

SŁOWNICTWO SYSTEMU TEZAURUSÓW BIBLIOTEKI SEJMOWEJ NA PRZYKŁADZIE MIKROTEZAURUSA „OŚWIATA”

System Tezaurusów Biblioteki Sejmowej (STEBIS) obejmuje 10 tezaurusów (tezaurus prawa, tezaurus zagadnień państwowych i administracyjnych, tezaurus polityki, tezaurus zagadnień społecznych, tezaurus gospodarki, tezaurus handlu, transportu i pracy, tezaurus ekonomii, tezaurus nauki, tezaurus organizacji i dokumentów międzynarodowych, tezaurus nazw geograficznych) scalonych w postaci elektronicznej w jeden zbiór. Mimo że każdy z nich może funkcjonować jako samodzielny słownik języka deskryptorowego, w rzeczywistości składają się one na spójny system terminów wyszukiwawczych.

System Tezaurusów Biblioteki Sejmowej związany jest silnie z tezaurem EUROVOC (wielojęzycznym tezaurem stworzonym z inicjatywy Parlamentu Europejskiego i Biura Publikacji Urzędowych Wspólnoty Europejskiej, który obejmuje swym zakresem wszystkie dziedziny zainteresowań instytucji Wspólnoty Europejskiej). Początki STEBIS wiążą się z decyzją o stworzeniu własnego systemu opracowania rzeczowego, podjętą w związku z przejściem Biblioteki Sejmowej z systemu MAK na system ALEPH. Jako przyszłe narzędzie indeksowania zdecydowano wykorzystać właśnie tezaurus EUROVOC. Atutami, które o tym zadecydowały było ukierunkowanie na problemy polityczne, społeczne i gospodarcze, możliwość skorzystania z doświadczeń wielu bibliotek parlamentarnych, które stosowały EUROVOC, możliwość wyszukiwania w bazach danych Wspólnoty oraz wielojęzyczność (9 wersji językowych – francuska, angielska, niemiecka, hiszpańska, duńska, grecka, włoska, holenderska i portugalska; później dokonano jeszcze 2 oficjalnych tłumaczeń na szwedzki i fiński oraz wielu lokalnych tłumaczeń), która umożliwia indeksowanie dokumentów i zapytań informacyjnych w jednym języku (języku użytkownika), a prowadzenie wyszukiwania i w odpowiedzi uzyskiwanie dokumentów zaindeksowanych w którymkolwiek z pozostałych języków tezaury.

Tłumaczenie tezaury EUROVOC (oparte na jego drugim wydaniu z 1987 r.) rozpoczęto w czerwcu 1993 r. Zdecydowano utrzymywać STEBIS w 4 wersjach językowych (polskiej, angielskiej, francuskiej i niemieckiej). Tłumaczenia dokonywano działami, rozpoczynając od najmniej kłopotliwych, tj. nazw geograficznych i organizacji międzynarodowych.

Wersję roboczą przekładu zakończono w 1994 r. W tym też roku przetestowano tezaurus, wprowadzając do systemu ALEPH ok. 150 opisów deskryptorowych. Następnie przystąpiono do indeksowania deskryptorami wybranych działów (jednocześnie, ze względu na konieczność utrzymania zgodności z trzecim wydaniem EUROVOC z połowy 1995 r., wykrywano i wprowadzano zmiany w stosunku do poprzedniej wersji).

W czasie testowania wprowadzono liczne zmiany, takie jak zmiana formy deskryptora, wzbogacenie o askryptory, wzbogacenie o dodatkowe relacje

kojarzeniowe, wzbogacenie o deskryptory nie występujące w EUROVOC (ze względu na specyficzne potrzeby Biblioteki Sejmowej). Oznaczało to w praktyce odchodzenie w coraz większym stopniu od pełnej zgodności z EUROVOC.

Wprowadzane zmiany doprowadziły do znacznego rozrostu słownictwa, przekształcenia struktury hierarchicznej tezaursu oraz dodania wielu nowych działów. Efektem tych zmian jest STEBIS, który można uznać za własny słownik Biblioteki Sejmowej, tylko w części kompatybilny z EUROVOC (głównie na poziomie terminologii).

STEBIS jest słownikiem dynamicznym, tzn. stale modyfikowanym i udoskonalanym w celu dostosowania go w jak największym stopniu do potrzeb użytkowników. Modyfikacje te polegają przede wszystkim na wzroście liczby deskryptorów, askryptorów, relacji i odpowiedników obcojęzycznych.

Przyrost deskryptorów jest stały, a w ostatnich latach coraz większy. Wynika to z odejścia od pełnej spójności z EUROVOC, pojawienia się nowych tematów i zmiany strategii wprowadzania deskryptorów do systemu. Początkowo przejmowano tylko te deskryptory, na które było zapotrzebowanie w katalogu. Później podjęto decyzję o wprowadzeniu do STEBIS wszystkich terminów z EUROVOC. Wynika to z popularności EUROVOC (oprócz baz danych Unii Europejskiej stosowany jest w wielu bibliotekach i systemach parlamentarnych i legislacyjnych) i możliwości wyszukiwania przy pomocy jego terminów w bazach danych dostępnych np. w Internecie.

Oprócz wprowadzania do STEBIS wszystkich deskryptorów EUROVOC oraz dodawania nowych deskryptorów prowadzone jest też „rozbijanie” deskryptorów, przy których liczba wyszukanych dokumentów jest zbyt duża, i dodawaniu nowych, bardziej szczegółowych.

Wielojęzyczność STEBIS polega na przydzielaniu każdemu deskryptorowi polskiemu odpowiedników w trzech pozostałych językach, nie są jednak dodawane dla wersji obcych ani askryptory, ani scope notes.

W celu ułatwienia nawigacji w obrębie poszczególnych tezaursów przyjęto jeden spójny system kodowania. Każdemu tezausowi STEBIS przyporządkowano jednocyfrowy kod. Tezaurus obejmuje szereg mikrotezaursów (nie więcej niż 10), z których każdy oznaczony jest dwucyfrowym kodem (pierwsza cyfra to kod tezaursu). Z kolei każdy z mikrotezaursów może mieć kilka działów (nie więcej niż 10) oznaczonych kodem trzycyfrowym (dwie pierwsze cyfry to kod mikrotezaursu).

Pomimo wciąż trwających prac nad tezausami STEBIS podjęto decyzję o sukcesywnym publikowaniu poszczególnych mikrotezaursów. Podstawową częścią każdego z nich jest tzw. część główna – alfabetyczny wykaz deskryptorów i askryptorów ze wskazaniem relacji zachodzących między nimi. Wprowadzono też trzy wykazy pomocnicze, przedstawiające ten sam zestaw deskryptorów w różnych układach i z wybranymi elementami pełnego artykułu deskryptorowego (pierwszy – hierarchiczny, zawierający kolejne działy mikrotezaursu wraz z przypisanymi im deskryptorami uporządkowanymi według ich hierarchicznej zależności, tj. terminy podrzędne wymienione są pod deskryptorem nadrzędnym; drugi – tematyczny, zawierający te same działy, ale deskryptory w obrębie działów uszeregowane są alfabetycznie; trzeci – wielojęzyczny, zawierający wszystkie deskryptory mikrotezaursu bez podziału na działy uporządkowane w jednym ciągu alfabetycznym wraz z ich odpowiednikami w pozostałych językach, przy czym

podstawą szeregowania w każdej wersji językowej jest alfabetyczna kolejność deskryptorów w języku danej wersji).

Terminy wyszukiwawcze występujące w poszczególnych tezaurusach wybierane są w sposób specyficzny (zagadnienie to poruszane jest często w recenzjach poszczególnych mikrotezaurusów)¹. Podstawowym kryterium jest występowanie danego terminu w tezaurusie EUROVOC lub obecność literatury na dany temat w zbiorach Biblioteki Sejmowej, jako że tezaurus nie opisuje stanu rzeczywistości, tylko odpowiada potrzebom użytkowników i zawartości treściowej piśmiennictwa. Ze względu na dostosowanie tezaurusów do potrzeb użytkowników Biblioteki Sejmowej, która gromadzi głównie zbiory z dziedzin takich jak prawo, polityka, zagadnienia parlamentarne i gospodarcze, zasób terminów wyszukiwawczych zgodny jest z tematyką zbiorów. Dlatego też STEBIS może okazać się niemożliwy do przeniesienia do bibliotek o innym profilu bez uzupełnienia zbioru terminów wyszukiwawczych o terminy dostosowane do ich potrzeb.

Tezaurus definiowany jest zazwyczaj jako słownik języka deskryptorowego służący do indeksowania i wyszukiwania informacji². Interesującym wydaje się porównanie terminów występujących w tezaurusie z terminami innego języka informacyjno-wyszukiwawczego w celu przetestowania, czy użytkownik mógłby wyszukiwać informacje w bazie danych indeksowanej przy pomocy języka deskryptorowego posługując się terminami wyszukiwawczymi innego języka. Dla potrzeb tego badania wybrany został mikrotezaurus „Oświata” w wersji drukowanej³. Wybór wersji drukowanej (która z związku z dynamicznym rozwojem słownika może być nie w pełni aktualna) związany jest z faktem, że ze względu na scalenie wszystkich tezaurusów (a co za tym idzie i mikrotezaurusów) w wersji elektronicznej w jeden zbiór, brak jest możliwości wybrania materiału leksykalnego pochodzącego z konkretnego mikrotezaurusa.

Do badania wybrane zostały tylko deskryptory jako terminy pełniące w tezaurusie rolę terminów wyszukiwawczych. Mikrotezaurus „Oświata” zawiera 162 deskryptory. Wszystkie one posiadają wymaganą w tezaurusach formę rzeczownika - deskryptory proste (24 terminy – 14,82% materiału) lub frazy rzeczownikowej - deskryptory złożone (138 terminów – 85,18% materiału)⁴. Tylko 24 deskryptory (14,82% wszystkich deskryptorów) mają formę liczby mnogiej, np.: **instytucje oświatowe, koszty oświaty, kursy, metody nauczania, materiały dydaktyczne**, itp.

Dla potrzeb badań porównawczych wybrano język haseł przedmiotowych jako jeden z najbardziej rozpowszechnionych języków informacyjno-wyszukiwawczych. Przy pomocy jednostek leksykalnych pełniących funkcję deskryptorów wyszukiwano informacje w bazie danych, indeksowanej hasłami przedmiotowymi. Przeważająca większość materiału leksykalnego występującego w mikrotezaurusie jako deskryptory (98 przypadków, tj. 60,5% materiału) nie znalazła odzwierciedlenia w hasłach przedmiotowych. Jednak 64 jednostki występujące w

¹ S. Kurek-Kokocińska, Terminologia z dziedziny oświaty w strukturze mikrotezaurusa. „Zagadnienia Informacji Naukowej” 1999 nr 1 s. 45-49.

² E. Chmielewska-Gorczyca, Funkcje tezaurusa w systemie informacyjno-wyszukiwawczym. „Zagadnienia Informacji Naukowej” 1995 nr 1-2 s. 3-17.

³ E. Chmielewska-Gorczyca, Oświata: mikrotezaurus. Warszawa 1998.

⁴ Norma PN-92/N-09018.

języku deskryptorowym (tj. 39,5% wszystkich deskryptorów) występują też w języku haseł przedmiotowych. Z grupy tej wyodrębnia się stosunkowo niewielka grupa 24 jednostek języka naturalnego (tj. 37,5% wszystkich jednostek, które występują zarówno w języku haseł przedmiotowych, jak i w języku deskryptorowym), które są strukturalnie tożsame w obu językach informacyjno-wyszukiwawczych, np.: **analfabetyzm**, **kształcenie**, **nauczanie**, **nauczyciel**, **oświata**, itp. W pozostałych 41 przypadkach występują różnego rodzaju zmiany. W 10 przykładach (6,17% materiału) występuje różnica w liczbie gramatycznej jednostek leksykalnych. Jednostkom, które jako deskryptory użyte są w liczbie pojedynczej, odpowiadają jednostki, które jako hasła przedmiotowe występują w liczbie mnogiej, np.:

egzamin (deskryptor) – **egzaminy** (hasło przedmiotowe)

żłobek (deskryptor) – **żłobki** (hasło przedmiotowe)

wycieczka szkolna (deskryptor) – **wycieczki szkolne** (hasło przedmiotowe)

W 22 przypadkach (13,5% materiału) występują różnice, wynikające ze specyfiki obu języków informacyjno-wyszukiwawczych. W 10 przypadkach (6,17% wszystkich deskryptorów) są to różnice proste – deskryptorowi złożonemu mającemu formę frazy rzeczownikowej odpowiada hasło przedmiotowe złożone z tematu i określnika, np.:

metody nauczania (deskryptor) – **nauczanie** – **metody** (hasło przedmiotowe)

reforma oświaty (deskryptor) – **oświata** – **reforma** (hasło przedmiotowe)

program nauczania (deskryptor) – **nauczanie** – **program**
(hasło przedmiotowe)

W 12 przypadkach (7,4% materiału) różnice te są bardziej złożone – jednostkom leksykalnym pełniącym funkcję deskryptorów odpowiadają jednostki, które w języku haseł przedmiotowych są bardziej złożone lub mają inną formę (przy zachowaniu tego samego rdzenia), lub deskryptorowi złożonemu odpowiada semantycznie zgodne hasło przedmiotowe, w którym tylko jeden z elementów występuje w języku deskryptorowym, np.:

zwalczanie analfabetyzmu (deskryptor) – **analfabetyzm** – **zapobieganie i zwalczanie** (hasło przedmiotowe)

poradnictwo szkolne (deskryptor) – **szkolnictwo** – **poradnictwo**
(hasło przedmiotowe)

materiały dydaktyczne (deskryptor) – **nauczanie** – **materiały**
(hasło przedmiotowe)

Interesującą grupę stanowią 4 jednostki leksykalne (4,64% materiału), które występują jako deskryptory w języku deskryptorowym, zaś w języku haseł przedmiotowych (po zmianach formy) pełnią funkcję semantycznie tożsamego terminu odrzuconego, np.:

szkolnictwo społeczne (deskryptor) - **szkolnictwo niepubliczne**
(termin odrzucony)

szkolnictwo prywatne (deskryptor) - **szkolnictwo niepubliczne**
(termin odrzucony)

szkolnictwo dla pracujących (deskryptor) – **szkolnictwo dla dorosłych**
(termin odrzucony)

inspekcja szkolna (deskryptor) – szkolnictwo – kontrola (termin odrzucony).

Jak wynika z przedstawionych wyników badań, większość jednostek leksykalnych, które występują w języku deskryptorowym pełnią funkcję deskryptorów, nie występuje w języku haseł przedmiotowych. Jedynie w 64 przypadkach istnieje możliwość formułowania zapytań w bazach danych indeksowanych terminami z jednego języka informacyjno-wyszukiawczego przy pomocy terminów z innego języka. Biorąc pod uwagę dodatkowe zmiany formy tych jednostek (zarówno zmiany proste typu zmiana liczby gramatycznej, jak i zmiany bardziej złożone, które w praktyce mogą uniemożliwić niewprawnemu użytkownikowi posłużenie się danym terminem) należy uznać, że tylko w niewielu przypadkach wyszukiwanie takie jest możliwe do przeprowadzenia.