

66

**PROPOZYCJE i MATERIAŁY**

Jadwiga Woźniak-Kasperek

**PODSTAWY  
BUDOWY TEZAURUSA**

**Poradnik**

WYDAWNICTWO  
SBP





**66**

PODSTAWY  
BUDOWY TEZAUROSA



STOWARZYSZENIE BIBLIOTEKARZY POLSKICH  
FEDERACJA BIBLIOTEK KOŚCIELNYCH – FIDES

# PROPOZYCJE I MATERIAŁY

Jadwiga Woźniak-Kasperek

## PODSTAWY BUDOWY TEZAUROSA

Poradnik

WYDAWNICTWO  
SBP



WARSZAWA 2005

Komitet Redakcyjny serii wydawniczej  
<<PROPOZYCJE I MATERIAŁY>>

Stanisław CZAJKA (przewodniczący), Piotr BIERCZYŃSKI, Lucjan BILIŃSKI,  
Marcin DRZEWIECKI, Janina JAGIELSKA, Janusz NOWICKI (sekretarz), Ewa  
STACHOWSKA-MUSIAŁ, Maria WASIK-ŚWIDERSKA, Elżbieta Barbara ZYBERT

**Publikacja dofinansowana przez Ministra Nauki i Informatyki  
z dotacji dla Federacji Bibliotek Kościelnych – FIDES**

Projekt graficzny okładki i strony tytułowej  
Wydawnictwo SBP

Redakcja techniczna i korekta  
Anna LIS

© Copyright by Stowarzyszenie Bibliotekarzy Polskich

ISBN 83-89316-53-6

CIP - Biblioteka Narodowa  
Woźniak-Kasperek Jadwiga

Podstawy budowy tezaurusa : poradnik / Jadwiga Woźniak-Kasperek ; Federacja  
Bibliotek Kościelnych FIDES. – Warszawa : Wydawnictwo Stowarzyszenia  
Bibliotekarzy Polskich, 2005. – (Propozycje i Materiały / Stowarzyszenie  
Bibliotekarzy Polskich ; 66)

## Spis treści

PRZEDMOWA .....	7
WPROWADZENIE .....	11
Część pierwsza	
ZWIĄZKI JĘZYKA INFORMACYJNEGO Z ELEMENTAMI SYSTEMU I JEGO OTOCZENIA .	15
Język informacyjno-wyszukiwawczy a zbiór informacyjny .....	17
Język informacyjno-wyszukiwawczy a użytkownicy informacji .....	33
Potrzeby informacyjne .....	35
Ważniejsze cechy potrzeb informacyjnych .....	41
Pytania użytkowników .....	44
Język informacyjno-wyszukiwawczy a środki techniczne .....	47
Część druga	
WAŻNIEJSZE WSKAZANIA DOTYCZĄCE BUDOWY JĘZYKA DESKRYPTOROWEGO A ZWŁASZCZA TEZAURUSA .....	65
Czym powinien charakteryzować się dobry język informacyjno-wyszukiwawczy, w tym język deskryptorowy? .....	68
Odpowiednio szeroki zakres tematyczny .....	69
Dostateczna szczegółowość wyrażań .....	72
Poprawna i odpowiednia siatka relacji paradygmatycznych .....	81
Jednoznaczność .....	82
Niesynonimiczność .....	85
Budowanie tezaury – najważniejsze działania .....	87
WAŻNIEJSZE LEKTURY .....	103





## PRZEDMOWA

Oddajemy do rąk czytelników publikację *Podstawy budowy tezaurusów*. Opracowanie to powstało z inicjatywy Federacji Bibliotek Kościelnych FIDES i Stowarzyszenia Bibliotekarzy Polskich, a jego realizację powierzono prof. dr hab. Jadwidze Woźniak-Kasperek, znając jej kompetencję w tej dziedzinie i talent popularyzatorski.

Inicjatywę tę podjęliśmy, gdyż wydaje się nam, że dziś środowisko bibliotekarzy bardzo potrzebuje praktycznych podręczników z zakresu tworzenia języków informacyjno-wyszukiwawczych. Odczuwa to Federacja FIDES, starając się od kilku lat opracować Tezaurus Nauk Kościelnych. Odczuwają to również bibliotekarze wielu innych sieci, szukając dobrego i efektywnego narzędzia dla rzeczowego opracowania książek w katalogach komputerowych i widząc wielkie zalety języka deskryptorowego lub przynajmniej charakterystyk rzeczowych opartych na słowach kluczowych<sup>1</sup>.

Opracowanie *Podstawy budowy tezaurusów* jest pomyślane jako praktyczna pomoc dla zespołów projektujących i opracowujących tezaursy z różnych dziedzin wiedzy oraz dla tych, którzy chcieliby coś zmienić, ulepszyć w stosowanym już języku informacyjnym.

---

<sup>1</sup> Swój osobisty pogląd „o wyższości słów kluczowych i deskryptorów nad hasłami przedmiotowymi” w komputerowych katalogach, autor tego tekstu wyraził w artykule: K. Gonet, *Dlaczego słowa kluczowe a nie hasła przedmiotowe. Co dalej z opracowaniem rzeczowym w bibliotekach FIDES? „FIDES-Biuletyn Bibliotek Kościelnych”* 2004 z. 1-2 s. 22-32.

Broszura ta będzie też, naszym zdaniem, nieocenioną pomocą dla studentów studiów na kierunku Informacja naukowa i bibliotekoznawstwo jako tymczasowy podręcznik do zajęć z przedmiotu „Projektowanie języków informacyjno-wyszukiwawczych”. Mamy uzasadnioną nadzieję, że wykładowcy tego przedmiotu chętnie będą zalecali tę lekturę swoim studentom. Sądzimy też, że studenci docenią logiczność, przejrzystość wykładu oraz praktyczność wskazówek zawartych w tej publikacji. Wskazówki te i konkretne instrukcje będą z pewnością mogli wykorzystać w swojej późniejszej pracy zawodowej w bibliotekach.

Chcielibyśmy zwrócić również uwagę na tę publikację jeszcze innym grupom czytelników. Widzimy użyteczność tego opracowania dla bibliotekarzy „z pierwszej linii frontu”, na co dzień borykających się z wielkim zadaniem opracowania rzeczowego zbiorów. Zanim zostaną wreszcie metodycznie opracowane kompletne tezaury lub przynajmniej słowniki słów kluczowych – często jesteśmy zmuszeni sami, spontanicznie tworzyć słowa kluczowe lub deskryptory. Tak od wielu lat czynią bibliotekarze z bibliotek członkowskich Federacji FIDES, zapewniając użytkownikom tymczasowy, niedoskonały, ale faktyczny dostęp rzeczowy do swoich zbiorów<sup>2</sup>.

Przedstawiana tu publikacja, szczególnie w swojej drugiej części, zawiera praktyczne instrukcje budowy deskryptorów. Na konkretnych przykładach pokazuje, kiedy hasło ma być jedno-wyrazowe, kiedy wielowyrazowe. W jakim ma być przypadku, i czy w liczbie pojedynczej czy mnogiej. Może to być wielką pomocą w codziennej pracy wspomnianych tu „frontowych” kolegów i koleżanek bibliotekarzy.

Kolejna grupa czytelników, którym chcielibyśmy polecić tę książkę to użytkownicy bibliotecznych systemów informacyjnych – czytelnicy naszych bibliotek. Większość z tych systemów informacyjnych daje możliwość wyszukiwania zaawansowanego – przeszukiwania bazy danych katalogowych lub bibliograficznych pod wieloma aspektami jednocześnie. Wykorzystuje się do tego właśnie deskryptory lub słowa kluczowe łączone operatorami logicznymi algebry Boole’a. Tak jest np. w najnowszym OPAC-

---

<sup>2</sup> Ks. K. Gonet lansuje od lat w Federacji FIDES żartobliwą zasadę mającą przełamać bibliotekarski perfekcjonizm: „Był niedoskonały ale istniejący, jest lepszy od bytu doskonałego, którego ciągle jeszcze nie ma”.

MAKWWW dla baz programu MAK. Na takiej metodzie oparte są również przeszukiwania w wyszukiwarkach korzystających z protokołu Z39.50: KARO i popularny FIDKAR<sup>3</sup>.

Użytkownicy przeżywają jednak wielkie rozczarowania otrzymując, niestety często, negatywną odpowiedź na swoje misternie zbudowane pytanie. Na ogół przyczyną jest nieznanym użytkownikom fakt, że różne bazy, katalogi itp. są opisane w różnych językach informacyjno-wyszukiwawczych, że nie ma żadnej globalnej unifikacji czy standaryzacji tych języków. Często jednak przyczyną niepowodzeń jest również nieznanostwo zasad budowy haseł (deskryptorów, słów kluczowych) używanych w tych językach. Dlatego właśnie gorąco zachęcamy tych użytkowników komputerowych systemów bibliotecznych, którzy chcą i mogą to uczynić, do zapoznania się z zawartymi w tej publikacji praktycznymi instrukcjami tworzenia haseł opracowania rzeczowego. Z pewnością pomoże to w efektywnym korzystaniu z aparatu wyszukiwawczego naszych bibliotecznych, komputerowych katalogów i baz danych bibliograficznych.

Autorce opracowania, prof. Jadwidze Woźniak-Kasperek, składamy wielkie podziękowania za podjęty trud opracowania tego użytecznego, praktycznego, a przez to bardzo cennej publikacji.

Stowarzyszeniu Bibliotekarzy Polskich pragniemy serdecznie podziękować za chętnie podjęcie idei wydania tej pozycji oraz za wspieranie, życzliwą współpracę z Federacją Bibliotek Kościelnych FIDES przy szybkiej jej realizacji.

ks. Krzysztof Gonet

---

<sup>3</sup> FIDKAR to popularna, tania multiwyszukiwarka do komputerowych, bibliotecznych baz danych opracowana przez Federację Bibliotek Kościelnych FIDES. Więcej na temat FIDKAR można przeczytać w artykułach ks. K. Goneta: *FIDKAR – a co to jest?* „Poradnik Bibliotekarza” 2005 z. 1 s. 8-9; *FIDKAR konkurencją dla KARO?* „Bibliotekarz 2005 z. 1 s. 34; Również w FIDES-Biuletynie Bibliotek Kościelnych 2004 z. 1-2 znajdujemy kilka artykułów ks. K. Goneta na ten temat: *O FIDKAR i FIDSERW mówiono pozytywnie...* s. 13-17; *FIDKAR a dalsze losy katalogu centralnego FIDES – zachęta do dyskusji* s. 18-21. W tym samym numerze Biuletynu FIDES znajdujemy też opinię o FIDKAR prof. Jadwigi Woźniak-Kasperek: *FIDKAR, czyli (bardzo krótka) pochwała rozsądku, umiaru i kompetencji.*



## WPROWADZENIE

Praca ta powstała na zamówienie Federacji Bibliotek Kościelnych FIDES jako materiał warsztatowy do tworzonego języka deskryptorowego z zakresu nauk kościelnych<sup>1</sup>. Zawarta w drugiej części parainstrukcja tworzenia języka deskryptorowego, zwłaszcza jego słownika zwanego tezauresem, ma za cel pomagać projektantom języka w rozwiązywaniu pewnych problemów praktycznych oraz przyczyniać się do większej spójności i jednorodności efektów pracy poszczególnych osób poprzez wskazanie jednego zestawu ważniejszych zasad i reguł postępowania.

Publikacja ma charakter materiału metodycznego, poradnikowego. Nie jest pracą naukową. Mimo to zdecydowałam się na włączenie doń wychodzących poza ramy instrukcji metodycznej rozważań o miejscu języka w systemie informacyjno-wyszukiwawczym (SIW) oraz związkach języka informacyjno-wyszukiwawczego (JIW) z elementami systemu. Jestem bowiem głęboko przekonana, że dobrze jest, gdy język informacyjny nie powstaje w oderwaniu od rzeczywistości, dla której jest tworzony i w której będzie funkcjonował. Bez wiedzy co tworzy kontekst projektowania i znajomości tego kontekstu nie da się zbudować dobrego

---

<sup>1</sup> W 2001 r. ukazało się wydanie studyjne *Słownika słów kluczowych z teologii* opracowanego przez ks. dra Piotra Klimka, który w tym samym roku został redaktorem naczelnym redakcji *Tezaurusu nauk kościelnych*. Prace nad TNK postępują nie tak szybko, jak wszyscy chcielibyśmy, ale jednak posuwają się naprzód. Obecnie są dostępne dwie pierwsze części – wersja studyjna *Tezaurusu biblistyki* i *Tezaurusu liturgiki*. W opracowaniu są Tezaurus teologii moralnej i Tezaurus nauk o rodzinie. Sukcesywnie mają powstawać kolejne części. Zob. bazy dostępne pod adresem [www.fides.org.pl](http://www.fides.org.pl)

języka. Nienaukowy charakter pracy jest też (w moim odczuciu) usprawiedliwieniem dla pewnej swobody, ale nie nonszalancji, terminologiczno-językowej, w rezultacie której tekst miał być lżejszy i łatwiejszy w czytaniu, np. na ogół wymiennie stosuję terminy biblioteka i system (informacyjny), automatyzacja i komputeryzacja, gdyż uważam, że w kontekście tej pracy precyzyjne rozdzielanie użyć nie jest konieczne. Staralam się pisać językiem prostym, na ile możliwe jest pisanie prosto o tak niekiedy skomplikowanych zagadnieniach. Zdaję sobie sprawę, że czytelnicy momentami mogą być zmęczeni stosowaną przeze mnie fachową terminologią. Staralam się kontrolować to zjawisko, ale z pewnością nie w pełni mi się udało, za co przepraszam.

Projektowanie języka informacyjno-wyszukiwawczego jest sztuką. Nie wszystko da się skodyfikować i przewidzieć, niemniej pewne sytuacje i problemy pojawiają się w trakcie tworzenia większości języków. Usiłowałam je tu opisać i usystematyzować. Z różnych powodów praca ma charakter skrótowy i wstępny. Moim pragnieniem byłoby po otrzymaniu informacji zwrotnej od czytelników, w tym projektantów języka, poprawić ją, rozwinąć i uaktualnić w kolejnym wydaniu. Być może edycja ta mogłaby być połączona z publikacją roboczej czy pierwszej wersji planowanego Tezaurusu Nauk Kościelnych (lub jego części), na materiale którego pewne zagadnienia mogłyby być lepiej objaśnione i sprecyzowane. Budowany od lat Katalog Centralny FIDES<sup>2</sup> (dziś również multiwyszukiwarka FIDKAR) zyskałby dodatkowo na wartości, gdyby w skład jego rekordów wchodziły charakterystyki rzeczowe utworzone we wspólnym kontrolowanym języku informacyjnym, nie tylko za pomocą swobodnych słów kluczowych. Osoby, które korzystają z Katalogu Centralnego FIDES, podobnie jak użytkownicy innych bibliotek i sieci, nie zdają sobie sprawy z mnogości istniejących i stosowanych systemów opracowania rzeczowego. Przeszukując tworzone przez nas bazy, przenoszą doświadczenia z jednej biblioteki czy bazy na drugą, dziwiąc się niekiedy, a częściej frustrując. Wspólnota języka i indeksowania jest tu ze wszech miar dobra i pożądana. Biblioteczno-informacyjna wieża Babel jest coraz większym ciężarem

---

<sup>2</sup> K. Gonet: *FIDKAR a dalsze losy Katalogu Centralnego FIDES – zachęta do dyskusji*. „FIDES Biuletyn Bibliotek Kościelnych” 2004 nr 1-2 s. 18-21.

i utrudzeniem. Od 1991 r. bazy FIDES oferują dostęp za pomocą typowanych przez bibliotekarzy słów kluczowych. Jednak, jak pisze ks. Krzysztof Gonet, „w przyszłości będziemy się starali te, spontanicznie tworzone słowa, uporządkować w formie słownika. Ten słownik miałby proponować używane przez FIDES formy słów kluczowych. Przewidywaliśmy również, że z czasem słownik ten określałby również podstawowe relacje pomiędzy słowami (synonimy, homonimy, terminy szersze i węższe, skojarzone) stając się powoli tezaurem”<sup>3</sup>. Ufam, że moment ten jest coraz bliższy.

Kończąc, pragnę podziękować Osobom, które pomogły mi nadać publikacji jej kształt i jakość. Dziękuję niestrudzonemu ks. Krzysztofowi Gonetowi, dyrektorowi Biura ds. Rozwoju i Komputeryzacji Federacji Bibliotek Kościelnych FIDES, za współpracę intelektualną oraz za zmobilizowanie mnie (by nie rzec, zmuszenie) do zajęcia się właśnie tym tematem. Dziękuję także pani Dorocie Szumilas za możliwość zapoznania się z wersją roboczą tworzonego przez Nią Tezaurusu Teologii Moralnej<sup>4</sup>, z którego zaczerpnęłam pewną część przykładów cytowanych w publikacji.

*Jadwiga Woźniak-Kasperek*

---

<sup>3</sup> K. Gonet: *Dlaczego słowa kluczowe a nie hasła przedmiotowe? Co dalej z opracowaniem rzeczowym w bibliotekach FIDES*. „FIDES Biuletyn Bibliotek Kościelnych” 2004 nr 1-2 s. 24.

<sup>4</sup> D. Szumilas: *Potrzeba opracowania polskiego tezaurusu teologii moralnej i etyki*. „FIDES Biuletyn Bibliotek Kościelnych” 2004 nr 1-2 s. 33-38.





## Część pierwsza

# ZWIĄZKI JĘZYKA INFORMACYJNEGO Z ELEMENTAMI SYSTEMU I JEGO OTOCZENIA

Języki deskryptorowe należą do rodziny języków informacyjnych (zwanym też językami informacyjno-wyszukiwawczymi «JIW» lub językami indeksowania). JIW są to sztuczne systemy znaków, zaprojektowane i utworzone przez człowieka. Za pomocą tych języków opisuje się dokumenty oraz „przekłada” na nie pytania czytelników (zwanym też użytkownikami), aby móc w stosownym czasie odszukać dokumenty odpowiadające aktualnemu zapotrzebowaniu użytkownika. Inaczej można powiedzieć, że języki informacyjne, w tym deskryptorowe, tworzone są w celu wyrażenia informacji, których nie dałoby się sformułować dostatecznie precyzyjnie w języku naturalnym i/lub gdy byłoby to nieekonomiczne, nieefektywne itp. Język naturalny jest, obrazowo mówiąc, zbyt bogaty z punktu widzenia potrzeb i możliwości informacyjno-wyszukiwawczych. JIW odwołując się do różnych cech i własności języków naturalnych, bogactwo

**Język  
informacyjny**

to poskramiają, próbując (lepiej lub gorzej) stać się narzędziem, za pomocą którego można sprawnie i skutecznie wyrażać pewne informacje, a następnie je odszukiwać ku pożytkowi i zadowoleniu czytelników.

Najczęstsze typy języków informacyjno-wyszukiwawczych spotykane w serwisach informacyjnych to: klasyfikacje, języki haseł przedmiotowych, języki słów kluczowych i języki deskryptorowe.

Największy wpływ na skuteczność i efektywność JIW ma ich jakość oraz poprawność wykorzystania. Szczególnie chodzi tu o jakość tworzonych opisów dokumentów oraz jakość „przekładu” pytań czytelników. Bardzo dużo zależy od kwalifikacji i rzetelności osób wykorzystujących JIW w codziennej pracy biblioteczno-informacyjnej.

Teoretycznie im bogatszy jest język informacyjno-wyszukiwawczy, im bardziej rozbudowany, skomplikowany, tym większe powinien mieć możliwości uzyskiwania wysokiej efektywności. W praktyce jest inaczej, gdyż trudności spowodowane czynnikami subiektywnymi, w tym trudność nauczania się języka zarówno przez bibliotekarzy, jak i użytkowników, wpływają negatywnie na skuteczność posługiwania się nim. Nie można ignorować faktu, że coraz więcej wyszukiwań realizowanych jest samodzielnie przez tzw. użytkowników finalnych (końcowych), którzy nie ko-

rzystają z pomocy pośredników, jakimi tradycyjnie byli bibliotekarze czy pracownicy informacji. Do zmiany tej muszą się dostosowywać także języki informacyjno-wyszukiwawcze, zarówno te nowe, dopiero projektowane, jak i już wykorzystywane, jeśli nie chcą się okazać bezużyteczne. Dzisiaj JIW musi być na tyle prosty i intuicyjny, żeby użytkownik mógł się nim względnie dobrze posługiwać bez pomocy osoby wykwalifikowanej. Obrazowo można powiedzieć, że ciężar pracy bibliotekarza zostaje dziś przeniesiony z wyszukiwania (z realizacji funkcji wyszukiwawczej JIW) na tworzenie dobrych, alternatywnych opisów dokumentów, spośród których użytkownik sam będzie dokonywał selekcji, wyszukiwania.

## JĘZYK INFORMACYJNO- -WYSZUKIWAWCZY A ZBIÓR INFORMACYJNY

Zbiór informacyjny tworzą dokumenty prymarne (pierwotne), na przykład książki, czasopisma, broszury, dokumenty życia społecznego, dokumenty kartograficzne, druki muzyczne, dokumenty dźwiękowe, niepublikowane prace magisterskie, doktorskie, rękopisy, fotografie, dokumenty elektroniczne, kasety wideo. W systemie dokumentacyjnym, a więc takim, któ-

Zbiór  
informacyjny



**Zbiór  
wyszukiwawczy**

ry działa nie na pełnych tekstach dokumentów, a na ich skrótowych reprezentacjach zwanych charakterystykami wyszukiwawczymi lub reprezentacjami dokumentów, zbiór ten poddawany jest przekształceniom, nazywanym w zależności od warunków, w jakich przekształcenia te mają miejsce i ich specyfiki, transformacjami metainformacyjnymi, indeksowaniem, katalogowaniem lub opracowaniem dokumentacyjnym zbiorów. W rezultacie przekształceń powstają nowe dokumenty – pochodne – na przykład opisy bibliograficzne czy rzeczowe, będące podstawą utworzenia zbioru wyszukiwawczego, w którym wyszukuje się informacje. W systemie dokumentacyjnym na zbiór wyszukiwawczy składają się charakterystyki wyszukiwawcze (reprezentacje) dokumentów, podczas gdy dokumenty pochodne mogą zawierać ponadto elementy, np. adnotacje czy abstrakty, na których w danym systemie nie realizuje się operacji wyszukiwania. Przekształceniem nas interesującym jest to, które prowadzi do utworzenia reprezentacji treściowej dokumentów prymarnych. Ono najlepiej wykorzystuje i nie może się obyć bez języka informacyjnego.

**JIW  
a treść  
dokumentów**

Między językiem informacyjnym a zbiorem dokumentów prymarnych istnieje bezpośredni i ścisły, choć często niedoceniany, związek. Na ogół

dostrzega się związek między treścią dokumentów prymarnych a zakresem tematycznym JIW, zapominając o innych ważnych parametrach, które bezwzględnie należałoby wziąć pod uwagę, tworząc JIW, takich jak wielkość zbioru (liczba dokumentów), tempo przyrostu czy tempo starzenia się informacji zawartej w dokumentach. Nie wolno też zapominać o cechach pozat treściowych dokumentów takich jak forma zawartości, forma wydawnicza, cechy fizyczne, w tym cechy nośnika. Są one ważne i często wykorzystywane przy wyszukiwaniu informacji.

Obecnie daje się zauważyć tendencja do zapisywania informacji o formie nie w charakterystyce wyszukiwawczej dokumentu, a w pewnych polach rekordu bibliograficznego. Na przykład, w rekordzie bibliograficznym MARC 21 dla książki jest to pole 008<sup>1</sup>, w którym można umieszczać m.in. kody wskazujące<sup>2</sup>:

- charakter zawartości dokumentu (ang. *nature of contents*), np. bibliografia, indeks;

**JIW  
a forma  
dokumentów**

---

<sup>1</sup> Większość oznaczeń dopuszczonych do stosowania w polu 008 nie została uwzględniona w polskiej wersji formatu MARC dla opisu bibliograficznego książki (M. Lenartowicz, A. Paluszkiwicz: *Format USMARC rekordu bibliograficznego dla książki*, Warszawa 2002, s. 15).

<sup>2</sup> *MARC 21 Concise Bibliographic : control field 008*  
<http://www.loc.gov/marc/bibliographic/ecbd008s.html>

- obecność i rodzaj materiałów graficznych, w tym ilustracyjnych dokumentu (ang. *illustrations*), np. mapy, portrety, tablice;
- formę literacką (ang. *literary form*);
- typ nośnika lub zapisu informacji na nośniku (ang. *form of item*), np. CD-ROM;
- inne informacje wskazujące, że mamy do czynienia na przykład z biografią, materiałami konferencyjnymi, księgą pamiątkową.

Konsekwentne wypełnianie pola 008 (lub jakiegokolwiek innego pola o takim przeznaczeniu) mogłoby być pomocne w oznaczaniu formy dokumentu, przyczyniłoby się do skrócenia charakterystyk wyszukiwawczych i zwiększenia precyzji wyszukiwań. Rozwiązanie to będzie jednak efektywne dopiero wtedy, gdy oprogramowanie, jakim dysponuje biblioteka, zostanie wyposażone w odpowiednio działające mechanizmy tworzenia indeksów i filtrowania wyszukiwanej informacji. W przeciwnym razie „w indeksie mogą pojawiać się hpr [hasła przedmiotowe rozwinięte – przypis mój] wyglądające tak samo, różniące się jedynie niewidocznym dla czytelnika kodem podpola, a oznaczające co innego, np. Słowacki, Juliusz (1809-1849) \$x korespondencja i Słowacki, Juliusz (1809-1849) \$v korespondencja. Tylko sięgając do rekordu zapisanego w formacie MARC 21, można odróżnić, czy w danym przypadku

Forma  
dokumentów  
a format

mamy do czynienia z określnikiem formy czy treściowym”<sup>3</sup>. Nie wystarczy jednak utworzyć indeksy i uruchomić filtry, trzeba jeszcze tak zbudować lub zmodyfikować interfejs katalogu, żeby użytkownik mógł wykorzystać informacje zawarte w polu 008. Czasami można spotkać się z argumentem, że wiedza użytkownika nie pozwala mu w pełni wykorzystać subtelności opisu rzeczowego. Ale to samo można powiedzieć w odniesieniu do informacji „odzyskiwanej” z różnych pól formatu. Użytkownik równie mało wie o formacie, co o językach informacyjnych. Jedynym rozwiązaniem jest tu przeniesienie pewnych ciężarów wyszukiwania z użytkownika na interfejs i możliwości selekcyjne systemu.

Dokumenty pochodne w systemach dokumentacyjnych powstają w rezultacie transformacji treści<sup>4</sup> i transfor-

---

<sup>3</sup>P. Rygiel: *Określenie formy dokumentu w językach haseł przedmiotowych i rekordach bibliograficznych formatu MARC 21*. „Przeгляд Biblioteczny” 2004 z. 3-4 s. 224.

<sup>4</sup> Terminy *treść* i *przedmiot* (także temat) nie mają w informacji naukowej jednoznacznych sensów. „Pojęcie treści i przedmiotu książki zwyczajnie mieszały się, co bardzo utrudniało uchwycenie istoty każdego z dwóch rodzajów katalogu rzeczowego. W bibliologii różniczenie to jest ważne dlatego, że treść i przedmiot bynajmniej nie zawsze idą równoległe do siebie. Są książki oparte na jednej zasadzie metodologicznej, mające zatem podobną treść, ale traktujące o różnych przedmiotach, np. praca psychologiczna „o zbrodni i karze”, „o powszechnych wyborach”, itp. Zdarza się też, że o jednym przedmiocie traktują książki różnej treści, zależnie od tego, z jakiego stanowiska go rozpatrują, na jakie cechy przedmiotu zwracają swoje badania, do jakiego działu piśmiennictwa lub nauki należą

**Informacje  
kontekstowe**

macji informacji pozatrześciowych, co do których zakłada się, że mogą być istotne, relewantne dla potencjalnego użytkownika. David Blair<sup>5</sup> nazwał te dwie kategorie informacji odpowiednio informacjami treściowymi (tematycznymi, przedmiotowymi, rzeczowymi; ang. *subject*) i kontekstowymi (ang. *context*). Mówiąc o informacji kontekstowej, zwykle wymienia się takie jej kategorie jak: autor, data publikacji, objętość dokumentu, jego język, forma wydawnicza itd. W jakimś sensie do kategorii informacji kontekstowej należą też informacje na temat środowiska wyszukiwawczego, na przykład mówiące o liczbie wyszukiwań danego dokumentu, jego cytowań (algo-

---

opinie o tym przedmiocie, np. o „powszechnych wyborach” oprócz psychologii, mówi historia, socjologia i publicystyka; „zbrodnia i kara” jest przedmiotem nauk prawnych, historii, socjologii, biologii, religii, a nawet tematem literatury pięknej”. Adam Łysakowski: *Katalog przedmiotowy. Cz. 1, Teoria*. Warszawa 2002, s. 44-45. Dokonując pewnego uproszczenia, można przyjąć, że przedmiotem dokumentu jest obiekt odpowiadający jakiemuś fragmentowi rzeczywistości; treścią zaś przynależne danemu przedmiotowi w danym dokumencie jego ujęcia, przedstawienia, interpretacje, lokalizacje itp., czyli jak pisał Ungurian, „zespół cech charakteryzujących tekst w sposób umożliwiający zakwalifikowanie dokumentu do określonej dyscypliny naukowej. Na cechy te składają się sądy i przedstawienia powzięte przez autora o danym przedmiocie (co i jak autor myśli), jego psychiczne i społeczne nastawienie, intencje: naukowe bądź publicystyczne, stawiane sobie cele i zadania, nazewnictwo specyficzne dla określonej dyscypliny oraz styl”. O. Ungurian: *Elementy teorii języków informacyjnych*. Warszawa 1976, s. 77.

<sup>5</sup>David C. Blair: *Language and representation in information retrieval*. Amsterdam 1990.



rytmy rankingowe w wyszukiwarkach internetowych często bazują właśnie na statystyce cytowań poszczególnych dokumentów sieciowych), wypożyczeń, datach wprowadzenia informacji o dokumencie do bazy itd.

Niekiedy projektując język informacyjno-wyszukiwawczy zapomina się, że różne dokumenty wymagają zróżnicowanej informacji kontekstowej. Na przykład, w systemie informacyjno-wyszukiwawczym dla rozgłośni radiowej zapewne największą część zbioru wyszukiwawczego będą stanowiły reprezentacje typowych dla tego miejsca dokumentów, tj. nagrań audycji radiowych i muzyki. Kategorie informacji kontekstowej przeniesione z bibliotek nie sprawdzą się w tym środowisku wyszukiwawczym, bo inna jest specyfika dokumentów tworzących tu zbiór informacyjny i inne potrzeby użytkowników tego systemu. Podobna sytuacja może mieć miejsce w bibliotece. Różne typy dokumentów gromadzonych przez biblioteki i różne oczekiwania użytkowników wymagają różnych kategorii kontekstu. Informacje kontekstowe są relatywnie proste do wyrażenia w charakterystyce dokumentu, gdyż cechy kontekstowe charakteryzują się binarnością, tzn. dokument albo posiada daną cechę, albo jej nie posiada. Przyczyn pojawiających się niekiedy problemów z określeniem wartości cechy w odniesieniu

do konkretnego dokumentu należałoby upatrywać raczej w nieostrości lub niewyraźności<sup>6</sup> wyrażenia cechy a nie w samych dokumentach. Trudniejsze jest przewidzenie, jaki wpływ na efektywność wyszukiwania informacji będą miały błędnie zaprojektowane lub pominięte kategorie informacji kontekstowej.

**Informacje kontekstowe, cd.**

Kontekst jest czynnikiem niedocenianym w tworzeniu reprezentacji dokumentów i w wyszukiwaniu. Wiele nawet komercyjnych systemów informacyjno-wyszukiwawczych, nie mówiąc o zautomatyzowanych katalogach bibliotecznych, jest budowanych w nawiązaniu do tradycyjnego modelu biblioteki i reprezentacji dokumentu, a przecież znaczna część podstaw aktywności językowych zawiera się w faktach pozajęzykowych. Wynikałoby z tego, że informacje dotyczące przynajmniej typowych działań, w wyniku których powstają lub w trakcie których wykorzystywane są określo-

---

<sup>6</sup> „Nieostrość wyrażenia powoduje kłopoty z jego zakresem, a ściślej kłopoty przy podejmowaniu decyzji czy dany przedmiot zaliczyć do tego zakresu. Tymczasem niewyraźność jest „mankamentem” znaczenia, polegającym na tym, że tylko o pewnych cechach wiadomo, iż do niego należą, i tylko o pewnych wiadomo, że nie należą; ale są i takie, co do których nie sposób rozstrzygnąć, czy stanowią elementy tego znaczenia, czy też nie stanowią. Nie sposób zaś rozstrzygnąć nie na skutek nieznamości języka, tego, iż użytkownik języka niewyraźnie rozumie dane wyrażenie, choć ma ono sprecyzowane znaczenie, lecz na skutek tego, iż owo znaczenie jest nie sprecyzowane”. J. Pelc: *Wstęp do semiotyki*. Warszawa 1984, s. 183-184.

ne dokumenty związane z tym czy innym tematem, mogą być użyteczne dla reprezentowania tych dokumentów w celu ich późniejszego odszukania. Informacje te najczęściej przybierają postać właśnie informacji kontekstowych. Należy przy tym pamiętać, że informacje kontekstowe nie zawsze muszą być jawnie zawarte w dokumencie. Czasami można je zidentyfikować (wydedukować), jedynie znając sytuację, w których dokument i zawarte w nim informacje znajdują właściwe zastosowanie. Tekst, który teraz Państwo czytacie, nie był pisany z myślą o studentach kierunku informacja naukowa i bibliotekoznawstwo. Nie jest ani podręcznikiem, ani skryptem akademickim. Mimo to z pewnością może być wykorzystywany w dydaktyce, zwłaszcza że nie ma podręcznika akademickiego poświęconego projektowaniu języków informacyjno-wyszukiwawczych. Informacja taka, stricte kontekstowa i nie wyrażona jawnie ani w tytule ani we *Wprowadzeniu*, moim zdaniem, powinna znaleźć się w charakterystyce wyszukiwawczej tej publikacji.

Zakres języka informacyjno-wyszukiwawczego jest nierozzerwalnie związany z tematyką (i specyfiką – cechy kontekstowe) zbioru informacyjnego, co nie znaczy, że organizacja tego pola zawsze jest odbiciem świata informacji, jaki wyłania się z dokumentów

**JIW  
a obraz  
świata**

(rzeczywistości dokumentacyjnej). Języki informacyjno-wyszukiwawcze wcale nie tak rzadko budowane są w oparciu albo o obraz samej rzeczywistości (realności), albo o obraz właściwy dla tego czy innego systemu nauk. JIW często odwzorowują naukowy obraz świata i klasyfikację nauk. Tymczasem wielu użytkowników, na przykład bibliotek publicznych, operuje na ogół inną „wizją” świata, opartą nie na taksonomiach naukowych, a raczej na obrazie potocznym lub „popularno-naukowym”. Najczęściej również dokumenty zgromadzone w tych bibliotekach nie są pracami naukowymi i zdają się być na siłę wtłoczone w ramy systematyki naukowej<sup>7</sup>. Pro-

---

<sup>7</sup> W informacji naukowej tym, którzy potrzebują dostępu do wiedzy eksperckiej poświęcano przez lata sporo uwagi. Można zaryzykować stwierdzenie, że badania zachowań w zakresie poszukiwania informacji (zapotrzebowania na informację) dałyby się w większości sprowadzić do badań zachowań naukowca poszukującego informacji. Można spierać się, czy poświęcanie uwagi naukowcom przez całe lata było dobrą czy złą przysługą dla informacji naukowej, zwłaszcza jeśli chodzi o jej behawioralne aspekty. Postępowanie w nauce jest kwestią wysoce podporządkowaną wzorcom wraz ze specyficznymi normami określającymi zachowania dopuszczalne (akceptowalne), zaś rola poszukiwania informacji i jej wykorzystania jest w tych normach raczej jasno zdefiniowana. Naukowcy mają do czynienia z problemami stosunkowo dobrze ustrukturalizowanymi, dla których mogą już istnieć drogi dochodzenia do rozwiązania (wzorcy), a wiele innych osób mogło tymi drogami wcześniej podążać przez całe lata. Analizowanie tych wcześniejszych prac, poszukiwanie pominiętych skojarzeń, przyjmowanie podejść wcześniej uważanych za ślepe uliczki, podejmowanie obiecujących wzmianek przeoczonych przez wcześniejszych badaczy jest zasadniczą częścią profesji naukowca. Wyrażanie uznania dla wcześniejszych

blem odwzorowania ogólnego<sup>8</sup> i naukowego obrazu świata nie tylko istnieje, ale jest jednym z tych, które do tej pory albo nie były zauważane albo – z różnych powodów – zredukowane do problematyki obrazu naukowego. A przecież nie wszystkie dokumenty gromadzone w bibliotekach i charakteryzowane za pomocą JIW odwzorowują naukowy obraz świata oraz nie zawsze (a w pewnych typach bibliotek na ogół) pytania formułowane przez użytkowników nawiązują do obrazu naukowego a właśnie do mniej lub bardziej zindywidualizowanego obrazu potocznego. Ignorowanie tej sytuacji nie wydaje się ani uprawnione, ani usprawiedliwione. Obraz naukowy kształtowany jest przede wszystkim przez odniesienia do rzeczywistości,

**JIW  
a obraz  
świata,  
cd.**

---

prac poprzez cytowanie, a przez to odnotowywanie użycia dokumentu również jest częścią reguł gry, lub inaczej mówiąc, częścią systemu relewancji i wzorców naukowca. W konsekwencji, identyfikacja informacji potencjalnie użytecznej, projektowanie systemów informacji oraz dostarczanie usług informacyjnych jest dla profesjonalisty w informacji naukowej stosunkowo prostą sprawą bynajmniej nie z powodu jego systemu relewancji, bo ten wiąże się z dokumentami i ich wykorzystaniem oraz pokrywa się z systemem relewancji naukowca – podobną wagę przykładają się do dokumentów, wyszukiwania oraz przekazywania informacji. Jednakże nauka jest tylko jedną ze sfer życia człowieka. Inne rodzaje pracy, w jaką angażują się ludzie, są zbyt liczne by je wymieniać, lecz każda jest odmienna i niewielu przysługują te same cechy co nauce. Traktowanie wszystkich potencjalnych użytkowników informacji tak, jakby byli naukowcami i patrzeć na wszystkie konteksty pracy jak na kontekst w nauce jest, powiedzmy, co najmniej nieuprawnione.

<sup>8</sup> Obraz ogólny, język ogólny bywają też nazywane potocznymi lub naiwnymi.

**Potoczny  
obraz  
świata**

dążąc do możliwie wiernego jej odwzorowania, do minimum redukując subiektywizm. Obraz świata zawarty w języku ogólnym uwzględnia zarówno obiektywnie dane cechy rzeczywistości, jak i inne uwarunkowania rządzące myśleniem i ocenami człowieka. Naukowe poznanie świata i pochodny od niego naukowy obraz świata w dużej mierze, choć niecałkowicie, są wspólne ludziom mówiącym różnymi językami, z kolei potoczny obraz świata jest utrwalony i odtwarzany poprzez język ogólny i dlatego każdy język etniczny ma sobie właściwy obraz świata, a właściwie obrazy świata. Na przykład, w języku ogólnym specjalistycznym terminowi motyl z taksonomii zoologicznej (entomologicznej) odpowiadają dwa słowa: motyl – najbliższy odpowiednik entomologicznego terminu motyl dzienny i ćma – ogólny odpowiednik terminu motyl nocny. U podstaw binarnego podziału tej grupy leżą nie tylko obiektywne właściwości świata owadów, co widoczne jest w taksonomii entomologicznej (motyl dzienny – motyl nocny), ale także ogólniejszy system kulturowych preferencji człowieka widocznych w tej grupie leksykalnej i rozciągających się również na inne sfery leksyki. W języku ogólnym nazwa motyl konotuje między innymi lekkość, radość i urodę, ćma zaś, na skutek etymologicznego

związku z ciemnością, nie budzi takich skojarzeń. Zbliżeniu języka informacyjno-wyszukiwawczego do świata użytkowników służą m. in. kategoryzacje stosowane w JIW. Kategoryzacja pozwala na ujawnienie explicite tych znaczeń, które wiążą znak języka ze „światem za słowami”, czyli z relewantnymi dla użytkownika elementami rzeczywistości pozajęzykowej.

Projektując JIW, należy pamiętać, że wprowadzicie użytkownik skierowujący pytanie do dokumentacyjnego systemu informacyjno-wyszukiwawczego nie może oczekiwać innej odpowiedzi, jak tylko takiej, która jest informacją o określonych dokumentach, to jednak na ogół interesuje się on nie samym dokumentem, lecz obiektem jakiejś rzeczywistości, o którym się mówi w dokumencie. Stąd można przyjąć, że nadawca określonego wyrażenia JIW zamierza poznać pewien obiekt rzeczywistości pozadokumentacyjnej, posługując się w tym celu dokumentem dotyczącym tego obiektu i posiadającym ponadto pewne określone własności związane ze sposobem prezentacji tego obiektu i samą formą dokumentu. Rzeczywistość, o której „mówi się” w językach informacyjno-wyszukiwawczych, składa się z dwóch płaszczyzn – rzeczywistości desygnowanej przez wyrażenia JIW, czyli rzeczywistości dokumentacyjnej oraz rzeczywistości desygnowa-

**Kategoryzacja**

**Denotacja  
i konotacja  
w JIW**

nej przez wyrażenia języków naturalnych, inaczej mówiąc rzeczywistości desygnowanej przez informacje zawarte w dokumentach. Na przykład, hasła przedmiotowe języka KABA:

Jan Paweł II (papież, 1920-2005)  
-- i komunizm;

Komunizm i chrześcijaństwo -- Kościół katolicki -- Europa Wschodnia -- historia -- 1970-...

desygnują bezpośrednio m.in. książkę Bernarda Lecomte'a *Prawda zawsze zwycięży*, zaś pośrednio wydarzenia historyczne, jakie miały miejsce w państwach tzw. Europy Wschodniej po 1978 r., oraz wpływ papieża Jana Pawła II na rozwój tych wydarzeń, zwłaszcza na upadek systemu komunistycznego.

Zagadnienie dualnego charakteru znaku języka informacyjno-wyszukiwawczego w literaturze polskiej wnikliwie przeanalizowała i opisała B.Sosińska-Kalata<sup>9</sup>, wyróżniając:

- dwa typy denotacji, czyli dwa typy obiektów, które desygnują wyrażenia JIW: desygnaty bezpośrednie – dokumenty opisywane przez język informacyjno-wyszukiwawczy oraz desygnaty pośrednie, czyli obiekty

---

<sup>9</sup> B. Sosińska: *Relacje między planem treści i planem wyrażania w językach informacyjno-wyszukiwawczych*. Rozprawa doktorska wykonana pod kierunkiem doc. dr hab. Bożenny Bojar w Katedrze Lingwistyki Formalnej Uniwersytetu Warszawskiego. Warszawa 1986.



rzeczywistości pozadokumentacyjnej opisywane w dokumentach;

- dwa typy konotacji: konotację bezpośrednią, którą stanowią zbiory cech desygnatów bezpośrednich, czyli zbiory własności dokumentów oraz konotację pośrednią – własności obiektów rzeczywistości pozadokumentacyjnej.

Istnieje ścisły związek między tymi czterema sferami interpretacji znaczenia znaku JIW. W dokumentacyjnych systemach informacyjno-wyszukiwawczych, żeby uzyskać informację o obiekcie rzeczywistości pozajęzykowej, trzeba odwołać się do rzeczywistości tekstowej. Z kolei możliwość informowania przez SIW o przedmiotach rzeczywistości pozadokumentacyjnej jest uzależniona od uprzedniego odwzorowania w jego zbiorach (dokumentach) pewnej wiedzy o tych przedmiotach. Do denotacji pośredniej sięga się przy określaniu w systemie leksykalnym JIW relacji mereologicznych, części relacji kojarzeniowych oraz przy konstruowaniu wyrażeń JIW o strukturze semantycznie umotywowanej. O relacjach, w jakie wchodzi wyrażenie JIW, tj. o miejscu wyrażenia w strukturze JIW, decyduje konotacja bezpośrednia i pośrednia, czyli zespół cech semantycznych, wynikających z postulatów znaczeniowych danego JIW, niezbędnych i wystarczających do identyfikacji bezpośrednich i pośrednich desygnatów wyrażeń.

Z powyższych rozważań wynikają dwa kolejne wnioski istotne dla tych, którzy korzystają z języków informacyjno-wyszukiwawczych lub je projektują. Jeśli jedną z płaszczyzn odniesienia dla języka informacyjno-wyszukiwawczego jest rzeczywistość dokumentów, to: (1) explicite powinna być wyrażana informacja o tym, w jakiej rzeczywistości dokumentacyjno-informacyjnej powstał dany JIW; (2) zaadaptowanie języka powinno być poprzedzone upewnieniem się, czy rzeczywistość, w jakiej (lub dla jakiej) powstał język, i ta, z którą mamy do czynienia w bibliotece, mają wystarczająco wiele cech wspólnych, by móc dany język przejąć i efektywnie wykorzystywać. Warto w tym miejscu przypomnieć, że historycznie biorąc, pewne języki informacyjno-wyszukiwawcze powstawały z myślą o konkretnych zbiorach dokumentów i raczej tylko na ich użytek, na przykład języki haseł przedmiotowych, inne zaś – z myślą o określonej dziedzinie, dyscyplinie do wykorzystania w dowolnym SIW, na przykład języki deskryptorowe. Dobry język informacyjno-wyszukiwawczy to nie taki, który w jakimś systemie informacyjno-wyszukiwawczym jest skutecznie stosowany, ale taki, który jest najlepiej jak się da dostosowany do warunków konkretnego SIW („rzeczy komunikowanej”, nadawców, odbiorców, uwarunkowań technicznych, finansowych itd.).

JĘZYK INFORMACYJNO-  
-WYSZUKIWAWCZY  
A UŻYTKOWNICY INFORMACJI

Problematyka użytkowników informacji, ich potrzeb i badań mimo jej fundamentalnego znaczenia dla działalności informacyjnej nie jest odpowiednio reprezentowana w teorii i w piśmiennictwie. W Polsce szczyt zainteresowania użytkownikami informacji przypadł na lata siedemdziesiąte i początek lat osiemdziesiątych. Problematykę tę badano m. in. w kierowanym przez Jerzego Szwalbego Zakładzie Badania Potrzeb Użytkowników i Efektywności Informacji Instytutu Informacji Naukowej, Technicznej i Ekonomicznej. Ukazały się wówczas liczne publikacje poświęcone problematyce użytkowników i ich potrzeb. Ostatnie lata mijają pod znakiem pogłębiającego się rozdźwięku między stopniem rozwoju kultury technicznej, informatycznej a jakością obsługi i poziomem usatysfakcjonowania użytkowników. Problematyka użytkowników i ich potrzeb staje się dziś jednym z ważniejszych i pilniejszych zagadnień teoretycznych i praktycznych. Wiedza o użytkowniku, umiejętność wspólnego z użytkownikiem rozwiązywania jego problemów, nawiązywania z nim kontaktu, wychodzenia mu naprzeciw, to fundamenty, bez których trudno bezpiecznie uprawiać

i perspektywicznie planować jakąkolwiek praktykę biblioteczno-informacyjną. Jedynym wiarygodnym sprawdzianem sensu i właściwego wymiaru wielu podejmowanych prac teoretycznych i działań praktycznych są użytkownicy, ich zaspokojone potrzeby i satysfakcja.

Analogicznie przedstawia się wpływ użytkowników na projektowanie JIW, ostateczny kształt języka, jego miejsce w systemie, związki z narzędziami wyszukiwawczymi. W sferze werbalnej projektowanie JIW jest ściśle związane z użytkownikami, w praktyce bywa różnie. I choć publikacja ta nie jest poświęcona użytkownikom informacji, warto przypomnieć ważniejsze zagadnienia, by stworzyć kontekst dla problematyki projektowania JIW w łączności z użytkownikami i ich oczekiwaniami.

#### **Użytkownik informacji**

W świadomości bibliotekarzy (pracowników informacji) użytkownik informacji to osoba (fizyczna lub prawna) korzystająca z usług systemu informacyjno-wyszukiwawczego, w tym z bibliotek i innych instytucji zajmujących się działalnością informacyjną. Jak jednak doskonale wiadomo, tylko pewna część osób „użytkujących” informację, czyni to za pośrednictwem instytucji. Znacząca (nikt nie wie, jak duża) część użytkowników informacji zaspokaja swoje potrzeby informacyjne kanałami nieformalnymi

poprzez indywidualne i zbiorowe kontakty osobiste. Czasami dzieje się i tak, że użytkownik zrażony, zniechęcony efektami kontaktów z biblioteką czy ich jakością, zrywa je, przechodząc do „szarej strefy” nieużytkowników (ang. *non-users*).

## POTRZEBY INFORMACYJNE

Rozczarowanie co do praktycznych walorów dużej części badań użytkowników informacji wydaje się mieć jedną z przyczyn w nieładzie i nieporozumieniach terminologicznych, jakie panują w tym obszarze. Nie wdając się w rozważanie problemu uniwersalnej definicji i typologii potrzeb, warto zwrócić uwagę na pewne, istotne dla analizy potrzeb informacyjnych, uogólnienia i obserwacje istniejące w nauce o informacji i w innych dyscyplinach naukowych, zwłaszcza w psychologii.

„Jeżeli mówimy, że X ma potrzebę p, znaczy to tyle co, jeżeli p jest niedostępne dla X, to następuje zachwianie homeostazy (równowagi układu), która jest podstawą prawidłowego funkcjonowania X”<sup>10</sup>. Inaczej mówiąc, gdy człowiek (podmiot działania) ma potrzebę, znaczy to, że dla normalnego jego funkcjonowania niezbędne są określone warunki, stwarzane przez

Pojęcie  
potrzeby

<sup>10</sup> K. Obuchowski: *Psychologia dążeń ludzkich*. Warszawa 1972, s. 105.

różnorodne bodźce. Niezaspokojenie potrzeby nie oznacza odzwierciedlenia w świadomości jakiegoś faktycznego braku, czyli nie „pragnienie”, ale obiektywną sytuację, w której człowiek jest pozbawiony czegoś istotnego do funkcjonowania. Sytuacja ta może (ale nie musi) znaleźć odbicie w jego świadomości. Każdej potrzebie daje się przyporządkować pewne bodźce jakościowe i ilościowe, które ją zaspokajają, jak również wymienić bodźce, które ją wywołują lub wzmagają. Istnieją dość znaczne różnice poglądów co do typów i hierarchii potrzeb. Niemniej panuje powszechna zgoda, że potrzeby wyższego rzędu wywodzą się z potrzeb podstawowych, oraz że potrzeby ludzkie nie są czymś stałym, lecz nieustannie się zmieniają.

### **Potrzeby informacyjne**

Wśród potrzeb wyższego rzędu lokują się te, które nas interesują, a które (trafnie lub nie) nazywane są potrzebami informacyjnymi. Analiza literatury przedmiotu prowadzi do wniosku, że informacja jest warunkiem (bezpośrednim lub pośrednim) zaspokojenia wielu potrzeb, które na pierwszy rzut oka nie są postrzegane jako potrzeby informacyjne. A zatem obszar, który należałoby spenetrować, szukając zakresu potrzeb informacyjnych jest dużo rozleglejszy niż ten, z którym potrzeby informacyjne tradycyjnie są kojarzone przez bibliotekarzy i pracowników informacji. Informacja może

być celem samym w sobie, czyli może zaspokajać autonomiczną potrzebę informacyjną. Znacznie jednak częściej informacja jest jedynie warunkiem, środkiem do zaspokojenia jakiejś innej potrzeby, której już nie jesteśmy skłonni nazwać informacyjną. Warto pamiętać, że realizowane od ponad pięćdziesięciu lat badania potrzeb informacyjnych w istocie nie były (lub były niezwykle rzadko) badaniami potrzeb. Częściej dotyczyły zachowań użytkowników w procesie szukania informacji, schematów korzystania z informacji, ocen samych systemów informacyjno-wyszukiwawczych itp.

#### **Informacja a potrzeby**

Większość ludzi nie powiedziałaaby o sobie, że ma potrzeby informacyjne jako takie. Ludzie raczej doświadczają trudności, mają problemy czy są pod jakąś presją, zaś ich potrzeby poznawcze i emocjonalne mogą być zaspokojone (lub częściowo zaspokojone) dzięki uzyskaniu informacji, a następnie jej wykorzystaniu. Potrzeby informacyjne (poza wspomnianymi wcześniej autonomicznymi potrzebami poznawczymi) powstają zatem z chęci zaspokojenia którejś z trzech podstawowych kategorii potrzeb ludzkich. Nie znaczy to, że potrzeby informacyjne są mniej ważne. Przeciwnie, skuteczne zaspokojenie potrzeby podstawowej może być uwarunkowane zaspokojeniem potrzeby informacyjnej, chociaż ta ostatnia nie jest zaliczana do potrzeb

**Sytuacja  
problemowa**

podstawowych. Dla pracowników informacji i dla projektantów JIW wypływa stąd wniosek, że powodzenie podejmowanych działań w części zależy od zrozumienia sytuacji, trudności i problemów, które dają początek potrzebom informacyjnym. Dla projektanta JIW wiedza o tym, jakiej informacji, kiedy, w jakiej postaci itd. potrzebują użytkownicy jest cennym źródłem do wykorzystania przy projektowaniu na przykład kategorii informacji kontekstowej.

Wartość operacyjną dla bibliotekarzy i pracowników informacji ma zatem znajomość sytuacji, w których osoba decyduje się na nawiązanie kontaktu z systemem informacyjno-wyszukiwawczym. Sytuacje te można „rejestrować”, gdy zostają ujawnione przez użytkownika lub próbować przewidzieć, aby nie być przez nie zaskoczonym. Nazywamy je sytuacjami problemowymi. Tadeusz Tomaszewski pisał następująco: „Sytuację, którą człowiek w wyniku wystąpienia procesów poznawczych, uznaje za trudną, czy niepożądaną i którą pragnie zmienić na inną nazywamy sytuacją zadaniową lub zadaniem”<sup>11</sup>. Za szczególną klasę sytuacji zadaniowych uważa Tomaszewski sytuacje problemowe. Wymienia następujące<sup>12</sup> cechy sytuacji

---

<sup>11</sup> T. Tomaszewski: *Ślady i wzorce*. Warszawa 1984 s. 129.

<sup>12</sup> Aby sytuacja mogła być uznana za problemową, nie wszystkie te cechy muszą wystąpić jednocześnie (naraz).



problemowej: brak precyzyjnego określenia stanu pożądanego (wyniku); nieznanostwo srodzkow prowadzacych do wyniku; nieznanostwo warunkow, w ktorych prowadzone beda czynnosci zmierzajace do wyniku. W literaturze psychologicznej wymienia sie jeszcze dwie inne cechy sytuacji problemowej: aktywnostwo osoby rozwiadzujacej problem i podmiotowy charakter sytuacji problemowej. Inaczej mowiac, sytuacja problemowa to taka, gdy podmiot dzialania uznaje swoje wewnetrzne struktury (wiedzy, postaw, osobowosci, otoczenia, problemu itd.) za niewystarczajace do jego (problemu) rozwiadzania, do osiagniecia celu. „Obraz sytuacji stworzony przez uzytkownika, ktory stanal przed problemem do rozwiadzania jest taki, ze wymagane sa zasilenia informacja pochodzaca ze zrodel zewnetrznych, aby osiagnac poziom pewnosci niezbedny do podjecia skutecznego dzialania. Sytuacja problemowa, jak kazda sytuacja, obejmuje zarowno czynniki mniej lub wiecej subiektywne i reakcje zachowan moga byc rozpatrywane tylko w zwiazkach z calym kontekstem, tj. sytuacja taka, jaka istnieje sprawdzalnie oraz jaka wydawala sie istniec zainteresowanej osobie. Tak wiec mozna przyjac, ze sytuacja problemowa (...) jest przyczyzna zachowan informacyjnych, w ktorych wiedza i cele uzytkownika przenikaja sie z efektami i wykorzy-

**Zachowania  
informacyjne**

staniem informacji”<sup>13</sup>. Problemy rodzą się w świadomości człowieka, w wyniku porównywania sytuacji aktualnej z pewnym stanem uznanym za docelowy oraz pojawienia się konieczności usunięcia dostrzeżonej rozbieżności. Ujmując to, co zostało wcześniej powiedziane w formę schematyczną, można stwierdzić, że:

- gdyby zignorować tzw. autonomiczne potrzeby informacyjne, to prawdziwe okazałoby się stwierdzenie, że nie istnieje taka podkategoria potrzeb ludzkich jak potrzeby informacyjne, ale różne potrzeby mogą być zaspokajane przez informację;
- potrzeby mogą dynamizować zachowania, ale reakcje na bodźce otoczenia oraz na bodźce wytwarzane przez samego człowieka mają miejsce tylko wtedy, gdy bodźce te działają na tle odpowiednich potrzeb;
- reakcją człowieka (mającego określone potrzeby) na bodźce jest stan zakłóconej homeostazy (sytuacja problemowa), której odzwierciedleniem w świadomości jest motywacja do działania;
- sformułowanie motywu, spełniającego określone warunki, umożliwia rozpoczęcie działania, poprzez wyznaczenie celu (kierunku działania) i programu. Przyczyną działania może być kilka motywów, nawet

---

<sup>13</sup> J. Woźniak: *O tak zwanych potrzebach informacyjnych*. „Zagadnienia Informatyki Naukowej” 1989 nr 1 s. 51-52.

sprzecznych ze sobą, co pozwala wyjaśnić pewne specyficzne sposoby działania, jego niekonsekwencje lub powstrzymanie się od działania. Natomiast wybór motywu jest warunkowany m.in. przez osobnicze i społeczne doświadczenie jednostki;

- działania, o których mowa wyżej, objawiają się w postaci różnych zachowań<sup>14</sup>, także informacyjnych.

W literaturze przedmiotu wśród ważniejszych czynników mogących wpływać na rozwiązywanie problemów informacyjnych wymienia się najczęściej: środowisko zawodowe, osobowość, typ intelektu<sup>15</sup>, świadomość, płeć, wiek, czas, przeciążenie informacją, bliskość źródeł informacji i jej koszt.

#### WAŻNIEJSZE CECHY POTRZEB INFORMACYJNYCH

Dla projektantów JIW ważne powinny być wypracowane przez praktykę kategorie cech potrzeb informacyjnych. Mając je na uwadze, należałoby przewidzieć w języku odpowiednie kategorie leksykalno-semantyczne, których słownictwo służyłoby do charakteryzowania zarówno pytań użytkowników,

<sup>14</sup> J. Woźniak: op. cit. s. 52-53.

<sup>15</sup> O związkach typu intelektu z rozwiązywaniem problemów informacyjnych pisała m.in. Maria Próchnicka w książce *Informacja a umysł* (Kraków 1991).

jak i dokumentów prymarnych. Warto zwrócić uwagę, że pewne z tych kategorii, jak przedmiot czy zasięg chronologiczny, nie są niczym nowym w kontekście projektowania JIW, inne – na pewno warte są rozważenia i ewentualnego uwzględnienia w języku.

Większość ujawnionych przez użytkowników potrzeb informacyjnych daje się scharakteryzować poprzez odniesienie do następujących jedenastu kategorii:

- Przedmiot** „Przedmiot. Najpowszechniej dostrzegana cecha potrzeb, obejmująca określenie zarówno liczby różnych przedmiotów interesujących użytkownika, jak i szczegółowość, z jaką przedmioty te powinny być „tropione”.
- Funkcja** Funkcja. Ludzie potrzebują informacji do realizacji pięciu funkcji. Są to: (1) dostarczanie odpowiedzi na szczegółowe pytania (funkcja poznawania faktów); (2) aktualność (funkcja bycia poinformowanym na bieżąco); (3) dogłębne rozpoznanie nowych obszarów (funkcja badawcza); (4) ogólne rozumienie zagadnienia/tematu (funkcja orientacyjna); (5) powstawanie nowych pomysłów lub bodźców (funkcja bodźcowa).
- Rodzaj** Rodzaj. Stwierdzenie, jakiego rodzaju informacja jest potrzebna – teoretyczna, historyczna, opisowa, statystyczna, metodologiczna czy inna.

<p>Poziom intelektualny. Zakładany minimalny poziom wiedzy (i czasami inteligencji) użytkownika, niezbędny do zrozumienia informacji.</p>	<b>Poziom</b>
<p>Punkt widzenia. Użytkownikowi może być potrzebna informacja związana z określoną szkołą myślenia, orientacją polityczną, pozytywną lub negatywną opinią publiczną, orientacją dyscyplinarną.</p>	<b>Punkt widzenia</b>
<p>Ilość. Chociaż każdy potrzebuje określonej ilości informacji, żeby wykonywać pracę, to rozmiary „apetytów” informacyjnych są bardzo zróżnicowane; są różne nie tylko u różnych osób i grup, ale zależą także od istoty zapytania.</p>	<b>Ilość</b>
<p>Jakość i wiarygodność. Wprawdzie ocena jakości i wiarygodności informacji jest subiektywna, mimo to jakość charakteryzuje się bardzo wysoką pozycją na liście priorytetów użytkowników informacji.</p>	<b>Jakość i wiarygodność</b>
<p>Zasięg chronologiczny i aktualność. Z tą cechą informacji łączą się dwa wzajemnie powiązane pytania: (1) jak dawno powstała informację należy odzyskać; (2) jak aktualna ma być ta informacja.</p>	<b>Zasięg chronologiczny i aktualność</b>
<p>Szybkość dostarczenia. Cecha ta jest ściśle związana z aktualnością informacji. Informacja nie powinna tracić „świeżości” w procesie przekazywania czy doręczania. Szybkość jest obsesją współczesności.</p>	<b>Szybkość</b>

<b>Miejsce</b>	Miejsce publikacji i/lub powstania. Należy rozważyć, czy miejsce/kraj, z którego informacja pochodzi, ma znaczenie dla użytkownika.
<b>Sposób prezentacji</b>	Przetworzenie i „opakowanie”. Przetwarzanie polega na różnych sposobach przedstawienia tych samych tematów, wyników badań. „Opakowanie” oznacza tu formę prezentacji lub formę fizyczną informacji, w jakiej informacja jest przechowywana i komunikowana <sup>16</sup> .

#### PYTANIA UŻYTKOWNIKÓW

Źródłem cennych informacji o problemach i potrzebach informacyjnych użytkowników są i powinny być pytania kierowane przez nich do systemów informacyjno-wyszukiwawczych. Chodzi tu zarówno o pytania „surowe”, w takim brzmieniu, w jakim zadają je użytkownicy, jak i o pytania „wynegocjowane”, będące rezultatem dialogu między pośrednikiem (bibliotekarzem, pracownikiem informacji) a użytkownikiem.

**Prostota** Dostosowanie języka informacyjno-wyszukiwawczego do użytkownika i jego potrzeb wyraża się nie tylko

---

<sup>16</sup> D. Nicholas: *Analiza potrzeb informacyjnych w dobie Internetu*. „Zagadnienia Informatyki Naukowej” 2000 nr 2 s. 24-25. Rozwinięcie tej problematyki można znaleźć w książce D. Nicholasa *Ocena potrzeb informacyjnych w dobie Internetu: idee, metody, środki*. Warszawa 2001.

w uwzględnieniu przy projektowaniu JIW tego, co wynika ze specyfiki jego potrzeb, ale także w trosce o łatwość tworzenia tekstów danego JIW i wyuczalność ich budowania. Stopień skomplikowania języka informacyjno-wyszukiwawczego ma istotny, choć powszechnie niedoceniany, wpływ na efektywność jego funkcjonowania. Jak wcześniej wspomniano, teoretycznie im bogatszy i bardziej wyrafinowany jest język informacyjno-wyszukiwawczy, tym większe ma on możliwości optymalnego funkcjonowania.

Odnosi się wrażenie, że zarówno w literaturze polskiej, jak i w praktyce bibliotekarskiej postawa przybliżania biblioteki użytkownikowi (a nie użytkownika bibliotece) nie cieszy się popularnością. Niektóre języki informacyjno-wyszukiwawcze prezentują się tak, jakby były projektowane dla satysfakcji tylko bibliotekarzy czy pracowników informacji, z pominięciem użytkowników. Wbrew pozorom nie chodzi mi tu o naiwny psychologizm wyrażający się na przykład w poszukiwaniu klucza do pojęcia przedmiotu tylko i wyłącznie w umyśle użytkownika. Systemy informacyjne muszą mimo wszystko opierać się na założeniu, że użytkownik w jakimś stopniu „nauczy” się systemu i języka. Z drugiej strony JIW powinien być możliwie łatwy do poznania. Deficycyjna sztuczność jego semantyki nie może

**Zrozumiałość  
JIW**

przekraczać granic zdrowego rozsądku. Postulat ten odnosi się także do (na ogół nie „odzyskiwanej” w automatycznym wyszukiwaniu) sztuczności, w tym pozycyjności, gramatyki. Język informacyjno-wyszukiwawczy powinien być integralną częścią SIW zespoloną z jego (systemu) narzędziami wyszukiwawczymi i interfejsem<sup>17</sup> (w przypadku systemów zautomatyzowanych). Projektując język informacyjno-wyszukiwawczy, polepszając ten, który już jest wykorzystywany, nie wolno zapominać, że robi się to dla użytkownika, ku jego pożytkowi i satysfakcji.

Wzgląd na użytkownika w projektowaniu JIW łączy warunek optymalności odwzorowania rzeczy komunikowanej, czyli informacji treściowych i kontekstowych, z warunkami zrozumiałości, dostępności i optymalności postrzeganej formy. Zrozumiałość tekstu JIW zależy od użycia w nim wyłącznie elementów niezbędnych, a zarazem wystarczających dla określonego użytkownika i zadania. Te i inne czynniki stwarzają nowy jakościowo kontekst, w którym przebiegają procesy projektowania i użytkowania języka. Kontekst ten nie może i nie powinien pozostać bez wpływu na oba te procesy, stając się wspólnym pro-

---

<sup>17</sup> Interfejsom poświęcona jest m.in. praca Marii Próchnickiej *Człowiek i komputer: dialogowy model wyszukiwania informacji*. Kraków 2004, Wydaw. Uniwersytetu Jagiellońskiego.



blemem projektujących język i wykorzystujących go w indeksowaniu i wyszukiwaniu dokumentów.

## JĘZYK INFORMACYJNO- -WYSZUKIWAWCZY A ŚRODKI TECHNICZNE

Nie tak rzadko można spotkać się z opinią, że obecne techniki informacyjne umożliwiają tak sprawne wyszukiwanie za pomocą słów w pełnym tekście, że koszty związane z tworzeniem i utrzymywaniem JIW oraz z indeksowaniem dokumentów nie są uzasadnione i opłacalne. Zapomina się jednak, że wyszukiwanie pełnotekstowe może być realizowane tylko w tekstach zapisanych w postaci cyfrowej. Niewiele bibliotek i systemów informacyjnych dysponuje obecnie tego typu zasobami, a nawet jeśli je mają, to nie znaczy, że zawsze pliki te można przeszukiwać pełnotekstowo. Dalszy rozwój dygitalizacji, zarówno tej pierwotnej, oryginalnej, jak i wtórnej, polegającej na zmianie sposobu zapisu już istniejących dokumentów, jest konieczny i ze wszech miar ważny. Jednak nie powinna to być jakakolwiek dygitalizacja, ale taka, która zapewni możliwość pracy, w tym przeszukiwania tekstu cyfrowego. Rzadko też biblioteki dysponują namiastką tekstów w postaci abstraktów lub streszczeń,

**Wyszukiwanie  
pełnotekstowe**

które dostępne są częściej w serwisach wąkospecjalistycznych. Co prawda można wyszukiwać według słów występujących w tytułach dokumentów, ale mankamentem jest duży szum informacyjny, powodowany nieprecyzyjnym odzwierciedlaniem treści w tytułach (często formułowanych w znaczeniu przenośnym), wieloznacznością słów, synonimią itp.

Wielojęzyczność i synonimiczność można częściowo wyeliminować przez wykorzystanie operacji sumowania logicznego odpowiedników słów w innych językach, np. objawy chorób OR symptomes OR symptoms, lub różnych form wyrazowych danego znaczenia, np. objawy chorób OR objawy chorobowe OR oznaki chorobowe OR oznaki chorób OR symptomy chorób OR symptomy chorobowe. Jeśli jednak trzeba uwzględnić i formy obce, i synonimiczne (często wielowyrazowe), utworzenie zapytania może być skomplikowane. Jeśli dodamy do tego problemy związane z fleksją i maskowaniem znaków, sprawa jeszcze bardziej się komplikuje.

**Kontrola  
w JIW**

Języki informacyjne o słownictwie kontrolowanym nie tylko eliminują pewne mankamenty indeksowania swobodnego, ale ponadto oferują dodatkowe narzędzia wspomagające użytkownika przy wyszukiwaniu, takie jak wskazywanie relacji hierarchicznych i kojarzeniowych, precyzowanie zna-

czeń, stosowanie różnego rodzaju odsyłaczy ułatwiających nawigację w systemie itp. Praktyka pokazuje, że w systemach stosujących indeksowanie swobodne, naturalnym procesem rozwojowym jest stopniowe wprowadzanie kontroli słownictwa w celu zmniejszenia szumu i ciszy informacyjnej. Kontrola ta dotyczy początkowo form synonimicznych, następnie eliminowania wieloznaczności i wskazywania pewnych relacji hierarchicznych (najczęściej jest to grupowanie pod jednym terminem szerszym znaczeń mu podporządkowanych, np. pod hasłem Rzeki w Polsce – nazw rzek). W ten sposób prosty wykaz słów kluczowych stopniowo przekształcany jest w słownik przypominający tezaurus o niezbyt rozbudowanym aparacie. Wyszukiwanie swobodne jest dobrym narzędziem wspomagającym JIW o kontrolowanym słownictwie. Pełni także rolę, jaką dawniej pełniły np. indeksy permutacyjne w rodzaju KWIC czy KWOC, tj. pozwala wyszukać opis poprzez słowo stojące na dowolnym miejscu w rekordzie, w tym w haśle wielowyrazowym lub w złożonym wyrażeniu JIW.

Dyskusje na temat wyższości jednych JIW nad pozostałymi mają długą historię i w zasadzie nie dają jednoznacznego wskazania ani najlepszego typu języka, ani konkretnej jego realizacji. Przeprowadzono pewną liczbę ba-

**Badania  
porównawcze  
JIW**

dań porównawczych JIW, z których najbardziej znane były badania Cranfieldskie. Wśród wniosków znalazło się m.in. stwierdzenie, że skuteczność języka zależy nie od tego, który z typów strukturalnych język ten reprezentuje, ale od zakresu i szczegółowości słownictwa, jakim dysponuje. Okazało się, że jeśli do różnych języków wprowadzimy wszystkie, odpowiednio szczegółowe potrzebne terminy, to języki będą działały z porównywalną efektywnością, niezależnie od tego, jaki typ strukturalny reprezentują (klasyfikacja, język haseł przedmiotowych, język deskryptorowy). Większe znaczenie ma stopień skomplikowania JIW i będący jego pochodną poziom trudności w posługiwaniu się językiem. Nie można też nie wziąć pod uwagę, że różnice pomiędzy efektywnością różnych JIW często wyływały z różnic w jakości poddawanych eksperymentowi reprezentantów poszczególnych typów JIW. Ocena typów JIW (wskazanie ich wad i zalet) nie jest tożsama z oceną poszczególnych realizacji tych typów, tj. konkretnych klasyfikacji, słowników tematów i określników czy tezaurusów. Jakość wielu języków pozostawia wiele do życzenia. Jednak wad poszczególnych realizacji nie można utożsamiać z wadami danego typu języka.

**Hybrydyzacja  
JIW**

Obecnie wyraźnie zaznacza się tendencja do łączenia w jednym języku informacyjnym cech i rozwiązań spe-

cyficznych dla różnych typów JIW. W literaturze nazywa się ją hybrydyzacją JIW. Szczególnie wyraźnie hybrydyzację widać w przypadku języków deskryptorowych i języków haseł przedmiotowych. Zbliżenie języków dotyczy zarówno formy terminów wyszukiwawczych, struktury słowników, zestawu odzwierciedlanych w nich relacji, sposobów oznaczania tych relacji, jak i metod wyszukiwawczych. Z jednej strony JHP coraz bardziej odchodzą od częściowo pozycyjnej gramatyki i, co za tym idzie, struktury haseł, częściej korzystają z tematów wielowyrazowych zamiast tematu z określnikiem, np. Prawo biblioteczne zamiast Biblioteki – prawo, z drugiej – języki deskryptorowe rezygnują z wysokiego stopnia postkoordynacji (unitermowości), stosując tzw. terminy prekoordynowane<sup>18</sup>, np. deskryptor etyka małżeńska zamiast rozbicia tego tematu na dwa deskryptory małżeństwo i etyka łączone dopiero przy wyszukiwaniu operatorem AND.

Obecnie funkcjonuje wiele typów języków informacyjnych w różnych wersjach i realizacjach. Każdy z nich związany jest ze specyficznym środowiskiem, narzędziami i technologiami.

---

<sup>18</sup> Hasło (termin) prekoordynowany to takie hasło wieloelementowe, którego części składowe istnieją w języku informacyjno-wyszukiwawczym jako odrębne i samodzielne wyrazy.

Można przypuszczać, że wzrastać będzie znaczenie języków deskryptorowych. „Analiza historyczna rozwoju JIW wykazuje ich coraz większe zbliżanie się do języka naturalnego, przynajmniej w funkcji wyszukiwawczej. Tendencję tworzenia systemów przyjaznych użytkownikowi (a więc łatwych, „naturalnych” i wygodnych) podchwyciły także JHP, przynajmniej w niektórych realizacjach”<sup>19</sup>. Efektem tego jest wyposażanie istniejących języków haseł przedmiotowych stosowanych w OPAC-ach w narzędzia pozwalające czytelnikom przeszukiwać opisy przedmiotowe i poszczególne hasła tak, jakby nie obowiązywały w nich sztuczne reguły budowy haseł przedmiotowych. Przykładem może być FAST (Faceted Application of Subject Terminology)<sup>20</sup>.

Biblioteki często stają przed problemem wyboru JIW do opracowania swych zbiorów. Związane to jest z rezygnacją z dotychczasowego systemu opracowania rzeczowego, który np. okazał się nieefektywny, stwarzający wiele problemów lub, którego twórca przestał go aktualizować i rozwijać. Czasami przyczyną jest komputeryza-

---

<sup>19</sup> E. Chmielewska-Gorczyca: *Przed wyborem języka informacyjno-wyszukiwawczego*. „Przegląd Informacyjno-Dokumentacyjny” 2000 R. 30 nr 1 s. 31.

<sup>20</sup> Na przykład: <http://www.oclc.org/research/projects/fast/default.htm>; [http://www.ifla.org/IV/ifla69/papers/010e-ONeill\\_Mai-Chan.pdf](http://www.ifla.org/IV/ifla69/papers/010e-ONeill_Mai-Chan.pdf) (dostęp 12 XI 2005).

cja biblioteki. Przy automatyzacji naturalnym zjawiskiem jest poddanie rewizji dotychczasowych rozwiązań, metod i technik, w tym wyszukiwawczych. Dodatkowym motywem do zmiany języka może być włączenie biblioteki do sieci, współpraca z innymi bibliotekami i serwisami. Coraz więcej dokumentów oferują biblioteki wirtualne, które z bibliotekami tradycyjnymi łączy przede wszystkim funkcja gromadzenia, porządkowania i udostępniania informacji. Ważnym narzędziem realizacji tej funkcji, szczególnie porządkowania, w dalszym ciągu są JIW. Zmianie uległ jednak sposób i charakter korzystania z języka. Obecnie użytkownik końcowy korzysta z niego zdalnie, nie może więc liczyć na pomoc bibliotekarza, na leżące przy katalogu tablice i instrukcje i inne „podpowiedzi”. Tym bardziej istotne jest tworzenie i stosowanie takich narzędzi, które tej pomocy nie wymagają, narzędzi przyjaznych użytkownikowi. Języki przyjazne muszą posługiwać się terminologią stosowaną przez użytkownika na co dzień, znaną mu z pracy lub wcześniejszych wyszukiwań, muszą także mieć jak najprostsze reguły gramatyczne. Użytkownik powinien odnieść wrażenie, że prowadzi z komputerem rozmowę w jakimś uproszczonym języku naturalnym. Z drugiej strony języki informacyjne powinny do pewnego stopnia „formalizować” język

**Przyjazność  
JIW**

**Wybór  
JIW**

naturalny, eliminując z niego m.in. synonimie i homonimie. Powinny też precyzować zakres stosowanych terminów, wskazywać na ich użycia i możliwości kombinacji z pozostałymi elementami słownika, podpowiadać, jak z nich korzystać, ułatwiać nawigację po rozległych zbiorach informacyjnych itp. Zdaniem Ewy Chmielewskiej-Gorczyca „językiem najbardziej zbliżonym do przedstawionego ideału jest język deskryptorowy”. I tu może paść pytanie: dlaczego w takim razie w dalszym ciągu, nawet w systemach komputerowych, tak popularne są inne języki, a szczególnie język przedmiotowy? Przyczyna tego (...) związana jest często z posiadaniem przez biblioteki katalogów rzeczowych rozpoczętych dawno temu (nieraz sięgających swymi korzeniami XIX w.), dla ogromnych nieraz zbiorów, których melioracja lub przeindeksowanie byłoby przedsięwzięciem nierealnym. Zaszłości te ciążyą nad decyzjami co do JIW i jego ewentualnej zmiany. Biblioteki nowe lub mniejsze, które nie są obciążone balastem tysięcy książek już opracowanych lub takie, które nie są odpowiedzialne za produkowanie opisów katalogowych dla całej sieci bibliotek (co jest zadaniem np. wielu bibliotek narodowych), nie powinny mieć co prawda tych zahamowań.

Z drugiej jednak strony możemy zaobserwować rozpaczliwe wręcz ogląda-



nie się bibliotek, szczególnie mniejszych, o ograniczonym personelu i środkach finansowych, na biblioteki większe (narodowe, uniwersyteckie), kurczowe naśladowanie ich działalności i rozwiązań, szukanie w nich gotowego wzorca dla własnych działań. Wynika to z obaw przed trudnościami związanymi z tworzeniem i utrzymywaniem własnego JIW, przed odpowiedzialnością za jego dobre funkcjonowanie oraz przed osamotnieniem i odizolowaniem biblioteki stosującej własny język.

Dodatkowym powodem może być opinia, wyrażana dość często w ostatnich latach w trakcie rozmów na temat przyszłości JIW, że tezauryusy są sprawą przeszłości, pojawiły się i „wymarły”. Wrażenie to może wynikać z coraz mniejszej, w porównaniu z latami siedemdziesiątymi i osiemdziesiątymi XX w., ilości piśmiennictwa na ten temat, szczególnie literatury podręcznikowej i przeglądowej. W rzeczywistości przyczyna tego zjawiska jest prosta. W tamtym czasie tezauryusy powstawały, rozwijały się, były nowością i dlatego dużo na ich temat pisano. Obecnie w teorii i metodyce tworzenia tezaursów nie ma wielkich zmian, stąd brak nowych publikacji (...). Nieliczne i niewiele znaczące zmiany i innowacje w tej dziedzinie są przedstawiane w artykułach poświęconych pewnemu wąskiemu zagadnie-

## Tezauryusy

niu i rozproszone w obszernej specjalistycznej literaturze z zakresu szeroko pojętej informacji naukowej, informatyki i bibliotekarstwa. Zmiany te nie mają jednak charakteru „rewolucyjnego”, dlatego mogą być niedostrzegane. Nie można natomiast nie dostrzec wzrastającej liczby utworzonych tezaurusów oraz baz danych, w których są stosowane do wyszukiwania. Podobna sytuacja panuje w teorii i praktyce klasyfikacji i JHP. Podkreślić natomiast należy, że zmiany i propozycje w zakresie budowy i formy tezaurusów uwzględniane są w stale aktualizowanych normach (narodowych i międzynarodowych) dotyczących budowy tezaurusów, podczas gdy prace nad normami międzynarodowymi poświęconymi klasyfikacji i JHP jak dotąd nie podjęto.

Tezaurusy  
cd.

Należy przy tym pamiętać, że utworzenie słownika JIW jest zajęciem czasochłonnym, a proces jego doskonalenia nigdy się nie kończy. Tworzenie schematów klasyfikacyjnych trwało najczęściej od kilku do kilkunastu lat, np. w przypadku klasyfikacji Dewey'a, Bliss'a czy Ranganathana – całe życie zawodowe ich twórców. Dlatego nie należy się dziwić, że tezaurusy nie powstają z dnia na dzień i że ich edycja, szczególnie pierwsze wydanie, dalekie jest często od doskonałości.

Przez długi czas współistnienia klasyfikacji i haseł przedmiotowych traktowano je jako zupełnie różne narzędzia. Świadczy o tym m.in. zróżnicowanie terminologii stosowanej w obu realizacjach, np. proces rzeczowego opracowania książek w pierwszej nazywano klasyfikowaniem, w drugiej – przedmiotowaniem. Nie istniał też termin generyczny obejmujący swym zakresem oba te języki. Dopiero w latach sześćdziesiątych XX w., wraz z wkroczeniem języków deskryptorowych, zaczęto dostrzegać pewne analogie i podobieństwa między tymi narzędziami.

Wtedy też wprowadzono pojęcie języka informacyjno-wyszukiwawczego obejmującego wszystkie typy języków służących przedstawianiu zawartości treściowej dokumentów w celu ich późniejszego odszukiwania. Nastąpiła też pewnego rodzaju unifikacja terminologii, np. termin „indeksowanie”, oznaczający początkowo proces opracowania rzeczowego w języku słów kluczowych lub deskryptorowym, stopniowo zaczął obejmować także proces klasyfikowania i przedmiotowania (jako nadrzędny, generyczny). Kolejny przełom w terminologii tej dziedziny zaobserwować można wraz z rozwojem i upowszechnieniem dokumentów elektronicznych i sieci komputerowych. Ogrom i różnorodność informacji dostępnej w Internecie oraz trudno-

Nawigacja

ści z jej rzeczowym uporządkowaniem i wyszukiwaniem stawiają nowe wymagania przed twórcami i użytkownikami narzędzi organizacyjno-wyszukiwawczych, zwanych często nawigacyjnymi.

Dodatkowym utrudnieniem jest forma prezentacji informacji elektronicznej, np. wprowadzenie nieliniarnej struktury danych (w postaci hipertekstów), łączenie informacji tekstowej z dźwiękową i obrazową (multimedia), połączenie oby tych technik (hipermedia), ale największym – niewymierność i nieuchwytność struktury informacji, znajdującej się w niewidocznych dla oka zbiorach elektronicznych (poza niewielkim ich fragmentem prezentowanym na ekranie), ograniczająca możliwości percepcji (odbioru) i prowadząca często do zagubienia się użytkownika w „przestrzeni cybernetycznej”. Zagadnienia strukturyzacji informacji elektronicznej okazały się na tyle istotne, że powstały nowe specjalizacje i gałęzie zajmujące się tymi problemami. Na stałe wkroczyły do teorii i praktyki działalności informacyjnej takie pojęcia, jak: inżynieria wiedzy, organizacja wiedzy, reprezentacja wiedzy, itp., zastępując niejako takie pojęcia jak: klasyfikacja piśmiennictwa, porządkowanie dokumentów, układ zbioru czy indeksowanie (...).

Mnogość i zróżnicowanie istniejących systemów informacyjnych, coraz

większa swoboda dostępu do nich, coraz większa potrzeba szybkiej i bardzo różnej informacji, coraz większe tempo życia, nie skłaniają użytkowników do korzystania z systemów, które wymagają zapoznania się (czytaj: nauczenia się i zapamiętania) ze strukturą stosowanych w nich JIW, z ich sztuczną symboliką i gramatyką, z regułami indeksowania i specyficznymi sposobami wyszukiwania. Rozwojem i rozpowszechnieniem JIW rządzą takie same prawa rynkowe, jak w przypadku innych towarów.

Zbytne skomplikowanie struktury i zasad korzystania prowadzić może do wyparcia tego „produktu” z rynku, czego dowodem może być Klasyfikacja Dwukropkowa czy kody semantyczne w rodzaju WRU lub do zmniejszenia jego roli, czego przykładem może być UKD. Popularność zyskują produkty nie tylko przydatne, lecz także wygodne w użyciu i proste w obsłudze (niewymagające specjalnych szkoleń lub czytania skomplikowanych instrukcji), nawet jeśli prostota ta wprowadzona jest częściowo kosztem jakości efektu końcowego. Zależność ta w większym stopniu jest dostrzegana i respektowana w systemach komercyjnych, bardziej zorientowanych na dostosowanie się do potrzeb rynku i zdobycie stałych klientów. O bibliotekach trudno powiedzieć, że „walczą o klienta”. Ich egzystencja nie zależy

bezpośrednio od liczby i częstotliwości zaspokojonych potrzeb i świadczonych usług, które do dzisiaj są w zasadzie bezpłatne.

#### Przyjazność

Stąd być może spotykana nieraz postawa bibliotekarzy: „jeśli chce korzystać z systemu (katalogu), to niech się go nauczy”. Pamiętać jednak należy, że Internet umożliwia dostęp do katalogów innych niż lokalna bibliotek oraz do innych niż katalogi źródeł informacji (komercyjnych i bezpłatnych). Dostępność wielu produktów oferujących podobne usługi i służących zaspokojeniu podobnych potrzeb informacyjnych uruchomiła (lub uruchomi w najbliższej przyszłości) mechanizmy znane z rynku innych towarów i usług: mając do wyboru wiele baz danych, użytkownik korzysta najczęściej z tej, która daje najlepsze wyniki i w której poszukiwanie jest najmniej „bolesne”. Na rynku biblioteczno-informacyjnym zaczęły działać zwykłe prawa konkurencji. Aby utrzymać się na tym rynku, także biblioteki muszą uatrakcyjnić swoje produkty i usługi, w tym zadbać o poprawę jakości i prostoty stosowanych narzędzi wyszukiwawczych, jakimi są m.in. JIW. Biblioteki, które praw tych nie zrozumieją lub nie będą chciały ich zaakceptować stracą rację bytu, szczególnie w sytuacji wolnego rynku i dofinansowywania instytucji kultural-

no-oświatowych przez zewnętrznych sponsorów”<sup>21</sup>.

Obserwując praktykę, można dojść do wniosku, że obecnie uwarunkowania techniczne systemu informacyjnego mają (choć też nie zawsze) wpływ na formy prezentacji języków informacyjno-wyszukiwawczych, sposoby ich realizacji, techniki wyszukiwania, nie oddziałują jednak znacząco na struktury leksykalno-gramatyczne JIW. Wydaje się, że ani projektanci JIW, ani projektanci SIW nie wykorzystują możliwości i szans stwarzanych przez technikę. Zbyt często JIW stosowane w zautomatyzowanych systemach informacyjno-wyszukiwawczych przypominają prymitywne elektroniczne kopie ich „papierowych” odpowiedników. Oczywiście konsekwencji stosowania niedobrego języka lub złego używania dobrego JIW nie zrekompensuje ani najdroższy sprzęt, ani najnowocześniejsze oprogramowanie, z drugiej strony ignorowanie środowiska technicznego, w którym język będzie stosowany i wykorzystywany, jest marnotrawstwem i błędem. Projektując nowy język informacyjno-wyszukiwawczy, należy wykorzystać możliwości, jakie stwarzają nowoczesne techniki informacyjne.

Dostępność i sprawność słownika JIW zależy od oprogramowania stosowanego w danym systemie i „przyjazności” samego słownika. Jeśli konsul-

**Dostępność  
słownika**

---

<sup>21</sup> E. Chmielewska-Gorczyca: op. cit. s. 37-41.

tacja ze słownikiem wymaga „wyjścia” z jednej bazy, otwarcia nowej sesji w innej i za chwilę powrotu do poszukiwań w pierwszej, to mało prawdopodobne wydaje się, żeby użytkownicy powszechnie korzystali z tej opcji. Wiele słowników JIW obecnie wykorzystywanych w zautomatyzowanych systemach informacyjno-wyszukiwawczych powiela strukturę i organizację słowników tradycyjnych, nie są więc w pełni dostosowane do możliwości i ograniczeń systemów komputerowych – z jednej strony nie wykorzystują w pełni zwiększonych możliwości wyszukiwawczych dostępnych dzięki systemom online, z drugiej zaś nie eliminują niedogodności wpływających z wad i ograniczeń prezentacji na ekranie (np. mniejsza czytelność i pojemność strony ekranowej w porównaniu z drukowaną, często prezentacja informacji w postaci formatowej, a nie słownikowej).

**Potrzeba  
standaryzacji  
i ujednolicenia**

Mając na myśli zautomatyzowane katalogi biblioteczne, dziś niemal powszechnie dostępne w postaci WWW, w których coraz częściej stosuje się powiązania (w formie tzw. linków) z innymi zbiorami informacyjno-wyszukiwawczymi, nie można zapominać o jeszcze jednym, pragmatycznym aspekcie nie tyle projektowania, co wykorzystywania JIW. Mając mnóstwo zalet, wirtualna agregacja różnych zbiorów może też być przyczyną fru-



stracji użytkowników, gdy każdy z tych zbiorów jest scharakteryzowany (zaindeksowany) za pomocą innego JIW. W takiej sytuacji użytkownicy często rezygnują z posługiwania się jakimkolwiek językiem informacyjno-wyszukiwawczym, decydując się na stosowanie słów kluczowych i algebry Boole'a. Podobnie dzieje się, gdy struktury semantyczne, leksykalne i gramatyczne JIW są dla użytkownika zbyt sztuczne, niezrozumiałe, nieprzyjazne. Wówczas reakcją użytkownika jest ucieczka od JIW, a często i od samego systemu. A do tego chyba nikt z pracowników informacji nie chciałby się przyczynić.



## Część druga

# WAŻNIEJSZE WSKAZANIA DOTYCZĄCE BUDOWY JĘZYKA DESKRYPTOROWEGO, A ZWŁASZCZA TEZAURUSA

Słownictwo tezauryusa stanowią deskryptory, askryptory i (opcjonalnie) modyfikatory. Deskryptory (terminy preferowane) to wyrazy języka deskryptorowego, służące do charakteryzowania dokumentów i pytań, mające na ogół budowę identyczną lub zbliżoną do wyrazów (wyrażeń) języka naturalnego (są równokształtne z wyrazami języka naturalnego), choć często nieco inne znaczenie. Askryptory (nondeskryptory) to terminy zakazane, nie używane w indeksowaniu, ale przydatne do sprawnej realizacji wyszukiwania; ich znaczenia reprezentują w języku deskryptorowym odpowiednie deskryptory. Modyfikatory (deskryptory aspektowe) to wyrazy pomocnicze, niesamodzielne składnikowo, najczęściej o ogólnym, szerokim znaczeniu, stosowane tylko i wyłącznie razem z deskryptorami (jak określniki w językach haseł przedmiotowych) w celu modyfikacji ich znaczeń.

Deskryptor

Askryptor

Modyfikator

<b>Wprowadzenie do tezaurusa</b>	Tezaurus powinien się składać z: wprowadzenia, części głównej oraz części i/lub indeksów pomocniczych. We wprowadzeniu do tezaurusa należy umieścić informacje o: przeznaczeniu i zakresie stosowania tezaurusa; zakresie tematycznym ze wskazaniem centralnych i peryferyjnych pól tematycznych; znaczeniu znaków i skrótów stosowanych w tezaurucie; ogólnej liczbie deskryptorów i askryptorów oraz innych wyrażen (modyfikatorów, identyfikatorów) <sup>1</sup> ; zasadach doboru deskryptorów i ustalania relacji między nimi; układzie i zasadach porządkowania deskryptorów i askryptorów w poszczególnych częściach tezaurusa; regułach aktualizacji tezaurusa; dacie wprowadzenia ostatniego deskryptora; szczegółowych zasadach indeksowania, jeśli nie przewiduje się odrębnej instrukcji indeksowania.
<b>Część główna</b>	Za część główną tezaurusa zwyczajowo uznaje się część alfabetyczno-hierarchiczną. Zaleca się, a właściwie zobowiązuje projektanta do uzupełnienia jej częścią systematyczną, graficzną lub indeksem hierarchicznym.
<b>Część pomocnicza</b>	W części pomocniczej tezaurusa dopuszcza się uwzględnienie tylko niektórych relacji między wyrażeniami tezaurusa lub wybranych rodzajów wyrażen, np.: wyłącznie deskryptorów

---

<sup>1</sup> Identyfikatory – deskryptory będące nazwami osób, ciał zbiorowych, miejsc, tytułami ujednoliconymi i formalnymi.

w części systematycznej lub indeksie hierarchicznym, wybranych deskryptorów w części graficznej, wyłącznie modyfikatorów, czy relacji ekwiwalencji w indeksie alfabetycznym lub wyłącznie relacji hierarchicznych w części systematycznej lub indeksie hierarchicznym. Części pomocnicze należy wybierać tak, aby zapewnić dodatkowe możliwości wyszukiwania w teaurusie, niedostępne za pomocą części głównej.

W części alfabetyczno-hierarchicznej należy zamieścić w porządku alfabetycznym artykuły deskryptorowe i artykuły askryptorowe. Porządek części alfabetyczno-hierarchicznej wyznacza abecadło deskryptorów tytułowych i askryptorów.

- Artykuł deskryptorowy składa się z:
- deskryptora tytułowego [z ew. objaśnieniem],
  - definicji poprzedzonej wskaźnikiem, np. DEF,
  - askryptorów poprzedzonych wskaźnikiem, np. NU,
  - (opcjonalnie) deskryptora głównego poprzedzonego wskaźnikiem, np. TG,
  - deskryptora lub deskryptorów szerszych poprzedzonych wskaźnikiem, np. TS,
  - deskryptora lub deskryptorów kojarzeniowych poprzedzonych wskaźnikiem, np. TK,
  - deskryptora lub deskryptorów węższych poprzedzonych wskaźnikiem, np. TW.

**Artykuł  
deskryptorowy**

Na przykład,  
Kapłaństwo  
[stosuje się tylko do jedyne-  
go kapłaństwa Jezusa Chrystusa; w pozosta-  
łych przypadkach używać hasła – Sa-  
krament kapłaństwa]

NU kapłaństwo Jezusa Chrystusa

TK Jezus Chrystus

Kapłan

Artykuł  
askryptorowy

W skład artykułu askryptorowego  
wchodzi: poprzedzony wskaźnikiem  
lub wyróżniony w jakiś sposób graficz-  
nie askryptor, wskaźnik U, deskryp-  
tor (lub kombinacja deskryptorów)  
reprezentujący znaczenie danego  
askryptora w projektowanym tezauru-  
sie, np.

kapłaństwo Jezusa Chrystusa U Ka-  
płaństwo.

### CZYM POWINIEN CHARAKTERYZOWAĆ SIĘ DOBRY JĘZYK INFORMACYJNO- -WYSZUKIWAWCZY, W TYM JĘZYK DESKRYPTOROWY?

- Najważniejsze cechy dobrego JIW to:
- odpowiednio szeroki zakres tematyczny,
  - dostateczna szczegółowość wyrażenia,
  - poprawna i odpowiednia siatka relacji,
  - jednoznaczność, czyli brak w języku wyrażenia wieloznacznych (homonimicznych),

- niesynonimiczność, czyli brak w języku synonimów,
- w miarę precyzyjne reguły stosowania.

#### ODPOWIEDNIO SZEROKI ZAKRES TEMATYCZNY

Zakres tematyczny (zwany też zakresem pola semantycznego lub zakresem systemu leksykalnego JIW) projektowanego JIW można określić albo poprzez zakres dokumentów, do opisywania i wyszukiwania których język będzie wykorzystywany, albo poprzez odwołanie się do zakresu tematycznego systemu informacyjno-wyszukiwawczego, w którym język będzie stosowany. Można się spodziewać, że w pierwszym przypadku projektowany język będzie bliższy rzeczywistości, w której ma funkcjonować. Projektanci nie będą marnować czasu, energii i środków na budowę tych fragmentów JIW, które nie mają swojego odbicia w dokumentach. Znajomość dokumentów pozwoli określić nie tylko zakres projektowanego JIW, ale także inne jego własności wynikające z cech materialnych dokumentów (jako obiektów fizycznych), liczebności zbioru dokumentów, tempa przyrostu i starzenia się informacji. Z drugiej strony nie można zapominać, że w bibliotece z różnych powodów mogą

się znajdować dokumenty przypadkowe lub że zbiór może w danym momencie nie być reprezentatywny, może się wreszcie przekształcać, zmieniać ze względu na zmiany w samym systemie, jak i jego otoczeniu. Można zatem powiedzieć, że na zakres języka informacyjno-wyszukiwawczego największy wpływ mają dokumenty widziane zarówno jednostkowo, jak i jako pewien zbiór (piśmiennictwo, literatura) mający cechy wspólne mogące być przedmiotem opisu i podstawą poszukiwań.

Zbiór dokumentów, do opisywania których projektujemy język, jest dynamiczny. Nie wszystkie zmiany dadzą się przewidzieć. Bywa też i tak, że język jest projektowany zanim zostanie zgromadzona cała kolekcja dokumentów – dokumenty na pewne tematy mogą się pojawić dopiero po zakończeniu projektowania języka. Dlatego zawsze dobrze jest potraktować dwa wymienione wyżej sposoby określania zakresu tematycznego JIW jako komplementarne. Wyjść na przykład od określenie zakresu dokumentów, a następnie porównać go z zakresem „wzorcowym” wynikającym z profilu biblioteki, jej specjalizacji, polityki gromadzenia zbiorów, czyli tego, co zostało wcześniej nazwane zakresem systemu informacyjno-wyszukiwawczego. Należy przy tym pamiętać, że określając zakres tematyczny JIW, należałoby wskazać obszary uznane za



najważniejsze, priorytetowe, stanowiące jądro zainteresowań biblioteki i czytelników (słownictwo z tych obszarów powinno być najliczniej i najdokładniej reprezentowane w JIW, najwnikliwiej opracowane), obszary, które mają charakter pochodny wobec pierwszego oraz obszary stanowiące tło, margines, obrzeża (można je jedynie „zasygnalizować” w teaurusie).

Warto w tym miejscu wspomnieć, że każdy język informacyjno-wyszukiawczy, każdy teaurus wymaga okresowej aktualizacji. Zmienia się profil biblioteki, terminologia, oczekiwania użytkowników, środki techniczne itd. – wszystkie te zmiany muszą być wnikliwie analizowane i delikatnie wprowadzane do JIW, mając na uwadze, że z jednej strony język musi być żywy, sprawny, elastyczny, z drugiej zaś, że wprowadzane zmiany (uaktualnienia) nie mogą być rewolucjami niszczącymi stan dotychczasowy. Mechanizmy i procedury aktualizacji muszą być w miarę precyzyjnie określone przed oddaniem języka do użytkowania.

Wprowadzenie (lub usunięcie) nowego wyrażenia do języka powoduje albo zmianę zakresu pola leksykalnego danego JIW, albo zmianę jego segmentacji (zmianę znaczeń wyrażen dotychczas istniejących w języku), gdy dana jednostka wchodzi w związki semantyczne. Możliwa jest też taka

**Konieczność aktualizacji**

**Konsekwencje wprowadzania nowych deskryptorów**

zmiana, która jest kombinacją obu tych przypadków. Na przykład, jeśli w jakimś języku informacyjno-wyszukiwawczym termin Miłosierdzie reprezentował znaczenia wyrażen języka naturalnego miłosierdzie Boże i miłosierdzie ludzkie, to po dodaniu do słownika tego języka terminu Miłosierdzie Boże całkowity zakres znaczeń możliwych do wyrażenia w tym JIW nie ulegnie zmianie. Ulegnie natomiast zmianie znaczenie wyrazu Miłosierdzie i wzrośnie wartość selekcyjna jednostek leksykalnych tego JIW. Zmieniają się także kryteria dokładności wyszukiwawczej, które dodatkowo będą zależeć od tego, czy po zmianie nastąpi także przeindeksowanie dokumentów według nowych zasad znaczeniowych języka, czy też nie.

#### DOSTATECZNA SZCZEGÓŁOWOŚĆ WYRAŻEŃ

Liczebność zbioru i tempo jego przyrostu oraz selekcji<sup>2</sup> mają znaczący wpływ na szczegółowość, zwaną też wartością selekcyjną, wyrażen JIW. Brak w języku wyrażen o odpowiedniej szczegółowości uniemożliwia zain-

---

<sup>2</sup> Selekcja w języku bibliotekarskim oznacza usuwanie dokumentów ze zbioru.

deksowanie (opisanie) dokumentów i pytań w sposób na tyle szczegółowy, na ile byłoby to pożądane. Konieczność używania wyrażenia o optymalnej w danej sytuacji wartości selekcyjnej jest uzasadniona z punktu widzenia wyszukiwania. Im więcej mamy, czy przewidujemy, że będziemy mieć, dokumentów z określonego obszaru tematycznego, tym bardziej szczegółowe w zasadzie powinno być słownictwo z tego obszaru. Chodzi o to, żeby użytkownik zadając pytanie, otrzymał w odpowiedzi opisy „rozsądnej” liczby dokumentów.

Nie jest możliwe podanie wartości bezwzględnych co do wielkości tezaury i szczegółowości zawartych w nim wyrażenia. W każdym przypadku ustala się to indywidualnie na podstawie znajomości dokumentów, przyjętych zasad indeksowania, potrzeb biblioteki, przewidywanych oczekiwań użytkowników, posiadanego sprzętu i oprogramowania itd. Warto jednak pamiętać, że dysponując tezaurem i zawartą w nim siatką relacji, dokumenty scharakteryzowane szczegółowymi deskryptorami można odszukać również w odpowiedzi na ogólniej sformułowane pytania, ale nie odwrotnie. Nie jest możliwe odszukanie dokumentu, który został zaindeksowany uogólniająco (ogólnie) w odpowiedzi na pytanie o charakterze wyszczególniającym (szczegółowym). Katalogi komputerowo-

Poziom  
szczegółowości

we są szansą uwolnienia użytkownika od konsekwencji nadmiernie uogólnionego indeksowania. Będzie to jednak możliwe wówczas, gdy język informacyjny będzie dysponował odpowiednio szczegółowym słownictwem.

#### **Pytania użytkowników**

Pewną wskazówką w ustalaniu podstawowego poziomu szczegółowości deskryptorów mogą służyć pytania formułowane przez użytkowników. Słownictwo tworzonego tezaurusa powinno być co najmniej tak szczegółowe, jak szczegółowa jest tematyka pytań użytkowników. Należy jednak mieć świadomość, że użytkownicy nie będący ekspertami w dziedzinie (a z takimi najczęściej mamy do czynienia) na ogół formułują pytania, posługując się pojęciami ogólniejszymi od faktycznego przedmiotu ich zainteresowania. Na przykład, szukając mapy jeziora Mamry, mogą pytać o mapę Mazur; będąc zainteresowani świętem Trzech Króli, mogą pytać o święta kościelne itd.

#### **Sztuka projektowania JIW**

Zarówno projektowanie, jak i użytkowanie JIW, jest nie tylko wiedzą, ale i swego rodzaju sztuką. Nie wszystko da się tu precyzyjnie określić, wymierzyć itp. Nie można podać „twardych” reguł, które wyrażenie należy wprowadzić do tezaurusa, a które odrzucić jako zbyt ogólne lub dla odmiany zbyt szczegółowe (czy z jakiegoś innego powodu). Nie zwalnia to jednak projektanta JIW z obowiązku

racjonalizowania decyzji i szukania złoto-  
tego środka między uogólnianiem i wy-  
szczególnianiem.

Zagadnienie wartości selekcyjnej  
wyrażeń JIW wiąże się ściśle z ustala-  
niem relacji odpowiedniości (ekwiwa-  
lencji) wyszukiwawczej, której człon  
reprezentują elementy uznane w danym  
systemie za równoważne z punktu  
widzenia przyjętej szczegółowości  
opisu dokumentów i wyszukiwania.  
W tradycyjnych katalogach rzeczowych  
relacja ta miała postać odsyłacza  
całkowitego (zob.).

Relacja odpowiedniości wyszuki-  
wawczej nie musi i często wcale nie  
łączy synonimów. Na przykład, wyra-  
żenie *Metabolizm* w języku haseł  
przedmiotowych (JHP) KABA repre-  
zentuje znaczenia wyrażane w języku  
polskim przez trzy wyrazy: *metabo-  
lizm*, *anabolizm*, *katabolizm*; w *Słowni-  
ku słów kluczowych*<sup>3</sup> z teologii de-  
skryptor *Potępienie* reprezentuje  
znaczenia dwóch wyrazów języka pol-  
skiego *piekło* i *potępienie*. Należy  
pamiętać, że jeśli w jakimś języku in-  
formacyjno-wyszukiwawczym zostanie  
ustanowiona relacja odpowiedniości na  
przykład pomiędzy terminami *mosty*  
i *wiadukty* (*wiadukty U Mosty*), to  
użytkownik szukający dokumentów  
o wiaduktach, będzie musiał wybrać  
z tych, którym przydzielono hasło *Mo-*

Relacja  
odpowiedniości  
wyszukiwawczej

---

<sup>3</sup> *Słownik słów kluczowych...*, op. cit. s. 21.

sty, oznaczające w danym JIW zarówno mosty jak i wiadukty, dokumenty dotyczące wiaduktów. W przypadku tych dwóch terminów jest wysoce prawdopodobne, że będzie mógł to uczynić na podstawie tytułów. Trudniej będzie w przypadku kwerend i dokumentów z tzw. dziedzin „miękkich” (społecznych i humanistycznych), w których tytuły najczęściej nie wystarczają do takich rozstrzygnięć.

Relacja  
odpowied-  
niości, cd.

Relacja odpowiedniości wyszukiwawczej przybiera w tezaurusach postać artykułu askryptorowego wiążącego askryptor z odpowiadającym mu deskryptorem, np. sakrament pojednania U sakrament pokuty i pojednania oraz odpowiedniego zapisu w artykule deskryptorowym, np.

sakrament pokuty i pojednania  
NU sakrament pojednania

Najczęściej relacją ekwiwalencji łączy się wyrażenia uznawane w danym języku za synonimy lub quasisynonimy, np.: tłumaczenie U Przekład; ociemniali U Niewidomi; kolor U Barwa; powołanie nadprzyrodzone U Powołanie Boże.

Wybierając deskryptor spośród grupy słów kluczowych, najczęściej uwzględnia się następujące kryteria w kolejności:

- rozpowszechnienie w tekstach pisanych z danej dziedziny (zakresu);
- poprawność,

- bieżące stosowanie,
- zwięzłość.

W szczególności relacją odpowiedniości łączy się:

- wyrażenia o tym samym znaczeniu, ale o różnej etymologii, np. lingwistyka – językoznawstwo; obmycie – ablucja.

Status deskryptora na ogół uzyskują wyrażenia o polskim źródłosłowie, chyba że: wyrażenie pochodzenia obcego jest lepiej, powszechniej znane, „poręczniejsze” itp., jak mamy do czynienia w przypadku na przykład irysów (pol. kosańce) czy hemofilii (pol. krwawiączka) lub gdy nie ma polskiego odpowiednika, np.: public relations, monitoring;

- wyrażenia należące do różnych podsystemów leksykalnych, np. do języka ogólnego czy potocznego i specjalistycznego czy naukowego, np.: sól (kuchenna) – chlorek sodu, łakomstwo – nieumiarkowanie w jedzeniu i picciu.

To, które wyrażenie uzyska status deskryptora, a które askryptora, zależy od tego, dla kogo budujemy język (dla użytkowników biblioteki naukowej, uczelnianej, publicznej typu naukowego, publicznej typu wypożyczalnia itd.);

- nazwy gatunkowe z nazwami firmowymi (także różne nazwy firmowe, handlowe, rynkowe między sobą), np. kwas acetylosalicylowy – aspiryna.

Najczęściej jako deskryptor przyjmuje się nazwę gatunkową, nazwom firmowym, handlowym itd., nadając status askryptorów;

- wyrażenia należące do słownictwa współczesnego i archaicznego, np. notariusz – rejent.

Preferencja jest dla słownictwa współczesnego:

- wyrażenia epinimiczne i odpowiadające im nazwy gatunkowe, np. choroba Parkinsona – drżączka porażna;
- wyrażenia o różnej pisowni, np.: system online – system online; Księga Ruth – Księga Rut; digitalizacja – dygitalizacja;
- wyrażenia różniące się liczbą gramatyczną, np.: człowiek – ludzie, asceci – asceta.

W językach deskryptorowych preferencję ma liczba pojedyncza z wyjątkiem: rzeczowników typu pluraria tantum (spodnie, drzwi, nożyczki); nazw rodów, rodzin, dynastii; nazw grup związków chemicznych i minerałów; nazw kategorii osób (dzieci, liberałowie, rolnicy, głusi); nazw wyższych grup zwierząt i roślin (tak, jak w systematyce biologicznej); wyrażen zwyczajowo, powszechnie stosowanych w liczbie mnogiej (czynności dowodowe, wody przybrzeżne);

- wyrażenia różniące się rodzajem gramatycznym, np. pisarki – pisarze.

Nie ma obowiązku sprowadzania jednej formy gramatycznej do drugiej,



ale często się to robi. Na ogół z punktu widzenia wielu dokumentów nie ma znaczenia, czy chodzi o żeńskich czy męskich przedstawicieli danej grupy. Jeśli jednak dojdziemy do wniosku, że będziemy potrzebować w języku na przykład zarówno terminu zakonnica, jak i terminu zakonnik, to obu tym wyrazom możemy nadać status deskryptorów;

- wyrażenia o różnym szyku elementów składowych, np. Wspólnota Anglikańska – Anglikańska Wspólnota.

Preferowany powinien być naturalny szyk wyrażen. W języku polskim w przypadku zdecydowanej większości wyrażen pospolitych, tj. nie nazw własnych, naturalny jest szyk porządkownikowy, np.: wiara katolicka, modlitwa poranna, historia powszechna;

- wyrażenia o różnej formie gramatycznej, np. duszpasterzowanie – duszpasterstwo;
- skróty i skrótowce z rozwinięciami, np. NMP – Najświętsza Maryja Panna; ISKO – International Society for Knowledge Organization;
- wyrażenia wieloelementowe i odpowiadające im ew. połączenia elementów składowych, np. normy etyki zawodowej U etyka zawodowa + norma;
- wyrażenia uznane za zbyt szczegółowe lub (przeciwnie) zbyt ogólne dla potrzeb indeksowania i wyszukiwa-

nia informacji w danym systemie, np. bazalt – skały, teologia moralna szczegółowa – teologia moralna.

#### Antonimy

Istotnym problemem przy ustalaniu relacji ekwiwalencji są antonimy. W teorii JIW rozróżnia się trzy podstawowe typy antonimii: komplementarną, konwersyjną i skalarną.

Antonimię komplementarną, czyli taką, w przypadku której jeden człon jest tożsamy znaczeniowo zaprzeczeniu drugiego, reprezentują np. następujące pary „noc – dzień”, „zdrowy – chory”, „szorstki – gładki”. O ile w przypadku antonimii konwersyjnej i skalarnej użycie jednego z antonimów powoduje automatyczne zanegowanie wszystkich pozostałych z danej klasy, to w przypadku antonimii komplementarnej użycie jednego antonimu implikuje zaprzeczenie tylko jego dopełnienia, np. „dzień” implikuje „nie noc”, „chory” implikuje „nie zdrowy” itd.

Antonimię konwersyjną reprezentują np. pary „kupić – sprzedać”, „mąż – żona”, czyli takie terminy, które są opozycjami związanymi, tzn. jeśli A jest mężem B, to B jest żoną A; jeśli C kupuje od D, to D sprzedaje C.

Przykładem antonimii skalarnej są z kolei pary: „dużo – mało”, „ciężko – lekko”, „suchy – mokry”. Antonimy tego typu pełnią funkcję porównawczą – jeżeli mówimy, że coś jest duże, to po prostu porównujemy z normalnym,

przeciętnym, typowym rozmiarem obiektów danej klasy.

Wśród teoretyków JIW panuje przekonanie, że antonimy komplementarne i konwersyjne powinny być reprezentowane w teaurusie (także w każdym innym słowniku JIW) przez dwa różne deskryptory połączone relacją skojarzeniową, natomiast antonimy skalarne – przez jeden deskryptor (drugi powinien być akryptorem). Uzasadnieniem jest argumentacja, że są to terminy przedstawiające dwa przeciwstawne punkty widzenia jednego ciągłego układu własności i na ogół użytkownicy poszukujący literatury na temat jednego z nich są zainteresowani także i drugim.

#### POPRAWNA I ODPOWIEDNIA SIATKA RELACJI PARADYGMATYCZNYCH

Relacje paradygmatyczne są to związki semantyczne (znaczeniowe) explicite zaznaczane w słowniku JIW, łączące wyrazy tego języka i odzwierciedlające stałe zależności między elementami rzeczywistości językowej i/lub pozajęzykowej.

Najczęściej w słownikach JIW wyróżnia się dwa podstawowe rodzaje relacji paradygmatycznych:

- hierarchię, obejmującą poza zwykłą hipero- i hiponią, czyli związkiem między rodzajem i gatunkiem (i od-

wrotnie), także związek mereologiczny, czyli pomiędzy całością i częścią (i odwrotnie), na przykład: Anabaptyści TS Protestanci; Diakon TS Duchowny; Ewangelia według św. Marka TS Nowy Testament; Jerozolima TS Palestyna;

- asocjacje (kojarzenia, skojarzenia), obejmujące zróżnicowaną klasę związków semantycznych istotnych dla efektywności wyszukiwania informacji, na przykład: Charyzmatyk TK dary Ducha Świętego; Kazanie TK Homilia; Opętanie TK Szatan; Wiara TK Nadzieja.

### JEDNOZNACZNOŚĆ

Ponieważ wyszukiwanie ma charakter, w mniejszym lub większym stopniu, mechaniczny, istotne jest zapewnienie jednoznaczności wszystkich wyrażeń JIW. Jednoznaczność będzie osiągnięta, jeśli nie tylko poszczególne poziomy znaki JIW, ale i ich złożenia będą spełniać warunek jednoznaczności. Homonimia, czyli brak jednoznaczności, w językach informacyjno-wyszukiwawczych najczęściej pojawia się wtedy, gdy wyrażenia JIW mają formę wyrażeń języka naturalnego, z którym to przypadkiem mamy do czynienia m.in. projektując tezaurus. Homonimia ta, zwana leksykalną, eliminowana jest przez stosowanie ob-

**Homonimia  
leksykalna**

jaśnienie czy definicji, na przykład: Analiza harmoniczna (matematyka); Krakowiak (taniec); czystość (higiena); czystość (etyka). Jeśli będziemy potrzebować w teaurusie pojęcia patriarchat w dwu znaczeniach (rodowy ustrój społeczny; godność i władza patriarchy), to musimy deskryptor patriarchat w znaczeniu ustrój odróżnić jakoś od deskryptora patriarchat w drugim znaczeniu. Można to zrobić na przykład za pomocą objaśnień semantycznych, tj. patriarchat (ustrój), patriarchat (władza patriarchy), lub „mechanicznie” – patriarchat1, patriarchat2 – podając w artykule deskryptora patriarchat1 jego definicję (rodowy ustrój społeczny), a w artykule deskryptora patriarchat2 drugą definicję (godność i władza patriarchy). W systemie łatwiej jest manipulować deskryptorami typu patriarchat1, patriarchat2, ale wtedy musielibyśmy zapewnić użytkownikowi dostęp do teaurusu i m.in. definicji znaczeń deskryptorów. Dlatego lepiej zdecydować się na stosowanie objaśnień, zakładając że będą one miały nieco umowny i skrótowy charakter, zaś obszerniejsze omówienia znaczeń i różnic znaczeniowych między deskryptorami wprowadzić do artykułów deskryptorowych.

Szczególnie kłopotliwe są przypadki, gdy poszczególne, zbliżone do sie-

bie, aczkolwiek nieidentyczne znaczenia trudno odróżnić. Obfitują w nie zwłaszcza niektóre nauki społeczne. Różnice te, choć często trudno uchwytne, są wystarczająco duże, aby stać się źródłem błędów i nieporozumień lub strat informacji w wyszukiwaniu. Wydaje się, że pomocnym narzędziem mogłyby tu być definicje, które pozwoliłyby na wyraźne nazywanie różnic, oraz zaangażowanie w tworzenie teaurusów osób, które nie tylko mają wiedzę narzędziową (wiedzą, jak to robić), ale i dobrze znają się na dziedzinie, której teaurus dotyczy.

*W Teaurusie technologii żywności i przemysłu spożywczego* tak rozwiązano problem wieloznaczności:

„a) zastępowanie wyrażenia wieloznacznego odpowiednimi terminami jednoznacznymi, np. askryptor «monitorowanie» reprezentują deskryptory KONTROLA PROCESU + STEROWANIE;

b) stosowanie objaśnienia, kwalifikatora ograniczającego zakres znaczenia danego deskryptora do określonej dziedziny i stanowiącego jego integralną część, np. KOMÓRKA (BIOL.), OWOCE (MORF);

c) umieszczenie pod deskryptorem określenia lub definicji zawężających jego znaczenie, np.:

- deskryptor PRODUKCJA. Zaopatrzone w określenie „ilość lub wartość”,
- deskryptor ŻYWNÓŚĆ POCHODZENIA MORSKIEGO. – w definicję „przetwory

- z organizmów morskich innych niż ryby”,
- deskryptor NORMALIZACJA MLEKA. – w definicję „doprowadzanie zawartości tłuszczu w mleku do ilości zgodnej z normą”,
  - deskryptor WODA. – w określenie „w sensie hydrologicznym”<sup>4</sup>.

### NIESYNONIMICZNOŚĆ

W językach informacyjno-wyszukiwawczych synonimia może (choć nie powinna) się pojawić zarówno na poziomie wyrazów, jak i na poziomie zdań i tekstów. Eliminacji synonimii leksykalnej (tj. na poziomie wyrazów) służy relacja odpowiedniości wyszukiwawczej. Jednemu z synonimów nadaje się status deskryptora, pozostałym – askryptorów. Na przykład:

**Synonimia  
leksykalna**

Eucharystia

NU komunija

Msza Święta

Wieczerza Pańska

Najświętszy Sakrament

sakrament Eucharystii

oraz odpowiednie artykuły askryptorowe:

komunija U Eucharystia

Msza Święta U Eucharystia

Wieczerza Pańska U Eucharystia

---

<sup>4</sup> *Tezaurus technologii żywności i przemysłu spożywczego: część systematyczna*. Warszawa 1996, s. VIII.

Najświętszy Sakrament U Eucharystia

sakrament Eucharystii U Eucharystia

**Synonimia  
międzypo-  
ziomowa**

W językach informacyjno-wyszukiwawczych istota synonimii sprowadza się do wyznaczania przez różne wyrażenia JIW tego samego obszaru rzeczywistości pozajęzykowej, tak więc, wyszukiwanie informacji o tym obszarze za pomocą tylko jednego z istniejących wyrażen synonimicznych prowadziłyby do utraty relewantnej informacji. Tego rodzaju synonimia może istnieć na każdym poziomie JIW. Na przykład, w Tezaurusie Medycznym<sup>5</sup> istnieją deskryptory TECHNOLOGIA PLANOWANIA ZDROWIA, TECHNOLOGIA i PLANOWANIE ZDROWIA. Można zatem przypuszczać, że literatura poświęcona technologii planowania zdrowia będzie rozproszona – część będzie zaindeksowana deskryptorem TECHNOLOGIA PLANOWANIA ZDROWIA, część deskryptorami TECHNOLOGIA i PLANOWANIE ZDROWIA. Jeśli użytkownik nie będzie świadom istniejącej synonimii, to realizując wyszukiwanie odnajdzie tylko część relewantnej literatury. Dlatego należy rozsądnie ograniczać liczbę wprowadzanych do tezauryusa deskryptorów złożonych (wielowyrazowych) do przypadków, kiedy rzeczywiście nie da się

---

<sup>5</sup> *Tezaurus Medyczny: polska wersja MeSH*. Warszawa 1995-1996.



wyrazić znaczenia deskryptora złożonego za pomocą „składanki” deskryptorów lub gdyby powodowało to istotne utrudnienie posługiwania się tezauresem. Wielką pomocą służą tu artykuły askryptorowe, dzięki którym utrzymana jest możliwość wyszukiwania za pomocą wyrażen wieloelementowych, bez obawy o wystąpienie synonimii między różnymi poziomami znaków JIW.

## BUDOWANIE TEZAURUSA – NAJWAŻNIEJSZE DZIAŁANIA

Jak wcześniej wspomniano, przystępując do projektowania języka informacyjno-wyszukiwawczego, trzeba nie tylko mieć wiedzę o przedmiocie projektowania (dziedzinie, dla której język jest projektowany) i sposobie realizacji zadania (wiedza z zakresu metodologii projektowania oraz teorii i praktyki języków informacyjnych), ale i o warunkach, w jakich projektowany język będzie funkcjonował.

Obowiązkiem projektanta JIW jest wnikliwe i rzetelne poznanie otoczenia języka – jego istotnych elementów składowych, ich funkcji, związków, praw nimi rządzących itp. Nie można jednak zapomnieć, że projektowanie tak „miękkich” obiektów, jak języki czy systemy informacyjno-wyszukiwawcze, zawsze łączy się ze szczególnym rodzajem niedookreśloności,

sztuką operowania projektowaną materią sprawiającą, że podobnie jak w twórczości artystycznej pozostaje zawsze do dyskusji obszar (raz większy, a raz mniejszy) „niekontrolowany” związany z kategoriami estetycznymi, myśleniem heurystycznym, wyczuwaniem itp.

W skład zespołu, który będzie opracowywał tezaurus, powinni wchodzić zarówno specjaliści w zakresie dyscypliny czy branży, która ma być objęta tezaurusem, jak i specjaliści od budowy języków informacyjnych i ich stosowania. Obecnie należałoby jeszcze zapewnić konsultacje ze strony „komputerowców”, aby z jednej strony usprawnić samo budowanie tezaury, z drugiej zaś stworzyć go w takiej postaci, która najlepiej będzie odpowiadać otoczeniu technologicznemu (software, wymagania sieciowe itp.). Liczebność zespołu uzależniona jest głównie od zakresu tematycznego tezaury i przewidywanej jego objętości. Niekiedy tezaury budowane są poprzez „sklejanie” tezaurusów częściowych (mikrotezaurusów), co ma tę zaletę, że do stworzenia mikrotezaury wystarczy niekiedy jeden projektant wspierany pomocą konsultantów. Należy jednak bezwzględnie zadbać, żeby wszystkie tezaury składowe docelowej, większej całości miały analogiczne struktury i powstawały według wskazań tej samej instrukcji.

DZIAŁANIA SKŁADAJĄCE SIĘ NA PROCES  
PROJEKTOWANIA JĘZYKA  
DESKRYPTOROWEGO

1. Określenie zakresu języka (por. s. 69-72).

2. Wskazanie źródeł, z których (lub na podstawie których) będzie pobierane słownictwo. Należałoby także zebrać informacje o istniejących polskich i zagranicznych tezaurusach lub słownikach innych języków informacyjnych o tym samym lub zbliżonym zakresie i rozważyć ich przydatność ze względu na cel podjętej pracy.

3. Robocze określenie struktury części rzeczowej (systematycznej) tezaurusu.

Na tym etapie może to być (bardziej lub mniej dokładny) schemat dziedzinowy czy dziedzinowo-zagadnieniowy. Chodzi o to, żeby osoby gromadzące słownictwo, mające „w ręku” terminy wstępnie przydzielały je do jakiejś dziedziny, klasy, obszaru tematycznego itp. Przydział ten nie musi być przydziałem jednokrotnym. Nawet wyrażenia (denotacyjnie) jednoznaczne mogą znaleźć się w więcej niż jednej klasie (kategorii, dziale) w zależności od punktu patrzenia na dane pojęcie, na przykład *kot* może znaleźć się w klasie „Zwierzęta domowe z rodziny *Felis domestica*”, ale także w klasie „Ulubione zwierzęta domowe”.

#### 4. Wybór metody gromadzenia słownictwa.

Słownictwo może być gromadzone albo metodą indukcyjną, albo dedukcyjną, albo mieszaną (indukcyjno-dedukcyjną). Z doświadczenia wynika, że najlepsza, choć czaso- i pracochłonna, jest metoda mieszana. W takim przypadku najpierw gromadzi się terminy metodą indukcyjną, a następnie uzupełnia powstały w ten sposób zbiór w celu eliminacji luk w słownictwie metodą dedukcyjną. Może też być odwrotnie: najpierw gromadzi się słownictwo metodą dedukcyjną, a później uzupełnia je metodą indukcyjną. Niezależnie od przyjętej metody, zarówno sam proces, jak i kryteria wyboru terminów, są w jakiejś mierze intuicyjne i subiektywne, często zależą od preferencji i uznania osoby, która buduje JIW.

#### **Metoda indukcyjna**

Metoda indukcyjna polega na indeksowaniu, tj. opisywaniu treści i formy wybranych dokumentów za pomocą swobodnie dobieranych słów kluczowych. Inaczej mówiąc, polega na gromadzeniu słów kluczowych, które trafnie charakteryzują treść i formę indeksowanych dokumentów. Chcąc podać jakąś wskazówkę praktyczną co do szczegółowości typowanych słów kluczowych, można powiedzieć, że podejmując decyzję, czy dane słowo kluczowe jest dobre, czy też za szczegółowe (lub za ogólne) powinno się odpowiedzieć na pytanie, czy gdybym

szukał literatury na temat, chciałbym otrzymać ten właśnie dokument w odpowiedzi. Słowa kluczowe niezarejestrowane na tym etapie pracy, zwykle pojawiają się przy uzupełnianiu leksyki metodą dedukcyjną. Jakies luki są też odkrywane i uzupełniane przy budowaniu części rzeczowej (systematycznej) tezauryusa.

Metoda indukcyjna wymaga określenia zbioru reprezentatywnego dokumentów, które będą indeksowane w celu zgromadzenia roboczego zestawu słownictwa do dalszej obróbki. Zbiór ten powinien proporcjonalnie odzwierciedlać cały zakres tematyczny tezauryusa. Na początku pracy przyrost liczby słów kluczowych jest znaczny i gwałtowny. Po jakimś czasie projektanci zauważają jednak, że przyrost ten wprawdzie nie uległ całkowitemu zahamowaniu, ale utrzymuje się przez dłuższy czas na stałym, niewysokim poziomie. W tym momencie gromadzenie słów kluczowych metodą indukcyjną można uznać za zakończone i przejść do fazy metody dedukcyjnej. Zbieranie słownictwa jest czasochłonne. Kłopoty sprawia również określenie wielkości zbioru reprezentatywnego, gdyż zależy ona zarówno od dziedziny wiedzy, jak i od tzw. czasookresu zbioru dokumentów<sup>6</sup> podda-

**Zbiór  
reprezentatywny**

---

<sup>6</sup> Czasookresem zbioru dokumentów nazywa się okres od daty wydania najwcześniej opublikowanego dokumentu, znajdującego się w zbiorze, do daty najpóźniej wydanego dokumentu.

**Metoda  
dedukcyjna**

nych obróbce oraz stosunku do względnej szybkości rozwoju danej dziedziny, a zatem i jej terminologii.

Metoda dedukcyjna wymaga z kolei sporządzenia wykazu źródeł, z których będzie pobierane słownictwo robocze do dalszej obróbki. Najczęściej są to: leksykony i encyklopedie, normy terminologiczne, tezauryusy o pokrewnej tematyce, klasyfikacje danej dziedziny, monografie i inne źródła (zwłaszcza ich spisy treści i indeksy rzeczowe). Oczywiście słownictwo przejmowane z tych źródeł będzie wymagać obróbki leksykograficznej zgodnie z regułami przyjętymi dla danego tezaurusa, ale na ogół ma się już gotowe wyrażenia ekwiwalentne oraz wyrażenia szersze, węższe i pokrewne.

**Technika  
pracy**

**5. Zgromadzenie słownictwa.**

Jeśli słownictwo jest gromadzone na kartach (np. formatu kart dokumentacyjnych), to zaleca się wypisywanie dla każdego słowa kluczowego dwóch kart, z których jedna jest przeznaczona do części alfabetyczno-hierarchicznej, a druga do systematycznej. Na karcie przeznaczonej do części alfabetycznej należy zapisać dane słowo kluczowe, jego znaczenie, ew. synonimy lub słowa bliskoznaczne, jeśli można je ustalić na podstawie danego dokumentu, odpowiedniki obcojęzyczne (w razie potrzeby i możliwości ustalenia), uwagi co do ew. używania termi-

nu oraz źródło, na podstawie którego zostało to słowo wytypowane. Źródło pochodzenia słowa kluczowego może mieć np. postać numeru dokumentu na liście dokumentów reprezentatywnych. Należy pamiętać, że na tym etapie pracy każde słowo kluczowe (nawet te, które później nie zostanie przyjęte i stanie się askryptorem lub w ogóle nie wejdzie do tezaurusa) powinno mieć własną kartę. Jeśli słowo kluczowe ma więcej niż jedno znaczenie, należy opracować odrębną kartę dla każdego znaczenia.

Na karcie do części systematycznej zaleca się podawanie słów kluczowych uporządkowanych w gałęzie lub mini-gałęzie hierarchiczne (słowa kluczowe szersze i węższe) z zaznaczeniem, do której klasy (klas) roboczej wersji części systematycznej tezaurusa należałoby dane słowo przydzielić. Na tym etapie lepiej jest nie odrzucać nawet bardzo szczegółowych słów kluczowych. Zawsze w późniejszych stadiach pracy nad tezaurusem mocą relacji ekwiwalencji można ich znaczenia zawrzeć w znaczeniach deskryptorów ogólniejszych.

Jeśli słownictwo będzie gromadzone od razu w komputerze, należy zakładać odpowiednie formularze i upewnić się, czy wykorzystywane oprogramowanie jest w stanie wykonać pożądane czynności, np. czy program potrafi utworzyć artykuły

Forma  
językowa  
deskryptorów

askryptorowe na podstawie artykułów deskryptorowych, oraz czy forma prezentacji gotowego teaurusu jest właściwa i przyjazna użytkownikowi.

Wybierane słowa kluczowe najlepiej jest zapisywać w formie rzeczowników w mianowniku liczby pojedynczej. Spóradycznie słowami kluczowymi mogą też być liczebniki i przymiotniki. W praktyce najczęściej przekształca się przymiotniki, imiesłowy i czasowniki w rzeczowniki. Przymiotniki zostają samodzielnymi słowami kluczowymi, jeśli nie istnieją odpowiadające im rzeczowniki, np. międzynarodowy.

Nie należy rozdzielać wielowyrzowych słów kluczowych, jeżeli są to:

- związki wyrazowe stałe, czyli takie, których znaczenie nie jest sumą znaczeń wyrazów składowych, np.: boża krówka, czarna skrzynka, żywe srebro, cesarskie cięcie;
- związki wyrazowe łączliwe, czyli takie, w których wyrazy składowe zachowują swoje podstawowe znaczenie, lecz stopień ich zespolenia jest znaczny, np.: opóźniacze korozji, współpraca międzynarodowa, grupa społeczna;
- wyrażenia użyte w znaczeniu przenośnym, np. pomarańczowa skórka;
- wyrażenia, w których jeden ze składników jest użyty w innym znaczeniu w porównaniu z tym, kiedy występuje samodzielnie, np.: przemysł lekki, zatrucie katody;



- wyrażenia, których co najmniej jeden element jest wyrazem wieloznacznym, np.: połączenia wieloznacznych wyrazów technika lub środek z wyrazami przetwarzania i transportu w wyrażeniach technika przetwarzania, środek transportu;
- wyrażenia będące nazwami pewnych całości fizycznych lub określonego materiału czy też wyrobu, np.: huta szkła, maszyna do pisania;
- nazwy własne, np.: Morze Północne, Kaplica Zygmuntowska, Wały Jagiellońskie.

Poza tym szczególną uwagę należy zwrócić na połączenia nazw czynności lub procesów z nazwami obiektów (przedmiotów), które są poddawane tym czynnościom, np. naprawa dachu. Zwykle tego typu wyrażenia należy rozdzielać na wyrazy składowe, a więc powinny to być deskryptory dach i naprawa.

Pojedynczymi słowami kluczowymi powinny też być zastępowane połączenia nazwy własności z nazwami obiektów, np.: trwałość żywności (trwałość, żywność), wartość białka (wartość, białko). Należy unikać deskryptorów będących połączeniami nazw całości i części, np.: apsyda nawy, apsyda prezbiterium (apsyda, nawa, prezbiterium). Analogicznie należy unikać połączeń nazw obiektów i ich przeznaczenia, np.: kwiaty do Grobu

Forma  
językowa  
deskryptorów,  
cd.

Pańskiego (Grób Pański, kwiaty), oraz nazw przedmiotów z formami ich prezentacji w dokumencie, np. nie encyklopedia Wrocławia, a encyklopedia, Wrocław.

Jednak słowa kluczowe o znaczeniu podstawowym dla zakresu tematycznego tezaury, które często będą używane, można wprowadzić do tezaury w formie połączeń wyrazowych nawet wtedy, gdy nie jest to zgodne z powyższymi wskazówkami.

6. Ustalenie zasad tworzenia ciągów synonimicznych, czyli podjęcie decyzji, co i jak będzie się łączyć relacją odpowiedniości wyszukiwawczej (por. s. 75-81).

7. Ustalenie, jak będą „oddzielane” różne znaczenia wyrażeń wieloznacznych. Czy zastosuje się objaśnienia? Jeśli tak, to jakie? Czy może będzie się preferować definicje?

8. Wybranie sposobu zapisu deskryptorów, askryptorów, ew. modyfikatorów; przyjęcie notacji (umownych wskaźników, symboli), za pomocą której będą oznaczane np. typy relacji łączących hasła w tezaursie.

9. Opracowanie kilku wzorcowych przykładów artykułów deskryptorowych i askryptorowych (z komentarzem).

## 10. Opracowanie tezaury, w tym części systematycznej.

Po wstępnym zgromadzeniu słownictwa przystępuje się do ich obróbki leksykograficznej i analizy ich wzajemnych powiązań. Na ogół analiza ta ma charakter intuicyjny i polega na porównywaniu poszczególnych terminów. Zgromadzona leksyka przydzielana jest do wcześniej zaprojektowanych roboczych kategorii semantycznych, stanowiących szkielet struktury języka. Gdy liczba zebranych słów kluczowych jest znacznie większa od planowanej liczby deskryptorów w tezaurucie, słowa kluczowe pokrewne znaczeniowo łączy się w tzw. gniazda (rodziny pojęć). Jedno ze słów otrzymuje status centrum pola. Pozostałe stanowią otoczenie tego centrum. Relacje łączące otoczenie z centrum są różne w różnych JIW. Wybór wyrazu jako centrum gniazda jest wypadkową znajomości terminologii danej dziedziny, przydatności wyszukiwawczej poszczególnych wyrazów, częstotliwości ich występowania, przyjętych zasad ustanawiania relacji odpowiedniości.

Część systematyczna tezaury może mieć różną organizację i postać. Może to być np. klasyfikacja dziedzinowa, fasetowa, dziedzinowo-fasetowa. Obecnie najwyższej cenie jest ostatnie z wymienionych podejść, przybierające postać kategoryzacji. Na pierw-

Część  
systematyczna

szym poziomie uniwersum, jeśli nie wyróżnia się tzw. wymiarów<sup>7</sup>, słownictwo dzielone jest na pewną liczbę kategorii semantycznych, po czym wyrazy wstępnie przyporządkowane do każdej kategorii porządkuje się według podziału fasetowego, dzieląc daną kategorię na podkategorie (fasety) zgodnie z przyjętymi zasadami podziału, np. „wg konstrukcji”, „wg zastosowania”, „wg gatunku” itp.

Klasyfikacje deskryptorów w ramach wszystkich faset przeprowadza się z wykorzystaniem tylko jednego typu relacji. Preferowana jest relacja hierarchii generycznej. Tworzenie podziałów w oparciu o różne relacje może powodować błędy w wyszukiwaniu informacji, a tworzenie oddzielnych faset według relacji poszczególnych typów (odpowiednio oznakowanych), utrudnia budowanie tezaurya i proces wyszukiwania informacji.

Wydaje się, że praktyczniej jest zbudować najpierw część systematyczną tezaurya, a następnie w oparciu o nią oraz zbiór deskryptorów utworzyć część alfabetyczno-hierarchiczną.

11. Opracowanie list identyfikatorów. Pożądane byłoby zachowanie

---

<sup>7</sup> O. Ungurian: *Kategoria semantyczna wymiaru*, „Przedmiot dokumentu”. „Zagadnienia Informacji Naukowej” 1983 nr 1 s. 29-54; Tenże: *Ogólna struktura języka informacyjnego: propozycja „szkieletu organizacyjnego” słownictwa*. „Zagadnienia Informacji Naukowej” 1982 nr 2 s. 21-40.

zgodności formy identyfikatorów z zaleceniami odpowiednich polskich norm (wybór norm w załączonym wykazie *Ważniejszych lektur*).

12. Opracowanie listy deskryptorów chronologicznych.

13. Określenie reguł gramatycznych. Należałoby się zastanowić, czy w projektowanym teaurusie będą istniały modyfikatory. Modyfikatory to wyrażenia, które mogą być użyte tylko i wyłącznie „w towarzystwie” deskryptorów (razem z deskryptorem), mające szeroki zakres łączliwości semantycznej i stosowania, np.: metody, teoria, organizacja, nauczanie itp.

**Modyfikatory**

14. Określenie zasad stosowania języka, czyli indeksowania dokumentów i pytań użytkowników. Istotny wpływ na funkcjonowanie języka informacyjno-wyszukiwawczego ma nie tylko stopień skomplikowania systemu JIW, ale także jednoznaczne i precyzyjne określenie zasad stosowania języka, czyli indeksowania dokumentów i pytań czytelników. Określanie zasad indeksowania często jest traktowane po macoszemu, gdy tymczasem kompletność i dokładność (zwane też szerokością, głębokością i szczegółowością indeksowania) tego procesu ma wpływ na efektywność funkcjonowania JIW równie duży jak cechy samego języka.

**Instrukcja  
indeksowania**

**Adekwatność  
indeksowania**

Wypracowanie uniwersalnych, jednoznacznych i obiektywnych sposobów ustalania kompletności i dokładności indeksowania nie wydaje się możliwe. Ta sama książka opracowana w różnych bibliotekach może mieć niejednakowe opisy i wcale nie musi to być wynikiem błędów katalogującego czy braków stosowanego języka informacyjno-wyszukiwawczego. Tymczasem w niewielu bibliotekach istnieją opracowania, z których nowo przyjęty pracownik mógłby się dowiedzieć, jak obszernie i szczegółowo opracowywać dokumenty. Zbyt często ufa się tu intuicji bibliotekarza, która, choć niezwykle cenna i ważna, nie powinna być jedyną i ostateczną instancją. Weźmy na przykład książkę *Idy Kurcz Pamięć. Uczenie się. Język. Czy stwierdzenie, że treścią książki są wybrane zagadnienia psychologii ogólnej jest adekwatne? A może należałoby powiedzieć, że treścią jest problematyka pamięci, uczenia się i mowy? A może zagadnienia rodzajów pamięci, metod badania, ich modeli, teorii pamięci, uczenia się, mowy i organizacji niektórych czynności psychicznych w mózgu? Nawet ten jeden przykład wystarczy, żeby pokazać względną naturę adekwatnej charakterystyki treści. Która bowiem z wymienionych charakterystyk jest adekwatna? Ta, która jest najbliższa zasadom indeksowania, w tym optymalnym parametrom kompletności*

i dokładności, obowiązującym w danym systemie.

Inną kwestią jest możliwość przełożenia adekwatnej charakterystyki treści i formy dokumentu na stosowany w bibliotece język informacyjno-wyszukiawczy. Nawet najlepiej opracowana charakterystyka dokumentu straci swą wartość, jeśli w JIW nie będzie leksyki i środków gramatycznych odpowiednich do wyrażenia pożądanej treści. Postulat niepominiania informacji relewantnych nie jest równoznaczny z koniecznością wprowadzania do charakterystyki dokumentu wszystkich jego wątków tematycznych w stopniu proporcjonalnym do tego, jaki przedstawia dokument. Pominiecie informacji relewantnej prowadzi wprawdzie do strat informacji, ale zbyt szerokie reprezentacje, uwzględniające informacje o znaczeniu marginalnym dla danej grupy użytkowników, mogą powodować wzrost szumu informacyjnego. Poziom szerokości, podobnie jak poziom szczegółowości, powinien być ustalany w efekcie badania systemu i oczekiwań użytkowników.

15. Wstępne wdrożenie i poprawienie zaobserwowanych błędów, niedociągnięć itp.

16. Określenie zasad aktualizacji oraz wskazanie osób, instytucji odpowiedzialnych za utrzymywanie języka.



i dokładności, obowiązującym w danym systemie.

Inną kwestią jest możliwość przełożenia adekwatnej charakterystyki treści i formy dokumentu na stosowany w bibliotece język informacyjno-wyszukiwawczy. Nawet najlepiej opracowana charakterystyka dokumentu straci swą wartość, jeśli w JIW nie będzie leksyki i środków gramatycznych odpowiednich do wyrażenia pożądanej treści. Postulat niepominiania informacji relewantnych nie jest równoznaczny z koniecznością wprowadzania do charakterystyki dokumentu wszystkich jego wątków tematycznych w stopniu proporcjonalnym do tego, jaki przedstawia dokument. Pominiecie informacji relewantnej prowadzi do strat informacji, ale zbyt szerokie reprezentacje, uwzględniające informacje o znaczeniu marginalnym dla danej grupy użytkowników, mogą powodować wzrost szumu informacyjnego. Poziom szerokości, podobnie jak poziom szczegółowości, powinien być ustalany w efekcie badania systemu i oczekiwań użytkowników.

15. Wstępne wdrożenie i poprawienie zaobserwowanych błędów, niedociągnięć itp.

16. Określenie zasad aktualizacji oraz wskazanie osób, instytucji odpowiedzialnych za utrzymywanie języka.







Wydawnictwo Stowarzyszenia Bibliotekarzy Polskich  
00-335 Warszawa, ul. Konopczyńskiego 5/7  
Wydanie I. Warszawa 2005 r. Ark. wyd. 4,5. Ark. druk. 6,5.  
Łamanie: Renard Hawryzko  
Druk i oprawa: Zakład Poligraficzny PRIMUM, Kozierki 17a  
05-825 Grodzisk Maz., tel. 724-18-76



ISBN 83-89316-53-6

WYDAWNICTWO  
SBP

