

Krystyna Dziewańska
Kraków

RODZINA GOOGLE W SŁUŻBIE TŁUMACZA. WYSZUKIWANIE TERMINOLOGII Z POMOCĄ INTERNETU

Zarys treści: Wychodząc od definicji pojęć *terminologia* i *termin*, artykuł omawia możliwości wykorzystania różnych typów narzędzi platformy Google do przeszukiwania zasobów Internetu w celu znajdowania słownictwa specjalistycznego, definicji pojęć oraz tekstów paralelnych. Szczegółowo scharakteryzowano modele i strategie wyszukiwania terminologicznego z zastosowaniem wyszukiwarki ogólnej Google Web Search, katalogu tematycznego Google Directory oraz wyszukiwarek specjalistycznych Google Scholar i Book Search na przykładzie takich dziedzin, jak edukacja, informacja naukowa i bibliotekoznawstwo.

Każde tłumaczenie na inny system językowy, w szczególności zaś tłumaczenie tekstów specjalistycznych, wiąże się ze znajomością specyficznej frazeologii i terminologii, której obecność w języku społeczeństw jest równie stara jak pierwsze nazwy narzędzi i zachowań typowych dla danej zawodowej działalności człowieka. Jednak dopiero stosunkowo niedawno przyspieszenie cywilizacyjne w obrębie nauki, techniki i technologii spowodowało wyłonienie się odrębnej nauki precyzującej zagadnienia przynależne do tej dziedziny ludzkiej aktywności.

W języku różnych badaczy opisujących kwestie związane z nazewnictwem charakterystycznym dla poszczególnych dyscyplin nauki czy działalności profesjonalnej człowieka panuje brak zgody co do ustaleń natury terminologicznej. Ponieważ wielość definicji i punktów widzenia w tej sprawie mogłaby z pewnością wypełnić niejeden artykuł, nas zaś bardziej zajmują kwestie praktyczne, porzeczamy na przytoczeniu dwu propozycji pozwalają-

jących na zarysowanie podstawowych kierunków myślenia o kwestiach terminologii.

Helmut Felber, uczeń jednego z twórców nauk terminologicznych – Eugena Wüster, wymienia trzy różne dziedziny zajmujące się badaniem działalności terminologicznej człowieka: terminoznawstwo, czyli teorię terminologii stanowiącą „pogranicze językoznawstwa, logiki, ontologii, teorii informacji i nauk empirycznych”, terminologię – rozumianą jako „uporządkowany zbiór pojęć określonej specjalności wraz z przypisanymi im znakami pojęć” oraz terminografię, czyli „teorię i praktykę gromadzenia, rejestrowania, magazynowania i prezentacji danych terminograficznych”, przy czym te ostatnie to dane terminologiczne wraz z towarzyszącym im opisem informacyjnym (Felber, Budin 1994: 23–28).

W obszernym rozdziale poświęconym kwestiom związanym z tą dziedziną wiedzy Jerzy Pieńkos stwierdza natomiast, że terminologia może być postrzegana zarówno jako zbiór terminów odnoszący się do danej dyscypliny naukowej lub technicznej, jak i systematyczne badania zmierzające do zbierania i analizowania terminów w kontekście i określania pojęć należących do różnych zakresów działalności zawodowej człowieka, podkreślając przy tym, że w polskiej praktyce pojęcie *terminologia* oznacza wciąż raczej zbiór terminów niż naukę o terminach (1993: 206). Tak zdefiniowanym zakresem pojęcia *terminologii* będziemy posługiwać się w niniejszym artykule.

Kolejnym pojęciem, które należałoby dookreślić, jest *termin*. Według Felbera termin to „znak pojęcia (czyli znak graficzny przyporządkowany na stałe do jednego lub kilku pojęć), składający się z jednego lub więcej ciągów znaków literowych” (Felber, Budin 1994: 26). Oprócz terminów znakami pojęć mogą być symbole cyfrowe (np. 25, XX), znaki abstrakcyjne (np. @, %, *), skróty (np. UKD 621.3 – elektronika, CO), nomeny (nazwy, np. watt jako jednostka miary). Pieńkos podaje znacznie bardziej uogólnioną definicję terminu, pisząc, iż „jest to nazwa mająca ustalone przez naukę znaczenie i dotyczy pojęcia wywodzącego się z określonej dziedziny nauki, techniki lub działalności zawodowej człowieka” (Pieńkos 1993: 209).

Obydwaj badacze zgodnie podkreślają prymat pojęcia wobec terminu, postulat jednoznaczności przyporządkowania terminu do pojęcia oraz odrębność terminów w stosunku do innych wyrazów języka ogólnego (Felber, Budin 1994: 27; Pieńkos 1993: 238). Dla tłumaczy wynika stąd w sposób oczywisty wniosek małej przydatności słowników ogólnych przy przekładzie tekstów fachowych oraz konieczność sięgania do słowników specjalistycznych. Te jednak – ze względu na długi cykl przygotowawczy poprzedzający wydanie – są w momencie wydania w dużej mierze nieaktualne, zwłaszcza w przypadku

dziedzin, w których rozwój technologiczny powoduje szybki przyrost nowego słownictwa. Wyjściem z sytuacji mogą być banki danych terminologicznych (banki terminologiczne), tworzone od lat 60. najpierw w postaci tradycyjnej, później w formacie elektronicznym, a obecnie dostępne w sieci za pośrednictwem stron WWW. Obserwując historię i problemy przeróżnych scentralizowanych inicjatyw w tej dziedzinie, już 20 lat temu Felber postulował, iż „przyszłość należy do małych banków danych terminologicznych w zakresie określonych dziedzin, prowadzonych przez instytuty uniwersyteckie lub inne organizacje zajmujące się daną dziedziną [...]. Wiarygodność danych gwarantuje współpraca specjalistów danej dziedziny. Sieć takich współpracujących ze sobą banków umożliwi użytkownikom dostęp do wiarygodnych danych z różnych dziedzin” (Felber, Budin 1994: 59). Nie wdając się w szerszą dyskusję na ten temat, można zaobserwować, że jest to stosunkowo nowa forma dostępu do terminologii i – jak wszystko w Internecie – podlegająca ustawicznym transformacjom oraz nie zawsze do końca sprawdzona i aktualna, zwłaszcza w przypadku bardzo dużych projektów¹.

Teoria informacji, informacja naukowa, bibliotekoznawstwo i bibliotekarstwo są dziedzinami, w których w chwili obecnej ogromne znaczenie odgrywają technologie informacyjne wykorzystywane w szerokim spektrum zagadnień, jak choćby automatyzacja bibliotek i usług bibliotecznych, w tym katalogów, digitalizacja zbiorów, tworzenie systemów i języków informacyjno-wyszukiwawczych, wykorzystanie Internetu jako nowego medium dla potrzeb informacji i edukacji społecznej oraz próby tworzenia modelu efektywnej organizacji zasobów sieciowych. W tych obszarach następuje wręcz lawinowy przyrost specjalistycznego słownictwa, składającego się w dużej mierze z żargonu komputerowego i internetowego. Ta produkcja quasi-terminotwórcza jest tak obszerna, że nie nadążają za nią nawet anglojęzyczne słowniki terminologiczne dostępne w sieci, nie mówiąc już o uznanych źródłach angielsko-polskich. Co ma w takiej sytuacji zrobić tłumacz pracujący w tej dziedzinie. Nawet jeśli nie zamierza tworzyć neologizmów i chce poprzestać na pozostawieniu terminu w jego brzmieniu oryginalnym, czyli posłużyć się techniką reprodukcji tłumaczeniowej, powinien przynajmniej odnieść czy przyporządkować opisywane zjawisko do bardziej ogólnej, zrozumiałej dla

¹ Przypadek firmowanej przez Komisję Europejską bazy EURODICAUTOM kilkakrotnie zmieniającej internetową lokalizację. Obecnie jest ona dostępna w wersji testowej z dodatkiem języków nowych krajów członkowskich jako Inter Active Terminology pod adresem <http://iate.europa.eu/iatediff/> (odczyt 27.03.2007). Tłumacze różnie oceniają przydatność tej bazy, zob. blog Margaret Marks, wpisy z 25.02.2005, <http://www.margaret-marks.com/Transblaw/archives/001296.html/> (odczyt 27.03.2007).

odbiorców kategorii pojęć. Stosując zaś metodę reprodukcji z objaśnieniem musi mieć wystarczającą wiedzę na temat tego, na czym dane zjawisko polega (Hejrowski 2006: 76–77).

W przypadku poszukiwania najświeższego słownictwa fachowego prócz wspomnianych już banków terminologicznych można posiłkować się przeglądaniem wszelkiego rodzaju słowników czy glosariuszy sieciowych, których adresy dostępne są w witrynach serwisów internetowych dla tłumaczy czy specjalistów z danej dziedziny. Zgromadzone tam słownictwo nie jest rzecz jasna terminologią *sensu stricto*, czyli terminami ściśle zdefiniowanymi. W tym przypadku mówimy o *terminach upowszechnionych*, czyli nazwach które uzyskały status terminu przez upowszechnienie występowania (Pieńkos 1993: 211). Bywa to jednak zajęcie bardzo czasochłonne, zwłaszcza w obliczu niestabilności adresów sieciowych powodującej częste zmiany lokalizacji dokumentów przenoszonych do innych plików lub na inne serwery. W tym momencie z pomocą przychodzi nam sam Internet potraktowany jako gigantyczna baza danych, także terminologicznych, w której tylko trzeba umiejętnie wyszukać potrzebne informacje.

Znawcy Internetu dzielą się na tych, którzy rekomendują równoległe korzystanie z wielu narzędzi wyszukiwawczych, przytaczając dowody w postaci wyników badań świadczących o rozłączności zakresów wyników zwracanych przez poszczególne wyszukiwarki, oraz takich, którzy zachęcają do posługiwania się tylko jednym narzędziem, argumentując, że ważniejsza jest gruntowna znajomość jego możliwości oraz skupienie się na doskonaleniu metod formułowania efektywnych kwerend wyszukiwawczych. Wśród tych ostatnich jest bardzo liczne grono zwolenników Google.com.

Za przyswojeniem sobie arkanów posługiwania się tą wyszukiwarką przemawia kilka jej cech. Jedną z nich jest używany przez Google, pilnie strzeżony i wciąż modyfikowany wskaźnik PageRank, czyli podstawowy algorytm określający pozycję danej witryny w rankingu listy wyników wyszukiwania. Jest on oparty na połączonej analizie liczby odsyłających do danej witryny stron i jakości tychże stron, czyli ich własnej wartości PageRank. Bazujący na tworzonych przez ludzi odnośnikach system pomiaru jakości przypomina zatem stosowaną od dawien dawna metodę analizy cytowań bibliograficznych, a tym samym dobrze współgra z ludzkim odczuciem relewancji. W połączeniu zaś z GoogleBotem, zaawansowanym technologicznie oprogramowaniem do przeszukiwania i indeksowania zawartości dokumentów internetowych, stanowi platformę dla bardzo sprawnego narzędzia wyszukiwawczego.

Do roku 2005 do głównych argumentów przemawiających za używaniem Google należała również ogromna baza zaindeksowanych dokumentów, li-

cząca według ostatniego oficjalnego komunikatu przedstawicieli firmy 1000 razy więcej elementów niż w momencie pojawienia się wyszukiwarki na rynku w roku 1998 (Paterson 2005), czyli około 25 miliardów stron². Ogłoszenie tego komunikatu szło w parze ze zdjęciem ze strony Google informacji na temat rozmiaru jej indeksu, co zamknęło okres wieloletniej rywalizacji w tym względzie z innymi wyszukiwarkami, na czele z Yahoo (Batelle 2005). Argument w postaci liczby zaindeksowanych stron przestał być brany pod uwagę także, a może nawet przede wszystkim, ze względu na fakt nieetycznej działalności redaktorów stron internetowych, inaczej zwanych pozycjonerami (*SEO spammers*), próbujących oszukać roboty indeksujące najważniejszych wyszukiwarek w celu podniesienia notowań własnej bądź zlecającej taką usługę witryny komercyjnej, co skutkowało przede wszystkim lawinowym wzrostem liczby małowartościowych lub nieczynnych odnośników. Dziś problem ten nie tylko nie zniknął, ale wręcz się nasila. W cytowanym w witrynie firmy Webinside wspólnym raporcie Microsoftu i Uniwersytetu Kalifornijskiego szacuje się, że na 100 wyników wyszukiwań co najmniej 12 stanowią zakamuflowane witryny komercyjne³.

Co więc nadal przemawia za posługiwaniem się Google poza względami czysto estetycznymi (prosta, wręcz ascetyczna strona startowa wyposażona, w zależności od nadarzającej się okazji, w przeróżne graficzne wersje nazwy wyszukiwarki, zwane *doodles*)? Ona także. Ale poza tym niespotykana gdzie indziej liczba pomysłowych inicjatyw, serwisów i produktów pozwalających na wykorzystanie wielorakich możliwości Internetu oraz na docieranie do pewnych zasobów sieci niedostępnych dla zwykłych narzędzi wyszukiwawczych, czyli do głębokiego bądź niewidzialnego Internetu. Nade wszystko zaś różnicowane metody konstruowania bardzo precyzyjnych zapytań wykorzystujące operatory czy modyfikatory (zwane zbiorczo składnią specjalną), czyli wyrazy o specjalnym znaczeniu, które zastosowane w wyszukiwaniu modyfikują je, a nawet całkowicie zmieniają jego typ. Posługiwanie się składnią specjalną umożliwia doprecyzowanie i zawężenie wyszukiwania już od samego początku, co w konsekwencji znacznie przyspiesza docieranie do poszukiwanych treści⁴.

² Wyliczenie na stronie <http://www.searchlores.org/google.htm/> (odczyt 04.04.2007).

³ „Najwięcej, bo aż do trzydziestu procent spamu zawartego jest w wynikach wyszukiwania dla haseł związanych z branżą medyczną i telefonią komórkową. Do tworzenia stron spamerskich pozycjonerzy wykorzystują głównie najtańsze dostępne domeny, a więc w zdecydowanej przeważającej części info., biz., a także net.” – Kulbaka 2007.

⁴ Pełna lista operatorów wyszukiwania zaawansowanego znajduje się na następujących stronach: <http://www.google.pl/help/operators.html/> (w języku polskim) oraz <http://www.google.com/help/operators.html/> i <http://www.google.com/help/features.html#definitions/> (w języku angielskim).

Kolej teraz na szczegółowe omówienie wykorzystania różnych strategii wyszukiwawczych oraz typów składni specjalnej kilku narzędzi platformy Google szczególnie przydatnych do wyszukiwania słownictwa specjalistycznego, definicji pojęć oraz tekstów paralelnych, a więc na wyszukiwarkę ogólną Google Web Search, katalog tematyczny Google Directory oraz wyszukiwarki specjalistyczne Google Scholar i Book Search. Przedstawione poniżej wyniki wyszukiwań pochodzą z takich dziedzin, jak edukacja, informacja naukowa i bibliotekoznawstwo, języki zaś, których dotyczą, to przede wszystkim język angielski i polski.

Zacznijmy od Google Web Search, dostępnej na stronie <http://www.google.com/> lub – w wersji polskiej – <http://www.google.pl/>.

Zanim przejdziemy do charakterystyki poszczególnych strategii wyszukiwawczych, warto przypomnieć pewne procedury ułatwiające posługiwanie się tym narzędziem, których opisy są dostępne także na stronach plików pomocowych⁵:

- Google nie rozróżnia wielkości liter i znaków przestankowych oraz pomija w wyszukiwaniu niektóre słowa często występujące w języku angielskim (np. zaimek *I*, przyimki, rodzajniki, zaimki pytające i pojedyncze cyfry). Aby pojawiły się one w wynikach trzeba albo postawić bezpośrednio przed nimi znak „+” (bez spacji), albo ująć wyrażenie w cudzysłów, tworząc frazę wyszukiwawczą;

- W wyrażeniach, mimo istnienia domyślnej koniunkcji między szukanymi słowami, warto także zastosować cudzysłów w przypadku nazw, w których istotna jest kolejność słów oraz fakt, że następują one jedno bezpośrednio po drugim, np. „terminology mining (wystarczający jest cudzysłów na początku frazy);

- Opcja „ustawienia” (*preferences*) daje możliwość wyświetlania większej niż 10 liczby rezultatów na stronie listy wyników wyszukiwania oraz wyświetlanie każdego wyniku w nowym oknie, co skraca czas przeglądania, bo witryny mogą być pobierane z sieci w tym samym czasie;

- Zanim otworzy się daną stronę na liście wyników, warto przeanalizować jej opis (*snippet*), bo już sam adres internetowy witryny daje nam pojęcie o tym, czy może ona zawierać wiarygodne informacje;

- Zawsze warto czytać pliki pomocowe danego narzędzia wyszukiwawczego – czas na to poświęcony zrekompensuje sprawniejsze posługiwanie się tym narzędziem w przyszłości.

⁵ Zob. literatura do niniejszego artykułu.

Sposób 1. Bezpośrednie wyszukiwanie odpowiedników w języku docelowym

Opcja „ustawienia” pozwala na wybór języka stron internetowych, w którym chcemy poszukiwać danego słowa czy wyrażenia. Zaznaczenie w polu „szukaj w innym języku” (*search language*) języka polskiego ogranicza wyszukiwanie do dokumentów w tym języku i powoduje, że wyszukiwarka po wpisaniu terminu w języku angielskim działa jak słownik dwujęzyczny. Analogicznie działa zastosowanie tej metody na stronie „szukania zaawansowanego” (*advanced search*), gdzie język polski należy zaznaczyć w polu „języki” (Eckstein, Sosnowski 2004: 89).

Wyszukiwanie terminu *wildcard* daje już w pierwszej i drugiej pozycji wyniki zawierające w samym opisie polski odpowiednik: znak zastępczy, symbol wieloznaczny, wieloznacznik. Warto nadmienić, że wyniki pochodzą z wiarygodnych źródeł – serwisu terminologii internetowej Digipedia.pl oraz polskiej witryny wolnej encyklopedii Wikipedia (odczyt 21.05.2007).

W ten sam sposób można wyszukiwać polskie odpowiedniki angielskojęzycznych skrótowców i wyrażen. Wpisanie akronimu *SEO* daje na drugiej pozycji link do artykułu z witryny polskiej Wikipedii na ten temat zatytułowany *Pozycjonowanie stron* (odczyt 11.04.2007). Z kolei wyszukując wyrażenie *tag cloud*, już w pierwszym wyniku otrzymujemy polski odpowiednik – chmura znaczników – również pochodzący z witryny polskiej Wikipedii (odczyt 11.04.2007).

Sposób 2. Wyszukiwanie odpowiedników w języku docelowym w specjalistycznych źródłach sieciowych z zastosowaniem operatorów Boole’a

Gdy zwykle wyszukiwanie nie przynosi wiarygodnych i szybko odnalezionych wyników, możemy pokusić się o wpisanie w okienku wyszukiwania nazw specjalistycznych źródeł sieciowych. W naszym przypadku były to sieciowe witryny zawierające encyklopedie, słowniki, glosariusze, m.in. terminologii internetowej i komputerowej, takie jak: Wikipedia, Wolna Encyklopedia, Algorytmy.pl, PC Format, Digipedia.pl, i-sloownik.pl. Słownik slangu informatycznego, Biuro Tłumaczeń Informatycznych BTInfo, International Data Group Poland. Zbudowanie kwerendy wyszukiwawczej wymagało zastosowania logiki boolowskiej. Była to koniunkcja szukanego terminu ze wszystkimi wymienionymi w nawiasie nazwami połączonymi operatorem OR („lub”): [snippet (algorytmy OR wikipedia OR digipedia OR btinfo OR pcformat OR idg OR „i-sloownik.pl”)] (odczyt 20.05.2007).

Jeszcze 2 lata temu skonstruowanie tak rozbudowanego zapytania nie byłoby możliwe ze względu na ograniczenie wyszukiwania do 10 początkowych

słów. Obecnie nie musimy się już obawiać przekroczenia limitu, gdyż od 2005 roku wynosi on nie 10, a 32 wyrazy (Lenssen 2005).

Niestety, w przypadku najnowszego nazewnictwa specjalistycznego najczęściej nie tylko nie udaje nam się znaleźć jego polskich odpowiedników, ale w dokumentach w języku polskim brak nawet jakichkolwiek opisów czy wzmianek na temat danego zjawiska. Wtedy pozostaje wyszukiwanie w zasobach Internetu definicji pojęć w języku oryginału i próby tłumaczenia opisowego. Ogromnym ułatwieniem zadania jest składnia specjalna do wyszukiwania terminologii, którą Google.com skonstruował w duchu swojej misji, określanej jako „uporządkowanie światowych zasobów informacji, aby stały się one powszechnie dostępne i użyteczne”⁶.

Sposób 3. Wyszukiwanie definicji pojęć z użyciem składni specjalnej *define*

Dla zawężenia zakresu wyszukiwania oraz przeprowadzenia wstępnej selekcji jakościowej dokumentów można ograniczyć wyszukiwanie do stron w języku angielskim.

Składnia dostępna jest w dwu wersjach. Po prefiksie *define* można użyć dwukropka i wpisać słowo kluczowe (uwaga na brak spacji przed szukanym słowem). Wtedy wyniki otrzymamy w postaci gotowych definicji znalezionych w Internecie. Wyszukiwanie [*define:pharming*] daje sześć definicji, z których pierwsza, trzecia i piąta odnoszą się do oszukańczego proceduru przekierowania użytkownika usługi internetowej na fałszywą stronę hakerską, gdzie odbywa się wyłudzenie poufnych danych osobowych, jak np. numer identyfikacyjny PESEL czy NIP, numer karty kredytowej czy numer konta bankowego, najczęściej w celu kradzieży pieniędzy z konta (odczyt 21.05.2007).

W innej formie zapisu składni pojawia się spacja przed szukanym słowem. Kwerenda [*define pharming*] przynosi 110 000 wyników, z czego pierwszych pięć podaje w opisie stron WWW definicję pojęcia (odczyt 21.05.2007).

W przypadku, gdy wyszukiwane jest wyrażenie, musi być ono poprzedzone znakiem cudzysłowu, np. [*define:"information retrieval*] lub [*define "information retrieval*].

⁶ Zob. Google. *Informacje o firmie 2007*, <http://www.google.pl/intl/pl/corporate/index.html/> (odczyt 20.05.2007).

Sposób 4. Wyszukiwanie definicji bądź opisu zjawiska z zastosowaniem wyrażenia z czasownikiem

Gdy wyszukiwanie z użyciem składni *define* nie daje rezultatów, można spróbować metody komponowania frazy wyszukiwawczej w oparciu o przewidywanie, jak będzie wyglądał konkretny fragment tekstu na stronie dokumentu internetowego zawierający opis interesującego nas zjawiska. Będzie to najprawdopodobniej wyrażenie z czasownikiem, najczęściej *is*, *are*, *means*, *mean* oraz *stands for* w przypadku skrótowców (Calishan, Dornfest 2003: 76).

Wyszukiwanie opisu zjawiska *podobook* przy pomocy kompozycji fraz [”*podobook is*” OR ”*what is podobook*” OR ”*podobooks are*] przynosi 30 wyników, z których czwarty zawiera opis wpisu w blogu organizacji Christianity 4 Life – można się z niego dowiedzieć, że *podobook* to rodzaj audycji internetowej (*podcast*), w trakcie której autor lub autorka czyta swój tekst (odczyt 14.04.2007).

Google Directory, <http://www.google.com/dirhp>

Kolejnym narzędziem mogącym służyć do poszukiwań terminologicznych jest Google Directory, sieciowy katalog tematyczny. Ze względu na pochodzenie tworzących go dokumentów – selekcyjonowanych i skrótowo charakteryzowanych przez ludzi, a nie programy indeksujące, oraz hierarchiczną budowę przypomina on poważniejsze źródła informacji będące skrzyżowaniem katalogu przedmiotowego i bibliografii adnotowanej.

Google Directory wykorzystuje zbiór dokumentów internetowych zestawiony przez redaktorów Open Directory Project. Stosuje jednak własny mechanizm szeregowania danych w obrębie określonej kategorii. Zawiera około 5 milionów adresów witryn internetowych sprawdzonych pod względem jakości prezentowanych informacji. Tworzą one zbiór 15 głównych kategorii tematycznych podzielony hierarchicznie na kategorie niższego rzędu i podkategorie. Każdej kategorii jest przyporządkowany zestaw korespondujących tematycznie witryn sieciowych z opisami charakteryzującymi ich zawartość.

Katalog tematyczny Google dostarcza gotowego spisu bibliograficznego posegregowanych dziedzinowo źródeł informacji, które można przeglądać według kategorii tematycznych. Z tego też względu nadaje się do szybkiego zorientowania się w terminologii i hierarchii pojęć występujących w danej dziedzinie wiedzy. Dzięki możliwości przeszukiwania z użyciem słów kluczowych może być przydatny do wyszukiwania tekstów paralelnych lub terminów wieloznacznych w stosunkowo niewielkiej (w porównaniu z bazą wyszukiwarki) bazie dokumentów. I tak np. wyszukiwanie występującego w bibliotekar-

stwie terminu *weeding* (selekcja, brakowanie) w podkatalogu LIBRARIES nie powoduje wyświetlenia stron dotyczących ogrodnictwa, które pojawiłyby się przy tym zapytaniu z użyciem wyszukiwarki ogólnej (odczyt 14.04.2007).

Wymienione powyżej cechy sprawiają, że katalog Google spełnia podstawowe kryteria źródła dostępu do informacji o wysokiej jakości. Natomiast rozległość tematyczna bazy danych oraz tryb indeksowania dokumentów poprzez zespół wolontariuszy powodują, że nie jest ona aktualizowana w szybkim tempie i nie obejmuje źródeł najnowszych.

Do znajdowania aktualnych dokumentów o wysokiej jakości merytorycznej można posłużyć się innym narzędziem Google – wyszukiwarką tekstów naukowych Scholar.

Google Scholar, <http://scholar.google.com/>

„Zadaniem tego narzędzia jest wyszukiwanie, sortowanie i udostępnianie użytkownikom wyłącznie materiałów o charakterze naukowym i dydaktycznym. W tym celu opracowany został zupełnie nowy algorytm wyszukiwania, który umożliwia precyzyjną selekcję i analizę znalezionych materiałów. Indeksuje wszelkiego rodzaju publikacje akademickie: książki, artykuły, streszczenia, reprinty czy raporty techniczne z wszystkich dziedzin nauki. Wyspecjalizowane roboty przeszukują zasoby instytutów naukowych, sieci uniwersyteckich i różnorodnych witryn specjalistycznych, tworząc w ten sposób wyselekcjonowaną bazę danych. Ponadto każdy z nich wyposażony został w odpowiedni program umożliwiający szczegółową analizę odnalezionych materiałów. Dzięki temu wszystkie cytaty, przypisy czy jakiegokolwiek informacje bibliograficzne zawarte w danym tekście zostają z niego wyodrębnione i odpowiednio sklasyfikowane” (Weryho 2005).

Przy pomocy tej wyszukiwarki można odnaleźć pełne teksty online dostępne w bazach zwanych repozytoriami preprintowymi, na stronach czasopism w wolnym dostępie oraz na stronach osobistych. W witrynach wydawnictw, towarzystw i komercyjnych instytucji naukowych można uzyskać dostęp do pełnych tekstów po opłaceniu jednorazowej bądź czasowej subskrypcji, natomiast zawsze istnieje dostęp do informacji bibliograficznych oraz – najczęściej – do streszczenia artykułu.

Google Scholar oferuje wyszukiwanie według następujących haseł:

- autor, np. [*”r rubin*] lub [*author:”r rubin*]
- tytuł, np. [*„personel management in libraries*]
- słowa kluczowe lub fraza, np. [*”terminology mining*].

Zakres tematyczny bazy dokumentów obejmuje główne dziedziny nauki z wyróżnieniem siedmiu działów pozwalających na zawężenie pola wyszukiwania.

Opcja *recent articles* daje możliwość uszeregowania listy wyników z uwzględnieniem materiałów najnowszych, z równoczesnym zachowaniem pozycjonowania ich według specyficznego algorytmu Scholar, na który mają wpływ m.in. liczba i jakość cytowań danego tekstu, autorytet naukowy autora, miejsce publikacji oraz zgodność z tematem wyszukiwania. Ponadto każde wyszukiwanie według słów kluczowych wyświetla także dodatkowo „artykuły podobne” (*related articles*), będące tematycznie powiązane z tematem przeszukiwania.

Wyniki wyszukiwania frazy [”*terminology mining*] zawierają odnośniki do witryn, na których można przeczytać streszczenia artykułów, a ich pełne teksty, głównie rozdziały książek, dostępne są online za opłatą lub dla subskrybentów serwisu. W dalszej kolejności jednak pojawiły się linki do pełnych tekstów artykułów pokonferencyjnych dostępnych bez ograniczeń (odczyt 11.04.2007).

Google Scholar z powodzeniem nadaje się do wyszukiwania tekstów modelowych. Dobrze jest zastosować formularz wyszukiwania zaawansowanego, który dla znalezienia modelu do tłumaczenia polskiego artykułu pt. *Komputeryzacja bibliotek naukowych w kontekście standardów oprogramowania bibliotecznego* wyglądał następująco: w polu *with all the words* została wpisana domyślna koniunkcja fraz [”*library automation*” *library software*], w polu *with at least one of the words* wpisano terminy [*academic science university*]. Zakres tematyczny wyszukiwania został ograniczony do nauk społecznych i humanistyki. Wśród 50 początkowych wyników, na 192 wskazanych przez wyszukiwarkę, pojawiło się 10 streszczeń oraz 6 pełnych tekstów artykułów na tak określony temat (odczyt 15.04.2007).

Google Book Search, <http://books.google.com/>

Ostatnim omawianym narzędziem jest Google Book Search, która wyszukuje teksty książek oraz artykułów z niektórych amerykańskich czasopism popularnonaukowych. Baza dokumentów powstała dzięki zakrojonemu na niespotykaną dotąd skalę procesowi digitalizacji zasobów przeprowadzanemu przez Google w ramach programów współpracy z wydawnictwami i największymi bibliotekami świata (obecnie oprócz bibliotek ważniejszych uniwersytetów Stanów Zjednoczonych są to biblioteki uniwersytetów w Oksfordzie i Madrycie oraz centralne biblioteki Katalonii i Bawarii). Teks-

ty są proponowane przez wydawców, bibliotekarzy oraz autorów posiadających prawa autorskie, w przypadku gdy nakład ich dzieł został wyczerpany (Barker 2007).

Wyszukiwarka znajduje następujące rodzaje tekstów:

- pełna wersja tekstowa książek, co do których wygasły prawa autorskie lub za zgodą wydawcy (opcja *full view*),
- obszerne fragmenty tekstu, spis treści i indeks oraz możliwość wyszukiwania w tekście, gdy książki są chronione prawem autorskim (opcja *limited preview*),
- adres bibliograficzny, ewentualnie z opisem treści, gdy kwestia prawa autorskiego jest niejasna (opcje: *no preview available* lub *snippet preview*).

Wyszukiwanie zaawansowane jest możliwe w trzech rodzajach zbiorów:

- wszystkie książki,
- pełne teksty książek,
- katalogi biblioteczne.

Według następujących haseł:

- słowa kluczowe/fraza,
- tytuł,
- autor,
- wydawca,
- data publikacji,
- ISBN.

Przykładowe wyszukiwanie dotyczyło najnowszego wydania poradnika *Google Hacks* omawiającego strategię wyszukiwawcze z zastosowaniem narzędzi Google. Po wypełnieniu pól *author*, *title* i *publication date* otrzymaliśmy wynik w postaci ekranu zawierającego obraz okładki, adres bibliograficzny, spis treści i streszczenie książki (odczyt 15.04.2007).

Dodatkowo na stronie Google Book Search są dostępne następujące opcje wyszukiwarki:

- *search in this book*, która pozwala na wyszukiwanie na stronach konkretnej książki,
- *selected pages*, pozwalająca na oglądania wyszukanej zawartości, gdy ma się zarejestrowane konto (*Google Account*),
- *related books* wyszukuje inne pozycje na ten sam temat,
- *other editions* może – gdy mamy szczęście – zawierać odnośnik do wydania, które jest dostępne w pełnej wersji tekstowej.

Wnioski

W Internecie można znaleźć wiele źródeł terminologicznych w postaci banków terminów, słowników specjalistycznych i glosariuszy. Wielokrotnie jednak okazywało się, że są one niewystarczające, gdy chodzi o najświeższe słownictwo z dynamicznie rozwijających się gałęzi wiedzy, zwłaszcza tych, w których znaczną rolę odgrywają nowoczesne technologie informacyjne. Zaprezentowane w niniejszym artykule badania wykazują, że w szczególności w przypadku terminów upowszechnionych znajdowanie wiarygodnych odpowiedników obcojęzycznych nazw bądź definicji pojęć może odbywać się bezpośrednio w Internecie za pomocą umiejętnie wykorzystanych strategii i narzędzi, jak na przykład omówione powyżej serwisy wyszukiwawcze platformy Google.

Cenne źródło terminologiczne mogą stanowić także artykuły naukowe oraz fragmenty książek wyszukiwane w bazach Google Scholar i Book Search. Ponadto te dwa ostatnie narzędzia szczególnie dobrze sprawdzają się przy wyszukiwaniu relewantnych tekstów paralelnych dla potrzeb tłumaczenia specjalistycznego.

Literatura

- Barker, J., 2007, *Googling to the Max*, <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/Google.html/>.
- Batelle, J., 2005, *Google Announces New Index Size. Shifts Focus from Counting*, <http://battellemedia.com/archives/001889.php/>.
- Calishan, T., Dornfest, R., 2003, *100 sposobów na Google*, Katowice.
- Cohen, L., 2006, *The Deep Web*, <http://www.internettutorials.net/deepweb.html/>.
- Cutts, M., 2006, *Google terminology*, <http://video.google.com/videoplay?docid=8475081922887713591/>, <http://www.viperchill.com/articles/matt-cutts-google-terminology.php/>.
- Eckstein, M., Sosnowski, R., 2004, *Komputer w pracy tłumacza*, Kraków.
- Felber, H., Budin, G., 1994, *Teoria i praktyka terminologii*, Warszawa.
- Grant, J., 2007, *Google Book Search. An Introduction*, http://www.google.com/librariancenter/articles/0606_01.html/.
- Grogg, J. E., Ashmore, B., 2007, *Google Book Search. Libraries and their Digital Copies*, [w:] *Information Today*, http://www.infotoday.com/searcher/apr07/Grogg_Ashmore.shtml/.

- Hejwowski, K., 2006, *Kognitywno-komunikacyjna teoria przekładu*, Warszawa.
- Hughes, T., 2006, *An interview with Anurag Acharya, Google Scholar lead engineer*, http://www.google.com/librariancenter/articles/0612_01.html/.
- Kulbaka, D. 2007, *Seo-spam coraz bardziej widoczny*, [w:] *Webinside*, <http://www.webinside.pl/news/3475/>.
- Lenssen, P. 2005, *Google raises word limit to 32 words*, <http://blog.outer-court.com/archive/2005-01-22-n48.html/>.
- Miller, M., *Searching the Google Directory*, [w:] *Googlopedia: The Ultimate Google Resource*, <http://www.quepublishing.com/articles/article.asp?p=606600&rl=1/>.
- Paterson, A., 2005, *We wanted something special for our birthday...*, http://googleblog.blogspot.com/2005_09_01_googleblog_archive.html/.
- Pieńkos, J., 1993, *Przekład i tłumacz we współczesnym świecie*, Warszawa.
- Pieńkos, J., 2003, *Podstawy przekładoznawstwa. Od teorii do praktyki*, Kraków–Zakamycze.
- Stoberski, Z., 1982, *Międzynarodowa terminologia naukowa*, Warszawa.
- Weryho, M., 2005, *Stań na ramionach gigantów, czyli Google Scholar*, [w:] *Biuletyn EBIB*, <http://ebib.oss.wroc.pl/2005/63/weryho.php/>.

Pliki pomocowe omawianych narzędzi Google

- About Google Scholar*, <http://scholar.google.com/intl/en/scholar/about.html/>.
- Google Book Search Help Center*, <http://books.google.com/support/>.
- Google Web Directory Help*, <http://www.google.com/dirhelp.html/>.
- Google Web Search Help Center*, <http://www.google.com/support/?ctx=web/>.

Sieciowe witryny zawierające encyklopedie, słowniki, glosariusze, m.in. terminologii internetowej i komputerowej

- Algorytmy.pl, <http://algorytmy.pl/>.
- Biuro tłumaczeń Informatycznych, <http://www.btinfo.pl/>.
- Digipedia.pl, <http://www.digipedia.pl/>.
- International Data Group Poland, <http://www.idg.pl/>.
- PC Format, <http://www.pcformat.pl/>.
- i-sloownik.pl, Słownik slangu informatycznego. <http://www.i-sloownik.pl/>.
- Wikipedia. Wolna Encyklopedia, <http://pl.wikipedia.org/>.

Google platform in service for translators. Terminology search in the Internet (summary)

Any specialised translation between languages requires knowledge of specific terminology. A variety of terminology sources can be found on the Internet, namely term banks, specialised dictionaries and glossaries. However, on many occasions they seem to be insufficient when it comes to the latest vocabulary related with the domains which utilize modern information technologies. The research shows that efficient use of search tools retrieves relevant equivalents of, at least, popular terms in the target language or definitions of concepts (when equivalents are not available) directly from the Internet viewed as an infinite term database itself.

Beginning with the definition of terminology, the paper discusses the use of several Google search tools for retrieving specialised vocabulary, definitions of concepts and parallel texts. By applying examples from such fields as library and information science, it gives a detailed characteristics of various terminology search strategies with the use of a general search engine Google Web Search, a web directory Google Directory and specialised search engines Google Scholar and Book Search. The following methods are discussed: using a search engine as a bilingual dictionary, search operator *define*: phrase searches with Boolean operators or a verb, searching and browsing a web directory, advanced search for the retrieval of parallel texts and various search options. The results show that all the examined tools prove to be useful in terminology search for specialised translation.