ literatura fachowa.

Analogicznie, temat artykułu "import czasopism naukowych z II obszaru płatniczego w bibliotekach polskich" wymaga następujących deskryptorów:

gromadzenie materiałów zagranicznych + gromadzenie czasopism +
+ czasopismo naukowe + Polska,
a temat "zautomatyzowany katalog centralny czasopism medycznych":
katalog centralny + katalog automatyczny + katalog czasopism +
+ czasopismo fachowe + informacja medyczna.

Niektóre spośród tego typu tematów złożonych znajdują się w TIN jako odsyłacze wskazujące odpowiednie połączenia deskryptorów, np.

centralny katalog czasopism zagranicznych U katalog centralny +
+ katalog czasopism zagranicznych
zautomatyzowany katalog centralny U katalog centralny +
+ katalog automatyczny.

Ma to na celu ułatwienie procesu indeksowania i wyszukiwania, szczególnie dla użytkowników nieobeznanych z systemem i zasadami wyszukiwania postkoordynowanego. W obecnej wersji tezauryasa odsyłacze tego typu wprowadzono niewiele, głównie dla tematów (połączeń deskryptorów) najczęściej spotykanych w piśmiennictwie i w zapytaniach użytkowników. Przed publicznym udostępnieniem systemu (umożliwieniem samodzielnego wyszukiwania przez użytkowników za pośrednictwem terminali) przewiduje się znaczne zwiększenie liczby tego typu odsyłaczy, szczególnie w wyszukiwawczej wersji tezauryasa (por. rozdz. o formie instrukcji).

Wieloznaczność charakterystyk wyszukiwawczych

Tezaurus Informacji Naukowej zawiera pewną liczbę deskryptorów dopuszczających wieloznaczność, tj. wykorzystywanych do indeksowania artykułów różniących się pod względem treści (poruszających różne tematy), np. deskryptor "koszt informacji" zawiera w swoim zakresie zarówno zagadnienia kosztów wytworzenia (wyprodukowania) informacji, jak i jej uzyskania (cena, jaką płaci za nią użytkownik końcowy, problem odpłatności informacji). Dopuszczenie tego typu niejednoznacznych deskryptorów podyktowane było próbą ograniczenia rozrostu słownictwa (szczególnie dla tematów niezbyt często reprezentowanych w literaturze przedmiotu), jak i trudnościami w rozgraniczeniu zakresów tych deskryptorów (zarówno w nazwie, jak i w piśmiennictwie).

Innego typu wieloznaczność ChWD spowodowana jest tzw. "fałszywą koordynacją". Na przykład artykuł omawiający międzynarodowy system informacyjny o pracach naukowo-badawczych z zakresu medycyny oraz o materiałach konferencyjnych z zakresu informacji naukowej posiada opis deskryptorowy: międzynarodowy system + prace naukowo-badawcze + informacja medyczna + materiały konferencyjne + informacja naukowa. Przy takim zestawie deskryptorów może on być wydany w odpowiedzi na pytanie o materiały konferencyjne z zakresu medycyny lub o prace naukowo-badawcze z informacji, choć żadnego z tych tematów nie dotyczy. W celu wyeliminowania (lub przynajmniej częściowego zmniejszenia) szumu informacyjnego spowodowanego błędną koordynacją (a więc łączeniem deskryptorów występujących w jednej charakterystyce wyszukiwawczej lecz nie odnoszących się do siebie wzajemnie) wprowadzono do IIN znaczną liczbę tzw. deskryptorów prekoordynowanych, szczególnie dla tematów, co do których istnieje duże prawdopodobieństwo, iż będą często wykorzystywane. Przykłady takich deskryptorów: literatura informacyjna (tj. z zakresu informacji naukowej), literatura bibliotekarska, literatura informatyczna; encyklopedia informacji naukowej, encyklopedia bibliotekarska, encyklopedia informatyki; terminologia informacji naukowej, terminologia informatyczna; norma bibliograficzna, norma biblioteczna, norma informacyjna, itp., dla pozostałych zagadnień danej grupy (niezbyt często stanowiących tematy artykułów) takich "gotowych" deskryptorów nie przewidziano. Artykuły te indeksujemy koordynując deskryptory z różnych faset, np. dla literatury medycznej deskryptory: literatura fachowa + informacja medyczna, dla encyklopedii technicznej: encyklopedia specjalna + informacja techniczna, dla terminologii z zakresu poligrafii: poligrafia + terminologia, dla norm z zakresu mikrografii: norma

poliska + mikrografia, itp. Informacje na temat łączenia deskryptorów z różnych faset często znajdują się pod odpowiednimi deskryptorami.

Deskryptory dziedzinowe

Jak już wspomniano, nazwy działów tezauryśa też s deskryptorami (pisane małymi literami), pełnią jednak bardziej funkcję modyfikatorów zawężających znaczenie niektórych terminów występujących w tych działach. Na przykład deskryptory "prace magisterskie" lub "prace naukowo-badawcze" poprzez dodanie deskryptorów dziedzinowych "informacja naukowa", "bibliotekarstwo", "czytelnictwo", itp. mają zawężony zakres oznaczając prace magisterskie (prace naukowo-badawcze) z informacji naukowej, z bibliotekarstwa lub czytelnictwa. Do samodzielnego indeksowania artykułów nie powinno się ich raczej wykorzystywać, starając się zastosować jeden (lub więcej) z deskryptorów węższych tezauryśa, np. zamiast deskryptora dziedzinowego "informacja naukowa" zastosować jeden (lub więcej) z deskryptorów węższych tego działu, np. informacja (jako wiadomość), nauka o informacji, teoria informacji, działalność informacyjna, itd. w zależności od zawartości treściowej indeksowanego artykułu. W nielicznych wypadkach deskryptory dziedzinowe mogą być wykorzystane do zaindeksowania artykułów bardzo ogólnych (zawierających wszystkie te zagadnienia), choć będzie to prawdopodobnie częstszy przypadek przy indeksowaniu książek (szczególnie podręczników) niż artykułów z czasopism.

Proces indeksowania

Każdy artykuł indeksujemy dwoma etapami: uogólniającym (całościowym) i wyszczególniającym.

Etap I. Indeksowanie całościowe

W pierwszym etapie przydzielamy deskryptory odwzorowujące w miarę dokładnie cały zakres treściowy artykułu, jego temat podstawowy. Czasami dla objęcia całego zakresu niezbędne będzie wybranie terminów szerszych, bardziej ogólnych, na ogół jednak temat ten wyrażony będzie adekwatnie przez jeden termin lub ich kombinację (połączenie kilku terminów). Po sprawdzeniu stopnia zgodności wybranych terminów z ewentualnymi zapotrzebowaniami użytkowników i faktyczną zawartością treściową artykułu dokonujemy zamiany wybranych słów kluczowych na formę zaakceptowaną w systemie, tj. zgodną z wykazem deskryptorów.

Zalecenia praktyczne. Po przeczytaniu całego artykułu formułujemy jego temat, możliwie w jednym zdaniu (najczęściej będzie to równoważnik zdania), np.:

1. Prace prowadzone przez Bibliotekę Narodową w Warszawie nad automatycznym tworzeniem słownika haseł przedmiotowych;
2. Metodyka tworzenia indeksów przedmiotowych do tablic klasyfikacyjnych;
3. Historia, stan obecny i ogólna charakterystyka "Klasyfikacji Bibliograficznej" utworzonej przez H.E. Bliss^{4/}.

Dużą pomocą przy formułowaniu podstawowego tematu artykułu może być jego tytuł, streszczenie, tytuły poszczególnych rozdziałów.

^{4/} Przykłady zaczerpnięte są z konkretnych artykułów z czasopisma "Zagadnienia Informacji Naukowej" zaindeksowanych w systemie WIST.

1. Sadowska J.: Komputerowa edycja słownika haseł przedmiotowych. - Zag. Inf. Nauk. 1987 nr 1(50), s.67-73.
2. Sadowska J.: Indeksy przedmiotowe do tablic klasyfikacyjnych. - Zag. Inf. Nauk. 1986 nr 1(48), s.83-96.
3. Chmielewska-Gorczyca E.: Klasyfikacja Bibliograficzna Bliss. - Zag. Inf. Nauk. 1986 nr 1(48), s.97-118.

Przestrzec jednak należy przed zbytnim sugerowaniem się tytułem artykułu i terminami stosowanymi przez autora; przy wyborze słów kluczowych staramy się możliwie adekwatnie (precyzyjnie) odzwierciedlić treść indeksowanego dokumentu. Wstępne indeksowanie wykazało tendencję do stosowania w tytułach terminów zbyt ogólnych w stosunku do faktycznego zakresu artykułu, np. artykuł o imporcie wydawnictw z II obszaru płatniczego w rzeczywistości dotyczył gromadzenia czasopism zagranicznych (bez wydawnictw książkowych) i tak go indeksujemy. Podobnie, artykuł o zautomatyzowanym systemie informacji o wydawnictwach zagranicznych dotyczył w rzeczywistości tylko automatycznego katalogu centralnego regionalnego; artykuł o doskonaleniu zawodowym bibliotekarzy mówi tylko o kursach doskonalących (z pominięciem innych form podnoszenia kwalifikacji zawodowych), itp.

Czasami może być odwrotnie, tytuł sugerować może węższy niż faktyczny zakres tematyczny artykułu, np. tytuł "przegląd systemów klasyfikacyjnych stosowanych w polskich bibliotekach społeczno-ekonomicznych" artykułu dotyczącego w rzeczywistości porównania różnych języków informacyjno-wyszukiwawczych (nie tylko klasyfikacyjnych).

Następną czynnością jest przyporządkowanie poszczególnym elementom wybranego tematu deskryptorów zaczerpniętych z tezauryusa, tj. rozbicie zdania na terminy proste i następnie odszukanie poprawnej ich formy w wykazie deskryptorów. Efektem tych prac będą następujące opisy deskryptorowe:

1. słownik haseł przedmiotowych

automatyczne tworzenie słownika JIW

Biblioteka Narodowa w Warszawie

2. indeks przedmiotowy

indeks do tablic

tworzenie indeksu

3. Klasyfikacja Bibliograficzna

Po wybraniu terminów przeprowadzamy ich selekcję oraz ewentualnie uzupełniamy ten zestaw dodatkowymi deskryptorami. Przy tej czynności należy się zastanowić, kogo może ten artykuł zainteresować, tj. jakim tematem zainteresowani użytkownicy chcieliby ten artykuł otrzymać, dla jakich zapytań informacyjnych byłby ten artykuł relewantny, jakie potrzeby informacyjne zaspokoi, itp. Analizując w ten sposób wybrane deskryptory stwierdzamy, że w pierwszym przykładzie trzeci deskryptor jest zdecydowanie za szeroki; użytkownik poszukujący dokumentów dotyczących ogólnych informacji na temat Biblioteki Narodowej z pewnością uzna ten artykuł za nieprzydatny. Dodanie modyfikatora "działalność" uściśli ten zakres w takim stopniu, iż możemy go zaakceptować w opisie, jako że artykuł rzeczywiście dotyczy pewnego wycinka działalności prowadzonej przez Bibliotekę.

W drugim przykładzie stwierdzamy zgodność wybranych deskryptorów z faktycznym tematem oraz ewentualnymi zainteresowaniami użytkowników, w trzecim przykładzie dodajemy deskryptor "klasyfikacja uniwersalna". Wynika to z reguł indeksowania zalecających przydzielenie deskryptora wskazującego typ omawianego języka informacyjnego oraz z faktu, iż użytkownicy poszukujący informacji na temat wszelkich klasyfikacji uniwersalnych funkcjonujących obecnie z pewnością zainteresowani będą otrzymaniem tego artykułu.

Etap II. Indeksowanie szczegółowe

Na tym etapie analizujemy poszczególne fragmenty (podrozdziały, części, akapity) artykułu, nadając każdemu fragmentowi stano-

wiacemu pewną całość tematyczną tytuł) (przydzielając mu słowa kluczowe odzwierciedlające jego zakres). Następnie wybrane terminy selekcjonujemy podobnie jak w poprzednim etapie i znajdujemy dla nich formę odwołującą w systemie (dopasujemy słowa kluczowe do deskryptorów). Przy indeksowaniu szczegółowym staramy się przydzielać poszczególnym fragmentom słowa kluczowe adekwatnie odzwierciedlające ich treść, unikamy uogólniania.

Zalecenia praktyczne. Czytamy jeszcze raz cały artykuł nie wyszukując tym razem ogólnego tematu, ale "kawałkując" go na stanowiące pewną całość tematyczną fragmenty. Każdemu fragmentowi przypisujemy tytuł (nazwę określającą jego zakres), najczęściej w postaci jednego słowa kluczowego, czasami dwóch, rzadziej - więcej. Dla pierwszego artykułu z przykładów wybieramy słowa kluczowe:

1. słownik JIW
2. komputerowe tworzenie słownika JIW
3. etapy tworzenia słownika JIW
4. słownik haseł przedmiotowych
5. funkcja słownika JIW
6. "Słownik języka haseł przedmiotowych Biblioteki Narodowej"
(jest to identyfikator)
7. struktura słownika haseł przedmiotowych
8. relacje w słowniku JIW
9. program komputerowy
10. aktualizacja JIW
11. wyszukiwanie automatyczne

Dokonujemy selekcji zgodnie z regułami podanymi przy omawianiu etapu pierwszego. Uznajemy, że artykuł nie będzie satysfakcjonować użytkowników poszukujących informacji na temat:

1. słownik JIW
2. etapy tworzenia słownika JIW
3. funkcja słownika JIW
4. program komputerowy
5. aktualizacja JIW
6. wyszukiwanie automatyczne

Usuujemy te słowa kluczowe z zestawu, dla pozostałych wybieramy z te-
zaurusu formy deskryptorowe. Efektem końcowym tych czynności będzie
zestaw deskryptorów:

1. automatyczne tworzenie słownika JIW
2. słownik haseł przedmiotowych
3. "Słownik języka haseł przedmiotowych BN"
4. struktura słownika haseł przedmiotowych
5. relacje słownikowe JIW

k który łączymy z deskryptorami przydzielonymi na etapie pierwszym.
Widzimy, że niektóre z nich powtarzają się, w związku z czym licz-
ba deskryptorów dodanych na etapie drugim ulega zmniejszeniu. W e-
fekte opis deskryptorowy artykułu będzie miał postać:

1. słownik haseł przedmiotowych
2. automatyczne tworzenie słownika JIW
3. "Słownik języka haseł przedmiotowych BN"
4. struktura słownika haseł przedmiotowych
5. relacje słownikowe JIW
6. Biblioteka Narodowa w Warszawie - działalność

Podobnie postępujemy z pozostałymi dwoma przykładowymi arty-
kułami. Poszczególne etapy indeksowania tych artykułów prześledzić
można w aneksie załączonym do instrukcji, w którym przytoczone są
pełne teksty artykułów wraz z wybranymi słowami kluczowymi i kome-
ntarzami do nich, ilustrującymi kolejne czynności i podejmowane de-

cyzje. Ostateczne opisy deskryptorowe tych artykułów mają postać:

2. indeks przedmiotowy

indeks do tablic

tworzenie indeksu

wybór tematu

postać tematu

temat złożony

określnik

struktura hasła przedmiotowego

skupienia tematyczne

odsyłacz całkowity

odsyłacz uzupełniający

struktura indeksu

3. klasyfikacja uniwersalna

"Klasyfikacja Bibliograficzna"

Bliss Henry E. (1870-1955)

JIW - osoby

struktura tablic

tablice pomocnicze

notacja mieszana

notacja mnemoniczna

indeks przedmiotowy

indeks do tablic

rewizja tablic

formuła fasetowa

Zalecenia powyższe są zaleceniami ogólnymi, czasami niemożliwymi do zrealizowania w całości; nie zawsze będą możliwe oba etapy. Dla artykułów omawiających ogólnie jeden problem (choć w wielu jego aspektach) niemożliwy do przeprowadzenia będzie etap drugi, np. dla

artykułu o ustawie w zakresie informacji naukowej. Z drugiej strony istnieją artykuły "pokawałkowane", omawiające równolegle kilka zagadnień z różnych działów tezaurusu, a więc takie, dla których nie ma terminu obejmującego swym zakresem wszystkie te zagadnienia. Istnieje zawsze możliwość przydzielenia terminu bardzo ogólnego, np. deskryptora dziedzinowego, nie byłoby to jednak zgodne z założeniami indeksowania szczegółowego systemu WIST.

ZALECENIA OGÓLNE

Rozdział drugi instrukcji stanowią zalecenia ogólne, a więc odnoszące się do deskryptorów z różnych działów tezaurusu. Poniżej wybrano kilka przykładów ilustrujących ten typ zaleceń.

Imprezy (konferencje, zjazdy, seminaria, targi, itp.)

Artykuł o imprezach indeksujemy deskryptorami wyrażającymi:

1. rodzaj imprezy, np. targi komputerowe, wystawa komputerowa, konferencja międzynarodowa, narada, itp.

2. tematykę imprezy, np. czasopismo bibliotekarskie, ochrona zbiorów, katalog centralny, czytelnictwo niepełnosprawnych, itp.

3. nazwę własną imprezy (np. Softarg) lub jej tytuł (w przypadku tytułu stosujemy cudzysłowy); jeśli nazwa kończy się cyframi wskazującymi rok odbycia się imprezy, np. Softarg'89 - opuszczamy tę część nazwy; nazwa może być tytułem konferencji odzwierciedlającym jej tematykę, jeśli oba deskryptory są równokształtne - jeden z nich opuszczamy, tj. pozostawiamy tylko deskryptor oznaczający tematykę (bez cudzysłowu); jeśli tytuł imprezy podany jest w języku obcym, podajemy jego tłumaczenie na język polski; tematykę imprezy wyraża deskryptor adekwatnie ją odzwierciedlający, tj. możliwie szczegółowy, w dziedzinach pokrewnych uogólniony do nazwy dziedziny; tema-

tyka imprezy może być wyrażona więcej niż jednym deskrytorem:

4. datę odbycia się imprezy wyrażoną rokiem;

5. miejsce odbycia się imprezy wyrażone nazwą miasta i państwa (rozdzielone przecinkiem i spacją); w przypadku Polski tylko nazwa miasta.

6. instytucję organizującą (sponsora, organizatora spotkania) zaopatrzoną w modyfikator "działalność"; jeśli organizatorów wielu - opuszczamy ten element opisu.

W przypadku braku w indeksowanym artykule informacji dotyczących niektórych z tych elementów mogą być one pominięte (jeśli nie ma możliwości zdobycia tych informacji poza tekstem); niektóre elementy może indeksator usunąć uznając je za nieistotne, np. zbyt ogólny temat konferencji, lub instytucję sponsorującą.

Przykłady:

1. seminarium międzynarodowe (charakter imprezy)
"Nowe techniki w działalności biblioteczno-informacyjnej" (tytuł)
automatyzacja biblioteki } (tematyka imprezy)
automatyzacja ośrodka informacji }
Londyn, Wielka Brytania (miejsce)
1987 (data)
2. seminarium międzynarodowe (charakter imprezy)
biblioteka szkoły wyższej } (tematyka imprezy)
działalność dydaktyczna biblioteki }
biblioteka ekonomiczna }
Biblioteka Główna Akademii Ekonomicznej w Poznaniu - działalność
(instytucja organizująca)
Poznań (miejsce)
1984 (data)
3. targi komputerowe
Softarg w rzeczywistości "Softarg'88"

oprogramowanie

Katowice

1988

4. konferencja międzynarodowa tytuł konferencji "Nowe horyzonty
IFLA - działalność bibliotekarstwa..." został opusz-
Tokio, Japonia czony jako nieistotny.
1986

Deskryptory geograficzne

Deskryptory tego typu służą do oznaczenia miejsca, w którym odbywa się lub którego dotyczy zagadnienie (zjawisko) omawiane w danym artykule; staramy się dodać je zawsze, jeśli temat artykułu zawężony jest do pewnego terytorium (jednostki geografii fizycznej, politycznej lub administracyjnej). Jeśli artykuł dotyczy więcej niż dwóch, trzech węższych jednostek geograficznych - uogólniamy do nazwy deskryptora geograficznego bezpośrednio szerszego, np. jeśli dotyczy bibliotek w trzech miastach uogólniamy do nazwy kraju.

Deskryptory geograficzne mogą występować w sąsiedztwie wszystkich pozostałych deskryptorów, niemniej pewna grupa deskryptorów jest na to bardziej podatna; charakter ich wymaga umieszczenia przy nich deskryptorów geograficznych w większości artykułów (poza nielicznymi artykułami teoretycznymi, definiującymi problem lub dotyczącymi wielu krajów). Przykładami tego typu deskryptorów są: polityka informacyjna (państwa), polityka biblioteczna, przemysł komputerowy, rynek komputerowy, edukacja komputerowa (społeczeństwa), organizacja bibliografii, itp. Deskryptory "podatne" na uściślenie geograficzne wymieniane są w poszczególnych częściach instrukcji szczegółowych. Uściślenie geograficzne implikują też niektóre modyfikatory, np. historia, stan obecny.

Pewne deskryptory wpisane mają niejako aspekt geograficzny w swoją strukturę, np. czasopisma polskie, czasopisma zagraniczne; polskie bazy danych, zagraniczne bazy danych, międzynarodowe bazy danych; bibliografia angielska, bibliografia francuska, bibliografia polska, itp. Podyktowane to było próbą wyeliminowania szumu informacyjnego przy takich tematach, jak: czasopisma polskie w obcych zbiorach, dostęp do zagranicznych baz danych w Polsce, bibliografia angielska w bibliotekach polskich, polskie bazy danych w systemach międzynarodowych, itp.

Z reguł indeksowania wynika konieczność umieszczania deskryptorów geograficznych przy wszelkich opisach imprez (por. "imprezy").

Modyfikatory

Modyfikatory umieszczamy po deskrypcorze, do którego się odnoszą, oddzielone od niego myślnikiem. Jeśli przy jednym deskrypcorze występuje więcej niż jeden modyfikator, powtarzamy w CHWD deskrypcor tyle razy, ile jest przy nim modyfikatorów. Wykaz modyfikatorów dodany jest do tezaurusa i będzie dalej uzupełniany.

Przy stosowaniu modyfikatorów ogólnych, np. tendencje rozwojowe, osoby, historia staramy się uogólnić deskrypcor wyrażający treść artykułu do nazwy całego działu, przynajmniej w działach innych niż "Informacja Naukowa" i "Bibliotekarstwo" (sugeruje to zresztą wykaz tych modyfikatorów umieszczony w części systematycznej na początku działu, tuż pod jego nazwą). Tak więc nawet jeśli w artykule mowa jest o historii komputerów tworzymy opis "informatyka - historia", a nie "komputer - historia". Jeśli artykuł dotyczy zdecydowanie jednego aspektu (zagadnienia węższego), możemy dodać dodatkowy deskrypcor precyzujący to zagadnienie, np. artykuł o przyszłości nośników danych indeksujemy:

informatyka - tendencje rozwojowe

nośnik danych

a o nowych kierunkach w zakresie tworzenia programów komputerowych:

informatyka - tendencje rozwojowe

oprogramowanie

Pozostałe modyfikatory (tj. modyfikatory szczegółowe) dodajemy do tych deskryptorów, które precyzują, np. artykuł o problemach związanych z użytkowaniem i konserwacją drukarek indeksujemy:

drukarka - eksploatacja

Niektóre modyfikatory implikują przydzielenie dodatkowego deskryptora geograficznego (por. "deskryptory geograficzne").

ZALECENIA SZCZEGÓŁOWE

Rozdział trzeci instrukcji zawiera zalecenia dotyczące tylko poszczególnych działów, poddziałów lub nawet pojedynczych deskryptorów. Jak już wspomniano, kolejność reguł jest zgodna z kolejnością deskryptorów, do których się odnoszą w systematycznej części tezauryusa. Oczywiście reguł tego typu jest znacznie więcej niż zaleceń ogólnych i ich liczba stale wzrasta w miarę prac nad tworzeniem zbioru wyszukiwawczego, dlatego ten rozdział instrukcji jest najobszerniejszy. Przykładowo wybrane fragmenty tej części instrukcji (zaczepnięte z działu "Informatyka") służyć mają pokazaniu ich charakteru i możliwości rozbudowy.

3.3. INFORMATYKA

3.3.2. Komputer

Artykuły o poszczególnych modelach komputera indeksujemy podając deskryptor wskazujący typ komputera i dwa identyfikatory: nazwę firmy produkującej komputer i nazwę modelu. Jeśli w jednym artykule omówionych jest kilka modeli jednej firmy (a więc na wysokim stopniu ogólności) możemy nazwy modeli opuścić pozostawiając tylko nazwę firmy.

Jeśli omawiany jest komputer specjalizowany, nastawiony na jeden rodzaj przetwarzania, przeznaczony do wykonywania określonych

zadań (celów) indeksujemy deskryptorem wskazującym typ komputera oraz dziedzinę zastosowania (przeznaczenia), np.

1. mikrokomputer

Compaq

Deskpro 386

projektowanie wspomagane komputerem

2. mikrokomputer

Amstrad

PC 8256

przetwarzanie tekstów

Jeśli przeznaczenie komputera jest na tyle wąskie lub odległe od dziedziny informacji, że nie ma deskryptora określającego jego zastosowanie indeksujemy wtedy deskryptorem "komputer specjalizowany" (oraz identyfikatorami).

Jeśli oprócz omówienia konstrukcji, parametrów, oceny komputera itp. podane są informacje na temat programów dostępnych dla tego komputera, do opisu dodajemy deskryptor "oprogramowanie".

3. mikrokomputer

IBM

PS/2

Model 50

4. mikrokomputer

Commodore

Amiga 500

oprogramowanie

3.3.3. Architektura komputera

Deskryptor ten przydzielamy artykułom omawiającym teoretycznie, ogólnie problemy budowy (architektury) komputerów (mikrokomputerów) niezależnie od typu komputera. Przy artykułach poświęconych architekturze (budowie) konkretnego modelu komputera stosujemy deskryptory odzwierciedlające typ komputera i jego nazwę handlową (firma i nazwa modelu), bez deskryptora "architektura komputera".

3.3.4 Pamięć komputera

Rodzaje pamięci indeksujemy zgodnie z podziałem tej fasety w tezaurusie; nie przydzielamy tej grupie deskryptorów żadnych nazw własnych (identyfikatorów).

3.3.5 Urządzenia peryferyjne

Artykuły o urządzeniach peryferyjnych indeksujemy podając tylko rodzaj urządzenia, np. modem, klawiatura, monitor dotykowy; wyjątek stanowią artykuły omawiające drukarki, które indeksujemy dodatkowo nazwą firmy i nazwą modelu, np.

1. drukarka mozaikowa

Star

NL-10

(. . .)

3.3.7 Język programowania

Jeśli artykuł omawia język programowania (jego historię, strukturę, zalety, zakres zastosowań, itp.) przydzielamy do opisu deskryptor "język programowania" oraz identyfikator będący nazwą własną języka, np.

1. język programowania

Basic

2. język programowania

Pascal

Jeśli artykuł pełni funkcję instrukcji programowania w tym języku (nauka, kurs programowania), indeksujemy go deskryptorem "programowanie" oraz identyfikatorem będącym nazwą języka, np.

3. programowanie

Basic

4. programowanie

Logo

Jeśli i pierwsze, i drugie ujęcie, tj. jeśli artykuł mówi o języku i o programowaniu w nim, wprowadzamy oba deskryptory wraz z identyfikatorem, np.

5. język programowania

programowanie

język C

Nazwy własne języków programowania są identyfikatorami, a więc pisane są dużą literą. Nazwy te wprowadzone są do tezaurusa. W nazwie pominięte jest słowo "język", z wyjątkiem języka C (por. przykłady). Artykuły omawiające mutacje poszczególnych języków programowania (wersje dostosowane lub opracowane dla konkretnych komputerów) indeksujemy deskrytorem sprowadzonym do nazwy ogólnej, np. zamiast "Basic XL" stosujemy "Basic".

3.3.8 Procesy informatyczne

Deskryptor **programowanie** stosujemy do indeksowania artykułów będących swego rodzaju nauką danego języka programowania (wraz z identyfikatorem nazwy języka, por. punkt 3.3.7) oraz do artykułów omawiających proces programowania teoretycznie, niezależnie od języka. Deskryptor **błędy programowania** obejmuje błędy popełniane przez programistów podczas procesu programowania, błędy (niedoskonałości) w istniejących już programach oraz problemy wynikające z błędów w programach ("pomyłki" komputerów).

Deskryptor **nietypowe liternictwo** obejmuje definiowanie (generowanie) znaków nietypowych dla alfabetu angielskiego, tj. dostosowywanie do języków narodowych, np. do języka japońskiego, chińskiego, znaków logiki i matematyki, itp. na monitorze, na klawiaturze i na drukarce (a więc także częściowo proces drukowania). Dla znaków języka polskiego zarezerwowany jest deskryptor węższy **polskie liternictwo**.

3.3.10 Zastosowania komputerów

Przy artykułach omawiających różne zastosowania komputerów stosujemy deskryptory węższe w stosunku do deskryptora "zastosowania

komputerów" adekwatnie odzwierciedlające treść artykułu. Jeśli artykuł dotyczy zastosowań zbyt odległych tematycznie dla informacji naukowej (a więc nie wyszczególnionych w teaurusie), np. wykorzystanie komputerów w medycynie, w produkcji przemysłowej, meteorologii, kinematografii, gospodarce magazynowej, itp. dajemy deskryptor ogólny "zastosowania komputerów". Podobnie postępujemy w przypadku, jeśli w słownictwie teaurusa brak wyszczególnienia danego rodzaju zastosowania, choć jest ono spokrewnione z działalnością informacyjną, wprowadzamy jednak wtedy dane słowo kluczowe do kartoteki terminów będących kandydatami na deskryptory.

Zastosowania komputerów charakterystyczne dla działalności informacyjnej, bibliotecznej i bibliograficznej znajdują się w innych działach IIN, w dziale "Informatyka" występują tylko jako odsyłacze kierujące do odpowiednich fragmentów części systematycznej teaurusa. Pamiętać jednak należy, że kierują one tylko do deskryptora najogólniejszego z danej fasety, dlatego zalecane jest odszukanie danego deskryptora w części systematycznej w celu sprawdzenia, jakie deskryptory węższe umieszczone są pod nim, np. jako odsyłacz występuje tylko deskryptor "automatyzacja biblioteki", natomiast wszelkie deskryptory bardziej szczegółowe, np. automatyzacja gromadzenia zbiorów, automatyzacja przechowywania zbiorów, itp. można odnaleźć dopiero w dziale "Bibliotekarstwo".

(. . .)

FORMA INSTRUKCJI

W obecnej postaci instrukcja jest maszynopisem. Planowane jest wydanie jej jako oddzielnej broszurki dołączonej do "Teaurusa Informacji Naukowej", dalej jednak w postaci tekstu pisanego stanowiącego jedną całość. Wspomniano już, że idealnym rozwiązaniem

byłoby "pokawałkowanie" instrukcji, to znaczy rozbitcie jej poszczególnych zaleceń na drobne teksty i umieszczenie ich w teaurusie przy tych deskryptorach, do których się odnoszą. Przy niektórych deskryptorach znajdowałoby się kilka reguł, pewne zalecenia należałoby umieścić przy kilku deskryptorach. Dla niektórych zaleceń jednak liczba deskryptorów, przy których należałoby je umieścić byłaby zbyt duża, aby można było uznać to za celowe (efektywne); wielokrotne powtarzanie tego samego tekstu przy wielu deskryptorach zwiększyłoby znacznie objętość teaurusu i zaciemniło obraz słownictwa języka. Reguły tego typu, jako zalecenia ogólne, pozostałyby poza tekstem teaurusu w postaci oddzielnego wykazu zasad indeksowania, do teaurusu wprowadzone zostałyby tylko reguły szczegółowe, dotyczące wązszych zagadnień i mniejszej liczby deskryptorów.

Rozwiązanie takie odciążałoby w znacznym stopniu podstawową część instrukcji, czyniąc ją bardziej przejrzystą, wygodniejszą i łatwiejszą w stosowaniu, szczególnie dla użytkownika końcowego. Obecna postać instrukcji nie jest odpowiednia dla użytkowników przypadkowych (korzystających z systemu od czasu do czasu, a więc nieobeznanych z systemem), prowadzących wyszukiwania samodzielnie bez pośrednictwa wykwalifikowanego, przeszkolonego dokumentalisty. Z doświadczenia bowiem wiadomo, że umieszczanie zbyt dużej liczby zaleceń w jednym bloku nie skłania tego typu użytkowników do ich studiowania (przeoglądania) w celu odnalezienia jednej, potrzebnej im w danej chwili reguły. Założyć można, że o ile indeksator zainteresowany będzie poznaniem (nauczeniem się) całej instrukcji, ponieważ będzie z niej korzystał w swojej pracy wielokrotnie, o tyle zwykły użytkownik nie uzna tej wiedzy za przydatną (potrzebną). W przypadkach wątpliwych najprawdopodobniej prowadzi będzie poszukiwania "po omacku", metodą prób i błędów,

a ponieważ błędów może być dużo, często odejście od systemu sfrustrowany, zniechęcony do korzystania z niego w przyszłości, uznając go za "nieдобry" (co należy rozumieć jako "za trudny") lub uważając, że nie ma w zbiorze wyszukiwawczym żadnych dokumentów relewantnych (zjawisko ciszy informacyjnej). Po instrukcję indeksowania, można przypuszczać, nie sięgnie.

Umieszczenie poszczególnych zaleceń przy deskryptorach, których dotyczą zwolniłoby użytkownika z konieczności studiowania całej instrukcji w celu odnalezienia interesującego go zagadnienia. Z drugiej strony, wprowadzenie zaleceń tego typu (i to w znacznej ich liczbie) zaciemni obraz słownictwa (schemat) tezaurusa, zwiększy jego i tak niemałą objętość i wprowadzi konieczność częstszej aktualizacji (szczegółowych reguł przybywa niewiele mniej niż nowych deskryptorów). Prosty wykaz deskryptorów i askryptorów, jakim jest obecna wersja tezaurusa, ma swoje zalety, szczególnie dla użytkownika korzystającego z systemu na co dzień (łatwość szybkiego odnajdowania potrzebnych terminów i powiązań między nimi), dlatego korzystnym rozwiązaniem byłoby utworzenie dwu wersji tezaurusa - skróconej (prostej) i rozbudowanej (wzbogaconej o instrukcje). Należy rozważyć możliwość utworzenia trzech niezależnych wykazów: prostego, indeksowego i wyszukiwawczego. Zależy to od tego, w jakim stopniu reguły indeksowania różnią się od reguł wyszukiwawczych, tj. na ile różniłaby się postać informacji w obu rozbudowanych wersjach (indeksowej i wyszukiwawczej). Porównania takiego dotychczas nie przeprowadzono, do prac tych potrzebny jest nie tylko zbiór wyszukiwawczy (który jest dopiero w trakcie tworzenia), ale i zbiór zapytań informacyjnych kierowanych do systemu oraz informacje na temat zachowań użytkowników, ich metod wyszukiwania, najczęściej popełnianych błędów, trudności, stopnia

przygotowania, itp. Te informacje można będzie uzyskać dopiero po uruchomieniu systemu i jego wykorzystywaniu przez użytkowników. Już teraz można jednak stwierdzić, że w części systematycznej tezaurusu niezbędne będzie umieszczenie przy deskryptorach wszystkich askryptorów kierujących do nich (z wyłączeniem różnych form gramatycznych, np. liczby mnogiej lub inwersyjnej postaci), askryptory te bowiem wskazują faktyczny zakres deskryptora, odbiegający czasami od jego zakresu znanego użytkownikowi z języka naturalnego.

Powyższe propozycje dotyczą wersji drukowanej tezaurusu (lub przynajmniej wydruków komputerowych); w jego wersji maszynowej problem może zostać rozwiązany przez wyświetlanie na ekranie tylko skróconej wersji tezaurusu, poszczególne instrukcje byłyby dostępne tylko na żądane, jako informacja pomocnicza dla użytkownika, po użyciu komendy typu HELP. Przy założeniu dwu wersji tych instrukcji (wyszukiawczej i indeksowej), w terminalach dla użytkowników wyświetlane byłyby tylko instrukcje wyszukiwania, w terminalach dla indeksatorów - tylko instrukcje indeksowania. Instrukcje wyszukiwania wyjaśniałyby znaczenie deskryptorów (zakres), ich użycie, podpowiadałyby użytkownikowi możliwe sposoby prowadzenia poszukiwań, wskazywały formę zapisu pytania przy wykorzystaniu algebry Boole'a, itp. Instrukcje te mogą być wielostopniowe, na różnych poziomach, w zależności od stopnia przygotowania użytkownika (stopnia znajomości systemu, metod wyszukiwawczych, budowania strategii itp.). Zgodne byłoby to z ogólną światową tendencją w zakresie wyszukiwania informacji i środków lingwistycznych temu służących, tj. zapewnienia jak największej "przyjazności" (user-friendliness) systemu. Dla instrukcji indeksowania rozróżnienie poziomów nie byłoby celowe.

Omawiane opracowanie jest projektem instrukcji. Instrukcja powstawała w trakcie procesu tworzenia faktycznych opisów deskryptorowych, stąd jej nierównomierna rozbudowa (precyzyjność, najwyraźniej widoczna w zaleceniach szczegółowych. Najpełniej opracowany jest dział informatyki, najmniej precyzyjnie - dział dziedzin pokrewnych, bibliografii i bibliotekarstwa. Spowodowane jest to faktem, że o ile dla artykułów z zakresu informatyki utworzono już ponad 1000 opisów (co dało wystarczający materiał do ustalenia stosunkowo szczegółowych reguł), o tyle z zakresu dziedzin pokrewnych nie opracowano jeszcze żadnych czasopism. Nieliczne zalecenia w działach takich, jak bibliotekarstwo, bibliografia i dziedziny pokrewne możliwe były do wprowadzenia dzięki występowaniu tych tematów (w znikomej ilości) w czasopismach z zakresu informacji naukowej już zaindeksowanych (APID i ZIN) lub dzięki analogicznym rozwiązaniom podobnych zagadnień występujących w regułach ogólnych instrukcji. Przewidywane jest dalsze uzupełnianie instrukcji zasadami szczegółowymi w miarę postępu prac nad tworzeniem zbioru wyszukiwawczego systemu WIST i poszerzania jego zakresu tematycznego (włączania nowych czasopism do indeksowanego zbioru).

Fakt tworzenia instrukcji w trakcie procesu indeksowania tłumaczy ewentualne niekonsekwencje w opisach dokumentów już zaindeksowanych znajdujących się w zbiorze wyszukiwawczym systemu WIST. Zmiana bowiem zasad w trakcie prac prowadzić mogła do zmian w opisach analogicznych (podobnych tematycznie) artykułów, a to z kolei do różnic w opisach powstałych na początku indeksowania i pod koniec, po sprecyzowaniu nowych zaleceń. Niekonsekwencje te będą w miarę możliwości eliminowane.

12.02.1990 r.

THE INSTRUCTION OF INDEXING TO THE "THESAURUS OF SCIENTIFIC
INFORMATION "

Summary

The article makes the survey of the instruction of indexing with using the vocabulary contained in the "Thesaurus of Scientific Information" (Pol. Tezaurus Informacji Naukowej - TIN).

There is discussed in detail the methodics of indexing (the form of the retrieval characteristics, the dept of indexing, coordination of descriptors, problem of ambiguity of retrieval characteristics, field descriptors, the process of indexing - the general and detailed indexing). There are also presented the general prescriptions concerning the use of geographic descriptors and of modifiers as well as detailed prescriptions about indexing. The present form of instruction and its assumed modification are characterized.

ИНСТРУКЦИЯ ИНДЕКСИРОВАНИЯ К ТЕЗАУРУСУ НАУЧНОЙ ИНФОРМАЦИИ

Резюме

Статья является обсуждением инструкции индексирования при использовании лексики содержимой в Тезаурусе научной информации (TIN).

Охарактеризована подробно методика индексирования (вид поисковой характеристики, подробность индексирования, координация дескрипторов, вопрос многозначности поисковых характеристик, отраслевые дескрипторы, процесс индексирования - целостное в детальное индексирование). Представлены подробные указания (географические дескрипторы, модификаторы), а также подробные указания относящиеся к индексированию. Обсуждается настоящая форма инструкции и предусматриваемые её модификации.

MATERIAŁY I PRZYCZYNKI

BARBARA SOSIŃSKA-KALATA

Instytut Bibliotekoznawstwa
i Informacji Naukowej UW

WYKORZYSTANIE TECHNIK SZTUCZNEJ INTELIGENCJI WE WSPÓŁCZESNYM BIBLIOTEKARSTWIE I SYSTEMACH WSPOMAGAJĄCYCH WYSZUKIWANIE ONLINE

Poszukiwanie możliwości wykorzystania technik sztucznej inteligencji i systemów ekspertowych w bibliotekarstwie i systemach wyszukiwania online. Systemy ekspertowe w pracach administracyjnych biblioteki. Systemy ekspertowe w tradycyjnych pracach bibliotecznych (katalogowanie, klasyfikowanie, indeksowanie, abstraktowanie). Systemy ekspertowe a szkolenie użytkowników. Techniki sztucznej inteligencji i systemy ekspertowe w wyszukiwaniu informacji w systemach online. Charakterystyka wybranych systemów wspomagających wyszukiwanie w systemach online.

Nadrzędnym celem funkcjonowania wszelkiego rodzaju serwisów informacyjnych i bibliotecznych, a więc także podstawowym zadaniem bibliotekarza i dokumentalisty jest ułatwienie użytkownikom dostępu do stale wzrastających zbiorów informacji i zapewnienie, aby otrzymywana przez nich informacja możliwie w najwyższym stopniu zaspokajała ich rzeczywiste potrzeby informacyjne. Rozwój tzw. technologii informacyjnej, czyli urządzeń i metod pracy wykorzystywanych przez bibliotekarzy i pracowników ośrodków informacji, zawsze zmierzał do ułatwienia wypełnienia tego zadania. Techniki

sztucznej inteligencji i systemy ekspertowe są jednym z ostatnich osiągnięć technologii informacyjnej, które otwierają znaczne możliwości poprawienia jakości usług informacyjnych, a także ułatwienia i usprawnienia pracy bibliotekarzy i dokumentalistów.

Z pierwszymi systemami ekspertowymi, które opracowano w latach sześćdziesiątych i siedemdziesiątych, wiązano ogromne, często jednak przesadne, nadzieje. Obecnie wokół tych systemów panuje już klimat bardziej realistyczny: większość ludzi pogodziła się z tym, że systemy ekspertowe nie mogą całkowicie wyeliminować ekspertów-ludzi i że nie są one panaceum na brak doświadczonych specjalistów w poszczególnych instytucjach czy organizacjach, czy też środkiem rekompensującym niskie kwalifikacje pracowników. Systemy ekspertowe przedstawia się i pojmuje obecnie - zgodnie z ich rzeczywistymi możliwościami - jako narzędzie wspomagające podejmowanie decyzji, które może udzielić rady czy informacji opartej na danych wprowadzonych do systemu i które przede wszystkim służy do rozwiązywania pewnych problemów praktycznych w określonych dziedzinach. Mogą więc one pomagać w podejmowaniu właściwszych i szybszych decyzji w określonych okolicznościach, zapewniają przy tym stosunkowo szeroką dystrybucję wiedzy praktycznej w danej specjalności w obrębie organizacji czy instytucji, w której są zainstalowane i wreszcie uwalniają wysoko wykwalifikowanych specjalistów od zadań trywialnych, rutynowych i często powtarzających się, a także zapewniają możliwość wykorzystania wiedzy ekspertów przy realizacji takich zadań, które są wyjątkowo uciążliwe a nawet niebezpieczne dla człowieka.

SYSTEMY EKSPERTOWE W BIBLIOTEKOWIEDZY I INFORMACJI NAUKOWEJ

Prace nad wykorzystaniem technik sztucznej inteligencji i systemów ekspertowych w pracach bibliotekarskich i dokumentacyj-

nych pojawiły się dopiero w ostatnich latach i właściwie trudno jeszcze mówić o jakimś ich rozpowszechnianiu. W większości efektem tych badań są projekty lub systemy eksperymentalne o bardzo wąskim zasięgu. Warto jednak odnotować, że badań takich podejmuje się coraz więcej. Prawie wszystkie szkoły bibliotekoznawstwa i informacji naukowej w Wielkiej Brytanii prowadzą prace nad sposobami wykorzystania systemów ekspertowych w bibliotekarstwie i serwisach informacyjnych. Badania te koncentrują się najczęściej wokół takich zadań, jak katalogowanie, klasyfikowanie, indeksowanie, wyszukiwanie informacji, obsługa bibliograficzna czytelników itp. /10/.

Współcześnie eksplorowane możliwości zastosowania systemów ekspertowych w bibliotekoznawstwie i informacji naukowej można najogólniej podzielić na trzy dziedziny:

1. prace administracyjne i podejmowanie decyzji związanych z zarządzaniem biblioteką,
2. rutynowe a zarazem czasochłonne prace biblioteczne,
3. udoskonalenie obsługi informacyjnej użytkowników.

Administracja i zarządzanie biblioteką

Na rynkach zachodnich dostępnych jest obecnie wiele małych i stosunkowo tanich pakietów programów przeznaczonych do tworzenia systemów ekspertowych na komputerach osobistych, które mogą być wykorzystane w takich pracach jak planowanie biblioteczne, selekcja czasopism, prowadzenie wypożyczania międzybibliotecznego, kontrola zbiorów czy statystyka biblioteczna. W najbliższym czasie zapowiedziana jest publikacja opracowanego w Department of Library and Information Studies w Loughborough międzynarodowego wykazu pakietów oprogramowania dla systemów ekspertowych, które można wykorzystywać w bibliotekach i serwisach informacji dokumentacyjnej /10/.

Rutynowe prace biblioteczne

Analizy wydatków bibliotek, które w ostatnich latach szczególnie licznie były przeprowadzane w bibliotekach angielskich, prowadzą do wniosku, że 60% budżetu biblioteki przeznacza się na opłacenie pracy personelu. Zdecydowaną większość czasu pochłaniają tradycyjne prace biblioteczne, takie jak katalogowanie materiałów bibliecznych, klasyfikowanie czy indeksowanie, tworzenie adnotacji czy abstraktów itp. Można przypuszczać, że wnioski te prawdziwe są również w odniesieniu do sytuacji bibliotek w innych krajach. Warto przy tym zwrócić uwagę na to, że większość tych tradycyjnych prac bibliecznych wymaga dużej koncentracji, dokładności i kwalifikacji zawodowych, a jednocześnie jest bardzo czasochłonna, w wysokim stopniu zrutynizowana, pełna powtarzających się operacji, a co za tym idzie, nużąca. Opracowywane w Wielkiej Brytanii i Stanach Zjednoczonych projekty systemów ekspertowych pozwalają zminimalizować ujemne cechy podstawowych operacji bibliecznych, uwalniając pracowników od powtarzalności działań.

Kilka projektów eksperymentalnych systemów ekspertowych zmierza do wspomagania operacji katalogowania zgodnie z anglo-amerykańskimi regułami katalogowania alfabetycznego (AACR-2), np. projekty opracowane w uniwersytetach w Exeter, Linköping i Strathclyde. Na uniwersytecie w Strathclyde prowadzone są też badania nad systemem ekspertowym w zakresie klasyfikowania, a w Bradford University przygotowano system ekspertowy zajmujący się kontrolą duplikatów. Wykorzystanie systemów ekspertowych w indeksowaniu i abstraktowaniu przyciąga uwagę także wielu wydawców i właścicieli systemów online. Najbardziej znanym systemem ekspertowym w zakresie indeksowania jest system opracowany w

amerykańskiej Narodowej Bibliotece Medycznej (National Library of Medicine), który współpracuje z indeksatorami systemu MEDLINE. Systemy ekspertowe w zakresie abstraktowania są bodaj najciekawsze i w najmniejszym stopniu opisane w literaturze. Obecnie nad takim właśnie systemem prowadzone są prace w Lanchester Polytechnic /10/.

Udoskonalanie obsługi informacyjnej użytkowników

Każdy, kto kiedykolwiek miał do czynienia z systemem wyszukiwania informacji online, czy to jako jego bezpośredni (tzw. "końcowy") użytkownik, czy też jako dokumentalista go obsługujący, zdaje sobie sprawę z konieczności podniesienia poziomu jakości tego rodzaju serwisów informacyjnych. Autentycznie efektywne korzystanie z systemu online wymaga przełamania szeregu trudnych barier, a to z kolei wymaga specyficznej wiedzy o konstrukcji sieci komputerowych, baz danych opisów dokumentów, języków informacyjno-wyszukiwawczych itp. Wymaganie tej wiedzy od użytkowników bezpośrednich byłoby podejściem nieracjonalnym i sprzecznym z ogólnym założeniem, że doskonalenie dostępu do informacji powinno być przede wszystkim realizowane przez doskonalenie samych systemów informacyjnych i ułatwianie kontaktu z nimi użytkownikom końcowym. Zwykle wiedzą o konstrukcji systemu informacyjnego oraz jej wpływie na proces wyszukiwania dysponują pracownicy systemów obsługujący użytkowników bezpośrednich, jednak ze względów zarówno ekonomicznych, jak i psychologicznych dąży się do wyeliminowania ich pośrednictwa w komunikacji między użytkownikiem i systemem. Tak więc właśnie ułatwianie kontaktu użytkowników z systemami online staje się szczególnie ważnym i interesującym polem wykorzystania systemów ekspertowych /2/.

Wśród projektów systemów ekspertowych, których zadaniem jest podniesienie jakości obsługi informacyjnej można wyróżnić

trzy główne grupy: systemy ułatwiające dostęp do źródeł informacyjnych przez przede wszystkim upraszczanie procedury połączenia z systemem, wyboru bazy danych, a także pomoc w tworzeniu strategii wyszukiwawczej, systemy doradcze oferujące informację praktyczną w określonych dziedzinach i systemy, które uczą właściwego korzystania z bibliotek i ich serwisów informacyjnych.

Ułatwianie dostępu do źródeł informacji, a więc upraszczanie procesu wyszukiwania informacji, jest bodaj najlepiej opisaną w literaturze dziedziną wykorzystania technik sztucznej inteligencji i systemów ekspertowych w bibliotekarstwie i informacji naukowej. Jej szczególnie intensywna w ostatnich latach eksploatacja wiąże się bezpośrednio z ideą budowania tzw. systemów "przyjaznych użytkownikowi" (user-friendly systems). Najczęściej przyjmują one postać systemów wyposażonych w specjalny fragment oprogramowania, zwany "inteligentnym urządzeniem kontaktu z użytkownikiem" (intelligent front end), który umieszczony jest pomiędzy użytkownikiem i innym pakietem programów realizujących wyszukiwanie informacji, a zadaniem jego jest ułatwienie i zarazem optymalizacja korzystania z tego pakietu programów. Tego rodzaju systemy przedstawimy szczegółowo w dalszej części artykułu, korzystając z najbardziej znanych przykładów.

Użytkownicy bibliotek i systemów informacyjnych nie zawsze potrzebują informacji bibliograficznej czy katalogowej, lecz niekiedy szukają informacji umożliwiającej im rozwiązanie pewnych konkretnych problemów. Tak więc nie tylko systemy informacji skierowującej, ale również faktograficzne systemy ekspertowe znajdują wykorzystanie we współczesnym bibliotekarstwie. Opracowane dotychczas systemy ekspertowe przeznaczone są do obsługi informacyjnej bardzo wąskich dziedzin czy szczegółowych problemów,

np. system PLEXUS doradza w zakresie ogrodnictwa, a system RESEDA odpowiada na pytania dotyczące istniejących baz danych o francuskiej historii mediewistycznej /10/. Wysoce specjalistyczne systemy ekspertowe obsługują użytkowników bibliotek naukowych lub fachowych, czy specjalistycznych, przewiduje się jednak również możliwość wykorzystania tego typu małych systemów ekspertowych w bibliotekach o bardzo zróżnicowanych kategoriach użytkowników, takich jak biblioteki publiczne. Mogłyby one obsługiwać pytania informacyjne dotyczące np. konserwacji samochodu, budżetu domowego, zasad pisania podań czy listów urzędowych, zasad racjonalnego odżywiania się itp.

Tworzenie ciekawych, łatwych w korzystaniu i niezmiernie pożytecznych komputerowych przewodników czy instrukcji dotyczących posługiwania się różnymi serwisami informacyjnymi biblioteki, korzystania z jej zbiorów, lokalizacji poszczególnych jej jednostek organizacyjnych, jest kolejną sferą współczesnego bibliotekarstwa, w której znajdują zastosowanie systemy ekspertowe. Tzw. "inteligentny katalog" używany jest już w niektórych bibliotekach w USA w postaci systemu o nazwie BiblioFile, korzystającego z technik sztucznej inteligencji i automatycznej analizy tekstu. BiblioFile kontaktuje się z użytkownikiem za pomocą generowanego ludzkiego głosu opowiadając o tym, jak posługiwać się katalogiem i pomagając w wyborze lektury do zgłoszonego przez czytelnika tematu /10/.

Wyszukiwanie informacji w systemach online

Analizując proces wyszukiwania informacji można wyodrębnić w nim kilka kolejnych faz, na które składają się pewne operacje o charakterze mechanicznym i pewne operacje o charakterze intelektualnym. Donald T. Hawkins charakteryzuje wyszukiwanie informacji jako sekwencję dziesięciu operacji /5/:

operacje mechaniczne

- 1/ połączenie z siecią komputerową,
- 2/ przekazanie hasła (password),
- 3/ wybór bazy danych,
- 4/ wyświetlenie wyników,
- 5/ polecenie wydruku wyników (offline).

operacje intelektualne

- 1/ konceptualizacja tematu wyszukiwania,
- 2/ utworzenie instrukcji wyszukiawczej,
- 3/ utworzenie strategii wyszukiwania.
- 4/ interpretacja wyników wyszukiwania,
- 5/ analiza, ocena i uzupełnienie wyników wyszukiwania.

Dotychczasowe działania zmierzające do ułatwienia wyszukiwania online koncentrowały się przede wszystkim na operacjach mechanicznych. Oprogramowania PC, takie jak np. Pro-Search czy systemy typu EasyNet zapewniają maksymalne ułatwienie połączenia z siecią komputerową czy przekazania hasła. EasyNet pozwala automatycznie wybrać najbardziej użyteczną dla zgłoszonego tematu wyszukiwania bazę danych. Cała więc seria operacji prowadzących do rozpoczęcia rzeczywistego wyszukiwania odbywa się tutaj bez ingerencji użytkownika. Niewątpliwie opracowanie tego typu środków upraszczania procedury wyszukiwania w systemach online stanowi duży postęp, jednakże nie rozwiązuje problemów, które według opinii wielu specjalistów, potwierdzonej przez liczne badania empiryczne, są dla użytkowników największym utrudnieniem posługiwania się dialogowymi systemami komputerowymi. Jak wykazały m.in. badania przeprowadzone w 1980 roku przez Carol Fenichel, zarówno dla doświadczonych, jak i dla niedoświadczonych użytkowników systemów online największe problemy wiążą się właśnie z prawidłowym wykonaniem koniecznych operacji intelektualnych. Znacznie trudniejsze od nauczania się np. języka komend systemu jest przyswojenie sobie zasad tworzenia strategii wyszukiwania informacji i właściwe sformułowanie instrukcji wyszukiwaw-

czej /2/. Wyszukiwanie w systemach online, zdaniem Daniela Grossa, wymaga zdolności kojarzenia informacji, umiejętności przewidywania sposobu jej reprezentacji językowej i lokalizacji w strukturze bazy danych /3/. Wyszukiwany temat rzadko można odnaleźć korzystając wyłącznie z najprostszego i bezpośredniego sposobu jego wyrażenia, zwykle jest on rozbity na pojęcia prostsze. Formułowanie instrukcji wymaga więc selekcji właściwego zespołu terminów reprezentujących poszczególne pojęcia, po której następuje konieczność zbudowania poprawnej i odpowiadającej potrzebom wyszukiwania formuły Boole'owskiej. Zasady konstruowania instrukcji wyszukiwawczej i planowania optymalnej strategii wyszukiwawczej są ściśle związane ze strukturą kartotek, w których wyszukiwanie jest przeprowadzane. W systemach online wyszukiwanie odbywa się za pomocą kartoteki inwersyjnej (odwróconej), czyli listy wszystkich wyrazów, mogących pełnić funkcję kluczy wyszukiwawczych (np. słów kluczowych czy deskryptorów), obok których umieszczony jest wykaz tzw. adresów, zwykle numerów rekordów w kartotece prostej, w których dane wyrażenie występuje. Kartoteka inwersyjna jest więc swego rodzaju słownikiem, który może być badany przez wyszukującego dzięki użyciu odpowiedniej komendy, np. w systemie DIALOG komenda **EXPAND** powoduje wyświetlenie na ekranie pewnego fragmentu kartoteki inwersyjnej, przy czym obok każdego terminu (deskryptora czy słowa kluczowego) podana jest liczba dokumentów, w których opisach termin ten występuje.

Kiedy użytkownik wprowadza do systemu pewien deskryptor czy słowo kluczowe, system przegląda kartotekę inwersyjną, aby stwierdzić, czy ciąg znaków składających się na wprowadzone wyrażenie pokrywa się z którymś z wyrazów umieszczonych w kartotece. Jeśli tak jest, zbiór adresów zostaje przez system sformułowany i informacja,

zwykle o ich liczbie, wysłana jest do wyszukiującego. Jeśli nie stwierdzi się zgodności formalnej wyrażenia wprowadzonego z żadnym wyrażeniem w kartotece inwersyjnej, użytkownik otrzymuje tzw. informację zerową, czyli wiadomość, że system nie posiada żadnego zapisu poszukiwanego wyrażenia. Tak więc, podstawą wyszukiwania, gdy odrzuci się zasady reagowania systemu na operatory logiczne i komendy wyszukiwawcze, jest porównywanie i stwierdzanie zgodności form wyrażeń wprowadzonych przez wyszukiującego i form wyrażeń umieszczonych w kartotece inwersyjnej. Jeśli wyrażenia wprowadzone nie będą odpowiadać żadnemu z wyrażeń w kartotece, wyszukiwanie zakończy się niepowodzeniem. Bazy danych obfitują jednak w błędy - częste przestawienia liter, różne formy ortograficzne zarówno poprawne, jak i niepoprawne tych samych terminów itp. Niektóre systemy próbują radzić sobie z błędnymi zapisami tworząc automatyczne słowniki najczęstszych błędów - takie rozwiązanie stosują np. biblioteczne katalogi online. W większości systemów online jednak zarówno poprawne, jak i błędne zapisy tego samego terminu wprowadzane są odrębnie do kartoteki inwersyjnej i traktowane jak całkowicie niezależne wyrażenia (np. w systemie INSPEC). Oznacza to, że jeśli formułujący instrukcję wyszukiwawczą nie przewidzi możliwości użycia pewnej błędnej formy danego wyrażenia, to nie odnajdzie tych informacji, które w systemie właśnie tą błędną formą są reprezentowane. Błędne zapisy danych są główną przyczyną, ze względu na którą konieczne jest korzystanie z truncacji (truncation), czyli maskowania, zasłaniania tych części wyrażenia, w których z dużym prawdopodobieństwem może być popełniony błąd, czy też które bywają różnie zapisywane w różnych konwencjach ortograficznych. Jest to niewątpliwie skuteczna technika eliminowania ciszy informacyjnej wynikłej z błędnych zapisów.

jednakże badania wykazują, że niewielu użytkowników potrafi ją stosować, co więcej - niewielu w ogóle uświadamia sobie, że maskowanie to nie sprytny chwyt techniczny, lecz obecnie bezwzględna konieczność bez której niemożliwe jest nawet względnie kompletne wyszukiwanie.

Wszystkie współcześnie funkcjonujące systemy wspomagające wyszukiwanie online, takie jak tzw. automatyczny logon, automatyczna selekcja baz danych, systemy typu front-end, ułatwiają kontakt użytkownika z systemem jedynie do momentu, kiedy należy rozpocząć rzeczywiste wyszukiwanie. Gdy na ekranie pojawi się polecenie ENTER YOUR SEARCH TOPIC (wprowadź swój temat wyszukiwania), użytkownik zdany jest właściwie wyłącznie na własne siły. Wprawdzie systemy wspomagające, takie jak wcześniej już wspomniany EasyNet czy Pro-Search oferują tzw. help-screens, czyli wyświetlane na ekranie monitora instrukcje pomocnicze dotyczące ogólnych zasad wyboru wyrażień, które mają reprezentować poszukiwany temat, jednakże taka metoda pomocy jest dyskusyjna w odniesieniu do bardzo skomplikowanych tematów najczęściej stanowiących przedmiot wyszukiwania. Doświadczenie wykazuje, że większość użytkowników w ogóle nie czyta tych instrukcji pomocniczych. Użytkownik potrzebuje pomocy innego rodzaju: takich, które ułatwiłyby mu wybór właściwego wyrażenia reprezentującego interesujący go temat, takich, które podpowiedziałyby mu w jaki sposób należy połączyć te wyrażenia w Boole'owską strukturę instrukcji wyszukiwawczej, takich, które wskazałyby, kiedy należy użyć techniki maskowania, takich, które prowadziłyby go poprzez gąszcz istniejących baz danych, wybierając najwłaściwszą, lokalizując w niej szukane dane i kierując je do użytkownika. Aby system mógł realizować tego rodzaju zadania, konieczne jest wyposażenie go w techniki sztucznej in-

teligencji. Poniżej przedstawionych zostanie trzynaście eksperymentalnych systemów wspomagających wyszukiwanie online właśnie w sferze operacji intelektualnych. Jedne z nich nazywane są przez autorów projektów systemami ekspertowymi, inne - systemami z pewnymi elementami systemów ekspertowych, jeszcze inne wreszcie - systemami wyposażonymi w tzw. programowy sprzęg do komunikacji w języku naturalnym (natural language interface). Wspólną cechą ich jest jednak to, że zmiierzają do ułatwienia użytkownikowi tworzenia efektywnych strategii wyszukiwania i formułowania instrukcji wyszukiwawczych /5/.

CHARAKTERYSTYKA WYBRANYCH SYSTEMÓW WSPOMAGAJĄCYCH WYSZUKIWANIE ONLINE

Wybór omówionych poniżej systemów dokonany został w oparciu o przeglądową pracę Donalda T. Hawkinsa /5/. Ponieważ szczegółowe charakteryzowanie budowy przedstawianych systemów jest niemożliwe w ramach niniejszego artykułu, przy uwagach o poszczególnych projektach podane odsyłacze do publikacji, w których znajdują się szczegółowe ich opisy. >

Answerman

Answerman jest gotowym systemem pomocniczym używanym w amerykańskiej National Agriculture Library /16/, który opracowano wykorzystując jeden z komercyjnych pakietów oprogramowania do projektowania systemów ekspertowych. Posiada on bazę wiedzy, do której wprowadzono dane zaczerpnięte z kilku podstawowych informatorów z zakresu nauk rolniczych. Techniki łańcuchowe Answermana prowadzą użytkownika poprzez serię coraz bardziej szczegółowych opcji (menu), aż do momentu odnalezienia poszukiwanej infor-

macji. Ciekawą cechą Answermana jest to, że może połączyć użytkownika bezpośrednio ze standardowymi faktograficznymi bazami danych z zakresu nauk rolniczych, np. z bazą AGRICOLA czy Superindex.

CANSEARCH

CANSEARCH/13/ jest systemem ekspertowym opartym na regułach produkcji, napisanym w PROLOG-u. Dzięki wykorzystaniu ekranu reagującego na dotyk, użytkownik nie musi nawet wiedzieć, jak korzystać z klawiatury komputera. System dysponuje pakietem programów do komunikacji z użytkownikiem sterowanych poprzez wybranie opcji z wyświetlanego menu: dotyk użytkownika przerywa matrycę promieni świetlnych rozmieszczonych na ekranie i w ten sposób dane wprowadzane są do systemu. Podobnie jak w innych systemach ekspertowych zakres tematyczny obsługiwany przez CANSEARCH jest ściśle określony i dosyć wąski - ograniczony do problematyki terapii raka, przy czym sam system przeznaczony jest do współpracy z jedną tylko bazą danych, tj. z MEDLINE. Wyniki przeprowadzonego wstępnego testu systemu CANSEARCH, w którym uczestniczyli zarówno lekarze, jak i specjaliści w wyszukiwaniu online, zostały bardzo wysoko ocenione. System budował dobre instrukcje wyszukiwawcze dostosowane do zasad obowiązujących w MEDLINE i prawidłowo przeprowadzał wyszukiwanie, dzięki czemu wyniki wyszukiwania wspomaganego przez CANSEARCH i przeprowadzanego przez niedoświadczonego użytkownika - lekarza były niekiedy lepsze niż wyniki wyszukiwania dokonanego przez obsługę MEDLINE. CANSEARCH uważany jest za jeden z najciekawszych i najbardziej obiecujących przykładów efektywnego wykorzystania systemów ekspertowych w systemach wyszukiwania informacji online.

Ekspertowy system konsultacyjny P. Shoala

Shoval /14/ opracował projekt systemu, którego zadaniem jest pomoc przy formułowaniu instrukcji wyszukiwawczych. W rzeczywistości system ten nie konstruuje samodzielnie instrukcji nie dokonuje podziału procedury wyszukiwania zgodnie z pewną strategią - nadal konieczna jest tutaj aktywna współpraca z człowiekiem. System Shoala tworzy natomiast sieć semantyczną dla wyrażeń wprowadzanych przez użytkownika, a następnie przeszukuje własną bazę wiedzy w celu odnalezienia wyrażeń, którymi wyszukiwający powinien rozszerzyć swoją instrukcję wyszukiwawczą. Jeśli użytkownik odrzuci sugerowane przez system wyrażenia, system ponownie analizuje zbudowaną sieć semantyczną i przeszukuje bazę wiedzy, aby znaleźć i zaproponować kolejne wyrażenia alternatywne lub modyfikujące znaczenie terminu pierwotnie przez użytkownika zgłoszonego. Wyrażenia zaakceptowane przez użytkownika system może rozwinąć, np. przez wyliczenie fraz, w których wyrażenia te mogą wystąpić, uzupełnić synonimami lub bliskoznacznikami, czy też podać terminy z nimi łączące się na mocy różnego rodzaju skojarzeń. Na przykład jeśli wyszukiwający wprowadzi termin DATA (dane), jedną z sugestii systemu będzie wyrażenie DATABASE (baza danych), co następnie może być rozwinięte do postaci fraz DATABASE MANAGEMENT (zarządzanie bazą danych) i DATA MANAGEMENT (zarządzanie danymi). System Shoala może współpracować ze skomplikowanymi, bardzo rozbudowanymi sieciami semantycznymi, natomiast jego komunikowanie się z użytkownikiem opiera się obecnie na bardzo prostych programach umożliwiających jedynie sekwencyjne wprowadzanie pojedynczych terminów przez użytkownika. W przyszłości przewiduje się wyposażenie systemu w bardziej rozbudowany i dający większe możliwości pakiet programów do komunikacji z użytkownikiem w języku naturalnym.

ORA

Opracowany przez Jamesa Parrota system ORA (Online Reference Assistance) /11/ realizuje podobne funkcje jak system Answerman: także tu użytkownik odpowiada na kolejno stawiane przez system pytania, a system wykorzystując techniki łańcuchowe i otrzymywane od użytkownika dane kieruje go do właściwego źródła informacji. ORA jest systemem doradzającym odpowiedni wybór wydawnictwa informacyjnego dla tematyki interesującej użytkownika. Dane o informato-
rach, ich zakresie, uwzględnionych w nich tematach pokrewnych oraz o ich stopniu trudności stanowią tzw. informacyjny moduł bazy wiedzy ORA. Obok modułu informacyjnego baza wiedzy systemu zawiera jeszcze dwa moduły wspomagające: moduł instruktażowy i moduł strategii wyszukiwawczej. Moduł instruktażowy oferuje użytkownikom informacje dotyczące usług biblioteki, w której system jest zainstalowany, a więc podpowiada, gdzie należy szukać poszczególnych agend biblioteki, prezentuje na mapach i schematach położenie różnych oddziałów biblioteki, lokalizację księgozbiorów, serwisów informacyjnych, a także informuje o innych usługach dostępnych na terenie campusu University of Waterloo. Moduł strategii wyszukiwawczej ma za zadanie współpracować z użytkownikiem przy tworzeniu instrukcji i strategii wyszukiwania prowadząc z nim pewnego rodzaju dialog, w trakcie którego użytkownik nakłaniany jest do możliwie precyzyjnego sformułowania poszukiwanego tematu. Powstająca w ten sposób strategia wyszukiwania jest stale wyświetlana na ekranie monitora, tak aby użytkownik mógł samodzielnie i na bieżąco oceniać podejmowane przez system decyzje. System ORA pełni więc funkcje podobne do tych, które realizuje pracownik oddziału informacji w bibliotece - zwany w Anglii i USA "reference librarian", od której to nazwy pochodzi nazwa samego systemu.

ORA utrzymywany jest na minikomputerze typu VAX, do którego podłączone są liczne terminale powszechnie dostępne na terenie biblioteki University of Waterloo. Do publicznego użytku system przekazano w marcu 1987 roku.

PLEXUS

System PLEXUS jest przykładem systemu stricte ekspertowego, oferującego bezpośrednio informację z dziedziny ogrodnictwa, który zainstalowano w jednej z londyńskich bibliotek publicznych. Został opracowany w University of London, napisany jest w języku Turbo-Pascal i utrzymywany na mikrokomputerze /15/. PLEXUS jest systemem opartym na regułach produkcji, jego baza wiedzy zawiera ponad 1000 reguł umożliwiających interpretację wprowadzanych przez użytkownika wyrażeń naturalnych i generowanie odpowiedzi systemu. Dostęp do niego zapewnia wprowadzenie wyrażeń naturalnych o dowolnej formie, np. ciągu wybranych terminów, frazy oznaczającej poszukiwany temat czy też pełnego zdania. Dane wejściowe użytkownika poddane są w systemie analizie, w wyniku której zostają wyodrębnione tzw. słowa znaczące (istotne z punktu widzenia wyszukiwania). Słowa te następnie wykorzystane są w automatycznie generowanej instrukcji wyszukiwawczej w postaci wyrażenia Boole'owskiego. Ponad 120 reguł wykorzystywanych w PLEXUSie przeznaczonych jest do identyfikacji znaczenia wyrażeń polisemicznych. Jeżeli pytanie użytkownika nie może być w pełni zinterpretowane (zrozumiane) przez system, PLEXUS konsultuje się z użytkownikiem stawiając mu szereg szczegółowych pytań, dzięki którym zostaje pozyskana dodatkowa informacja stanowiąca podstawę prawidłowego zinterpretowania pytania zanim system rozpocznie przeszukiwanie swojej bazy wiedzy.

Prototyp systemu PLEXUS był z powodzeniem testowany zarówno przez bibliotekarzy zatrudnionych w bibliotekach publicznych, jak i przez czytelników. Przewiduje się dalsze rozbudowanie systemu, m.in. przez wyposażenie go w dodatkowy pakiet programów umożliwiających samoistne uczenie się systemu dzięki wykorzystaniu informacji dostarczanych przez użytkowników. *

System informacji bibliotecznej wspomagany tezauresem

Micco i Smith opracowali projekt systemu, którego zadaniem jest odnajdywanie wydawnictw informacyjnych odpowiednich dla poszukiwań informacji dotyczących zgłoszonego przez użytkownika tematu /9/. Projekt nosi nazwę Ready Reference Thesaurus System, a powstał na podstawie analizy niepowodzeń wyszukiwania informacji za pomocą systemu Search Helper. Obserwując przeprowadzane przez studentów wyszukiwanie, Micco i Smith stwierdzili, że częstym powodem nieodnalezienia szukanej informacji jest mała wiedza studentów o źródłach informacji, toteż użytecznym narzędziem wspomagającym wyszukiwanie może być system podpowiadający, z jakich źródeł najlepiej korzystać, aby trafić do tej informacji, która interesuje szukającego. Ready Reference Thesaurus System jest więc urządzeniem, które ma w pewnym sensie zastąpić bibliotekarza działu informacji - realizuje podobne zadania, jak moduł informacyjny wcześniej opisanego systemu ORA.

Budując swój system Micco i Smith wykorzystali powszechnie dostępne komercyjne pakiety programów przeznaczone dla systemów ekspertowych - MicroPROLOG i TurboLightning. Szczególną cechą systemu jest wyposażenie go w tezaurs online, wykorzystywany w trakcie wyszukiwania w celu ustalenia właściwych form wyrażen reprezentujących potrzebną informację. Pakiet programów zapewniających komunikację z użytkownikiem (interfejs system) prowadzi

go od chwili wprowadzenia pierwszego słowa kluczowego, przy czym na każdym etapie pracy wyszukiwający może odwołać się do pomocy tezaurusu. Po wybraniu wyrażen określających temat poszukiwań, użytkownik może wprowadzić tzw. ograniczniki zawężające w miarę potrzeby pole wyszukiwania, a następnie może sprawdzić menu bazy danych, do której kieruje go system. Projekt Micco i Smitha przewiduje możliwość ciągłej obserwacji przez użytkownika toku rozumowania przeprowadzonego przez system, tj. wyświetlanie wyników podejmowanych przez system kolejnych decyzji oraz reguł, na podstawie których zostały podjęte. Opcja ta zapewnia dużą swobodę działania użytkownika, konsultacji i kontroli wyszukiwania.

RUBRIC

System RUBRIC (RUle-Base Retrieval of Information by Computer) określany jest przez swoich twórców jako system z bazą wiedzy wykorzystujący zbiór reguł, na podstawie których budowane jest heurystyczne drzewo decyzyjne zawierające wzorce terminów wyszukiwawczych /7/. System ten może po dokonaniu wyboru wyrażen reprezentujących pewien temat nadać im odpowiednie wagi, zgodnie ze stopniem ich istotności dla poszukiwanego tematu, jeśli jest on wskazany przez użytkownika. Ponadto RUBRIC wyposażony jest w tzw. reguły modyfikacji, pozwalające automatycznie ustalić synonimy wykorzystywanych w wyszukiwaniu wyrażen, a także zidentyfikować użyte w określonym kontekście znaczenie wyrażenia homonimicznego. Podobnie jak w przypadku poprzednio omówionego systemu Micco i Smitha, także RUBRIC na żądanie użytkownika może zademonstrować reguły, którymi posłużył się w celu odnalezienia informacji na określony temat.

System RUBRIC testowany był w symulowanym wyszukiwaniu w pełnotekstowych systemach online, dotychczas jego działania nie skonfrontowano z rzeczywistymi warunkami wyszukiwania informacji.

ARGON

ARGON /12/ jest systemem reprezentacji wiedzy wykorzystującym zasady organizacji informacji w tzw. ramach (frame-based knowledge representation system). Współpracuje on z systemem typu faktograficznego KANDOR, który zawiera bazę wiedzy i mechanizm wnioskowania opracowany dla procedur wyszukiwania informacji. Działanie ARGON prezentowane było na przykładzie wyszukiwania w bazie danych o członkach pewnej organizacji, w której umieszczone były takie informacje, jak numer organizacyjny, numer telefonu, zainteresowania badawcze, zwierzchnicy w hierarchii organizacyjnej, związki rodzinne itp. W trakcie tworzenia instrukcji wyszukiwawczej na ekranie monitora ARGON wyświetla przebieg przeprowadzanego przez siebie wyszukiwania, tj. zasady na podstawie których podejmuje określone decyzje oraz wyniki realizacji tych decyzji. Jeśli jest to konieczne, system prosi użytkownika o dodatkowe informacje. Dzięki takiemu właśnie ciągiemu pokazywaniu zasad rozumowania prowadzących do zbudowania optymalnej instrukcji wyszukiwawczej system spełnia równocześnie ważną funkcję instruktażową.

ARGON nie był dotychczas testowany w żadnej bazie danych bibliograficznych, zdaniem jego autorów jednak nie ma powodu przypuszczać, że zastosowanie go w tego typu systemach wyszukiwania informacji online byłoby niemożliwe /5/. ARGON i KANDOR napisane są w języku LISP i wykorzystują takie techniki sztucznej inteligencji, jak zasady inferencji i wyszukiwania łańcuchowe (chaining). W dalszym ciągu prowadzone są prace nad rozbudową i udoskonaleniem obu tych systemów.

CONIT i IIDA

CONIT (COnnector for Networked Information Transfer) /6/ i IIDA (Individualized Instruction for Data Access) /8/ są systemami typu front-end przeznaczonymi dla systemów online, wykorzystującymi podstawowe oprogramowanie i utrzymywanymi na tym samym komputerze centralnym. CONIT może współpracować z systemami DIALOG, ORBIT i NLN, zaś IIDA umożliwia jedynie dostęp do systemu DIALOG. CONIT opracowano w Massachusetts Institute of Technology (MIT), gdzie w dalszym ciągu prowadzone są prace nad jego rozbudową, natomiast IIDA utworzona została w Drexel University i obecnie nie funkcjonuje. Systemy te monitorują współpracę z użytkownikiem, analizują podejmowane przez niego decyzje i w przypadkach, kiedy stwierdzą, że użytkownik potrzebuje pomocy - proponują określone rozwiązania zaistniałego problemu.

Zarówno CONIT, jak i IIDA w tworzeniu instrukcji wyszukiwawczych wykorzystują logikę Boole'a: Temat wyszukiwania zgłoszony przez użytkownika dzielony jest na elementy proste, czyli zostają w jego strukturze wyodrębnione tzw. pojęcia podstawowe, tj. terminy reprezentujące poszczególne składniki kompleksowego tematu, które następnie za pomocą operatorów łączone są zgodnie z pewną strategią wyszukiwania. Pomocnicze instrukcje wyświetlane na monitorze podpowiadają użytkownikowi, jak poprawnie należy sformułować odpowiednie wyrażenie Boole'owskie. Po utworzeniu wstępnej instrukcji wyszukiwawczej i przesłaniu jej do bazy danych, w której przeprowadzane jest wyszukiwanie, CONIT może sugerować modyfikacje tej instrukcji, jeśli użytkownik zażąda poszerzenia lub zawężenia pola wyszukiwania. IIDA natomiast analizuje komendy wprowadzane przez użytkownika w trakcie wyszukiwania i w przypadkach, gdy użytkownik zaczyna nadmiernie powtarzać te same polecenia, czy

też wykazuje inne zachowania świadczące o tym, że konieczną jest pomoc, proponuje określone strategie.

IT

IT (Information Transfer) jest brytyjskim systemem typu front-end, opracowanym przez Philipa Williama na podstawie jego badań nad automatyzacją dostępu do systemów informacyjnych i baz danych /17/. Powstał on w dwóch etapach - w pierwszym etapie Williams skonstruował model systemu ułatwiającego połączenie z systemem informacyjnym' (a więc zapewniającego automatyczny logon), który wzbogacił wykorzystującym reguły produkcji systemem ekspertowym ustalającym poziom szczegółowości i kompletności wyszukiwania, zgodnie z danymi wprowadzonymi przez użytkownika. W drugim etapie /18/ Williams rozwinął swój model zapewniając automatyzację całego procesu wyszukiwania. Pełny system IT umożliwia więc automatyczne połączenie z systemem informacyjno-wyszukiwawczym, automatyczny wybór odpowiedniej dla zgłoszonego tematu bazy danych oraz ułatwia pomocniczymi instrukcjami i informacjami o słownictwie, które użytkownik może wykorzystać tworząc instrukcję wyszukiwawczą, proces formułowania strategii wyszukiwania i odpowiednich wyrażeń Boole'owskich. IT nie posiada żadnej własnej bazy wiedzy, wykorzystuje natomiast zalecenia producenta bazy danych dotyczące łączenia wyrażeń, które mają być zastosowane w wyszukiwaniu.

IT zapewnia dostęp do systemu ESA-IRS i dla ułatwienia wyboru wyrażeń, które można zastosować przy wyszukiwaniu informacji na określony temat, wykorzystuje komendę tego systemu 200M, pozwalającą generować uporządkowaną według stopnia istotności dla danego tematu listę wyrażeń należących do kontrolowanego słownictwa języka informacyjnego systemu. Lista ta prezentowana jest użytkownik-

kowi, którego prosi się o wskazanie spośród proponowanych wyrażen jego zdaniem najlepiej charakteryzujących interesujący go temat. Wybrane wyrażenia wprowadzane są do instrukcji wyszukiwawczej, a następnie przeprowadzone zostaje wyszukiwanie i odpowiednio do jego wyników instrukcja w ten sam sposób może być modyfikowana. Wysoko ocenianą własnością systemu IT jest strategia pozwalająca użytkownikowi wybierać alternatywne wyrażenia wyszukiwawcze, zastępująca stosowaną w innych systemach pełną algorytmizację selekcji słownictwa /5/.

CITE

Jednym z najlepiej znanych systemów przetwarzania języka naturalnego opracowanych dla wspomagania wyszukiwania w systemach online jest CITE (Computerized Information Transfer in English) Thomasa Doszkocsa /1/. Wykorzystywany jest on w amerykańskiej Narodowej Bibliotece Medycznej (National Library of Medicine, NLM) i pierwotnie został napisany jako tzw. programowy sprzęg do komunikacji w języku naturalnym dla systemu MEDLINE. Później zastosowano go także w katalogu online CATLINE i przetestowano w czytelni NLM. Następnym pomyślnych wyników tego testu było włączenie CITE do stałej współpracy z publicznie dostępnym katalogiem online NLM, gdzie wykorzystywany jest z powodzeniem.

Zapytania użytkowników mogą być wprowadzane do CITE w postaci angielskich zdań lub związków wyrazowych. W odpowiedzi system proponuje odpowiednie hasła przedmiotowe, identyfikuje ewentualne błędy ortograficzne korygując je i formułuje instrukcję wyszukiwawczą dla CATLINE. Wyszukane opisy dokumentów zostają uporządkowane według malejącego stopnia ich zgodności z instrukcją i wyświetlone na monitorze. Użytkownik może zgłosić systemowi zadowolenie lub niezadowolenie z wyników wyszukiwania, przy czym w drugim przypadku

podejmowana jest kolejna próba wyszukiwania w oparciu o zmodyfikowaną instrukcję.

Doszkoćs uważa, że jego CITE nie wykorzystuje technik sztucznej inteligencji, gdyż nie interpretuje znaczenia wprowadzonych przez użytkownika wyrażeń naturalnych, a jedynie koryguje ich formę ortograficzną i niejako przekłada na język haseł przedmiotowych. CITE nie potrafi także odpowiadać na żadne pytania z zakresu medycyny. Tym niemniej, ponieważ system ten korzysta ze ściśle związanych ze sztuczną inteligencją technik przetwarzania języka naturalnego, wydaje się, że można traktować go jako przykład zastosowania szeroko pojętej sztucznej inteligencji w udoskonalaniu dostępu i komunikacji z systemami online.

IR-NLI

Twórcy systemu IR-NLI (Information Retrieval - Natural Language Interface) wykorzystali techniki przetwarzania języka naturalnego w dwóch aspektach - rozumowania i wnioskowania oraz interpretacji i dialogu /4/. Zakładając, że przeprowadzenie wyszukiwania online opiera się na swoistym rozumowaniu w języku naturalnym (ustalenie właściwych terminów, wyrażeń wobec nich znaczeniowo szerszych lub węższych czy, z nimi skojarzonych itp.), wyposażyli system w tzw. moduł rozumowania współpracujący z bazą wiedzy IR-NLI, którego zadaniem jest przede wszystkim rozszerzanie pytania użytkownika. Drugiemu aspektowi przetwarzania języka odpowiada moduł rozumienia i dialogu, który przeprowadza rozbiór gramatyczny zdań wejściowych, ustala ich znaczenie oraz przekazuje do modułu rozumowania. Trzecim elementem IR-NLI jest tzw. moduł formalizowania, który odpowiada za konstrukcję instrukcji wyszukiwawczej. W bazie wiedzy systemu IR-NLI przechowywane są informacje o zakresie tematycznym i strukturze bazy danych, do której system ma

umożliwiać dostęp; relacje między wyrażeniami języka, w którym reprezentowane są charakterystyki dokumentów czy innego typu obiektów, przedstawione są w postaci sieci semantycznej. System IR-NLI wyposażony jest również w tzw. "generator strategii", kontrolujący i wspomagający proces wyszukiwania przy wykorzystaniu tzw. taktyk Batesa opracowanych przez Marcję J. Bates i przedstawionych w Journal of the American Society for Information Science, 1979 nr 5(30) s.281-289.

Ze względu na konieczność tworzenia baz wiedzy dla każdej bazy danych, w której miałyby przebiegać wyszukiwanie, zakres wykorzystania IR-NLI jest ograniczony. Mimo to, system zasługuje na uwagę, gdyż jest jednym z nielicznych systemów przetwarzania języka naturalnego, opracowanych specjalnie w celu wspomaganie wyszukiwania informacji w systemach online.

Przedstawione wyżej systemy wspomaganie wyszukiwania online - jak wcześniej już zostało powiedziane i co należy podkreślić - w większości przypadków są albo systemami eksperymentalnymi, albo projektami, które nie wyszły jeszcze poza etap wstępnego testowania. Są więc to rozwiązania, które otwierają dopiero bardzo ciekawy i ogromnie ważny dla praktycznego funkcjonowania systemów informacyjno-wyszukiwawczych online okres w rozwoju technik informacyjnych. Informacje o poszczególnych systemach z konieczności w artykule przedstawione są w ogromnym skrócie, który uniemożliwia pełne zrozumienie ich zasad budowy i działania, dlatego też poniżej podana jest literatura, w której znaleźć można obszernie i szczegółowe charakterystyki tych systemów, często uzupełnione także charakterystyką i wynikami testów.

Literatura

1. DOSZKOCZ I.E. CITE NLM: Natural language searching in an online catalog. *Information Technology and Libraries* 1983 nr 2(4) s.364-380 [cyt. za 5]
2. FENICHEL C.H. Online searching: measures that discriminate among users with different types of experiences. *Journal of the American Society for Information Science* 1981 nr 1(32) s. 23-32
3. GROSS D. Improving online via expert systems. *Information Today* 1987 nr 7(4) s. 7, 17.
4. GUIDA G., TASSO C. IR-NLI: An expert natural language interface to online databases. *Automatica* 1983 nr 19(6) s. 759-766 [cyt. za 5]
5. HOWKINS D.T. Applications of artificial intelligence (AI) and expert systems for online searching. *ONLINE* 1988 vol. 12 nr 1 s. 31-38
6. MARCUS R.S. Design questions in the development of expert systems for retrieval assistance. *Proceedings of the 49th ASIS Annual Meeting, Chicago 1986*. Ed. by J.M. Hurd. Medford, N.J.: Learned Information 1986 s.185-189 [cyt. za 5]
7. McCUNE B.P., TONG R.M., DEAN J.S., SHAPIRO D.G. RUBRIC: A system for rule-based information retrieval. *Proceedings of the 7th International Computer Software and Applications Conference (COMPSAC'83), Chicago: IEEE 1983* s.166-172 [cyt. za 5]
8. MEADOW Ch.I., HEWETT T.T., AVERSA E.S. A Computer Intermediary for Interactive Database Searching. Part I: Design, Part II: Evaluation. *Journal of the American Society for Information Science* 1982 nr 5(33) s. 325-332, nr 6(33) s. 357-364 [cyt. za 5]
9. MICCO H.M., SMITH I. Designing an expert system for the reference function subject access to information. *Proceedings of the 49th ASIS Annual Meeting, Chicago 1986*. Ed. by J.M. Hurd. Medford, N.J. Learned Information 1986 s.204-210 [cyt. za 5]
10. O'NEILL M. Expert systems and library and information science. *Aslib Information* 1988 vol. 16 nr 9 s. 224-226.

11. PARROTT J.R. Expert systems for reference work. Microcomputers for Information Management 1986 vol. 3 nr 3 s, 155-171
12. PATEL-SCHNEIDER P.F., BRACHMAN R.J., LEVESQUE H.J. ARGON: knowledge representation meets information retrieval. Proceedings of the First Conference on Artificial Intelligence Applications, Denver, C.O. Dec. 5-7 1984. Silver Spring, DM: IEEE Computer Society Press 1984 s.280-286 [cyt. za 5]
13. POLLITT S. CANSEARCH: An expert systems approach to document retrieval. Information Processing and Management 1987 vol. 23 nr 2 s. 119-138
14. SHOVAL P. Principles, procedures, and rules in an expert system for information retrieval. Information Processing and Management 1985 vol. 21 nr 6 s. 475-487
15. VICKERY A., BROOKS H.M. PLEXUS - the expert system for referral. Information Processing and Management 1987 vol. 23 nr 2 s. 99-117
16. WATERS S.T. Answerman, the expert information specialist: An expert system for retrieval of information from library reference books. Information Technology and Libraries 1986 vol. 5 nr 3 s. 204-212 [cyt. za 5]
17. WILLIAMS P.W. A model for an expert system for automated information retrieval. Proceedings of the 8th International Online Information Meeting, London Dec. 4-6 1984. Oxford: Learned Information Ltd. 1984 s. 139-149
18. WILLIAMS P.W. The design of an expert system for access to information. Proceedings of the 9th International Online Information Meeting. London Dec. 3-5 1985. Oxford: Learned Information Ltd. 1985 s. 23-29

25.01.1990 r.

USING OF THE TECHNIQUES OF ARTIFICIAL INTELLIGENCE IN MODERN
LIBRARIANSHIP AND IN THE SYSTEMS ASSISTING THE ONLINE RETRIEVAL

Summary

There are discusses inquiries about possibilities of using the techniques of artificial intelligence and of expert systems in librarianship and in the online systems (in administration and management of library, routine library works such as: cataloguing, classifying, indexing, annotating and abstracting, facilitating the access to the sources of information, improvement of the information service for users).

There is presented the detailed characteristics of the selected systems which assist information retrieval in online systems: Answerman, CANSEARCH, expert consulting system of P.Showal, ORA, PLEXUS, library information system thesaurus aided of Micco and Smith, RUBRIC, ARCON, CONIT, IIDA, IT, CITE and IR-NLI.

ИСПОЛЬЗОВАНИЕ ТЕХНИК ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В СОВРЕМЕННОМ
БИБЛИОТЕКОВЕДЕНИИ И В СИСТЕМАХ ВСПОМОГАЮЩИХ ПОИСК В РЕЖИМЕ
ДИАЛОГА

Резюме

В статье обсуждаются попытки нахождения возможности использования техники искусственного интеллекта и систем-эксперт в библиотековедении, и также в диалоговых системах (в администрации и управлении библиотекой рутинных библиотечных работах таких как каталогизация, классифицирование, индексирование, подготовка аннотации и рефератов, облегчение доступа к источникам информации, совершенствование информационного обслуживания потребителей).

Подробно представлена характеристика избранных систем
вспомогательных поисков информации: Answerman, CANSEARCH консультационная система - эксперт П.Новля, ORA, PLEXUS, вспомогательная тезаурусом система библиотечной информации Микко и Смиса, ARCON CONIT, FIDA, IT, EITE, IR-NLI.

EWA CHMIELEWSKA-GORCZYCA

Instytut Bibliotekoznawstwa
i Informacji Naukowej UW

WYSZUKIWANIE RZECZOWE W PUBLICZNIEDOSTĘPNYCH KATALOGACH ONLINE

Historia powstania i podstawowe zalety publiczniedostępnych katalogów online (Online Public Access Catalogue - OPAC). Prace Biblioteki Kongresu nad nowymi rozwiązaniami OPAC. Opracowanie rzeczowe, środki wyszukiwawcze, rola słownika i udogodnienia dla użytkownika stosowane w tych katalogach.

Automatyzację prac katalogowych rozpoczęto z chwilą wprowadzenia komputerów do bibliotek, jeszcze w połowie lat sześćdziesiątych. Początkowo była to jednak tylko próba wykorzystania komputerów do przetwarzania danych katalogowych w celu wydrukowywania katalogów książkowych, automatycznego drukowania kart katalogowych lub tworzenia katalogów COM. Komputery wykorzystywane były głównie jako inteligentne maszyny do pisania; postać, struktura i zawartość katalogu nie uległa zmianie. Szybko jednak zdano sobie sprawę z korzyści, jakie mogą wynieść biblioteki ze wspólnego katalogowania swoich zbiorów. Rozwój współpracy w zakresie katalogowania związany był głównie z powstawaniem komputerowych sieci bibliotecznych oraz katalogów centralnych dostępnych w trybie online. Biblioteki wchodzące w skład sieci z założenia mogą być zwolnione z ka-

talogowania części własnych zbiorów, mając możliwość automatycznego kopiowania gotowych rekordów z innych bibliotek sieci. Oczywiście, kopiowane rekordy stanowią najczęściej tylko część nabytków danej biblioteki. Konsekwencją wspólnego katalogowania stała się konieczność wypracowania, a następnie przejęcia przez biblioteki członkowskie, pewnych standardów w zakresie opisu katalogowego; sprawne funkcjonowanie sieci wymaga bowiem pewnego stopnia zgodności (kompatybilności), możliwej do osiągnięcia tylko dzięki normalizacji, obejmującej typy (rodzaje) katalogów funkcjonujących w ramach sieci, opis katalogowy oraz procedury wyszukiwawcze.

Następnym krokiem było wypracowanie idei OPAC (Online Public Access Catalogue) - katalogu online publicznie dostępnego, czyli przeznaczonego dla użytkownika niewykwalifikowanego. Z licznych zalet OPAC wymienić można kilka, które mają w pewnym stopniu wpływ na wyszukiwanie rzeczowe:

1. **Bardzo łatwy, tani i szybki proces aktualizacji (melioracji).** Wprowadzanie nowych opisów lub zmian do opisów już istniejących, (np. usuwanie starych haseł, zastępowanie ich nowymi, likwidowanie zbędnych odsyłaczy, itp.) może być przy odpowiednim oprogramowaniu dokonywane w sposób bardzo prosty i na bieżąco.

2. **Ogromna szybkość wyszukiwania,** zapewniająca natychmiastowy dostęp do informacji nawet w bardzo dużych zbiorach i według wielu kluczy wyszukiwawczych jednocześnie.

3. **Łatwy fizyczny dostęp do katalogu.** Katalog online może być fizycznie "rozproszony" za pośrednictwem terminali, umieszczonych np. w różnych czytelniach lub w różnych działach biblioteki, a także poza budynkiem biblioteki, i to - dzięki wykorzystaniu środków teletransmisji - w miejscach bardzo odległych, np. w poszczególnych wydziałach macierzystej uczelni, w domach studenckich, w bibliotekach zakładowych, itp. Użytkownicy mogą mieć

dostęp do katalogu bez konieczności przyjeżdżania do budynku biblioteki i niezależnie od godzin jej otwarcia.

4. **Większa elastyczność systemu.** W katalogach tradycyjnych informacja przedstawiana użytkownikowi miała niezmienną formę i raz ustaloną strukturę. W OPAC użytkownik ma do dyspozycji tę samą informację przedstawioną w wielu układach i w wielu formatach do wyboru.

5. **Interakcyjny charakter wyszukiwania.** Interakcyjność umożliwia użytkownikowi reagowanie na uzyskiwane informacje i modyfikowanie swoich poleceń w zależności od otrzymywanych odpowiedzi. Przy korzystaniu z katalogu użytkownik może liczyć na pomoc, instrukcje i "podpowiedzi" w trakcie pracy z systemem.

6. **Większy zestaw środków wyszukiwawczych.** Poprzez odpowiednie oprogramowanie OPAC stwarza znacznie bogatsze możliwości wyszukiwawcze, nieosiągalne w katalogach tradycyjnych, np.: znacznie większa liczba kluczy wyszukiwawczych, możliwość ich kombinowania, umożliwienie wyszukiwania pełnotekstowego (tj. według wszystkich słów kluczowych zawartych w tytułach dokumentów, hasłach przedmiotowych, abstraktach czy adnotacjach). Innym udogodnieniem jest maskowanie, tj. wyszukiwanie według fragmentów słów oraz wskazywanie pożądanego bliskości słów. Środki te szerzej będą omówione później.

Dodatkową korzyścią stosowania OPAC jest możliwość bardziej wiarygodnego badania potrzeb, przyzwyczajęń i oczekiwań użytkowników, a dzięki temu - lepsze dostosowanie katalogu do tych potrzeb. Zastosowanie odpowiedniego oprogramowania daje możliwość "podglądania" sposobów wykorzystania katalogów przez użytkowników, co ma szczególne znaczenie przy katalogach rzeczowych. Bibliotekarz po raz pierwszy ma doskonałe narzędzie badania faktycznych po-

trzeb użytkowników, możliwość śledzenia ich dróg poszukiwań, przeglądania się szczegółom formułowania instrukcji i strategii wyszukiwawczej (no. doborowi terminów, strukturze wyrażen złożonych, rodzajowi poszukiwanych tematów), włącznie z danymi statystycznymi dotyczącymi częstości wykorzystania poszczególnych haseł. Istnieje możliwość "wglądu" w publiczne użycie katalogu, oceny jego przydatności i wykorzystania, analizy przyczyn niepowodzeń i błędów przy wyszukiwaniu, możliwość poprawy i stałego doskonalenia katalogu.

W krótkiej historii OPAC wyodrębnić można (ze względu na metody wyszukiwawcze stosowane w tych katalogach) dwie kategorie (generacje):

1. OPAC pierwszej generacji wzorowane na tradycyjnych katalogach kartkowych lub mikrofizycznych i przeznaczone głównie do poszukiwań konkretnego dokumentu.

2. OPAC drugiej generacji wzorowane często na zautomatyzowanych systemach wyszukiwania informacji bibliograficznej (typu DIALOG, BRS lub MEDLARS), bazujące na wyszukiwaniu według słów kluczowych (lub deskryptorów) i przeznaczone głównie do wyszukiwania artykułów z czasopism według ich cech treściowych, bez konieczności lokalizowania wyszukanych dokumentów w konkretnym zbiorze bibliotecznym. Ze względu na rozbudowany aparat wyszukiwawczy i skomplikowaną strategię niezbędną była fachowa pomoc osoby pośredniczącej (bibliotekarza, dokumentalisty), mającej odpowiednie przygotowanie specjalistyczne w tym zakresie.

Z punktu widzenia wyszukiwania rzeczowego istotna jest różnica między tymi dwiema generacjami na poziomie środków komunikacji z użytkownikiem, tj. języka do wydawania poleceń oraz wykorzystywanego słownictwa (gotowe, "prekoordynowane" frazy, czy swobodny dobór i kombinacja słów kluczowych).

Wyszukiwanie według fraz (pierwsza generacja) lub według słów kluczowych (druga generacja) ma pewien wpływ na strukturę dialogu, a szczególnie na sposób, w jaki wybór kryteriów wyszukiwawczych jest prezentowany przez system na ekranie. W OPAC wykorzystujących frazy, użytkownika prosi się, aby wybrał potrzebne mu pole rekordu lub hasło w wykazie gotowych formatów lub haseł. Po wprowadzeniu (lub wybraniu) terminów wyszukiwawczych zapisy indeksowe najbliższe w ciągu alfabetycznym do podanych przez użytkownika są wyświetlane na monitorze. W OPAC wykorzystujących słowa kluczowe i algebrę Boole'a użytkownik musi nauczyć się samemu wyszczególniać pola wybrane do przeszukiwania, musi zrozumieć i nauczyć się stosować operatory logiczne do łączenia wybranych terminów, itp.

Obecnie jesteśmy świadkami powstawania OPAC trzeciej generacji wykorzystującej zalety grupy pierwszej i drugiej, a więc dające duże możliwości wyszukiwawcze i łatwe do wykorzystania.

W katalogach online pierwszej generacji często nie było możliwości wyszukiwania rzeczowego (tak było w najstarszym i najbardziej rozpowszechnionym do niedawna systemie OCLC). Funkcję katalogu rozumiano bardzo tradycyjnie - jako narzędzie odnajdowania znanych lub wcześniej wyszukanych (np. w bibliografiach) dokumentów. Biblioteki wychodziły z założenia, że istnienie wielu bibliograficznych i faktograficznych baz danych dostępnych w trybie online zwalnia je z obowiązku zaspokajania bardziej skomplikowanych zapytań użytkowników, wymagających czasochłonnej analizy treściowej katalogowanych dokumentów. Szybko okazało się to jednak błędnym założeniem. Podjęto więc prace zmierzające do wprowadzenia nowych elementów opisu rzeczowego, dostosowania go do zmienionych warunków oraz udoskonalenia, dzięki możliwościom stworzonym przez systemy komputerowe. Prace te rozpoczęto od analizy i oceny do-

tychczasowych metod katalogowania rzeczowego. Badania efektywności katalogów przedmiotowych w bibliotekach różnych typów i różnej wielkości wykazały, że tylko 1/3 poszukiwań zakończona jest powodzeniem /3, 21/. Analiza tych niepowodzeń ujawniła niedoskonałości istniejących systemów, np.: nieadekwatne odzwierciedlenie zawartości dokumentów przez źle dobrane hasła, nieodpowiedni wybór terminów na hasła (szczególnie w naukach o szybko zmieniającej się terminologii), brak słowników, a w słownikach istniejących - brak odpowiedniej szczegółowości, zbyt ubogi aparat odsyłaczowy, itp. Badania te ujawniły, że **opracowanie treściowe jest najsłabszym punktem całego systemu katalogowego**, zarówno tradycyjnego, jak i automatycznego. Aby poprawić jakość tego opracowania postanowiono wykorzystać osiągnięcia i doświadczenia zdobyte przy tworzeniu komercyjnych baz danych typu bibliograficznego. Jednocześnie rozpoczęto całą serię prac i badań zmierzających do wypracowania własnych, bardziej odpowiednich dla funkcji katalogu, metod i form opracowywania i wyszukiwania rzeczowego. Prace nad poszukiwaniem nowych języków informacyjnych i ich wariantów oraz nowych rozwiązań w językach już istniejących podjęło lub sponsorowało szereg bibliotek, wśród nich Biblioteka Kongresu w Waszyngtonie, przy czym akcent położony był na samodzielne wyszukiwanie przez użytkownika a nie przez wyszkolonego dokumentalistę pośredniczącego w tym procesie /m.in. 18, 17, 5/.

PRACE BIBLIOTEKI KONGRESU

W momencie przystąpienia do prac tradycyjnym, podstawowym elementem opisu służącym do wyszukiwania rzeczowego były hasła przedmiotowe wybierane z dwutomowego wykazu LCSH (Library of Congress Subject Headings). Jednemu dokumentowi przeciętnie przy-

dzielano 2,13 haseł przedmiotowych (z wyłączeniem działu "Literatura", który tę statystykę zanizła). Innym elementem dostępu rzeczowego w opisach Biblioteki Kongresu były symbole KBK (Klasyfikacji Biblioteki Kongresu), odzwierciedlające jednocześnie uporządkowanie dokumentów na półkach. Oprócz tego książki z zakresu medycyny posiadały poza symbolami KBK także symbole NLM (National Library of Medicine, Narodowej Biblioteki Medycznej), a około 100 000 książek rocznie katalogowanych przez Bibliotekę posiadało dodatkowo symbole KDD (Klasyfikacji Dziesiątej Deweya). Innym kluczem wyszukiwawczym mógł być dodawany do opisów katalogowych kod jednostki geograficznej, której dotyczy dany dokument.

Pomimo tych kilku możliwości dostępu rzeczowego sytuacja została oceniona jako niezadowolająca, czego dowodem było m.in. zbyt rzadkie korzystanie czytelników z katalogów rzeczowych oraz wspomniany już wysoki wskaźnik niepowodzeń przy ich stosowaniu. Biblioteka rozpoczęła prace nad wprowadzeniem nowych, dodatkowych kluczy wyszukiwawczych oraz ulepszeniem metod wyszukiwania według kluczy już istniejących. Pierwszym posunięciem było wprowadzenie możliwości wyszukiwania według słów kluczowych w obrębie hasła przedmiotowego, co znacznie rozszerzyło liczbę punktów dostępu w każdym opisie (w tradycyjnym katalogu wyszukiwać można było tylko przez pierwszy wyraz w każdym hasle przedmiotowym). Następnym etapem było rozszerzenie wyszukiwania według słów kluczowych na pozostałe wybrane elementy opisu, np. tytuły dokumentów, nazwy serii lub adnotacje (uwagi). Postanowiono też zwiększyć liczbę adnotacji (poprzednio występowały one tylko w 2-5% wszystkich opisów), umieszczając w tym polu m.in. spisy treści dokumentów (a więc adnotacje zawartościowe), abstrakty (głównie streszczenia autorskie) lub nawet gotowe indeksy przedmiotowe zaczerpnięte z katalo-

gowanych ksiązek. Planowano umożliwić w przyszłości wyszukiwanie według wszystkich słów kluczowych znajdujących się w opisach lub w polach dodatkach do tych opisów, łącznie nawet z całymi tekstami dokumentów. W przypadku pól symboli klasyfikacyjnych (zarówno KBK jak i KDD) postanowiono umożliwić wyszukiwanie według słów kluczowych znajdujących się w tekstach przedstawiających znaczenie symboli w języku naturalnym, tj. w odpowiednikach słownych symboli.

Poza rozszerzaniem opisów o dodatkowe elementy opracowania rzeczowego postanowiono udoskonalić metody i strategie wyszukiwawcze, którymi może dysponować użytkownik, czyli podnieść wartość wyszukiwawczą opisów już utworzonych. Postulowano wykorzystanie techniki online do wyszukiwania przy użyciu m.in. operacji logicznych, wprowadzenie maskowania słów zaczerpniętych z tytułów, nazw serii, spisów treści (a więc z pól zawierających teksty języka naturalnego, stosującego różne formy gramatyczne i graficzne tych samych terminów), itp. Zwrócono uwagę na niedoskonałość dotychczasowych metod przedmiotowania, postulowano rozbudowanie pól haseł przedmiotowych LCSH (co ma przyczynić się do rozwiązania problemu braku wyczerpywalności indeksowania, tak często syngalizowanego przez czytelników), możliwe do wykonania kilkoma sposobami:

- poprzez przydzielenie dodatkowych haseł przedmiotowych poszczególnym pozycjom, tzn. zwiększenie liczby haseł przedmiotowych przypadających na jeden tytuł;

- wprowadzanie terminów niekontrolowanych ze spisów treści i tekstów samych dokumentów;

- dodawanie do poszczególnych haseł deskryptorów z tezaursów specjalistycznych przy wykorzystaniu specjalnie w tym celu opracowanych słowników przejścia;

- dodawanie haseł przedmiotowych zaczerpniętych z innych wykazów (systemów, bibliotek), np. z Hennepic County Subject List lub z PRECIS.

Zwrócono także uwagę na konieczność udoskonalenia słowników (wykazu haseł, tablic klasyfikacyjnych i tezaurusów) wykorzystywanych do opracowania rzeczowego, wprowadzenia ich na nośnik maszynowy i wyświetlania odpowiednich ich fragmentów w trakcie wyszukiwania w przypadku udzielenia odpowiedzi podającej zbyt małą lub zbyt dużą liczbę pozycji.

Po wprowadzeniu niektórych z proponowanych rozwiązań i podjęciu prac nad pozostałymi Biblioteka przystąpiła do analizy i oceny efektywności tych udoskonaleń. Przeprowadzone przez P. Atherton badania wykazały, że wzbogacenie pól LCSH terminami zaczerpniętymi ze spisów treści i indeksów rzeczywiście znacznie podniosło efektywność wyszukiwania; wskaźnik trafności wyszukiwania wzrósł z 35% przy tradycyjnych opisach MARC do 45% przy opisach nowych, poszerzonych. Ocena kompletności wyszukiwania dała jeszcze lepsze rezultaty - w naukach społecznych system wydał ponad dwukrotnie większą liczbę dokumentów relewantnych, a w naukach humanistycznych - trzykrotnie więcej. Ceną za tego rodzaju ulepszenie było jednak podwojenie długości rekordów MARC oraz zwiększenie czasu katalogowania rzeczowego średnio o 12 minut na jeden tytuł. Podniosło to znacznie koszt sporządzania opisów (dotychczasowy koszt utworzenia jednego opisu w Bibliotece Kongresu oceniany był na 100 dolarów). Zaproponowano więc rozwiązanie kompromisowe - wzbogacanie opisów przedmiotowych tylko dla wydzielonego podzbioru dokumentów biblioteki, uznanego za najcenniejszy z punktu widzenia wartości wyszukiwawczej (np. w bibliotece Iowa State University uznano za taki podzbiór wszystkie wydawnictwa informa-

cyjne - encyklopedie, słowniki, bibliografie, tabele, itp.)^{1/}. Dla pozostałych pozycji zarekomendowano przejmowanie gotowych indeksów z instytucji wydawniczych stosujących automatyczną technikę drukowania, co nie obciąża biblioteki kosztami bardziej szczegółowej analizy treści. Ocena przydatności spisów treści i indeksów przedmiotowych wykazała znacznie wyższą efektywność tych drugich; spis treści zawiera stosunkowo dużo terminów zbyt ogólnych dla potrzeb wyszukiwawczych. Stwierdzono, że aby spis treści mógł być włączony jako element opisu rzeczowego musi spełniać pewne warunki, które przedstawiono w formie sformalizowanych zaleceń.

OPRACOWANIE RZECZOWE W KATALOGACH ONLINE

Przykład prac prowadzonych przez Bibliotekę Kongresu jest tylko jednym z wielu. Jak wspomniano, prace takie były prowadzone także przez wiele innych bibliotek, stowarzyszeń bibliotekarskich, sieci bibliotecznych i serwisów informacyjnych w wielu krajach. Wyniki badań oraz widoczne tendencje w zakresie rozwoju metod wyszukiwania rzeczowego są najczęściej bardzo podobne, choć dochodzono do nich różnymi drogami.

Coraz częściej w działających katalogach online rezygnuje się z ograniczenia wyszukiwania przedmiotowego tylko do elementów rekordu specjalnie do tego celu przeznaczonych (np. pole haseł przedmiotowych, symboli klasyfikacyjnych, deskryptorów); bardziej efektywne okazuje się zapewnienie możliwości przeszukiwania wszyst-

^{1/} Podobną praktykę stosują wielobazowe serwisy informacyjne, np. Lockheed reklamuje możliwość korzystania z tzw. Superindeksu, który jest po prostu bazą danych zawierającą indeksy przedmiotowe i inne spisy pomocnicze zamieszczone na końcu kilkuset dokumentów, ocenionych jako szczególnie cenne wydawnictwa informacyjne.

kich lub przynajmniej wielu pól rekordu w celu odnalezienia słowa kluczowego podanego przez użytkownika jako "termin wyjściowy". Liczba pól przeszukiwanych metodą słów kluczowych jest różna w różnych bibliotekach, niektóre systemy przeznaczają na to wszystkie pola, inne - tylko pola wybrane, najczęściej pole tytułu, haseł przedmiotowych, adnotacji (uwag) lub abstraktu.

Inną widoczną tendencją w wielu katalogach jest wzbogacanie rekordów o takie elementy, jak adnotacja (lub abstrakt), deskryptory lub automatycznie generowane terminy ze spisów treści i indeksów przedmiotowych katalogowanych książek. Zbliża to opisy katalogowe do formatów i strategii stosowanych w systemach wyszukiwawczych typu bibliograficznego (w "niekatalogowych" bazach danych). Dążenie do wzbogacenia opisów o wymienione elementy jest odbiciem wzrastających oczekiwań i potrzeb użytkowników w stosunku do bibliotek zapewnienia pełniejszej i bardziej szczegółowej informacji o zawartości treściowej dokumentów. Problemem dla bibliotek jest tu koszt, zarówno utworzenia tych danych, jak i ich utrzymywania w komputerze (z reguły katalogowe bazy danych są objętościowo większe od baz bibliograficznych). Indeksowanie na poziomie oczekiwanym przez użytkowników wymaga dokładnej i szczegółowej analizy przedmiotowej, wykraczającej poza możliwości przeciętnej biblioteki (z wyjątkiem nielicznych bibliotek specjalistycznych). Wzbogacanie opisów o wymienione elementy musiałyby być prowadzone na szczeblu krajowym (np. przez bibliotekę narodową lub bibliografię narodową) lub wymagałyby ścisłej współpracy ze specjalistycznymi serwisami informacyjnymi, co też zwiększyłoby znacznie koszty opracowania rzeczowego (usługi te byłyby płatne). Jako jedyne możliwe do przyjęcia rozwiązanie w obecnej sytuacji uznano wykorzystywanie gotowych spisów pomocniczych istnieją-

cych w wielu dokumentach, to jest indeksów i spisów treści. Doświadczenia Biblioteki Kongresu wykazały, że nawet taki wysiłek może być nie do zaakceptowania dla wielu bibliotek. Rozwiązaniem może być tu nawiązanie współpracy z instytucjami wydawniczymi i przejmowanie od nich gotowych wykazów już w formie maszynowej. Planuje się opracowanie specjalnych programów "przerabiających" te wykazy dla potrzeb katalogów bibliotecznych, np. eliminujących terminy zbyt ogólne (tzw. stoplisty), szeregujące wybrane terminy w kolejności alfabetycznej, itp.

Dodawanie abstraktów do opisów katalogowych ma jeszcze jeden aspekt: umożliwia lepszy wybór dokumentów relewantnych, co jest szczególnie ważne w sieciach bibliotecznych dysponujących katalogiem centralnym dającym dostęp do zbiorów odległych fizycznie, tj. do zbiorów innych bibliotek, np. zakładowych, znajdujących się w innych budynkach, innych częściach miasta lub nawet w innych miastach (filie). W sieciach tego typu obserwowano ogromny wzrost wypożyczeń międzybibliotecznych, obciążających w dużym stopniu wszystkie biblioteki. Umieszczenie abstraktów lub spisów treści wyeliminuje trud związany ze sprowadzaniem przez biblioteki wielu książek wybranych przez użytkownika z katalogu na podstawie krótkich elementów opisu treściowego, które to książki po przejrzaniu okazują się nierелеwantne.

Konsekwencją wprowadzania nowych elementów opisu i metod wyszukiwawczych, potwierdzających słuszność przyjętych rozwiązań, było zwiększenie liczby wyszukiwań rzeczowych w katalogach online w stosunku do katalogów tradycyjnych oraz lepsze rezultaty osiągnięte nawet przez użytkowników niewykwalifikowanych. W badaniach prowadzonych w USA /13/ stwierdzono, że po niedługim okresie stosowania OPAC ponad 53% poszukiwań dotyczyło cech treściowych dokumentów,

a co było zaskakujące - wskaźniki procentowe nie różniły się zbyt-
nio w różnych typach bibliotek. Inne badania /6/, przeprowadzone
w 31 bibliotekach amerykańskich wykazały, że o ile w katalogach
tradycyjnych wyszukiwanie rzeczowe było prowadzone przez użytkow-
ników stosunkowo rzadko (z wyraźną "niechęcią"), to w katalogach
online liczba wyszukiwań tego typu znacznie wzrosła, a co najważ-
niejsze - zyskują one u użytkowników coraz większą aprobatę.
Procentowo, w ogólnej liczbie wyszukiwań, w bibliotekach uniwersy-
teckich 38% dotyczyło cech treściowych, a w bibliotekach koledżów
aż 62%. Dla porównania: według tytułów od 15 do 40%, według autora
- nie więcej niż 15%. Z drugiej jednak strony trzeba przyznać,
że ten rodzaj wyszukiwania dostarczał użytkownikom i bibliotekarzom
najwięcej problemów i pytań. Analiza tych problemów wykazała,
że spowodowane one były w większości wypadków niedoskonałościami
istniejących systemów, np. brakiem słownika, zbyt małą ilością
odsyłaczy, itp. możliwych do usunięcia stosunkowo niewielkim
kosztem.

ŚRODKI WYSZUKIWAWCZE STOSOWANE W KATALOGACH ONLINE

Poza rozszerzaniem opisu rzeczowego o nowe elementy prowadzone
są próby i prace zmierzające do lepszego wykorzystywania "danych
treściowych" znajdujących się już w rekordach. Jak już wspomniano,
w obecnych OPAC stosowane są różnorodne środki i techniki wyszuki-
wania; różnice te są czasami trywialne, czasami jednak bardzo
istotne. Nawet te same możliwości oferowane przez system (np.
maskowanie) mogą być prezentowane na wiele sposobów, a sam sposób
prezentacji ma duży wpływ na efektywność ich stosowania.

W obecnych katalogach online dysponujących rozbudowanym
aparatem wyszukiwawczym spotkać można następujące środki:

1. **Operatory algebry Boole'a.** Większość programów wyszukiwawczych stwarza tę możliwość, choć pod różnymi postaciami. Niektóre systemy wymagają, aby terminy wprowadzane były oddzielnie (jako pojedyncze zapytania) i dopiero wyniki tych poszczególnych wyszukiwań są następnie kombinowane. Inne pozwalają od razu na wprowadzanie wielu terminów jednocześnie (w jednym zapytaniu); niektóre nie pozwalają na grupowanie terminów w zapytaniu, inne dysponują wielopoziomową strukturą wyodrębniania zbiorów i podzbiorów (gniazd) przy pomocy nawiasów lub innych oznaczeń; niektóre posiadają tylko operatory AND i OR, inne dodają do tego NOT. Stopień łatwości korzystania (wspomagania użytkownika) oraz elastyczności logiki wyszukiwawczej też nie jest jednakowy w istniejących systemach.

2. **Operatory arytmetyczne i pozycyjne.** Oprócz operatorów boolowskich wiele systemów ma też operatory innego typu, np. nakładające warunek, aby wartość danego pola opisu była równa, mniejsza, nie większa, większa, itp. od wartości wyszczególnionej w zapytaniu użytkownika (dotyczy to głównie pól zawierających dane liczbowe, np. dokumenty wydane po roku 1985). Operatory pozycyjne typu SAME, WITH, NEAR, ADJ służą do określania odległości między wybranymi terminami (ich położenia względem siebie) w poszukiwanych tekstach, np. w tytułach, abstraktach, hasłach przedmiotowych.

3. **Maskowanie.** Terminy użyte do wyszukiwania mogą być maskowane prawo- lub lewostronnie, o jedną literę lub o dowolną liczbę znaków, co umożliwia wyszukiwanie terminów różniących się tylko końcówkami (np. różne przypadki, formy przymiotnikowe) lub przedrostkami. Niektóre systemy stwarzają możliwość maskowania wewnątrz wyrazu. Maskowanie jest szczególnie przydatne w systemach, w których prowadzi się wyszukiwanie według swobodnych słów kluczowych.

wych, podnosi bowiem w ogromnym stopniu wskaźnik kompletności. Niestety, użytkownicy niezbyt często wykorzystują tę opcję, nie rozumiejąc jakby samej idei i korzyści z jej stosowania. Stwierdzono, że na ogół tylko 5-6 pierwszych liter wystarczy, aby doprowadzić użytkownika do tego miejsca w wykazie (liście indeksowej), w którym znajduje (lub znajdowałby) się dany termin (tj. po wpisaniu kilku pierwszych liter i wciśnięciu klawisza Enter pojawi się na ekranie ten fragment ciągu alfabetycznego, który jest faktycznie potrzebny użytkownikowi). Niestety, czytelnicy rzadko zdają sobie z tego sprawę i najczęściej "wystukują" pieczołowicie długie ciągi liter, niepotrzebnie przedłużając czas wyszukiwania i zwiększając ryzyko popełnienia błędu. Z tego powodu w wielu systemach wprowadza się wyświetlanie odpowiedniego fragmentu listy wzorcowej (zapisów indeksowych w kolejności alfabetycznej) w trakcie wpisywania polecenia. Użytkownik ma możliwość przeglądania (wertowania) listy w górę i w dół z różną prędkością, np. linia po linii, rekord po rekordzie, ekran po ekranie lub "przeskakując" np. co 30. zapis. Możliwość szybkiego przeglądania zbioru jest wysoko oceniana przez użytkowników.

W celu przybliżenia użytkownikom mechanizmu maskowania wprowadzono w wielu systemach tzw. opcję "Q" - Quick Search (Szybkie Przeszukiwanie), która spotkała się z bardzo przychylnym przyjęciem i ogromnie zyskuje na popularności. Do wyszukania danego dokumentu użytkownik musi "wypisać" tylko 8 liter, np. 4 pierwsze litery nazwiska autora i 4 pierwsze litery tytułu z wyłączeniem przedimków (czasami oddzielone przecinkiem). Wbrew obawom wyszukiwanie tego typu nie prowadzi do szumu informacyjnego (pomyłkowych trafień), np. w bazie danych Uniwersytetu w Sussex /22/ liczącej 290 000 woluminów opcja Quick Search w większości wypadków dawała jedno

trafienie rzadziej dwa, a tylko w nielicznych wypadkach więcej. Opcja "Q" została bardzo przychylnie przyjęta przez użytkowników (szczególnie nie mających wprawy w korzystaniu z klawiatury), ponieważ znacznie skraca czas wpisywania poleceń i zmniejsza ryzyko błędnego wpisania terminu. Badania przeprowadzone przez wspomnianą bibliotekę w zakresie wykorzystania katalogu online wykazały, że w jednym dniu na 6750 poszukiwań (sesji wyszukiwawczych) aż 2089 (30,9%) dokonanych było według opcji "Q", pozostałe to: 1868 (27,7%) według autora, 2109 (31,2%) według tytułu, 501 (7,4%) według słów kluczowych, 183 (2,7%) według symbolu klasyfikacyjnego. Warto tu zwrócić uwagę, że dla 70% swych poszukiwań użytkownicy wybrali klucze wyszukiwawcze, które w tradycyjnych katalogach w ogóle nie występowały (Quick Search, Tytuł i Słowa kluczowe). Świadczy to o popularności nowych metod oraz braku niechęci użytkowników w stosunku do nowych technik wyszukiwawczych, nie stosowanych w tradycyjnych katalogach (czego nie można powiedzieć o środowisku bibliotekarskim).

ROLA SŁOWNIKA

Wprowadzenie niekontrolowanego słownictwa jako bazy wyszukiwawczej stwarza konieczność opracowania słownika eliminującego niekorzystne cechy takiego rozwiązania i umożliwiającego wzbogacenie zapytań informacyjnych o synonimy, quasisynonimy lub terminy pokrewne. Słowniki takie istniały już od dawna w systemach opracowania rzeczowego (np. tezaury, wykazy haseł przedmiotowych), lecz pełniły dotychczas inną funkcję. W systemach tradycyjnych wykorzystywane były jako normatywne wykazy terminów dozwolonych przy indeksowaniu, obecnie mają pełnić rolę narzędzi pomocniczych ułatwiających formułowanie instrukcji wyszukiwawczych, a w szcze-

gólności dobór jednostek leksykalnych powiązanych różnego typu relacjami z jednostką leksykalną charakteryzującą główny temat zawarty w zapytaniu.

Zmienia się też miejsce słownika w systemie. W pierwszych zautomatyzowanych katalogach słownik funkcjonował niezależnie od zbioru wyszukiwawczego; można było z niego korzystać, ale tylko na specjalne życzenie. Jeśli był w formie drukowanej, znajdował się w pokoju katalogowym i użytkownik mógł po niego sięgnąć, aby odszukać odpowiedni fragment (hasło); jeśli był w postaci komputerowej czytelnik musiał wywołać odpowiedni plik i ponownie odszukać odpowiedni element. Obecnie dąży się do wbudowania słownika (wraz z odsyłaczami) w bazę danych katalogowych lub w interfejs użytkownika. Słownik jest w ten sposób niejako zawsze "obecny" przy wyszukiwaniu. Użytkownik może z niego korzystać nie przerywając poszukiwań, przy czym system sam wybiera odpowiednie fragmenty (pozycje) słownika i wyświetla je na ekranie nie usuwając danych dotychczasowych (np. w specjalnie do tego przeznaczonym oknie). Zwiększa się przy tym "aktywność" słownika, system sam powinien kierować użytkownikiem, podpowiadać mu, podsuwać alternatywne terminy lub rozwiązania. Sytuacja taka może wystąpić, jeśli użytkownik otrzymał zbyt dużą liczbę pozycji w odpowiedzi na swoje pytanie początkowe (system może zasugerować terminy węższe, bardziej szczegółowe, lub dodatkowe, zawężające zapytanie;) lub przeciwnie - kiedy "trafień" jest zbyt mało albo wcale (zero dokumentów wyszukanych); system sugerować może uogólnienie zapytania lub dodatkowe terminy pokrewne.

Decyzja co do liczby wyszukanych dokumentów zaakceptowanej przez system (uznanej za odpowiednią i wyświetlonej na ekranie w odpowiedzi na pytanie) zależy od biblioteki. W bibliotece w

Newcastle /11/ przyjęto jako górną granicę liczbę 30 opisów dokumentów. Jeśli w odpowiedzi zostaje wyszukana większa liczba rekordów system sam proponuje użytkownikowi trzy sposoby dalszego postępowania (w postaci menu):

- zawężenie zapytania,
- przejrzanie wszystkich opisów po kolei, zaczynając od najnowszych,
- rozpoczęcie poszukiwań od nowa.

Jeśli użytkownik wybierze opcję pierwszą - zawężenie zapytania, system ponownie przedstawi mu kilka możliwości, tym razem sześć; ograniczenie ilości dokumentów poprzez:

- rok wydania;
- język dokumentu;
- identyfikator typu dokumentu;
- tytuł;
- temat;
- słowo kluczowe

Dla zilustrowania dalszego działania systemu przyjmijmy, że użytkownik wybrał ograniczenie poprzez dodatkowe słowo kluczowe (jest to jednocześnie przykład ilustrujący wpisanie mechanizmów logiki Boole'a w system). Użytkownik podał jako termin wyjściowy RAILROADS (koleje) i system poinformował go, że otrzymał 120 trafień, jednocześnie proponując ograniczenie zapytania. Użytkownik wybiera ograniczenie poprzez słowo kluczowe wpisując termin BRITAIN. System po raz drugi przeszukuje cały zbiór i wyświetla wszystkie słowa pasujące do zadanego pytania (spełniające wymienione kryteria), przy czym wykonywane jest to na zasadzie prawostronnego maskowania wpisanego implicite w program. Na ekranie pojawiają się takie terminy, jak: Britaine, Britains, Britain's, Britain-Cuba itp. Użytkownik z podanej listy terminów (wraz z liczbą dokumentów

im przypisanych) wybiera ten, który mu odpowiada, np. "Britain" i na ekranie uzyskuje wszystkie zapisy indeksowe danego słowa, wraz z informacją, w jakiego typu polu wystąpiły, np.

Termin BRITAIN pojawił się w dokumentach

- R1 8 razy jako nazwa autora
- R2 5786 - jako temat
- R3 12226 - jako określnik geograficzny
- R4 1379 - jako nazwa korporatywna
- R5 19 - jako nazwa konferencji
- R6 1751 - jako tytuł.

Użytkownik ponownie wybiera jedną z tych możliwości w celu przejrzania opisów wyszukanych dokumentów. System kolejny raz przeszukuje cały zbiór automatycznie łącząc za pomocą boolowskiego AND rezultaty obu wyszukiwań. Jeśli znów uzyskał więcej niż 30 opisów system kolejny raz prosi o zawężenie pytania, choć zawsze daje możliwość przejrzania wszystkich rekordów.

Także pozostałe informacje z tezaury, jak synonimy, terminy pokrewne, itp. mogą być wyświetlane w trakcie prowadzenia wyszukiwań. I tu też system powinien sam prowadzić użytkownika po wszystkich możliwych terminach, dając mu tylko szansę wyboru jednego z nich, nie przerzucając na użytkownika ciężaru ich odszukania.

W systemach stosujących słownictwo znormalizowane system odsyłaczy od różnych wariantów terminów wbudowany powinien być w interfejs użytkownika. Coraz częściej stosowane jest rozwiązanie, w którym użytkownik po wybraniu jednej z form (jednego z synonimów) nie jest "karany" koniecznością odszukania i użycia formy przyjętej (obowiązującej), lecz zostaje mu udzielona odpowiedź według formy poprawnej, jedynie z delikatnym napomknieniem, jaka jest ta forma (w celu uniknięcia nieporozumień); termin jest więc jakby automatycznie tłumaczony na słowa JIW.

SYSTEMY PRZYJAZNE UŻYTKOWNIKOWI

Inną charakterystyczną tendencją widoczną w katalogach online, mającą ogromny wpływ na efektywność i jakość wyszukiwania rzeczowe- go jest projektowanie interfejsów użytkownika dostosowanych stop- niem trudności i możliwościami do zindywidualizowanych potrzeb i umiejętności użytkowników. W pierwszych katalogach online system "podpowiadania" użytkownikowi oraz program instruktażowy były bardzo słabo rozbudowane. W konsekwencji katalogi te wymagały wykwalifikowanego, przeszkolonego użytkownika lub pośrednictwa bibliotekarza w prowadzeniu poszukiwań. Szybko jednak stwierdzono, że zalety OPAC, takie jak szybkość wyszukiwania giną przy systemie trudnym do nauczenia się i stosowania, gdy użytkownikowi dużo czasu zajmuje samo sformułowanie zapytania (zdefiniowanie instruk- cji wyszukiwawczej). Typowy język komend wykorzystywany w większoś- ci pakietów w połączeniu ze skomplikowaną procedurą stosowania operatorów boolowskich sprawia, że poprawne sformułowanie instruk- cji wyszukiwawczej dla nieprzeszkolonego użytkownika jest zadaniem przerastającym jego możliwości i prowadzi do zniechęcenia i nieza- dowolenia z systemu. Jedną z form sformułowania nawet bardzo prostego zapytania, np. dotyczącego książek na temat programowania w języku Pascal może być zapis (wg pakietu STATUS):

Q Pascal + Program & + Data > 1982 ?

W systemach typu publicznego (dla użytkowników nieprzygotowa- nych) komendy typu "Q" lub "Data" muszą być zastąpione prostymi znakami zachęty (prompts) lub pytaniami; użytkownik "wypełnia" tylko puste miejsca odpowiednimi danymi.

Ze względu na technikę prowadzenia wyszukiwań i stopień jej trudności wyróżnić można dwie kategorie OPAC:

1. OPAC stosujące język komend (proces wyszukiwania kontrolowany przez użytkownika), gdzie użytkownik musi znać odpowiednie rozkazy i ich składnię, aby spowodować wykonywanie poleceń przez system. Możliwości systemu są ukryte przed użytkownikiem. System taki jest sprawny i szybki, ale przeznaczony tylko dla użytkowników posiadających pewne przeszkolenie i doświadczenie w korzystaniu z systemu.

2. OPAC oparte na technice menu (proces wyszukiwania kontrolowany przez system), gdzie poszczególne możliwości (opcje) są zawsze przedstawiane użytkownikowi na ekranie do wyboru; użytkownik posuwa się "krok po kroku" po poszczególnych etapach wyszukiwania. Katalogi tego typu są wygodne dla użytkownika początkującego, niewprawnego i nie wymagają uprzedniego przeszkolenia, są jednak zbyt wolne i czasochłonne przy częstym stosowaniu i irytujące dla użytkownika doświadczonego. Systemy tego typu są z reguły mniej elastyczne i dysponują skromniejszym zestawem środków wyszukiwawczych (mają ograniczone funkcje wyszukiwawcze).

Tak więc, o ile systemy stosujące język komend są lepsze dla osób zawodowo zajmujących się wyszukiwaniem lub przynajmniej często korzystających z systemu, to katalogi stosujące menu są jedynym rozwiązaniem dla użytkowników początkujących lub korzystających z systemu sporadycznie. Dlatego coraz częściej projektuje się systemy dysponujące obiema formami, każda z nich przeznaczona dla innej kategorii użytkowników, przy założeniu, że użytkownicy początkujący w miarę zdobywania wprawy w korzystaniu z systemu będą stopniowo przechodzić na język komend.

Obecnie uważa się, że łatwość i szybkość korzystania z systemu przy zapewnieniu bogatego aparatu wyszukiwawczego są najważniejszymi aspektami wyszukiwania w katalogach online. Interfejs użytkownika powinien być:

- przystosowany do użytkowników o różnym poziomie znajomości
 - a) biblioteki i praktyki katalogowej
 - b) komputerów
 - c) danego programu wyszukiwawczego
- łatwy do poznania bez konieczności przechodzenia odpowiedniego przeszkolenia.

Widoczna jest tendencja odchodzenia od tradycyjnych form szkolenia (kursy, pokazy, materiały drukowane), a przechodzenia na formy instruktażowe "wpisane" niejako w system.

Systemy dysponujące dwoma (lub więcej) poziomami wyszukiwania, jednym bardziej rozbudowanym, z większą gamą środków "podpowiadających" i kierujących użytkownikiem, drugim - skromniejszym, okazały się najbardziej efektywne ze względu na możliwość stosowania zróżnicowanego zestawu środków i strategii wyszukiwawczych.

Wspomniane poziomy mogą dotyczyć nie tylko technik i łatwości sposobu prowadzenia wyszukiwania, ale także samych możliwości wyszukiwawczych. Np. w systemie LS/2000 (wykorzystywanym m.in. w OCLC) podstawowe menu ukazujące się na ekranie daje następujące możliwości wyszukiwania /1/:

1. Autor
2. Tytuł
3. Hasło przedmiotowe lub słowo kluczowe
4. Inne

Opcja czwarta (inne) wprowadza dalsze możliwości, wśród nich: nazwę konferencji, nazwę serii, nazwę geograficzną (jako temat dokumentu), hasła z MeSH (Medical Subject Headings, wykaz haseł z zakresu medycyny), nazwy osobowe (jako temat dokumentu), nazwę konferencji (jako temat dokumentu), itp. Daje to jakby drugi poziom wyszukiwania z przeznaczeniem dla użytkowników o spe-

czajnym przygotowaniu (o czym świadczy m.in. fachowa terminologia stosowana w tych opisach, np. nazwa korporatywna) lub do bardziej skomplikowanych wyszukiwań.

Literatura

1. BAGNALL J., JEFFREYS A. First experience with OCLC's LS/2000 OPAC at Newcastle University Library. - Program 1986 vol. 20 No 2, s.151-163
2. BIELICKA L.A., ŚCIBOR E. Tendencje rozwoju języków informacyjnych w Polsce do 1990 i 2000 roku. APID 1987 nr 2, s. 8-15.
3. The CARD catalog: current issues. Readings and selected bibliography. Ed. by C. Ryans. The Scarecrow Press, Metuchen 1981
4. COCHRANE P. Redesign of catalogs and indexes for improved on-line subject access. DRYX PRESS 1985 Phoenix
5. COCHRANE P, ARRET L. Online library users - the library comes of age. - Online 1985 vol. 9 No 5, s.58-70
6. CONGREVE J. Problems of subject access (I): automatic generation of printed indexes and online thesaural control. - Program 1986 vol. 20 No 2, s.204-210
7. COWLEY R. Use of ALS Browser terminal for OPAC searching at Bromley Public Library. - Program 1986 vol.20 No 2, s.164-177
8. CROUCHER Ch. Problems of subject access (II): user studies and interface design. - Program 1986 vol. 20 No 2, s.211-224
9. FAYEN E.G.: The online catalog. Improving public access to library materials. Knowledge Industry Publications, Inc. 1983 White Plains
10. HUNTER E.J., BAKEWELL K.G.B. Cataloguing. 2nd ed. Clive Bingley 1983 London
11. LUNDEEN G., TENOPIR C. Microcomputer Software for in-house databases... Four top packages under \$ 2000. - Online 1985 vol.9 No 5, s.30-38

12. MITEV N.N. Users and ease of use: online catalogues raison d'etre. - Program vol. 20 1986, No 2, s.11-119
13. ONLINE Public Access to Library Files: Conference Proceedings. The proceedings of a conference held at the University of Bath 3-5 September 1984. (Ed. by J. Kinsella). Elsevier Internat. Bull. 1985 Mayfield House
14. POTTER W.G. Online catalogues in North America: an overview. -Program 1986 vo¹ 20 No 2, s.120-130
15. PROWSE S.G. Use of BRS/Search in OPAC experiments. - Program 1986 vol. 20 No 2, s.178-195
16. SEAL A., BRYANT P., HALL C. Full and short entry catalogues. Centre for Catalogue Research 1982 Bath University Library
17. SETTEL B., COCHRANE P. Augmenting subject descriptions for books in online catalogs. Database 1982 vol. 5 No 4, s.29-37
18. SUBJECT Access. Report of a meeting sponsored by the Council on Library Resources. Dublin, Ohio, June 7-9 1982. Compiled and edited by K. W. Russel. Council on Library Resources, Washington, D.C., December 1982
19. SUBJECT access and the MARC record. Report of a questionnaire conducted by the Marc Users' Group. Reding, Marc Users' Group, 1985
20. TESKEY F.N. Information retrieval systems for the future, Library and Information Research Report 26; The British Library Board 1984, London
21. VIGIL P.J. The software interface. Annual Review of Information Science and Technology 1986 vol.21, s.63-86
22. WHITSED N. Using an information retrieval Package CAIRS as an OPAC in a small, university medical library. - Program 1986 vol. 20 No 2 s.196-203
23. YOUNG R.C., LEE S.R.: The Geac-based OPAC at the University of Sussex Library. - Program 1986 vol.20 No 2, s.138-150.

12.02.1990 r.

SUBJECT RETRIEVAL IN THE ONLINE PUBLIC ACCESS CATALOGUES

Summary

The article deals with the history and the main advantages of the Online Public Access Catalogue (OPAC). There are presented the works carried out by the Library of Congress in Washington on the new treatments of the problem as well as the aspects of subject working out of documents in online catalogues. There are presented in detail the retrieval tools used in online catalogues: Boolean operators, arithmetic and positional operators, truncation. The role of the dictionary in these catalogues which should facilitate the formulation of the search strategy as well as other solutions making the use of the catalogue easier for an untrained user are discussed.

ПРЕДМЕТНЫЙ ПОИСК В ОБЩЕДОСТУПНЫХ ДИАЛОГОВЫХ КАТАЛОГАХ

(OPAC)

Резюме

Статья посвящена истории возникновения и основным преимуществам диалоговых каталогов (Online Public Access Catalogue - OPAC). Представлены работы Библиотеки конгресса в Вашингтоне по новым решениям в этой области, а также вопросы предметной обработки в каталогах online. Подробно обсуждены средства поиска используемые в каталогах online: булевы операторы, арифметические и позиционные операторы, усечение. Представлена роль словаря в этих каталогах, который должен облегчать формулировку поисковых инструкций и другие решения облегчающее использование этих каталогов неподготовленным потребителем.

BOGDAN TRAWIŃSKI
ARTUR ROZWADOWSKI
MIROSLAW ZIOLEK

Biblioteka Główna i OINT
Politechniki Wrocławskiej

**MODELE KOMPUTERYZACJI
PROCESÓW GROMADZENIA I OPRACOWANIA WYDAWNICTW CIĄGLYCH
W BIBLIOTECE NAUKOWEJ**

Funkcje mikrokomputerowego systemu gromadzenia i opracowania wydawnictw ciągłych. Trzy wariantowe modele komputeryzacji tego systemu: dla małej, średniej i dużej biblioteki. Model skomputeryzowanego Oddziału Gromadzenia i Opracowania Wydawnictw Ciągłych Biblioteki Głównej i OINT Politechniki Wrocławskiej.

Komputeryzacja procesów bibliotecznych związanych z wydawnictwami ciągłymi jest zagadnieniem złożonym i bardzo trudnym. Różnorodność wydawnictw uznawanych za ciągłe, różnorodność formatów i rozmieszczenia elementów tytułatury, nieregularność ukazowania się, liczne i częste zmiany tytułów, złożoność związanych z tymi wydawnictwami funkcji bibliotecznych, sprawiają, że do tej pory zagadnienie komputeryzacji nie zostało rozwiązane w sposób kompleksowy /5/. Na rynku amerykańskim oferowanych jest obecnie kilkadziesiąt systemów kontroli wydawnictw ciągłych takich, jak np. CHECKMATE, EBSCONET, LINX SC-10, OCLC Serials

"Zagadnienia Informacji Naukowej" 1990 nr 1(56)

Control Subsystem, NOTIS /7/.

Jednak nie w każdym wypadku uwzględniają one wszystkie czynności związane z gromadzeniem i opracowaniem wydawnictw ciągłych. Systemy te mają liczne ograniczenia i uproszczenia, np. w zakresie opisu bibliograficznego.

W Polsce, poza kilkoma stosunkowo prostymi systemami realizującymi pojedyncze funkcje, nie opracowano komputerowego systemu kontroli wydawnictw ciągłych, który obejmowałby podstawowe czynności procesów gromadzenia i opracowania takie, jak zamawianie prenumeraty, akcesja, reklamowanie, prowadzenie inwentarza. Co więcej nie stworzono dostatecznej bazy do projektowania takich systemów, a w tym przede wszystkim nie opracowano norm dotyczących wydawnictw ciągłych, określających zasady sporządzania opisu bibliograficznego, zasady utrzymywania katalogu i ewidencji zasobów biblioteki. W przeglądzie dokumentów normalizacyjnych z zakresu działalności informacyjnej sporządzonym przez H. Kłodnicką /3/ nie znajdzie się polskich odpowiedników takich norm jak International Standard Bibliographic Description for Serials /2, 5/ czy American National Standards for Information Sciences - Serial Holdings Statements /1/.

W Bibliotece Głównej i DINT Politechniki Wrocławskiej prowadzone są prace nad stworzeniem mikrokomputerowego systemu gromadzenia i opracowania wydawnictw ciągłych w bibliotece naukowej.

Zamierzeniem projektantów jest opracowanie systemu modelowego, to znaczy takiego systemu, aby możliwe było adaptowanie go do różnych warunków i wymagań. Pozwoliłoby to na wdrażanie systemu w bibliotekach naukowych o różnej wielkości i odmiennej organizacji procesów gromadzenia i opracowania wydawnictw ciągłych.

Projekt modelowego systemu należy oprzeć na głębokiej analizie środowiska, w którym system ten będzie funkcjonował. W wyniku tej analizy należy określić wymagania i funkcje, jakie system będzie musiał spełniać, a następnie opracować modele skomputeryzowanego środowiska, w którym system wykonując założone funkcje, spełniał będzie określone potrzeby użytkowników.

Przedmiotem niniejszej pracy jest prezentacja ogólnych modeli komputeryzacji procesów bibliotecznych związanych z wydawnictwami ciągłymi oraz rozważenie konkretnego przypadku-projektu komputeryzacji Oddziału Gromadzenia i Opracowania Wydawnictw Ciągłych w Bibliotece Głównej i OINT Politechniki Wrocławskiej.

Przedstawione tu modele powstały w wyniku analizy procesów gromadzenia i opracowania wydawnictw ciągłych w kilkunastu bibliotekach naukowych w kraju /8/ oraz przeglądu światowych rozwiązań w zakresie automatyzacji wydawnictw ciągłych /6/.

FUNKCJE MIKROKOMPUTEROWEGO SYSTEMU GROMADZENIA I OPRACOWANIA WYDAWNICTW CIĄGŁYCH

Przegląd komputerowych systemów kontroli wydawnictw ciągłych eksploatowanych w USA pokazał, że systemy te posiadają wiele funkcji i możliwości /6/. Ukształtowany w Polsce system kolportażu i gromadzenia wydawnictw ciągłych, utrwalone organizacyjne zasady opracowania i udostępniania oraz brak ogólnokrajowej sieci komputerowej, znikoma komputeryzacja i automatyzacja procesów bibliotecznych sprawiają, że nie wszystkie funkcje systemów amerykańskich mogłyby być realizowane w Polsce.

Poniżej wyszczególniono te funkcje, w które może zostać wyposażony mikrokomputerowy system gromadzenia i opracowania

wydawnictw ciągłych przeznaczony do wykorzystania w warunkach krajowych.

- Wspomaganie selekcji czasopism, a w tym utrzymywanie zbioru dezyderatów, tj. tytułów czasopism zgłoszonych do zaprenumerowania.

- Wspomaganie prenumeraty czasopism, a w tym utrzymywanie zbioru tytułów czasopism aktualnie prenumerowanych, utrzymywanie zbioru kolporterów, wspomaganie aktualizowania wydruków dostarczanych przez kolporterów poprzez wyświetlanie lub drukowanie listy czasopism prenumerowanych za pośrednictwem danego kolportera, sporządzanie pism z zamówieniami.

- Wspomaganie obsługi finansowej gromadzenia, a w tym ewidencja rachunków, ewidencja opłat za prenumeratę, za zakup, sporządzanie pism finansowych.

- Wspomaganie akcesji czasopism, a w tym utrzymywanie zbiorów akcesyjnych, utrzymywanie rekordów z opisami bibliograficznymi akcesjonowanych tytułów, predykcja numerów zeszytów do akcesji, wskazywanie dubletów.

- Automatyczne reklamowanie opóźnień i braków w realizacji prenumeraty, a w tym generowanie list reklamowanych zeszytów, sporządzanie i drukowanie pism z reklamacjami, ewidencja reklamacji.

- Wspomaganie uzupełniania zbiorów, a w tym sporządzanie wykazów brakujących numerów, ewidencja wniosków o uzupełnienie brakujących numerów.

- Wspomaganie prowadzenia wymiany, a w tym utrzymywanie zbiorów kontrahentów, utrzymywanie zbioru wysyłanych tytułów, utrzymywanie zbioru otrzymywanych tytułów, drukowanie danych adresowych kontrahentów dla wysyłanych tytułów, sporządzanie i drukowanie pism.

- Wspomaganie sygnowania, a w tym automatyczne generowanie sygnatur dla nowych tytułów.

- Automatyczne tworzenie inwentarza i katalogu, a w tym generowanie rekordów zbioru inwentarza (katalogu) na podstawie rekordów zbiorów akcesyjnych.

- Wspomaganie przygotowywania woluminów do oprawy, a w tym generowanie list czasopism gotowych do oprawy, sporządzanie opisów oprawianych woluminów.

- Utrzymywanie katalogu on-line wydawnictw ciągłych, a w tym umożliwienie wyszukiwania według różnych elementów opisu bibliograficznego, umożliwienie przeglądania tytułów w układzie alfabetycznym, umożliwienie uzyskania informacji o posiadaniu określonego zeszytu czasopisma z określonego roku (tomu), umożliwienie uzyskania informacji adresowych o bibliotekach instytutowych, zakładowych itp.

- Sporządzanie wydawnictw informacyjnych, a w tym sporządzanie wykazów zasobów biblioteki w różnych przekrojach, sporządzanie wykazów nabytków.

- Sporządzanie i drukowanie zestawień statystycznych.

- Kontrola obiegu wydawnictw ciągłych w jednostkach organizacyjnych do momentu przekazania do magazynu, a w tym utrzymywanie i prezentowanie użytkownikowi informacji o aktualnym miejscu przebywania określonego zeszytu.

Wykonując większość z powyższych funkcji system korzystać będzie z tego samego zbioru opisów bibliograficznych wydawnictw ciągłych na dysku magnetycznym. Dzięki temu uniknie się dublowania opisów, co jest nieodłączną cechą systemów tradycyjnych, w których manualnie utrzymuje się różnorodne kartoteki i katalogi kartkowe.

WARIANTY KOMPUTERYZACJI PROCESÓW GROMADZENIA I OPRACOWANIA

Proponowane przez autorów warianty komputeryzacji procesów gromadzenia i opracowania wydawnictw ciągłych w bibliotekach naukowych uwzględniają następujące przesłanki:

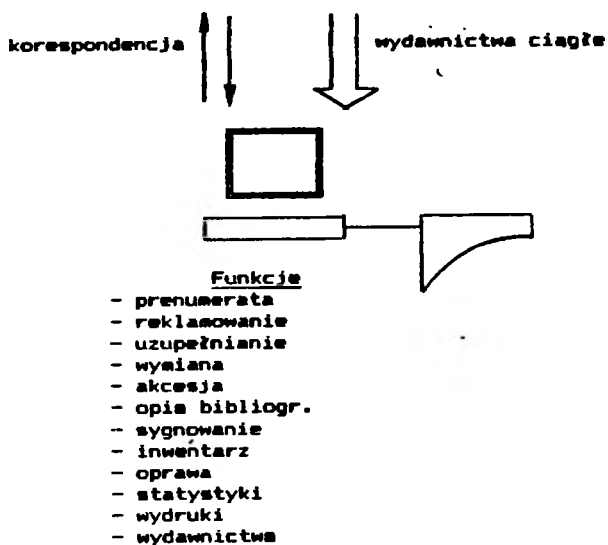
1/ Wyniki analizy procesów gromadzenia i opracowania wydawnictw ciągłych przeprowadzonej w wybranych bibliotekach naukowych wskazują, że nie ma jednolitego modelu ich organizacji. Czynności biblioteczne związane z tymi procesami mogą być wykonywane w jednej lub być rozproszone w kilku jednostkach organizacyjnych biblioteki.

Jednostki te mogą mieć różne usytuowanie terytorialne np. znajdować się na różnych piętrach czy nawet w różnych budynkach. Przy planowaniu komputeryzacji powinno się brać pod uwagę ukształtowany w danej bibliotece model organizacji procesów gromadzenia i opracowania. Nie oznacza to jednak, że komputeryzacja nie może prowadzić do zmian liczby stanowisk pracy oraz do innego podziału zadań wewnątrz jednostek organizacyjnych.

2/ Komputeryzację procesów bibliotecznych (w dużych bibliotekach) w obecnych warunkach można przeprowadzić przy pomocy mikrokomputerów kompatybilnych z IBM PC połączonych w lokalną sieć. Za przyjęciem takiego rozwiązania przemawia duża dostępność tego typu sprzętu, elastyczność budowy konfiguracji dostosowanej do potrzeb konkretnej biblioteki, duża znajomość i umiejętności posługiwania się oprogramowaniem narzędziowym tych komputerów. Jednakże, ze względu na stosunkowo wysokie koszty instalacji sieciowej, średnie i małe biblioteki zapewne nie będą miały wystarczających środków na jej zakup. Stąd też należy zaproponować również takie warianty komputeryzacji, które będą uwzględniać jednostanowiskowe, tańsze rozwiązania.

3/ W konkretnej bibliotece przy gromadzeniu i opracowaniu wydawnictw ciągłych pracują ludzie o określonych kwalifikacjach, doświadczeniu, wieku i stażu pracy. Należy to również brać pod uwagę.

Poniżej przedstawiono trzy warianty modelowe komputeryzacji uwzględniające wielkość biblioteki i związane z tym możliwości zakupu lokalnej sieci mikrokomputerowej oraz model organizacji procesów gromadzenia i opracowania wydawnictw ciągłych w bibliotece. Na rysunkach zostały uwidocznione jedynie stanowiska pracy wyposażone w mikrokomputery. Oprócz nich, w danej jednostce organizacyjnej biblioteki będą występować inne, równie ważne stanowiska, nie wymagające bezpośredniego dostępu do sprzętu mikrokomputerowego. Przydział funkcji do poszczególnych stanowisk należy traktować jako przykładowy - może on ulec zmianie w zależności od potrzeb i warunków danej biblioteki. Również pokazana na rysunkach liczba mikrokomputerowych stanowisk pracy w danej jednostce organizacyjnej stanowi jedno z możliwych rozwiązań.

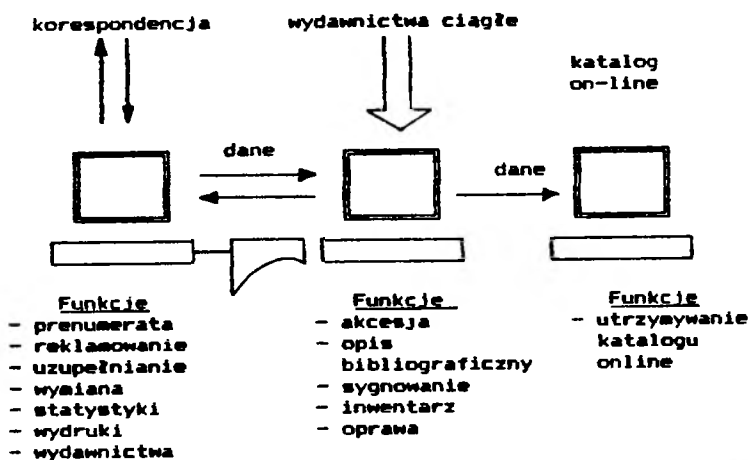


Rys.1. Schemat ideowy wariantu jedno stanowiskowego

Wariant jednostanowiskowy przeznaczony jest dla małej biblioteki, gdzie możliwe jest zrealizowanie wszystkich czynności na jednym mikrokomputerze. Procesy gromadzenia i opracowania wydawnictw ciągłych muszą być scentralizowane, to jest realizowane w jednym oddziale. W wypadku, gdy wymianę prowadzi inny oddział np. oddział druków zwartych - również skomputeryzowany, to akcesja wydawnictw ciągłych pochodzących z wymiany może być wykonywana w tym oddziale, a dane do oddziału wydawnictw ciągłych będą przekazywane na dyskietkach.

Nie przewiduje się utrzymywania katalogu on-line przeznaczonego dla użytkowników biblioteki, a jako dodatkowe narzędzia wspomagające tradycyjny katalog kartkowy, komputer będzie drukował różnorodne wykazy zasobów biblioteki.

Wariant wymaga stosunkowo taniego sprzętu: 1 IBM PC z napędem dysków elastycznych 360 kB, twardym dyskiem 20 MB oraz drukarką.



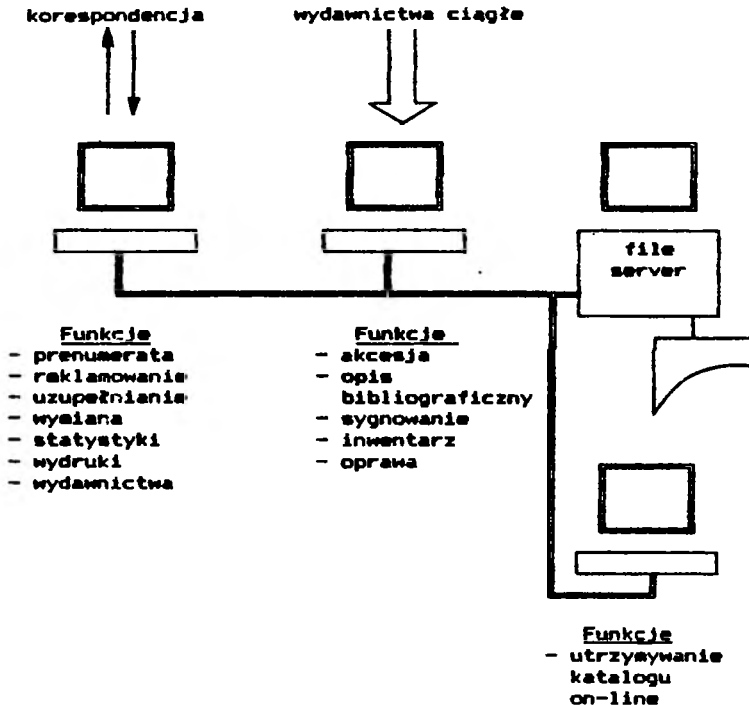
Rys.2. Schemat ideowy wariantu B dla biblioteki średniej wielkości

Wariant przeznaczony jest dla biblioteki średniej wielkości, w której jeden mikrokomputer nie wystarczy na zrealizowanie wszystkich funkcji, a zakup instalacji sieciowej nie jest uzasadniony lub możliwy w danym momencie ze względu na szczupłość środków finansowych.

W wariacie tym oddzielono funkcje opracowania od funkcji gromadzenia. Dlatego też jest możliwe zastosowanie tego wariantu zarówno w bibliotekach, w których procesy gromadzenia i opracowania wydawnictw ciągłych skupione są w jednym oddziale, jak i w bibliotekach, gdzie procesy te realizowane są w osobnych oddziałach. Przekazywanie danych pomiędzy mikrokomputerami odbywa się na dyskietkach.

Przewiduje się możliwość zainstalowania na osobnym komputerze katalogu on-line wydawnictw ciągłych i ewentualnie katalogu druków zwartych, jeżeli oddział druków zwartych jest skomputeryzowany. Przenoszenie danych do aktualizacji katalogu lub przenoszenie zaktualizowanego katalogu następuje za pośrednictwem dyskietek.

Wariant ten wymaga stosunkowo taniego sprzętu: 2 (3) komputerów IBM IBM PC XT/AT z napędami dysków elastycznych 360 KB, twardymi dyskami 20 KB oraz jedną drukarką.



Rys.3. Schemat ideowy wariantu dla dużej biblioteki

Przeznaczony jest dla dużej biblioteki, w której jest uzasadniony oraz możliwy finansowo zakup lokalnej sieci mikrokomputerowej. Jest to sieciowa wersja wariantu B. Procesy gromadzenia i opracowania są rozdzielone na różne stanowiska pracy, które mogą znajdować się w jednym oddziale (model scentralizowany procesów) lub w osobnych oddziałach (model rozproszony procesów).

W wariantcie tym wymagany jest drogi sprzęt:

- jeden mikrokomputer IBM AT286 lub AT386, napędy dysków elastycznych 360 kB i 1,2 MB, twardy dysk 80 MB, drukarka, streamer 60 MB
- minimum 3 mikrokomputery IBM XT z napędami 360 kB i ewentualnie twardym dyskiem 20 MB

- karty sieciowe Arc-Net
- oprogramowanie sieciowe Novell.

Podstawowymi czynnikami, jakie należy uwzględnić przy wyborze wariantu komputeryzacji są: wielkość biblioteki, posiadane przez bibliotekę środki finansowe, struktura organizacyjna biblioteki oraz kwalifikacje zatrudnionego personelu.

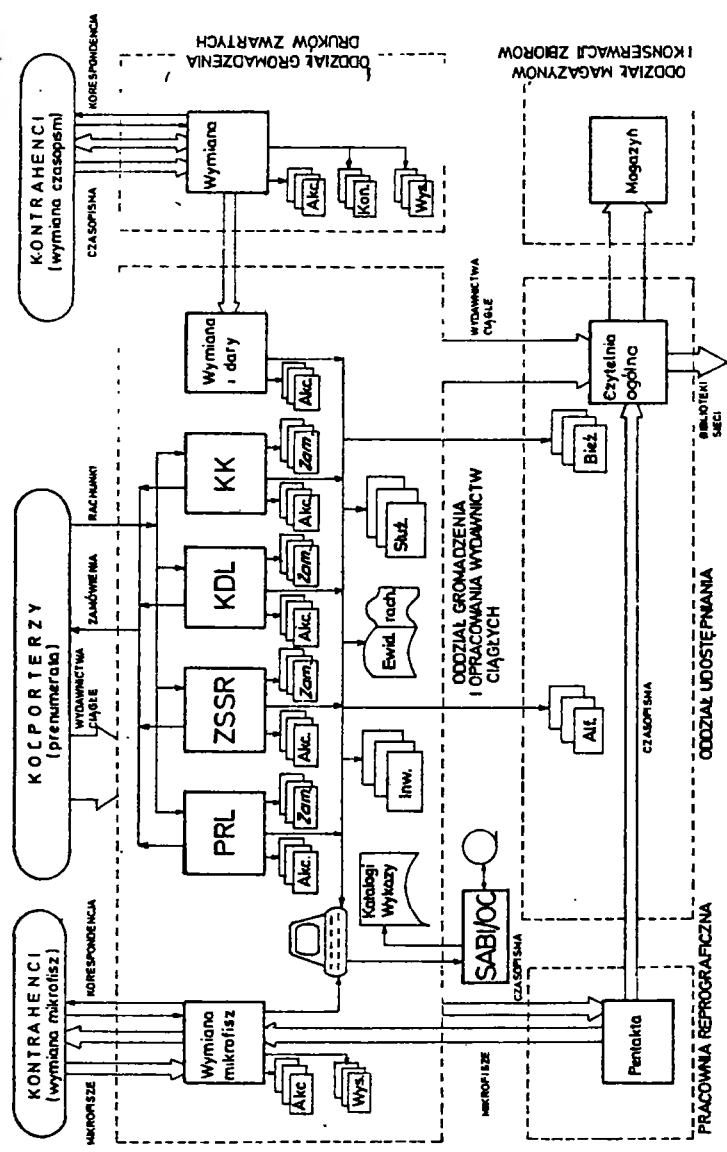
MODEL SKOMPUTERYZOWANEGO ODDZIAŁU GROMADZENIA I OPRACOWANIA WYDAWNICTW CIĄGŁYCH W BIBLIOTECE GŁÓWNEJ I OINT POLITECHNIKI WROCŁAWSKIEJ

Organizacja procesów gromadzenia i opracowania wydawnictw ciągłych

W Bibliotece Głównej i OINT Politechniki Wrocławskiej (BG i OINT PWr.) procesy biblioteczne dotyczące wydawnictw ciągłych są scentralizowane w jednej jednostce organizacyjnej, w Oddziale Gromadzenia i Opracowania Wydawnictw Ciągłych (OGiOWC). Z kolei, procesy dotyczące druków zwartych realizowane są w odrębnych jednostkach organizacyjnych - w Oddziale Gromadzenia Druków Zwartych (OGDZ) oraz w Oddziale Opracowania Druków Zwartych (OODZ).

Na rys. 4 przedstawiono schemat organizacji procesów gromadzenia i opracowania wydawnictw ciągłych w BG i OINT PWr. Na rysunku tym do oznaczenia kartotek użyto następujących skrótów:

- Akc. - kartoteka akcesyjna
- Zam. - kartoteka fiszek zamówieniowych
- Inw. - inwentarz roczników w układzie topograficznym
- Służb. - katalog służbowy w układzie alfabetycznym
- Alf. - centralny katalog alfabetyczny
- Bież. - katalog alfabetyczny czasopism bieżących
- Kon. - kartoteka kontrahentów
- Wys. - kartoteka wydawnictw ciągłych (mikrofisz) wysyłanych
- Ewid. rach. - księga ewidencji rachunków



Rys. 4. Schemat organizacji procesów gromadzenia i opracowania wydawnictw ciągłych w Bibliotece Główniej i OIIN (Politechniki Wrocławskiej)

Z procesami gromadzenia i opracowania wydawnictw ciągłych w OGiOWC bezpośrednio związanych jest 6 stanowisk pracy. Cztery stanowiska obsługują prenumeratę, przy czym zadania podzielone są na 4 grupy krajów: PRL, ZSRR, KDL oraz KK. Na każdym z tych czterech stanowisk wykonywane są wszystkie czynności gromadzenia i opracowania, począwszy od przygotowania zamówień na prenumeratę, skończywszy na aktualizowaniu centralnego katalogu alfabetycznego. Pozostałe stanowiska to opracowanie wydawnictw ciągłych otrzymywanych drogą wymiany i w formie darów oraz obsługa wymiany mikrofilmów.

Obieg wydawnictw ciągłych

Wydawnictwa ciągle zaprenumerowane przychodzą bezpośrednio do OGiOWC, wydawnictwa otrzymywane drogą wymiany i darów trafiają najpierw do OGDZ. Wydawnictwa pochodzące z wymiany po akcesji, a przekazane w formie darów - bezpośrednio, są przekazywane z OGDZ do OGiOWC. Po opracowaniu część czasopism, około 400 tytułów z KK, jest przekazywana do Pracowni Reprograficznej w celu mikrofilmowania. Pozostałe wydawnictwa ciągle są przekazywane do Oddziału Udostępniania - do Czytelni Ogólnej. Tam wydawnictwa ciągle są przechowywane i udostępniane przez okres co najmniej jednego roku do czasu skompletowania rocznika, po czym przekazywane są do Oddziału Magazynów i Konserwacji Zbiorów.

W BG i OINT PWr., odmiennie aniżeli w wielu innych bibliotekach naukowych w Polsce, czasopisma po skompletowaniu roczników nie wracają do OGiOWC, w celu dokończenia ich opracowania. Inwentarz sporządzany jest na podstawie danych zawartych w kartotekach akcesyjnych oraz w dokumentach finansowych.

Model komputeryzacji procesów gromadzenia i opracowania wydawnictw ciągłych w BG i OINT PWr.

Podstawą proponowanego modelu skomputeryzowanego Oddziału Gromadzenia i Opracowania Wydawnictw Ciągłych jest następująca ocena organizacji procesów gromadzenia i opracowania wydawnictw ciągłych:

1. Większość czynności związanych z gromadzeniem i opracowaniem wydawnictw ciągłych skupionych jest w jednej jednostce organizacyjnej - OGiOWC - sprzyja to wprowadzaniu komputeryzacji.

2. Organizacja stanowisk obsługi prenumeraty wydawnictw ciągłych wg grup krajów (PRL, ZSRR, KDL, KK) nie wydaje się właściwa. Skupienie zadań związanych z zamawianiem i rozliczaniem prenumeraty oraz z opracowaniem wydawnictw ciągłych na tym samym stanowisku nie sprzyja, zdaniem autorów, wydajności pracy. W kilku nawet dużych bibliotekach czynność akcesji wszystkich wydawnictw ciągłych wykonywana jest na jednym wydzielonym stanowisku /8/. Rozdzielenie funkcji gromadzenia i opracowania na odrębne stanowiska wydaje się być bardziej efektywnym rozwiązaniem.

3. Dublowanie czynności akcesji wydawnictw ciągłych otrzymywanych drogą wymiany nie jest właściwe, chociaż przy obecnej organizacji procesów w BG i OINT PWr. jest konieczne. Prowadzenie wymiany mikrofisz oraz wymiany wydawnictw w formie drukowanej ma wiele wspólnych cech. M.in. wymaga prowadzenia podobnej korespondencji oraz utrzymywania podobnych zbiorów danych (kartotek). Wydaje się, że zlikwidowanie w OGDZ stanowiska wymiany i połączenie go ze stanowiskiem wymiany mikrofisz w OGiOWC mogłoby okazać się rozwiązaniem bardziej efektywnym.

Biorąc pod uwagę powyższe oceny zaproponowano następujący wariant komputeryzacji procesów gromadzenia i opracowania wydaw-

nictw ciągłych w BG i OINT. Został on przedstawiony na rysunku 5. Dla tego wariantu zaproponowano strukturę funkcjonalną Oddziału. Strukturę tę przedstawiono na rysunku 6.

Opis modelu

- Na odrębnych stanowiskach skupiono czynności związane z prowadzeniem korespondencji i rozliczeń finansowych 2/ oraz czynności związane z obsługą spływających zeszytów wydawnictw ciągłych 3/ i 4/.

- Pracownicy na stanowiskach 3/ i 4/ dzielić będą między sobą czas pracy przy mikrokomputerach. W czasie, gdy jedna osoba będzie prowadziła akcesję, druga wypełniać będzie formularze pomocnicze do wprowadzania nowych opisów bibliograficznych lub poprawek do istniejących.

- Prowadzenie wymiany druków i mikrofisz skupiono na jednym stanowisku 5/.

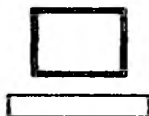
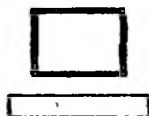
- Proponuje się zlikwidowanie wszystkich kartotek łącznie z centralnym katalogiem alfabetycznym - zastąpią je odpowiednie zbiory na nośnikach maszynowych, przede wszystkim na dyskach twardej. Katalogi on-line będą generowane automatycznie przez system.

- Wszelkie zestawienia statystyczne, wydawnictwa informacyjne, podręczne wykazy i katalogi będą generowane automatycznie przez system.

- Prowadzenie korespondencji, m.in. związanej z zamawianiem i reklamowaniem prenumeraty oraz z wymianą będzie w znacznym stopniu wspomagane przez system.

Podane na rysunku 6 liczby mikrokomputerowych stanowisk pracy są orientacyjne. Dopiero wdrożenie i przetestowanie wersji pilotowej systemu pozwoli ocenić wydajność pracy na poszczególnych

- 1) KIEROWNIK
- zarządzanie Oddziałem
 - autoryzowanie opisów bibliograficznych
 - klasyfikowania wydawnictw ciągłych



- 2)
- selekcja
 - prenumerata
 - reklamowanie
 - uzupełnianie
 - rachunki
 - inwentarz
 - korespondencja
 - statystyki
 - wydruki
- 3)
- akcesja
 - sygnowanie
 - opis bibliograficzny
- 4)
- akcesja
 - sygnowanie
 - opis bibliograficzny
- 5)
- wymiana mikrofisz
 - wymiana druków
 - akcesja wydawnictw wysyłanych
 - wydawnictwa informacyjna
- 6) PRACE TECHNICZNE
- rozpakowywanie paczek
 - selekcja materiałów
 - pieczętowanie zeszytów
 - wysyłka dokumentów
 - pakowanie mikrofisz
 - pakowanie druków do wysłania
- 7) ADMINISTRATOR SYSTEMU
- opieka nad sprzętem
 - opieka nad oprogramowaniem
 - rozwijanie oprogramowania

Rys. 6. Struktura funkcjonalna skomputeryzowanego OBIOWC.

stanowiskach pracy i określić liczbę mikrokomputerów. Również liczba stanowisk pomocniczych może po wdrożeniu systemu, w zależności od potrzeb, ulec zmianie.

W artykule określono miejsce i rolę mikrokomputerowego systemu gromadzenia i opracowania wydawnictw ciągłych w bibliotece naukowej. Przedstawione rozważania stanowią podstawę do opracowania projektu systemu.

Trudno na tym etapie w pełni określić spodziewane efekty wdrożenia systemu, tym bardziej że brak jest w kraju większych doświadczeń na tym polu. Podstawowym założeniem komputeryzacji jest polepszenie sprawności przetwarzania poprzez umożliwienie szybszego i dokładniejszego wykonywania operacji oraz poprzez wyeliminowanie redundancji charakterystycznej dla przetwarzania manualnego.

Zastosowanie technik automatycznych do operacji bibliotecznych powinno również polepszyć dostęp użytkowników do materiałów bibliotecznych. Osiągnie się to poprzez zwiększenie dokładności i aktualności danych w zbiorach, zastosowanie elastycznych strategii wyszukiwania, zmniejszenie czasu wymaganego na przetwarzanie.

Dużą rolę odegra zapewne eksperymentalne wdrożenie wersji pilotowej systemu w Bibliotece Głównej i OINT Politechniki Wrocławskiej. Pozwoli ono na:

- wszechstronne testowanie wszystkich modułów systemu na rzeczywistych zbiorach danych, w warunkach zbliżonych do docelowej eksploatacji,
- nabycie doświadczeń w szkoleniu personelu i przezwyciężaniu barier psychologicznych przy przejściu z manualnej technologii pracy na technologię mikrokomputerową,

- określenie "wydajności pracy" na poszczególnych stanowiskach pracy przy realizacji każdej z funkcji wspomaganej komputerowo,

- określenie i wprowadzenie niezbędnych zmian organizacyjnych zapewniających najbardziej wydajne funkcjonowanie skomputeryzowanego oddziału,

- nabycie doświadczenia w zakresie obsługi systemu, utrzymywania i zabezpieczania zbiorów.

Dzięki temu możliwe będzie opracowanie ulepszonych, bardziej precyzyjnych wersji systemu.

Literatura

1. AMERICAN National Standards for Information Sciences - Serial Holdings Statements. ANSI Z39.44-1986
2. ISBD(S): International Standard Bibliographic Description for Serials. Revised edition. London: IFLA 1988
3. KŁODNICKA H. Baza normatywno-techniczna w działalności informacyjnej. Przegląd dokumentów normalizacyjnych według stanu na 1987-11-30. Materiały metodyczne SINTO 1988 nr 32
4. MIĘDZYARODOWY Znormalizowany Opis Bibliograficzny Wydawnictw Ciągłych. ISBD(S). International Standard Bibliographic Description for Serials. Tłum. K. Pierkowska, red. M. Lenartowicz. Przegląd Biblioteczny 1979 R.47 z.1 s.19-77
5. ROZWADOWSKI A., ZIÓŁEK M., TRAWIŃSKI B.. Problemy automatyzacji zarządzania wydawnictwami ciągłymi. Zagadnienia Informatyki Naukowej 1989 nr 2(55) s.155-175
6. ROZWADOWSKI A., ZIÓŁEK M., TRAWIŃSKI B. Przegląd literaturowy rozwiązań w zakresie komputeryzacji procesów gromadzenia i opracowania czasopism na świecie. Raporty Biblioteki Głównej i OINT Politechniki Wrocławskiej 1989, Seria SPR nr 120
7. SCHMIDT N.P. Choosing an Automated Serials Control System. Serials Librarian 1984 vol. 9 no 1 s.65-86

8. ZIÓŁEK M., TRAWIŃSKI B., ROZWADOWSKI A. Analiza procesów gromadzenia i opracowania wydawnictw ciągłych w bibliotece naukowej (na przykładzie 14 wybranych bibliotek naukowych). Raporty Biblioteki Głównej i OINT Politechniki Wrocławskiej 1989, Seria SPR nr 119.

2.11.1989

MODELS OF COMPUTERIZATION OF THE PROCESSES OF ACQUISITION AND
OF ELABORATION OF SERIALS AT THE SCIENTIFIC LIBRARY

Summary

The microcomputer-base system of acquisition and of elaboration of serials should fulfill some basic functions comprising such a activities as: ordering of subscription, accession, advertising, maintaining of the accession register. There has been not created in Poland any basis for designing of such systems view of the lack of norms concerning serials (i.e. of the rules of preparing of bibliographic descriptions, of maintaining of catalogue and of keeping of the records of library resources). The purpose of the designers is to work out the model system in three variants: for a small, a middle and a great scientific library.

There are presented in the article the general models of computerization of the library processes related to the serials and elaborated for these three types of library. The models were made in the result of the analysis of the processes of acquisition and of elaboration of serials at the over ten scientific libraries in Poland and as a result of the review of the world-wide treatment in this field.

The is discussed in detail the model of the computerized Department of Acquisition and of the Department of Elaboration of Serials created for the needs of the Main Library and of the Scientific Information Centre of the Technical University in Wrocław.

**МОДЕЛИ АВТОМАТИЗАЦИИ ПРОЦЕССОВ НАКОПЛЕНИЯ И ОБРАБОТКИ
ПРОДОЛЖАЮЩИХСЯ ИЗДАНИЙ В НАУЧНОЙ БИБЛИОТЕКЕ**

Резюме

Микро-компьютерная система накопления и обработки продолжающихся изданий должна исполнять некоторые основные функции, к которым принадлежат такие действия как: заказы подписки, регистрация, реклам, ведение инвентаря. В Польше до сих пор не был создан базис для проектирования таких систем аз-за отсутствия норм относящихся к продолжающимся изданиям (принципов подготовки библиографического описания, принципов ведения каталога и учёта библиотечных фондов). Проектанты намерены разработать образцовую систему в трёх вариантах: для малой, средней и большой научной библиотеки.

В статье представлены общие модели автоматизации библиотечных процессов связанных с продолжающимися изданиями разработанных для этих трёх типов библиотек. Эти модели возникли как результат анализа процессов накопления и обработки продолжающихся изданий в нескольких научных библиотеках в Польше, а также обзора решений в этой области в мировом масштабе.

Подробно обсуждена модель автоматизированного Отделения накопления и обработки продолжающихся изданий разработана для целей Главной библиотеки и Центра научно-технической информации Вроцлавского политехнического института.

ELŻBIETA BARBARA ZYBERT

Instytut Bibliotekoznawstwa i
Informacji Naukowej UW

AMERYKAŃSKI SYSTEM INFORMACJI Z DZIEDZINY OŚWIATY

I WYCHOWANIA - ERIC

Działalność Educational Resource Information Center (ERIC). Struktura organizacyjna systemu i sieć specjalistycznych domów clearingowych. Podstawowe wydawnictwa systemu oraz baza danych ERIC.

Educational Resource Information Center - ERIC - (Centrum Informacji o Źródłach Oświatowych) jest na terenie Stanów Zjednoczonych Ameryki Płn. głównym ośrodkiem informacji o dokumentach publikowanych i niepublikowanych dotyczących wszystkich zagadnień z zakresu oświaty i wychowania. Centrum gromadzi informacje dotyczące rozwoju oświaty, badań oświatowych i kierunków poszukiwań pedagogicznych prowadzonych w Stanach Zjednoczonych oraz o najważniejszych osiągnięciach z zakresu pedagogiki, oświaty i wychowania poza USA (opisanych w języku angielskim); streszczenia różnorodnych publikacji z tej dziedziny; informacje o czasopiśmie pedagogicznych i oświatowych; umożliwia także reprodukcję materiałów zamówionych przez użytkowników.

Dzięki działalności ERIC użytkownicy stosunkowo łatwo docierają do takich cennych, a nie publikowanych i trudno dostępnych

"Zagadnienia Informacji Naukowej" 1990 nr 1(56)

materiałów, jak programy badań, założenia podejmowanych prób badawczych i ich wyników, raporty komisji rządowych i pozarządowych.

ERIC nie jest systemem komercyjnym i nie zarabia sam na siebie. Cała praca domów clearingowych, ich utrzymanie i opłacanie personelu pokrywane jest z funduszy federalnych. Jedynie, gdy w grę wchodzi kosztów reprodukcji dokumentów, taśm czy drukowania wydawnictw działalność ta jest odpłatna.

W kręgu tematów, którymi zajmuje się ERIC znajdują się: oświata i wychowanie na terenach wiejskich i miejskich, szkolnictwo podstawowe, średnie i wyższe, problemy dzieci specjalnej troski (odbiegających od normy tak w górę, jak i w dół), sprawy zawodu nauczycielskiego, informacji naukowej i bibliotekoznawstwa.

Swoją działalnością ERIC nie zastępuje instytucji specjalistycznych. Celem jego działania jest niesienie pomocy wszystkim osobom i instytucjom zainteresowanym sprawami oświaty i wychowania. Informacje gromadzone przez ten system mają przyczynić się do ulepszenia programów nauczania w szkołach i podniesienia jakości kształcenia.

STRUKTURA ORGANIZACYJNA ERIC

ERIC został zaprojektowany w 1965 roku i utrzymywany jest przez Ministerstwo Oświaty (Department of Education). Jego centrala mieści się obecnie w utworzonym w 1972 roku Narodowym Instytucie Oświaty (National Institute of Education) w Waszyngtonie.

Rozwój ERIC następował w trzech etapach. W początkowym okresie najwięcej czasu pochłaniało zdobycie i wybór najodpowiedniejszych materiałów. Prace drugiego etapu koncentrowały się wokół analizy zebranych materiałów. Etap trzeci obejmował wypracowanie najlepszych metod dotarcia do użytkowników.

Ze względu na to, że system szkolnictwa amerykańskiego jest zdecentralizowany, struktura ERIC jest oparta na zdecentralizowanej sieci domów clearingowych, bo ma służyć ściślejszemu powiązaniu tych placówek z lokalnymi organizacjami oświatowymi.

Na strukturę organizacyjną Centrum ERIC składa się 6 powiązanych ze sobą elementów:

- **Narodowy Instytut Oświaty** - pełniący funkcje koordynacyjne dla całego systemu ERIC. Działa w ramach Biura Badań i Rozwoju Oświaty (Office of Education Research and Development) Ministerstwa Oświaty.

- **Komórka ds. Ułatwień Informacyjnych i Przetwarzania Danych** (ERIC Processing and Reference Facility - ERIC PRF) pełniąca funkcje centralnego, ośrodka przetwarzania danych i informacji o całokształcie prac ERIC. Znajduje się w Bethesda (Maryland).

- **Służba Dokumentacji i Reprodukacji** (ERIC Documentation and Reproduction Service - EDRS) działająca przy Reproduction Computer Microform International Corporation w Arlington w Wirginii. Zajmuje się reprodukowaniem dokumentów, przygotowaniem i wydawaniem "Resource in Education" w wersji tradycyjnej i na taśmach magnetycznych.

- **Biuro Wydawnicze Rządu Stanów Zjednoczonych** (US Government Printing Office) drukujące "Resource in Education" i większość rządowych dokumentów z dziedziny oświaty.

- **Wydawnictwo Oryx Press** w Phoenix (Arizona), które wydaje "Current Index to Journals in Education" - drugie podstawowe wydawnictwo ERIC.

- **sieć specjalistycznych domów clearingowych**, znajdujących się przy ważniejszych organizacjach i uniwersytetach.

Obecna przedstawiona wyżej struktura uległa w ciągu minionych 20 lat pewnym zmianom. Miejsce obecnego Narodowego Instytutu Oświaty zajmował zarząd pracowników Biura Oświaty Stanów Zjednoczonych (US Office of Education), a poprzednikiem Oryx Press było wydawnictwo Crowell, Collier, Macmillan Corporation.

SIEĆ SPECJALISTYCZNYCH DOMÓW CLEARINGOWYCH ERIC

Główny ciężar działalności systemu spoczywa na sieci specjalistycznych domów clearingowych, spośród których 11 pierwszych powstało pomiędzy marcem a czerwcem 1966. Każdy z nich gromadzi materiały zgodnie z własnym profilem zainteresowań, opracowuje je, przechowuje i udostępnia. Rzadko działają one jako samodzielne instytucje. Jednym z zadań domów clearingowych jest gromadzenie materiałów dla dwóch podstawowych wydawnictw ERIC: "Resource in Education" i "Current Index to Journals in Education". Poza gromadzeniem i opracowaniem materiałów dla tych wydawnictw domy clearingowe wydają własne bibliografie i syntetyczne opracowania stanu badań. Czasem powstają tam oryginalne prace naukowe. Co roku placówki te organizują seminaria i spotkania dla osób związanych z oświatą i wychowaniem. W czasie tych spotkań tysiące specjalistów zostało przeszkolonych w zakresie korzystania z systemu.

Działalność domów clearingowych zakłócona jest jednak przez zbyt częste zmiany profilu ich zainteresowań, wynikające z zmiany zainteresowań instytucji macierzystych przy których zostały zlokalizowane. Zdarza się, że na skutek likwidacji lub reorientacji zainteresowań jakiejś placówki działalność domu clearingowego ulega zawieszeniu. W 1966 roku było ich 11, w 1970 - 19, a w 1988 - 16. W chwili obecnej są to następujące ośrodki:

ERIC Clearinghouse on Reading and Communication Skills (ERIC CRCS)

- Clearinghouse ds. Sprawności Czytania i Komunikacji. Działa przy Radzie Nauczycieli Języka Angielskiego w Urbanie (Illinois). W kręgu problemów interesujących ten ośrodek znajdują się: nauczanie języka angielskiego, umiejętność czytania, czytelnictwo, dziennikarstwo, krasomówstwo, interpretacja teatralna, radiowa, filmowa i telewizyjna.

Wydaje: "Language Art", English Journal".

ERIC Clearinghouse for Social Science Education (ERIC SSE) - Clearinghouse ds. Nauczania Nauk Społecznych. Działa w Boulder (Colorado) i ściśle współpracuje z Towarzystwem Nauczania Nauk Społecznych. Zajmuje się problematyką badań społecznych i nauczania nauk społecznych.

Wydaje:

- "Reference Sheet"-serię biuletynów informacyjnych zawierającą takie tytuły jak: "Career Education", "Competency Based Education", "Curriculum Development", "Legal Education", "Slower Learners", "Women Studies", "Values Education".

- "Looking at" - serię informacyjną obejmującą tytuły: "Teaching about aging", "Staff Development", "International Education", "Sex Education", "Teaching ERIC".

ERIC Clearinghouse on Information Resource (ERIC IR) - Clearinghouse ds. Źródeł Informacji. Działa przy School of Education w Syracuse (New York). Gromadzi materiały dotyczące źródeł informacji, nauki o informacji naukowej, bibliotekoznawstwa i bibliotek, a także technik i środków nauczania.

Wydaje:

- "ERIC/IR Update" - kwartalnik

Ukazały się także:

- An Alerting Service Bibliography on Educational Technology. 1980 r.
- Computer Based Education: the best of ERIC June 1976-Aug. 1980 . 1980 r.
- An overview of Videodisc Technology and Some Potential Applications in the Library Information and Instructional Sciences. 19880 r.

ERIC Clearinghouse on Language and Linguistics (ERIC CLL) - Clearinghouse ds. Języka i Lingwistyki. Działą przy Centrum Lingwistyki Stosowanej w Arlington (Wirginia). Gromadzi informacje o nauczaniu języków obcych, nauczaniu w środowiskach dwujęzycznych, dwukulturowych, a także o nauczaniu języków mało znanych. Zbiera materiały dotyczące nauczania języka angielskiego jako języka obcego, a także psycholingwistyki. Współpracuje z National Indo-chinese Clearinghouse on Technical Assistance Center, który zajmuje się problemami oświatowymi, językowymi i kultralnymi uchodźców z Wietnamu, Laosu, Kambodży.

Wydaje:

- The Language in the Education - serię, w której ukazują się: Directory of Foreign Language Service Organizations; Games and Stimulations in the Foreign Language Classroom; Graduate Theses and Dissertations in English; Intensive Foreign Language Courses.
- Bilingual Education - nieregularnie
- Linguistic Reporter - 9 razy w roku.

ERIC Clearinghouse on Test, Measurement, and Evaluation (ERIC TM)

- Clearinghouse ds. Testowania, Pomiarów i Oceny. Działą w Princeton (New Jersey) przy Służbie Testów w Nauczaniu. Zajmuje się gromadzeniem informacji o sposobach pomiaru i oceny procesów, technik i efektywności nauczania.

Wydaje:

- ERIC/TM Notes - nieregularnie
- ERIC/TM Update Series, w której ukazują się: Adversary Evaluation; Coaching for Tests; Selection and Evaluation of Evaluators.

ERIC Clearinghouse on Science, Mathematics, and Environmental Education (ERIC/SMEE) - Clearinghouse ds. nauczania Nauk Ścisłych, Matematyki i Oświaty Środowiskowej. Działają przy Uniwersytecie stanu Ohio w Columbus. Gromadzi informacje dotyczące nauczania w zakresie nauk ścisłych, matematyczno-przyrodniczych i techniki.

Wydaje:

- ERIC/SMEE Publication - nieregularnie

ERIC Clearinghouse on Junior College (ERIC JC) - Clearinghouse ds. Kolegiów Dwuletnich. Zlokalizowany jest przy Uniwersytecie Kalifornijskim w Los Angeles. Zajmuje się wszelkimi problemami związanymi z kształceniem na poziomie powyżej średniego i średnim.

Wydaje:

- Junior College Resource Review - nieregularnie

ERIC Clearinghouse in Adult, Career, and Vocational Education (ERIC/ACVE) - Clearinghouse ds. Oświaty Dorosłych i Kształcenia Zawodowego. Działają przy Uniwersytecie Stanowym w Columbus (Ohio). Zakres tematyczny to: zagadnienia edukacji zawodowej dorosłych, kształcenie ustawiczne, problemy dotyczące nowych specjalizacji zawodowych, umiejętności podejmowania decyzji i ról życiowych.

Wydaje szereg bibliografii i informatorów, m.in.:

- Energy Conservation Resource Materials. 1978,
- Agricultural Education. Review and Synthesis of the Research. 1978,
- Metric Education. An annotated bibliography. 1982.

ERIC Clearinghouse on Counseling and Personnel Service (ERIC/CPS) - Clearinghouse ds. Poradnictwa i Obsługi Kadr. Działą przy Uniwersytecie Stanowym w Ann Arbor (Michigan). Gromadzi materiały dotyczące poradnictwa w nauczaniu we wszystkich typach i poziomach szkolnictwa, teorii poradnictwa, przeprowadzania testów i korzystania z ich wyników, poradnictwa rodzinnego, charakterystyk studentów i pracowników oraz sposobów aktywizacji studentów.

Wydaje:

- CPS Capsule - kwartalnik
- Searchlight computer search of the ERIC data base - nieregularnie
- Searchlight Plus computer search of ERIC data base including analysis of trends and issues - nieregularnie.

ERIC Clearinghouse on Elementary and Early Childhood Education (ERIC/EECE) - Clearinghouse ds. Szkolnictwa Podstawowego i Nauczania we wczesnym Dzieciństwie. Działą w Urbanie przy Uniwersytecie stanu Illinois. Zajmuje się problemami kształcenia we wczesnym dzieciństwie, rozwojem dziecka, teorią i praktyką nauczania dziecka od urodzenia do 12 roku życia.

Wydaje:

- ERIC/EECE News Letter

ERIC Clearinghouse on Education Management (ERIC/CEM) - Clearinghouse ds. Zarządzania Oświatą. Działą przy Bibliotece Uniwersytetu Stanowego w Eugene (Oregon). W zakres zainteresowań tego domu clearingowego wchodzi wszelkie sprawy związane z organizowaniem oświaty, zarządzaniem, planowaniem i jej finansowaniem.

Wydaje:

- ERIC CEM News Letter

- serię Research Action Briefs (wspólnie z National Association of Secondary Schools Principals).

ERIC Clearinghouse on Handicapped and Gifted (ERIC/CHG) - Clearinghouse ds. Upośledzonych i Uzdolnionych. Zlokalizowany jest przy Radzie ds. Dzieci Wyjątkowych w Reston (Wirginia). Gromadzi informacje dotyczące wszelkich zagadnień nauczania specjalnego w odniesieniu do wszystkich kategorii upośledzeń (umysłowego, fizycznego, wzroku, mowy, przewlekle chorych). Interesuje się także problemami dotyczącymi dzieci wybitnie utalentowanych i genialnych.

Wydaje:

- 'Exceptional Child Education Resource' - kwartalnik

- 'Insight' - 10 razy w roku

- 'Teaching Exceptional Child' - nieregularnie

oraz informatory bibliograficzne:

- Being at Ease with a Handicapped Child

- Career Education for Handicapped.

ERIC Clearinghouse on Higher Education (ERIC/HE) - Clearinghouse ds. Szkolnictwa Wyższego. Działa przy Uniwersytecie J. Waszyngtona w Waszyngtonie. Zakres tematyczny to wszystkie zagadnienia związane ze szkolnictwem wyższym z wyjątkiem spraw dotyczących kształcenia nauczycieli na tym poziomie. Jest to bowiem przedmiotem zainteresowań innej placówki.

Wydaje:

- 'Higher Education Research Report Series' (10 razy w roku)

- 'Research Current Series' (8 razy w roku)

- 'Administrator Update' (3 razy w roku)

ERIC Clearinghouse on Teacher Education (ERIC/TE) - Clearinghouse Kształcenia Nauczycieli. Działa jako samodzielna instytucja w Waszyngtonie. Zajmuje się problemami kształcenia i doksztalcenia się nauczycieli, a także sprawami dotyczącymi zawodu nauczycielskiego i sytuacji społeczno-ekonomicznej tej grupy pracowników.

Wydał m.in.:

- On Retirement Planing. A resource books for teacher retirement. 1980.
- Teacher Stress: Selected ERIC resource.1984
- Retrenchment in Education: Selected ERIC resources.1985

ERIC Clearinghouse on Rural Education and Small Schools (ERIC CRESS)

- Clearinghouse ds. Oświaty Wiejskiej i Małych Szkół. Działa w Las Cruces przy Uniwersytecie stanu Nowy Meksyk. Gromadzi informacje dotyczące nauczania w małych szkołach, głównie wiejskich. Uwzględnia problemy oświatowe mniejszości narodowych i dzieci emigrantów.

Wydaje:

- "CRESS Notes News Letter" - (kwartalnik)

oraz nieregularnie:

- American Indian Education Legislation,
- Outdoor Education and the Handicapped,
- Reading Achievement in Rural Areas,
- American Indian Children's Literature,
- Multicultural Education and Mexican Americas.

ERIC Clearinghouse on Urban Education (ERIC CUE) - Clearinghouse ds.

Oświaty w Miastach. Działa w Nowym Jorku przy Teacher College Uniwersytetu Columbia. W kręgu zainteresowań tego ośrodka znajdują się zagadnienia dotyczące nauczania dzieci wielkomiejskich. Uwzględnia także problemy dzieci upośledzonych ekonomicznie i socjalnie ze względu na ich odmienność kulturową, religijną i etniczną.

Wydaje:

- Information Retrieval Center for Dissadvantaged Bulletin (4 razy w roku)
- ERIC/CUE Urban Diversity Series,
- Equal Opportunity Review, który obejmuje zestawienia bibliograficzne na różne tematy m.in.: ERIC Reference on Urban and Minority Education.

WYDAWNICTWA I BAZA DANYCH ERIC

Najważniejsze drukowane wydawnictwa ERIC to:

- "Resource in Education" - RIE - miesięcznik wydawany od 1966 roku (w latach 1966-79 - "Research in Education"). Zawiera abstrakty dokumentów trudnodostępnych i o ograniczonej dystrybucji, a znajdujących się w posiadaniu ERIC. Selekcji i opracowania informacji dla potrzeb RIE dokonują poszczególne domy clearingowe korzystając z najnowszej wersji tezaury ERIC. Całościowego opracowania tego wydawnictwa dokonuje Komórka ds. Ułatwień informacyjnych i przetwarzania (ERIC PRF), a drukuje je Biuro Wydawnicze Rządu USA.
- "Current Index to Journals in Education" - CIJE - jest to także miesięcznik, a ukazuje się od 1966 roku. Zawiera abstrakty z ponad 750 czasopism z zakresu oświaty i wychowania. Podobnie jak w przypadku RIE materiały do tego czasopisma przygotowują domy clearingowe, a wydaje je wydawnictwo Oryx Press (w latach 1969-1979 ukazywało się w wydawnictwie Crowell, Collier, Macmillan Corporation). Publikowane są także półroczne kumulacje.

Informacja o zasobach ERIC dostępna jest w dwóch formach:

1. W postaci tradycyjnych wydruków RIE i CIJE, z katalogów centrali, lub z danego clearinghouse. Z materiałów ERIC można także skorzystać w ponad 580 ośrodkach informacji na terenie Stanów Zjedno-

czonych i Kanady. Ponadto clearinghouses rozsyłają do ponad 90 tys. specjalistów w całym kraju wykazy nowości, które zawierają krótkie przeglądy i zestawienia bibliograficzne;

2. Na taśmach magnetycznych do wyszukiwania komputerowego.

Z chwilą powstania w 1967 roku serwisu DIALOG stworzona została w oparciu o RIE i CIJE i na bieżąco jest uzupełniana, baza danych ERIC. Weszła ona w skład DIALOG jako baza oznaczona numerem 1. Dostęp do tej i innych baz danych DIALOG możliwy jest w trybie on-line. Polscy użytkownicy mogą korzystać z baz oferowanych przez DIALOG za pośrednictwem serwisu BRIOLIS uruchomionego przez British Council w Warszawie i eksploatowanego od 1983 roku.

ERIC jest systemem ciągle rozwijającym się. Planuje się zwiększenie liczby istniejących domów clearingowych, by możliwe stało się uwzględnienie w jego pracach wszelkich, także bardzo szczegółowych zagadnień z dziedziny oświaty i wychowania. Z drugiej jednak strony czynione są starania, by ustabilizować istnienie obecnie działających ośrodków informacji ERIC. Przewiduje się w większym niż dotychczas stopniu wykorzystanie literatury anglojęzycznej spoza Stanów Zjednoczonych.

Literatura

1. ALL about ERIC. Journal Education Data Processing 7(2)1970. Entire issue devoted to ERIC system.
2. EDUCATION Resource Information Center. Encyclopedia of Library and Information t. 7 New York 1972 s.546-556
3. DRAZAN J.G. The unknown ERIC. A selection of document for the general library. Scaracrow 1981

4. AN ANNOTATED bibliography of ERIC bibliographies 1966-80. Compiled by J.G. Drazen. Westport 1983
5. HANDBOOK on Research on Teaching. ed. M.C. Wittrock. New York 1985
6. HESSONG R., WEEKS T.: Introduction to education. New York 1989
7. JOHANNIGMEIER E.R. Foundation on Contemporary American Education. Scottsdale 1987.

15.01.1990 r.

THE AMERICAN INFORMATION SYSTEM IN THE FIELD
OF EDUCATION - ERIC

Summary

There are discussed the tasks and the activity of the Educational Resource Information Center (ERIC) in the USA and its organizational structure. There is presented with full particulars the activity of the clearing houses the network of which makes the information basis of the system. Their interests and publications are characterized, especially the main system's publications such as: "Resource in Education", "Current Index to Journals in Education" which are made available also on magnetic tapes making the data base of ERIC.

АМЕРИКАНСКАЯ ИНФОРМАЦИОННАЯ СИСТЕМА В ОБЛАСТИ ПРОСВЕЩЕНИЯ
И ВОСПИТАНИЯ (ERIC)

Резюме

Обсуждены задачи и деятельность Центра ресурсов по воспитательской информации (Educational Resource Information Center) в США и его организационную структуру. Подробно представлена деятельность "clearing houses", которых сеть является основной информационной базой системы. Охарактеризован объем их интересов и издания, а в частности важнейшие издания системы: "Resource in Education", "Current Index to Journals in Education" * которые делается доступными также на магнитных лентах являющихся базой данных ERIC.

RECENZJE I OMÓWIENIA

PROJEKTOWANIE JĘZYKA INFORMACYJNO-WYSZUKIWAWCZEGO JAKO ELEMENTU SYSTEMU WYSZUKIWANIA INFORMACJI*

Przedmiotem rozprawy były języki informacyjno-wyszukiawcze stosowane w niepełnotekstowych^{1/} systemach dokumentacyjnych. Przedmiot ten poddany został analizie z punktu widzenia tylko pewnych jego własności i w określonym ujęciu. Tak więc treścią pracy były zagadnienia tworzenia języka informacyjno-wyszukiawczego (jiw) jako elementu systemu wyszukiwania informacji^{2/} badane przy użyciu aparatu współczesnej metodologii projektowania. Dotychczas jiw tworzone doraźnie podejmując decyzje co do ich struktury i dość stereotypowo traktując pewne czynniki zewnętrzne, takie

*/ Jadwiga WOŹNIAK: Projektowanie języka informacyjno-wyszukiawczego jako elementu systemu wyszukiwania informacji. Rozprawa doktorska wykonana pod kierunkiem doc.dr hab. Bożenny Bojar. Obroniona na Uniwersytecie Warszawskim, Wydz. Neofilologii 15 stycznia 1990 r. Recenzenci: prof.dr O.A. Nejtasiwicz, doc.dr hab. E. Ścibor, Warszawa 1989, 295 s., maszyn.

1/ Ograniczenie analizy do systemów niepełnotekstowych wydaje się w pewnym stopniu naturalne, gdyż systemy pełnotekstowe nie są traktowane jako systemy dokumentacyjne. Zalicza się je raczej do systemów faktograficznych. Wyłączenie z zakresu analizy języków stosowanych w systemach informacji faktograficznej odpowiada również pewnej tradycji terminologicznej, gdyż języki te zwykle nie są zaliczane do klasy języków informacyjno-wyszukiawczych.

2/ Termin system wyszukiwania informacji może być rozumiany albo jako synonim terminu system informacyjno-wyszukiawczy SIW, albo jako nazwa systemu o zakresie czynności ograniczonym (w porównaniu z SIW) tylko do transformacji wyszukiawczych. W pracy terminy system wyszukiwania informacji i system informacyjno-wyszukiawczy były traktowane jako równoznaczniki.

jak na przykład oczekiwania przyszłych użytkowników języka. Przyjmowaniu nie zawsze słusznych rozwiązań sprzyjał uproszczony sposób ujmowania tej problematyki w literaturze normatywnej i metodycznej. Często popełnianymi błędami było łączne traktowanie etapów przygotowywania koncepcji jiw i jego praktycznej budowy oraz ograniczanie konstrukcji języka tylko do jednego z elementów, na przykład leksyki. Dorobek metodologii projektowania pozwala na zweryfikowanie pewnych dotychczasowych ujęć problematyki budowy języków informacyjno-wyszukiwawczych zarówno w teorii, jak i w praktyce.

Na użytek pracy przyjęto, iż proces projektowania można podzielić na trzy podstawowe etapy: formułowanie problemu, określanie rozwiązania wstępnego i rozwiązania szczegółowego^{3/}. Rozwiązaniem szczegółowym w wypadku projektowania języków informacyjno-wyszukiwawczych jest prezentacja konkretnego zbudowanego języka. Ten etap nie był w pracy rozważany, gdyż celem rozprawy nie było przedstawienie metodyki tworzenia określonego języka jako elementu danego SIW, lecz stworzenie punktu odniesienia dla działań konstrukcyjnych poprzez wskazanie systemu przesłanek metodologicznych niezbędnych przy formułowaniu koncepcji języka informacyjno-wyszukiwawczego. Projektowanie jiw przedstawiono w pracy nie jako algorytm postępowania, lecz jako proces identyfikowania, analizowania, syntetyzowania i prezentowania problemów, które trzeba rozważyć przed przystąpieniem do budowy języka.

Ogromny wpływ na tok i efekty działań ma pierwszy etap projektowania, czyli formułowanie problemu (zadania) projektowego.

^{3/}Tak określona struktura procesu projektowania była oparta na propozycji Norrisa, przedstawionej w pracy Dorosińskiego, Gasparskiego i Wrony. (W. Dorosiński, W. Gasparski, S. Wrona: Zarys metodyki projektowania, Warszawa 1981).

Nieprawidłowe określenie problemu prowadzi do rozwiązania innego zadania niż było to w zamierzeniu, czyli w tym wypadku do zbudowania takiego jiw, który nie spełni oczekiwań użytkowników, niezależnie od stopnia pomysłowości i trudu poniesionego przez projektanta w dalszych fazach procesu.

Pierwszym krokiem w kierunku sformułowania problemu projektowego jest zebranie informacji o projektowanym obiekcie oraz żądaniach przyszłych jego użytkowników. Podstawowym obowiązkiem projektujących jiw jest zatem wnikliwa i rzetelna analiza sytuacji, w jakiej będzie funkcjonować projektowany język, czyli zidentyfikowanie relacji jiw z otoczeniem oraz wyodrębnienie istotnych elementów składowych języka, ich korelacji, funkcji, zasad i praw nimi rządzących. Rozważania na temat tworzenia koncepcji języka informacyjno-wyszukiwawczego nie mogą zatem obyć się bez zidentyfikowania i poznania relacji jiw z otoczeniem, a także bez wnikliwego, analitycznego opisu przedmiotu projektowania, czyli jiw. Chcąc bowiem zaprojektować sprawny jiw, lub dokonać wyboru spośród już istniejących języków, trzeba najpierw poznać możliwości, jakie stworzył rozwój teorii i praktyki, trzeba wiedzieć, jakie cechy języków i jak należy porównywać oraz oceniać. Inaczej mówiąc, trzeba dysponować wzorcem jakościowym jiw w postaci zestawu cech dobrego języka.

Zespół postulatów optymalności języka sformułował T. Wójcik^{4/}. Dostosowując je do rzeczywistości funkcjonowania języków informacyjno-wyszukiwawczych można przed jiw postawić następujące wymagania.

^{4/}T. Wójcik: Prakseosemiotyka. Zarys teorii optymalnego znaku. Warszawa 1966

Optymalne dostosowanie do rzeczy komunikowanej.

Możliwości formułowania w języku informacyjno-wyszukiwawczym komunikatów optymalnych z punktu widzenia rzeczy komunikowanej zależą od wielu własności jiw. Jedną z nich jest odpowiednia siła semantyczna języka, osiągnięta na osi paradygmatycznej i syntagmatycznej. Teksty języka informacyjno-wyszukiwawczego będą maksymalnie wiernie odwzorowywały komunikowaną rzeczywistość, z optymalną, ze względu na wyznaczone zadania, szczegółowością, jeśli jiw będzie charakteryzował się odpowiednio szerokim zakresem systemu leksykalnego, właściwą jego segmentacją na pola znaczeniowe, poprawną i odpowiednią strukturą relacyjną i hierarchiczną oraz dostateczną wartością selekcyjną jednostek leksykalnych. Wymagana jednoznaczność strukturalna i informacyjna komunikatów języka informacyjno-wyszukiwawczego będzie osiągnięta, jeśli projektant języka uczyni jednoznanymi nie tylko poszczególne poziomy znaków jiw, ale i ich złożenia. W wypadku języków informacyjno-wyszukiwawczych wiedza językowa nie jest wystarczającym warunkiem optymalnego dostosowania komunikatu do rzeczy komunikowanej. Niezbędna jest także wiedza pozajęzykowa dotycząca dwóch płaszczyzn odniesienia - rzeczywistości desygnowanej przez wyrażenie jiw, czyli rzeczywistości dokumentacyjnej oraz rzeczywistości desygnowanej przez wyrażenia języków naturalnych, przez informacje zawarte w dokumentach i pytaniach użytkowników.

Z postulatu optymalnego dostosowania komunikatu jiw do rzeczy komunikowanej wynika, iż aby projektowany język informacyjno-wyszukiwawczy był maksymalnie operatywny, należy formułując problem projektowy wnikliwie rozważyć teoretyczne i praktyczne możliwości oraz ograniczenia wynikające z: funkcji języków informacyjno-wyszukiwawczych, desygnatów ich wyrażen, poziomów jiw, zakresu

i segmentacji systemu leksykalnego, reguł produktywności, warunków jednoznaczności języka i jego tekstów. Projektowanie języka informacyjno-wyszukiwawczego lub poznawanie systemu istniejącego języka powinno być łącznym rozważaniem trzech jego podsystemów - leksykalnego, gramatycznego i notacyjnego, z co najmniej dwóch punktów widzenia: przedmiotowego i czynnościowego. Pewne zależności między elementami struktury jiw łatwiej bowiem dają się zidentyfikować i poznać przy opisie statycznym języka, inne - w trakcie strukturalizowania procesu działań projektotwórczych.

Drugim postulatem optymalności języka informacyjno-wyszukiwawczego jest warunek maksymalnego dostosowania komunikatu do jego twórcy, wyrażający się łatwością tworzenia tekstów i wyuczalnością ich budowania. Ma on istotny, chociaż nie doceniany, wpływ na sprawność funkcjonowania języków i systemów informacyjno-wyszukiwawczych. Jak wynika z wielu badań jiw, zasadniczy wpływ na efektywność wyszukiwania ma stopień skomplikowania systemu języka. Teoretycznie im bogatszy i bardziej wyrafinowany jest język informacyjno-wyszukiwawczy, tym większe ma on możliwości optymalnego funkcjonowania. Teza ta na ogół nie potwierdza się w praktyce, gdyż wraz ze wzrostem trudności tworzenia tekstów jiw i wyuczalności ich budowania rośnie prawdopodobieństwo wystąpienia błędów w indeksowaniu i zamiast oczekiwanego wzrostu efektywności wyszukiwania może nastąpić jej spadek.

Na maksymalnie operatywne funkcjonowanie języka informacyjno-wyszukiwawczego z punktu widzenia nadawcy komunikatu wpływ ma nie tylko stopień skomplikowania systemu jiw, ale także jednoznaczne i precyzyjne określenie zasad tworzenia tekstów jiw, zarówno tych reguł, które dotyczą posługiwania się językiem, jak i tych, które odnoszą się do realizowania procesów transformacji metainformacyjnych w systemie.

Język informacyjno-wyszukiwawczy, żeby optymalnie realizować swoje funkcje, musi być nie tylko poprawnie skonstruowany, ale i operatywnie wykorzystywany. Etap precyzowania zasad tworzenia tekstów jiw, czyli indeksowania dokumentów i pytań użytkowników, często jest niestety pomijany w działaniach projektotwórczych. Tymczasem dwie cechy tego procesu - szerokość i szczegółowość indeksowania - mają równie duży wpływ na efektywność funkcjonowania jiw, jak i cechy strukturalne tego języka. Do tej pory nie wypracowano jednoznacznych i obiektywnych sposobów ustalania szerokości i szczegółowości indeksowania. W zamian arbitralnie formułowano miary np. szerokości indeksowania, które jednak w żaden sposób nie dają się obronić.

Trzecim warunkiem optymalności języka informacyjno-wyszukiwawczego jest jego dostosowanie do użytkowników informacji. Warunek ten łączy optymalność odwzorowywania rzeczy komunikowanej, czyli informacji zawartych w pytaniach użytkowników, z warunkami zrozumiałości, dostępności i optymalnej postrzeganej formy komunikatu. Do tej pory nie osiągnięto zadowalających rezultatów co do możliwości wywodzenia opytalizacji dostosowywania się języka informacyjno-wyszukiwawczego do użytkowników z badań zachowań czy tzw. potrzeb informacyjnych. Istnieje zatem potrzeba kompleksowych badań użytkowników i zadawanych przez nich pytań informacyjnych, żeby móc w pełni realizować postulaty z tej grupy.

Ważny teraz pod uwagę drugi z wymienionych wcześniej warunków poprawnego sformułowania problemu projektowego. Do tej pory albo w ogóle nie zajmowano się problemem relacji jiw z otoczeniem, albo sprowadzano go do, nie zawsze rzetelnej, analizy wybranych pojedynczych zagadnień. Tymczasem relacje jiw z otoczeniem konstruują wyraźny system elementów i zależności. Najbliższym otoczeniem języka informacyjno-wyszukiwawczego jest system wyszukiwania

206

informacji. SIW jest układem o hierarchicznej budowie, co znaczy, że jego elementy mogą być z kolei traktowane jako system i organizowane według schematów różnych sprzężeń i relacji. Tak więc w obrębie SIW należałoby, w aspekcie projektowania jiw, wyróżnić podsystemy: dokumentów, języków (w tym języków informacyjno-wyszukiwawczych), transformacji. Każdy z nich powinien być dalej dzielony, tak żeby można było uchwycić wszystkie istotne i tylko te cechy ich związków z jiw. Dotychczas nie robiono tego, zadowolano się stwierdzeniem, że, na przykład, istnieje zależność między wielkością, charakterem i formą zbiorów dokumentów a cechami języka. Nie szły za tym jednak głębsze analizy i refleksje. Stosunkowo najwnikliwiej zbadano wpływ tematyki dokumentów na system leksykalny jiw i zasady transformacji metainformacyjnych oraz (pośrednio) wyszukiwawczych. Rzadko uwzględniano fakt, iż do klasy dokumentów należą także pytania kierowane przez użytkowników do systemu, choć oczywiście są one specyficznymi dokumentami. Dopiero w ostatnich latach na popularności zyskuje na przykład postulat określania systemu paradygmatycznego jiw przy wykorzystaniu asocjacji wyrażen w tekstach pytań.

W literaturze rozpowszechniona jest opinia, iż istnieje zasadniczy związek między stosowanymi w systemie środkami technicznymi a strukturą języka informacyjno-wyszukiwawczego. Przeprowadzone na użytek pracy analizy upoważniają do stwierdzenia, iż jak do tej pory uwarunkowania techniczne zdają się mieć wpływ na formy prezentacji oraz sposoby realizacji języków informacyjno-wyszukiwawczych, nie determinują jednak znacząco ich struktury semantycznej, tak w planie paradygmatyki, jak i syntagmatyki. Dlatego w pracy aspekt techniczny został uznany za cechę dystynktywną systemów, a nie języków informacyjno-wyszukiwawczych i wzięty pod

uwagę tylko tam, gdzie rzeczywiście zaznacza się jego wpływ.

Relacji języka informacyjno-wyszukiwawczego z otoczeniem nie można ograniczać i sprowadzać tylko do związków z systemem wyszukiwania informacji. Do tego obszaru problemowego należą także zagadnienia wiążące się z użytkownikami informacji. Nie wydaje się słuszne umieszczanie użytkowników w obrębie SIW, bowiem ich potrzeby i zachowania ujawniające się w kontaktach z systemami są tylko częścią całościowo widzianej problematyki użytkowników. Analiza wpływu użytkowników na kształt projektowanego jiw powinna opierać się na dorobku teoretycznym w tym zakresie i wynikach badań empirycznych. Niestety postęp w kierunku refleksji teoretycznej był dość wolny i nadal wymaga ona wielu uzupełnień i uściśleń, między innymi w sferze podstawowej terminologii. Terminologii nie można oddzielić od nauki, a nauki od terminologii, dlatego dopóki nie zostanie zaprowadzony pewien porządek w języku nauki o użytkownikach, dopóty trudno będzie scalić refleksję teoretyczną, zaś różne badania praktyczne będą budzić zastrzeżenia i wątpliwości. Definicje nawet tak podstawowego tu terminu, jak potrzeba informacyjna, nie wydają się ani poprawne teoretycznie, ani (w większości wypadków) użyteczne praktycznie. Dlatego wiele badań realizowanych według różnych schamtów strukturalnych i metodologicznych, ale wywodzących się z tych niepoprawnych założeń terminologicznych, ma niższą wartość niż przewidywana. Chcąc, żeby badania użytkowników dostarczały wiarygodnych przesłanek do projektowania języków informacyjno-wyszukiwawczych, należy także dokonać zasadniczej zmiany w ich orientacji - z pozornie skierowanej na użytkowników na istotnie na nich skierowaną. Większość przeprowadzonych do tej pory badań użytkowników była tylko z nazwy badaniami użytkowników. Faktycznie były to badania na przykład pewnych cech strukturalnych czy sprawnościowych systemów informacyjno-wyszukiwawczych.

Przeobrażenia, jakim ulegał model obsługi użytkowników, nie miały wiele wspólnego z rzeczywistymi zmianami ich potrzeb. Najczęściej były one rezultatami zastosowań nowych urządzeń i technik przetwarzania informacji.

Sformułowanie tematu pracy "Projektowanie języka informacyjno-wyszukiwawczego jako elementu systemu wyszukiwania informacji" mimo przyjętych ograniczeń, wyznacza dość rozległy obszar rzeczywistości, który powinien być eksplorowany przy szukaniu rozwiązania. Zbadanie całego tego obszaru nie jest ani możliwe, ani potrzebne. Natomiast konieczne jest uwzględnienie wszystkich ważnych elementów. Poszukiwać ich można różnymi drogami, których przydatność i sprawność zależy od konkretnej sytuacji praktycznej, rozwiązywanego problemu, cech indywidualnych projektanta i innych uwarunkowań. Jednym z możliwych podejść jest podejście systemowe.

Z przyjętej metody rozwiązywania problemu sformułowanego w temacie pracy (podejście systemowe) oraz wybranego ujęcia przedmiotu (metodologia projektowania), po uwzględnieniu wymienionych wcześniej ograniczeń, wynikła następująca struktura pracy.

Rozdział pierwszy ma za zadanie przedstawienie wybranego modelu fragmentu sytuacji praktycznej projektowania języka informacyjno-wyszukiwawczego poprzez wskazanie związków jiw z systemem informacyjno-wyszukiwawczym i określenie jego udziału w formułowaniu zadania projektowego. Zawiera charakterystykę SIW jako najbliższego otoczenia języka. W rozdziale znalazły się między innymi informacje o elementach składowych SIW, ich funkcjach, wzajemnych powiązaniach, zasadach nimi kierujących i relacjach jiw z systemem informacyjno-wyszukiwawczym. Dokonując kategoryzacji wejść i wyjść oraz zbiorów dokumentów i informacji w systemie zrezygnowano z przedstawienia procesów transformujących wejścia w wyjścia i jedne zbiory dokumentów w inne, uznając, iż silniej skojarzony

przedmiotowo z tą problematyką będzie rozdział dotyczący charakterystyki funkcjonalno-strukturalnej jiw.

W rozdziale drugim podjęto próbę scharakteryzowania innego czynnika konstruującego sytuację praktyczną projektowania jiw - użytkowników SIW, ściślej ich potrzeb, zwłaszcza tych, które ujawniają się w kontaktach z systemami informacyjno-wyszukiwawczymi oraz możliwości ich zaspokojenia przez odpowiedni język informacyjno-wyszukiwawczy. Poddano krytycznej analizie dotychczasowe poglądy na tak zwane potrzeby informacyjne użytkowników i konwencje realizacji badań. Wskazano na konieczność zmiany orientacji badań użytkowników oraz na prawdopodobne (nie można mówić o pewności ze względu na brak doświadczeń w tym zakresie) obszary i sposoby oddziaływania żądań i życzeń użytkowników na kształt projektowanego języka informacyjno-wyszukiwawczego.

Rozdziały trzeci i czwarty mają za zadanie dostarczyć informacji niezbędnych do zakończenia etapu formułowania problemu projektowego. Rozdział trzeci jest charakterystyką funkcjonalno-strukturalną modelu operatywnego języka informacyjno-wyszukiwawczego w ujęciu przedmiotowym. Są w nim rozważane teoretyczne i praktyczne możliwości oraz ograniczenia jiw wynikające z: funkcji tych języków, desygnacji ich wyrażeń, struktury znaków, zakresu i segmentacji systemu leksykalnego, reguł produktywności, warunków jednoznaczności i niesynonimiczności. W rozdziale tym przedstawiono także dwa główne typy transformacji dokonywanych w systemach informacyjno-wyszukiwawczych - transformacje metainformacyjne i wyszukiwawcze. Język informacyjno-wyszukiwawczy, żeby optymalnie realizować swoje funkcje musi być bowiem nie tylko poprawnie skonstruowany, ale i operatywnie wykorzystywany.

Rozdział czwarty rozbudowuje problematykę rozdziału 3, charakteryzując język informacyjno-wyszukiwawczy w aspekcie dyna-

micznym. Zwrócono w nim uwagę na formułowanie podstawowych zasobów struktury języka na wszystkich jego poziomach z uwzględnieniem zależności między koncepcją gramatyki a zasobem i ustrojem systemu leksykalnego oraz na projektowanie systemu notacyjnego jiw jako planu wyrażania struktury semantycznej tego języka. W rozdziale tym zawarte są ponadto pewne postulaty czynnościowe do wykorzystania przy określaniu rozwiązania wstępnego projektu jiw.

Jednym z celów pracy było pokazanie, jak wykorzystać dorobek współczesnej metodologii projektowania w tworzeniu języków informacyjno-wyszukiwawczych. Na obecnym etapie rozwoju nauk systemowych i metodologii nauk indukcyjnych nie jest osiągalny pełny opis zasobów wiedzy i działań odpowiadających różnym fazom projektowania systemów językowych. Dlatego też zawarte w pracy analizy mają charakter propozycji wybranych składników procesu projektowania poprzedzonego działaniami analitycznymi. Przy obecnym braku oparcia w metodologii nauk, umożliwiającego uniwersalne rozwiązanie problemów budowy języków informacyjno-wyszukiwawczych dla zróżnicowanych systemów wyszukiwania informacji, konieczne jest zatem skoncentrowanie uwagi na opracowywaniu wariantów rozwiązań konkretnych problemów w określonych sytuacjach praktycznych, oczywiście w nawiązaniu do zasad i środków metodologii projektowania oraz systemowego charakteru jiw i jego otoczenia. Mamy nadzieję, że dzięki zastosowaniu podejścia systemowego, jako metody badania wyznaczonego przez temat pracy fragmentu rzeczywistości, udało się rozszerzyć i pogłębić wiedzę o relacjach języka informacyjno-wyszukiwawczego z otaczającymi go systemami i elementami oraz wskazać te spośród nich, które istotnie mają wpływ na kształt języka.

Jadwiga Woźniak

16.01.1990 r.

211

**EKSPLOATACJA ZAUTOMATYZOWANYCH WIELOBZIEOZINOWYCH
BAZ DANYCH W OŚRODKU INFORMACJI NAUKOWEJ PAN**

Nauki przyrodnicze i ścisłe

W latach osiemdziesiątych w ramach Systemu Informacji o Naukach Ścisłych i Przyrodniczych Ośrodek Informacji Naukowej PAN rozpoczął eksperymentalną eksploatację zautomatyzowanej bazy danych o naukach przyrodniczych. Nawiązano kontakt z BioSciences Information Service z Filadelfii, skąd dostarczane były taśmy magnetyczne zawierające kolejne wydania bazy BIOSIS. Dokumenty zawarte w tej bazie pochodziły z Biological Abstracts oraz Biological Abstracts (RRM). W bazie BIOSIS od 1976 r. dokumentowane są artykuły z czasopism, monografie, sympozja naukowe, przeglądy, raporty, czasopisma popularnonaukowe oraz wszelkie inne doniesienia z dziedziny nauk biologicznych i medycznych. W bazie BIOSIS zgromadzonych jest ponad 6 mln dokumentów. Aktualizacja bazy następuje dwa razy w miesiącu. Baza ta zawiera takie zagadnienia jak: cytologia, ekologia, biochemia, biofizyka, anatomia, histologia, fizjologia, patologia, metabolizm, technologia żywności, psychiatria, farmakologia, toksykologia, mikrobiologia, wirusologia, immunologia, parazytologia, chemoterapia, paleobotanika, morfologia, anatomia i embriologia roślin, agronomia, ogrodnictwo, leśnictwo i dziewiarstwo, fitopatologia, systematyka strunowców, kręgowców, paleozoologia.

Eksperymentalnym wyszukiwaniem informacji bibliograficznej objętych zostało szereg placówek PAN oraz placówek uniwersyteckich, instytutów z akademii medycznych oraz rolniczych.

Opracowano metodykę przygotowania profilu wyszukiwawczego. Wyszukiwanie miało miejsce w Zakładzie Elektronicznej Techniki Obliczeniowej. Profile wyszukiwawcze testowane były zarówno w trybie off-line, jak i on-line. Obecnie OIN PAN posiada w dalszym ciągu dostęp do zasobów bazy BIOSIS. Poprzez węzeł VNIIPAS (w Moskwie) można uzyskać łączność teletransmisyjną z bazą BIOSIS zlokalizowaną w Centralnym Instytucie Informacji Naukowej i Technicznej w Sofii.

Radzieckie bazy WINITI

Za pośrednictwem OIN PAN istnieje możliwość korzystania z zasobów zautomatyzowanych baz danych tworzonych we Wszechzwiązkowym Instytucie Informacji Naukowej i Technicznej w Moskwie. WINITI rozpowszechnia następujące bazy danych:

1) Automatyka i radioelektronika

(Automatyka i telemekhanika, urządzenia półprzewodnikowe, materiały techniki elektronicznej, technika obliczeniowa, elektroakustyka, radiolokacja, radionawigacja, radiotechnika teoretyczna, elektrołączność).

2) Biologia

(Cytologia, genetyka, ekologia, biocenologia, hydrobiologia, wirusologia, mikrobiologia, botanika, fizjologia i biochemia roślin, zoologia, ichtiologia, entomologia, morfologia człowieka i zwierząt, onkologia, genetyka i selekcja roślin, uprawa roli i agrochemia, fitopatologia, leśnictwo i drzewiarstwo, farmakologia, toksykologia, genetyka człowieka, immunologia i alergologia, bionika, biocybernetyka, bioinżynieria).

3) Geologia

(Stratografia, paleobiologia, geomorfologia lądu i dna morskiego, geologiczne i geochemiczne metody poszukiwań kopalin, technika prac geologicznych, geochemia, mineralogia, petrografia, hydrogeologia, geologia inżynierska).

4) Przemysł wydobywczy

(Złóża ropy i gazu - problemy ogólne, przemysł, ekonomika, budownictwo).

5) Informatyka

(Ogólne problemy informatyki, organizacja działalności informacyjnej, źródła informacji, wyszukiwanie informacyjne, obsługa informacyjna, technika informacyjna).

6) Korozja i ochrona przed korozją

(Teoria korozji, korozja gazowa, odporność na korozję metali i stopów metali, elektrochemiczne metody ochrony, ochrona pokrycia i chemiczna obróbka powierzchni, niemetaliczne materiały odporne na korozję, metody badania korozji).

7) Budowa maszyn

(maszyny i sprzęt dla przemysłu tekstylnego, technologia i organizacja przedsiębiorstwa przemysłu tekstylnego, cięcie materiałów, ciągniki i maszyny rolnicze, osprzętowanie przemysłu spożywczego, budowa, materiały i konstrukcja detali maszyn, maszyny budowlane i drogowe, budowa rakiet i technika kosmiczna, podnośniki samolotowe i raketowe, budowa turbin, fotokinetyka).

8) Metalurgia

(technologia procesów metalurgicznych, analiza techniczna

w metalurgii, metalurgiczna technika cieplna, sprzęt, kontrola i automatyzacja w przedsiębiorstwie metalurgicznym, metalurgia metali kolorowych).

9) Mechanika

(mechanika ogólna, mechanika płynów i gazów, trwałość konstrukcji i materiałów).

10) Ochrona przyrody i zasobów naturalnych

(teoria i metody badania ochrony środowiska naturalnego i zasobów naturalnych, zanieczyszczenie środowiska i kontrola zanieczyszczenia, tereny chronione, użytkowanie zasobów naturalnych).

11) Transport

(budowa lokomotyw i wagonów, eksploatacja torowiska, budowa torowiska, automatyka, telemechanika i łączność na kolei, budowa samochodów, transport samochodowy, budowa samolotów, eksploatacja i remont samolotów i innych aparatów latających, lotniska, budowa maszyn transportowych, wzajemne oddziaływanie różnych rodzajów transportu i przewozów kontenerowych).

12) Fizyka

(akustyka, optyka, radiofizyka i fizyczne podstawy elektroniki, fizyka atomu i molekuł, fizyka gazów i płynów, termodynamika i fizyka statystyczna, fizyka plazmy, fizyka ciał stałych, fizyka jądrowa).

13) Fizyko-chemiczna biologia i biotechnologia

(chemia bioorganiczna, biologia molekularna, biofizyka, biochemia, immunologia komórkowa i molekularna, biotechno-

logia, ogólne problemy fizyko-chemicznej biologii i biotechnologii).

14) Chemia

(Chemia analityczna, wyposażenie laboratoriów, chemia nieorganiczna, związki kompleksowe, chemia fizyczna, budowa molekuł, kinetyka, kataliza, fotochemia, plazmochemia, radiochemia, chemia ciała stałego, gazów, płynów, ciała amorficzne, chemia koloidów, chemiczna termodynamika, analiza fizykochemiczna, roztwór, elektrochemia, chemia i technologia produktów spożywczych, masy plastyczne, materiały silikatowe).

15) Ekonomia przemysłu

(Ekonomia światowa, ekonomiczne położenie krajów, zastosowanie metod matematycznych w badaniach ekonomicznych i planowaniu, organizacja i finansowania prac naukowo-badawczych, kadry, zatrudnienie, ogólnoteoretyczne problemy ekonomiki i organizacja przemysłu, ekonomika i organizacja przemysłu ciężkiego, lekkiego, spożywczego, poligraficznego, organizacja kierowania).

16) Elektrotechnika

(Ogólne problemy elektrotechniki, aparaty elektrotechniczne, maszyny elektrotechniczne i transformatory, materiały elektrotechniczne, kondensatory elektroniczne, elektryfikacja i automatyzacja rolnictwa).

17) Energetyka

(Energetyka atomowa, stacje i sieci energetyczne, systemy energetyczne i ich automatyzacja, teoretyczne problemy techniki cieplnej, przemysłowa technika cieplna, hydroenergetyka).

Wyszukiwanie w bazach danych WINITI można prowadzić w języku rosyjskim. OIN PAN jest w posiadaniu materiałów pomocniczych służących do pracy z bazami (rubrykator, słownik słów kluczowych). Dotyczy to baz: Biologia i Fizyko-chemiczna biologia i Biotechnologia).

OIN PAN pośredniczy również w korzystaniu z radzieckich baz danych Międzynarodowego Centrum Informacji Naukowej i Technicznej w Moskwie. Dostępne są następujące bazy:

1) Baza danych o wydawnictwach seryjnych krajów - członków Systemu

(nauki ścisłe, przyrodnicze, techniczne, stosowane, społeczne).

2) INIS - Międzynarodowy System Informacji Nuklearnej (baza wydawana w języku angielskim)

(fizyka ogólna, fizyka jądrowa, chemia, materiały, nauki o ziemi, nauki biologiczne stosowane, ochrona zdrowia, środowisko naturalne, izotopy, technika jądrowa, reaktory jądrowe, centrastacje atomowe, aparatura do pomiarów jądrowych, odpady radioaktywne).

3) STANDARDY - baza danych o standardach międzynarodowych (terminy i określenia, materiały, półfabrykaty, produkty gotowe i produkcja, typy, rodzaje, marki, parametry, warunki techniczne, prawa marketingu, prawa transportu, prawa eksploatacji, procesy technologiczne, klasyfikacje, konstrukcje, dokumentacja techniczna).

4) INSPEC (baza w języku angielskim)

(Fizyka cząstek elementarnych i pól, fizyka jądrowa, fizyka atomowa, fizyka molekularna, fizyka płynów, fizyka plazmy, fizyka ciała stałego, geofizyka, astronomia, astrofizyka, komponenty, urządzenia i materiały optyczne, optyka elektro-

niczna, fale elektromagnetyczne, łączność, przyrządy pomiarowe, systemy energetyczne).

- 5) INSPEC-CD (baza w języku angielskim)
(Techniczne środki i metody kierowania, elektronizacja, technika obliczeniowa, mikrokomputery, sieci lokalne).
- 6) EMIS-ELECTRONIC MATERIALS INFORMATION SERVICE (baza w języku angielskim)
(fizyczne i elektryczne właściwości materiałów, metody pomiarowe).
- 7) Technika, Technologia (COMPENDEX) (baza w języku angielskim)
(Agrotechnika, technika budownictwa, budownictwo miejskie, elektrotechnika, budowa maszyn, technika jądrowa, radiotechnika, energetyka, technika obliczeniowa, przemysł wydobywczy, metalurgia, biotechnologia, łączność, fizyka stosowana).
- 8) Przemysł węglowy
(geologia złóż węgla, technika i technologia rozpracowania złóż węgla, budowa maszyn dla górnictwa).
- 9) Elektrotechnika - INEL
(Elektrotechnika teoretyczna, materiały elektrotechniczne, maszyny elektroniczne, aparaty elektryczne, transformatory, reaktory elektryczne, elektryczne kondensatory siłowe, źródła ruchu, przewody i kable, izotopy elektryczne).
- 10) TEI - Informacja handlowo-ekonomiczna
(Ogólne problemy technologii chemicznej, technologia związków nieorganicznych i środków farmaceutycznych, technologia związków organicznych, technologia mas plastycznych, przemysł leśny i drzewiarski).

Nauki społeczne

Ośrodek Informacji Naukowej PAN posiada dostęp do zautomatyzowanych baz danych z dziedziny nauk społecznych. Bazy te tworzone są wspólnie w oparciu o porozumienie między Akademiami Nauk ZSRR, Bułgarii, Węgier NRD, Czechosłowacji, Polski, Mongolii, Kuby i Wietnamu w ramach Automatycznego Systemu Informacyjnego o Naukach Społecznych (AIS MISON). Zlokalizowane są one w Instytucie Informacji Naukowej Nauk Społecznych Akademii Nauk ZSRR (INION AN).

Bibliograficzne bazy danych mają charakter tematyczno-problemowy i regionalno-krajowy.

Do bibliograficznych baz danych tematyczno-problemowych wprowadzane są każdego roku informacje:

- o nowej zagranicznej literaturze z dziedziny nauk społecznych (około 60 tys. opisów książek, artykułów z czasopism, materiałów z konferencji i recenzji).

Informacja o zagranicznych wydawnictwach zaopatrzona jest w przetłumaczony tytuł i adnotację:

- o nowej radzieckiej literaturze z dziedziny nauk społecznych
- Do systemu wchodzi około 60 tys. opisów książek, artykułów i czasopism, materiałów ze zbiorów i czasopism, materiałów konferencyjnych, sympozjów, autoreferatów, dysertacji, recenzji i rękopisów deponowanych w INION AN.

Do bibliograficznych baz danych o charakterze regionalno-krajowym wprowadzane są każdego roku informacje stanowiące około 80 tys. opisów dokumentów - z książek, artykułów z czasopism, wydawnictw informacyjnych, publikacji historycznych i recenzji (przeszło 85% stanowią wydawnictwa zagraniczne). Roczny przyrost wszystkich bibliograficznych baz danych wynosi 200 tys. dokumentów.

Bibliograficzne bazy danych funkcjonują w systemie operacyjnym MRE-IV na komputerze "Hewlett Packard (HP) 3000". Pojemność pamięci dyskowej 700 Mb. Bazy zajmują 462 Mb dyskowej pamięci (dane z 01.01.1989) i funkcjonują pod kontrolą systemu "POISK-4" (adaptowany analog. pakietu "MINISIS 0.02"). Zawierają kompleks środków lingwistycznych (tezaurusów, spisów znormalizowanej leksyki, rubrykatorów).

Dostęp do baz danych systemu odbywa się przy pomocy terminali (lub personalnych komputerów) przez komutowaną sieć telefoniczną. Każda baza zawiera bibliograficzne opisy dokumentów z następującymi elementami:

- porządkowy (systemowy) numer w bazie danych (ISN)
- szyfr ochrony dokumentu
- numer dokumentu w buforowej bazie danych
- opis bibliograficzny (autor, redaktor, tytuł, dane wydawnicze itp.)
- kody rubryk rubrykatora (systemu klasyfikacyjnego), do których odnosi się dokument
- deskryptory i słowa kluczowe oddające podstawową zawartość dokumentu
- adnotacje
- rodzaj dokumentu
- kod języka

W rezultacie wyszukiwania w bazie danych opis dokumentu pokazany jest na ekranie terminala i na wydruku w następującej formie:

*** ISN = 238043 ШИФР = 16427643 #15583 ЯЗЫК: 045

CARTER M.P.

A METHODOLOGY FOR THE ECONOMIC APPRAISAL OF
MANAGEMENT INFORMATION.

// INTERN. J. OF INFORM. MANAGEMENT. -
GUILDFORD, 1986. - VOL. 6. N 4. - P. 193-201.

АННОТАЦИЯ: ОПРЕДЕЛЕНИЕ ИНФОРМАЦИОННЫХ ПОТРЕБНОСТЕЙ
МЕНЕДЖМЕНТА И МЕТОДОЛОГИЯ ЭКОНОМИЧЕСКОЙ ОЦЕНКИ
УПРАВЛЕНЧЕСКОЙ ИНФОРМАЦИИ (ЧЕТЫРЕ ПОСЛЕДОВАТЕЛЬНЫХ
СТАДИИ ОЦЕНКИ).

КОДЫ РУБРИК: А06811712:

ТЕРМИНЫ: ЭКОНОМИЧЕСКАЯ ИНФОРМАЦИЯ; ВНУТРИФОРМЕННОЕ
УПРАВЛЕНИЕ; КАПИТАЛИСТИЧЕСКОЕ ПРЕДПРИЯТИЕ;
ПРИНЯТИЕ РЕШЕНИЙ; МЕНЕДЖМЕНТ;

Deskryptory i adnotacje są w języku rosyjskim. Adnotowana jest literatura zagraniczna i radziecka, jeśli tytuł dokumentu nie oddaje w stopniu dostatecznym jego zawartości. Ogólna objętość bibliograficznych baz danych wynosi przeszło 1 milion dokumentów (dane na dzień 01.01.1990).

Tabela zamieszczona poniżej zawiera wykaz bibliograficznych baz danych AIS MISON, ich objętość, profil tematyczny, roczny przyrost i procentowy udział dokumentów w podziale językowym (Zał. 1). Wyszukiwanie dialogowe informacji bibliograficznej oparte jest na zasadach zawartych w "Instrukcji wyszukiwania bibliograficznej informacji w formie dialogu w bazach danych AIS MISON". Opisane są tam możliwości systemu i metody pracy z bazami danych. System zapewnia szybkie wyszukiwanie informacji bibliograficznej przy posługiwaniu się deskryptorami, indeksem rubrykatora, nazwiskami autorów.

Lp.	Nazwa bazy danych	Objętość bazy w tys. dok.	Profil tematyczny	Zespolony czasowy	Roczny przyrost (w tys. dok.)	Język dokumentu				
						Rab.	Angl.	Niem.	Franc.	Por.
1.	ECON	52,0	Ekonomika i demografia	1988	30,0	42	33	7,3	5,2	12,5
2.	HISTORY	88,8	Historia, archeologia i etnografia	1986	27,0	45	23	12,3	5	16,7
3.	PHILOS	73,0 37,7 11,2 13,1 11,0	Ogólna BD Filozofia Socjologia Ateizm i religia Kommunizm naukowy	1987	33,2 16,0 10,0 4,2 3,0	45	30	10	5	10
4.	LABOUR	20,8	Miedzynarodowy ruch robotniczy	1986	5,5	35	30	15	15	5
5.	LAW	45,5	Państwo i prawo	1986	18,0	35	30	15	15	5
6.	SCIENCE	58,8	Naukoznawstwo	1986	16,0	41	16	12	5	5
7.	PHILOLOG	106,0 54,0 52,0	Ogólna BD Literaturoznawstwo Językoznawstwo	1986	32,7 16,2 16,5	35	24	11,3	4,7	25
8.	BIBSOC	143,4	Problemy europejskich krajów socjalistycznych	1984	30,5			12,3		37,7
9.	AFRASI	39,3	Problemy krajów Azji i Afryki	1986	11,6	30	44,6	5,5	7	18,3
10.	BIBNEW	239,0 142,2 78,5 18,3	Ogólna BD Ekonomika i demografia Filozofia i socjologia Kommunizm naukowy	1982 -1987						
11.	BIBLIO	100,0 60,0 40,0	Ogólna BD Ekonomika i demografia Filozofia i socjologia	1979- 1982						
12.	LATAMAR	9,0	Problemy krajów Ameryki Łacińskiej	1985	2,0		34,5	5,0	5,5	5,3
13.	EAST	6,0	Daleki wschód	1986	1,6					
14.	SOCIAL	8,0	Socjologia	1985	1,5					

Jest możliwe wyszukiwanie z ograniczeniem wyszukiwawczego terminu z prawej strony (maskowanie) i wyszukiwanie tekstowe (według adnotacji i elementów opisu bibliograficznego). Użytkownik ma możliwość odczytania na ekranie terminala znalezionych w czasie wyszukiwania opisów dokumentów, wybranie najbardziej interesujących i wydrukowanie, podając ich numery systemowe.

W Zespole Automatyzacji OIN PAN znajdują się materiały potrzebne do pracy z bazami AIS MISON. Są to środki lingwistyczne w postaci spisów znormalizowanej leksyki i rubrykatorów.

Użytkownikowi zainteresowanemu jedną z wcześniej wymienionych bibliograficznych baz danych może być udostępniony "Katalog bibliograficznych baz danych AIS MISON". Zawiera on szczegółowy opis zawartości poszczególnych baz danych i w ten sposób ułatwia prawidłowy wybór bazy danych. W procesie rozwoju systemu przewidziana jest budowa wielojęzycznych spisów znormalizowanej leksyki i tezaurusów MISON, pozwalających na indeksowanie dokumentów i wyszukiwanie zapytań w ojczystych językach uczestników systemu.

Bazy serwisu BRIOLIS

W 1989 roku OIN PAN uzyskał prawo pośrednictwa w korzystaniu z komputerowego serwisu informacyjnego zlokalizowanego w Instytucie Brytyjskim. Celem jest obsługa informacyjna badań naukowych prowadzonych głównie w placówkach PAN.

Komputerowy serwis informacyjny BRIOLIS pozwala na uzyskiwanie najbardziej aktualnych informacji bibliograficznych o monografiach, artykułach, sprawozdaniach, materiałach konferencyjnych na dowolny temat ze wszystkich dziedzin wiedzy - w drodze przeszukiwania części brytyjskich i międzynarodowych baz danych dostępnych w systemach informacyjnych BLAISE i DIALOG.

System BLAISE (British Library Automated Information Service) udostępnia:

Zbiory danych MARC (Machine Readable Cataloguing) zawierające pełne opisy bibliograficzne wszystkich książek opublikowanych w Wielkiej Brytanii od roku 1950 i wszystkich publikacji nabytych przez Bibliotekę Kongresu USA od roku 1968, jak również zbiór danych AVMARC na temat materiałów audiowizualnych ze wszystkich dziedzin.

Zbiory danych MEDLARS-MEDLINE, TOXLINE, CHEMLINE, CANCERLIT, CANCERPROJ, CLINPROT i RTECS - zawierające dane bibliograficzne o artykułach, referatach i materiałach konferencyjnych, sprawozdaniach z prac badawczych i o innych publikacjach z medycyny, toksykologii i dziedzin pokrewnych z całego świata.

System DIALOG udostępnia ponad 300 z obecnie dostępnych za jego pośrednictwem baz zawierających informacje bibliograficzne o najnowszych publikacjach i wynikach badań z szerokiego zakresu tematyki naukowej, technicznej, społecznej i gospodarczej oraz humanistycznej, a w niektórych dziedzinach również bazy danych faktograficznych. Istnieje możliwość przeszukiwania według wybranego schematu (na to samo hasło) szeregu baz danych albo regularnego powtarzania tej operacji na określonej bazie danych w celu okresowej aktualizacji informacji na dany temat.

Jadwiga Oziugieł

Elżbieta Gorzkoś

2.04.1990 r.

KRONIKA

SEMINARIA, KONFERENCJE I SZKOLENIA Z ZAKRESU INFORMACJI NAUKOWEJ Z UDZIAŁEM PRACOWNIKÓW OIN PAN W 1989 r.

Seminaria naukowe OIN PAN

Opracowany program seminariów na II półrocze 1989 r. zawierał 8 tematów, ze względu na przedłużający się remont sali bibliotecznej Ośrodka, zostały zrealizowane tylko cztery tematy seminaryjne w listopadzie i grudniu 1989 r.

W referacie "Etyczno-prawne problemy wykorzystywania informacji" dr Jerzy Kozakiewicz przedstawił te zagadnienia, wiążące się ściśle z zagadnieniem humanizacji warsztatu pracy projektantów i użytkowników systemów informacyjnych. Dokonał także przeglądu niektórych sposobów zapewnienia ochrony banków danych przed niepożądanymi użytkownikami oraz sposobów ograniczania dostępu do określonych zespołów informacji; należy do nich m.in. macierzowy układ ochrony dostępu do danych.

Referat dr. Józefa Marońskiego na temat informatyzacji i automatyzacji procesów informacji zaznajomił słuchaczy z przedsięwzięciami organizacyjno-technicznymi, zapoczątkowanymi w Polsce pod koniec lat 70-tych, mającymi na celu stworzenie krajowej sieci połączeń informatycznych, powiązanej z międzynarodowymi systemami i centrami informacji. Sieć ta będzie m.in. umożliwiała funkcjonowanie poczty elektronicznej, przekazywanie i wymianę informacji, transmisję zbiorów, organizację systemu telekonferencji. Jednym z najważniejszych ogniw krajowej sieci połączeń

informatycznych jest i będzie nadal Instytut Podstaw Informatyki PAN.

Tematyka kolejnego seminarium dotyczyła działalności Pracowni Organizacji Informacji OIN PAN w Krakowie w zakresie informatyzacji procesów informacji bibliotecznej. Omówili ją kolejno: mgr Wiesław Babik i mgr Marek Nahotko. Projekty zautomatyzowanych systemów informacji bibliotecznej opracowane dla kilku instytucji regionu krakowskiego zawierają m.in.: bazy danych opisów formatów książek, bazy danych wypożyczeń książek, bazy danych rejestracji własnych wydawnictw, bazy danych wypożyczeń (czytelników), bazy danych firm komputerowych.

W drugim referacie w ramach tego seminarium przedstawiono zastosowanie minikomputera w informacji na przykładzie ZOINTE przy "Energoprojekcie" w Krakowie.

Na zebraniu seminaryjnym w grudniu 1989 r. zreferowany został temat "Systemy informacyjne o środowisku ze szczególnym uwzględnieniem systemu INFOTERRA", który przygotowali i wygłosili mgr inż. Teresa Gałczyńska i dr Andrzej Witowski. Przedstawiono w nim zagrożenie ekologiczne świata, które zmusza zarówno kraje wysokoprzemysłowe, jak i rozwijające się do wzmocnienia wysiłków na rzecz ochrony środowiska. Popierając i jednocząc te wysiłki, Organizacja Narodów Zjednoczonych utworzyła wielotematyczny Międzynarodowy System Informacji o środowisku INFOTERRA. W referacie omówiono strukturę organizacyjną i zasady jego funkcjonowania na świecie i w Polsce^{1/}.

^{1/}Zob. T. Gałczyńska, A. Witowski System Informacji o Środowisku - INFOTERRA. ZIN 1989 nr 2(55) s. 177-187.

Udział pracowników OIN PAN w konferencjach i sympozjach naukowych

1. Sesja Naukowa Polskiego Stowarzyszenia Filmu Naukowego,
Warszawa 9-10 maja 1989 r.

Sesja została zorganizowana z okazji 30 rocznicy powstania Stowarzyszenia. Wzięli w niej udział twórcy, działacze, krytycy i inni specjaliści zainteresowani filmem naukowym. Zostały wygłoszone referaty naukowe i okolicznościowe przemówienia, w których dokonano oceny dorobku Stowarzyszenia oraz nakreślono zadania na przyszłość w obecnej trudnej sytuacji kraju. Odznaczono zasłużonych działaczy i twórców. W sesji jako pracownik OIN PAN i jako członek Polskiego Stowarzyszenia Filmu Naukowego uczestniczył mgr A. Pietrzak.

2. II Międzynarodowe Sympozjum "Zastosowanie mikrokomputerów w inte", Katowice 4-6 października 1989 r.

Organizatorem Sympozjum były: Instytut Metali Nieżelaznych w Gliwicach, Ośrodek Postępu Technicznego w Katowicach, Centrum Informacji Naukowej Technicznej i Ekonomicznej w Warszawie oraz komisja ds. inte SITPH. Celem Sympozjum była ocena stanu wiedzy i osiągnięć praktycznych w tej dziedzinie, wymiana doświadczeń i informacji. Ze strony Ośrodka Informacji Naukowej PAN w sympozjum wzięli udział: mgr Wiesław Babik z referatem "Języki informacyjno-wyszukiwawcze stosowane do opisu zbiorów kartograficznych map", mgr Marek Nahotko z referatem "Wykorzystanie mikrokomputera w zakładowym OINTE Energoprojekt Kraków". Program sympozjum obejmował wygłaszanie referatów i dyskusję na sesjach plenarnych i w sekcjach, pokazy stosowanych programów, wystawę sprzętu technicznego wspomagającego inte: mikrokomputerów i kserografów, składopisów, urządzeń do mikrofilmowania, materiałów drukarskich, korektorskich i innych.

3. Ogólnopolska konferencja naukowa "Nowe metody rozpowszechniania informacji", Jachranka, 17-18 października 1989 r.

Organizatorem konferencji był Komitet Informacji Naukowej przy Prezydium Polskiej Akademii Nauk, a w skład Komitetu Programowego weszli przedstawiciele kilku ważniejszych placówek krajowych z zakresu informacji naukowej. Celem konferencji było dokonanie przeglądu nowych metod projektowania, realizacji i wdrażania systemów informacyjnych różnych typów^{2/}.

Ze strony Ośrodka Informacji Naukowej PAN udział w konferencji wzięło 6 osób, w tym 4 z referatami.

Szkolenia poza OIN PAN

Pracownicy Ośrodka wzięli udział w następujących kursach, seminariach i kursokonferencjach szkoleniowych organizowanych poza OIN PAN:

Lp.	Nazwa kursu, seminarium	Organizator	Liczba uczestników z OIN PAN
1	2	3	4
1.	Kurs administratorów baz danych micro CDS/ISIS dla mikrokomputerów kompatybilnych z IBMPC	IINTE	3
2.	Kurs podstawowy int	CINTE	2
3.	Kurs: Nowoczesne metody zarządzania i funkcjonowania ośrodków informacji	CINTE	2
4.	Kursokonferencje: Funkcjonowanie Centralnego Specjalistycznego Systemu Informacji o Pracach Naukowo-Badawczych, Rozwojowych i Ekspertyzach Naukowych SYNABA - ze szczególnym uwzględnieniem informacji naukowej	CINTE	3

^{2/} Szczegółowe omówienie konferencji - ZIN 1989 nr 2(55) s. 205-214

1	2	3	4
5.	Kursokonferencje: Kierunki polityki informacyjnej w Świecie	CINTE	1
6.	Kursokonferencja: Zakres i sposób udostępniania informacji z bazy danych systemu SYNABA	CINTE	2
7.	Kursokonferencje: Usługi i zasady finansowania działalności informacyjnej w ramach Krajowego Centrum Przetwarzania i Udostępniania Baz Danych	CINTE	1
8.	Kursokonferencja: Systemy ekspertowe. Cel, zasady budowy i zastosowanie systemów	CINTE	3
9.	Kurs: Zautomatyzowane systemy biblioteczno-informacyjne	Bibl.Gł. i Ośrodek Informacji N-Tech.Poli- techniki Wrocławskiej	3
10.	Cykle wykładów w ramach XLIII Kursu Zastosowań Matematyki	Instytut Matematyczny PAN	6
11.	Poczta elektroniczna i jej zastosowanie	Instytut Pod- staw Informa- tyki	2

W wymienionych wyżej imprezach szkoleniowych wzięło udział łącznie 28 pracowników OIN PAN.

Andrzej Pietrzak

2.01.1990 r.

S P I S T R E Ś C I

1. Wereszczyńska-Cisło B.: Definiowanie języka deskryptorowego 3
2. Artowicz E.: Udostępnianie wiedzy w dialogowym systemie informacyjno-wyszukiwawczym (on-line) 27
3. Chmielewska-Gorczyca E.: Instrukcja indeksowania do "Tezaurusza Informacji Naukowej" 79

Materiały i przyczynki

1. Sosińska-Kalata B.: Wykorzystanie technik sztucznej inteligencji we współczesnym bibliotekarstwie i systemach wspomagających wyszukiwanie online 111
2. Chmielewska-Gorczyca E.: Wyszukiwanie rzeczowe w publicznie niedostępnych katalogach online 139
3. Trawiński B., Rozwadowski A., Ziółek M.: Modele komputeryzacji procesów gromadzenia i opracowania wydawnictw ciągłych w bibliotece naukowej 165
4. Zybert E.B.: Amerykański System Informacji z Dziedziny Oświaty i Wychowania - ERIC 187

Recenzje i omówienia

1. Projektowanie języka informacyjno-wyszukiwawczego jako elementu systemu wyszukiwania informacji - Woźniak J. ... 201
2. Eksploatacja zautomatyzowanych wielodziedzinowych baz danych w Ośrodku Informacji Naukowej PAN - Dziugieł J., Gorzkoś E. 212

Kronika

- Seminaria, konferencje i szkolenia z zakresu informacji naukowej z udziałem pracowników OIN PAN w 1989 r. -
Pietrzak A. 225

CONTENTS

1. B. Wereszczyńska-Cisło	Defining of the descriptor language	3
2. E. Artowicz	Access to the knowledge in the on-line information retrieval system	27
3. E. Chmielewska-Gorczyca	The instruction of indexing to the "Thesaurus of scientific information"	79

Materials and Contributions

1. B. Sosińska-Kaląta	Using of the techniques of artificial intelligence in modern librarianship and in the systems assisting the online retrieval	111
2. E. Chmielewska-Gorczyca	Subject retrieval in the online public access catalogues	139
3. B. Trawiński, A. Rozwadowski, M. Ziółek	Models of computerization of the processes of acquisition and of elaboration of serials at the scientific library	165
4. E. B. Zybert	The american information system in the field of education - ERIC	187

Reviews and Surveys

1.	Designing of the information retrieval language as the element of information retrieval system - J. Woźniak	201
2.	Exploitation of the automatized multidiscipline data bases at the Scientific Information Centre of the Polish Academy of Sciences - J. Dziugieł, E. Gorzkoś	212

Chronicles	225
------------	-------	-----

СО Д Е Р Ж А Н И Е

1. Б. Верещиньска-Цисло **Определение дескрипторного языка** 3
2. Э. Аргович **Доступ к знаниям в диалоговой информационно-поисковой системе (on-line)** 27
3. Я. Хмелевска - Горчица **Инструкция индексирования и Тезаурусу научной информации** 79

М а т е р и а л ы и з м е ч а н и я

1. Б. Сосниьска - Калити **Использование техники искусственного интеллекта в современном библиотековедении и в системах вспомогательных поисков в режиме диалога**111
2. Э. Хмелевска - Горчица **Предметный поиск в общедоступных диалоговых каталогах (OPAC)** 139
3. Б. Травниьска, А. Розвадовски, М. Злехи **Модели автоматизации процессов накопления и обработки продолжающихся изданий в научной библиотеке** 165
4. Э.Б. Зиберт **Американская информационная система в области просвещения и воспитания (ERIC)** 187

Р е ц е н з и и и о б з о р ы

1. **Проектирование информационно - поискового языка как элемента информационно - поисковой системы** - Я. Вольник 201
 2. **Эксплуатация автоматизированных политематических баз данных в Центре научной информации польской академии наук** - Я. Дзюгал, Э. Гоклось 212
- Х р о в и к в 225

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 autor-
skie ok. 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.

Maszynopisy artykułów nie zamówionych nie przyjętych do druku redakcja nie zwraca.