STOWARZYSZENIE BIBLIOTEKARZY POLSKICH

INSTYTUT INFORMACJI NAUKOWEJ
I STUDIÓW BIBLIOLOGICZNYCH
UNIWERSYTETU WARSZAWSKIEGO

ZAGADNIENIA INFORMACJI NAUKOWEJ

WYDAWNICTWO SBP

WARSZAWA 2010 NR 2 (96)

STOWARZYSZENIE BIBLIOTEKARZY POLSKICH

INSTYTUT INFORMACJI NAUKOWEJ
I STUDIÓW BIBLIOLOGICZNYCH
UNIWERSYTETU WARSZAWSKIEGO

ZAGADNIENIA INFORMACJI NAUKOWEJ

WYDAWNICTWO SBP

WARSZAWA 2010 NR 2 (96)

PL ISSN 0324-8194

WYDAWNICTWO SBP

Dyrektor
Janusz NOWICKI

Zawartość tego czasopisma jest dokumentowana m.in. w „Library and Information Science Abstracts" oraz „Knowledge Organization"

Adres Wydawnictwa
ul. Konopczyńskiego 5/7
00-335 Warszawa, tel. 22 827-52-96

PRENUMERATA I SPRZEDAŻ
Dział Promocji i Kolportażu SBP
Al. Niepodległości 213, 02-086 Warszawa, tel. 22 825-50-24

Wydawnictwo SBP – Warszawa 2010. Nakład 400 egz.
Ark. wyd. 6,0. Ark. druk. 7,0
Skład i łamanie: Renard Hawryszko
Druk i oprawa: Zakład Poligraficzny PRIMUM s.c., Kozerki,
ul. Marsa 20, 05-825 Grodzisk Mazowiecki

KRYZYS WARTOŚCI WIEDZY?

Jadwiga Woźniak-Kasperek
Instytut Informacji Naukowej
i Studiów Bibliologicznych
Uniwersytetu Warszawskiego

Informacja, prawda, system aksjologiczny, wartości, wiedza jako wartość, zarządzanie wiedzą

Celem artykułu[1] jest zwrócenie uwagi na znaczenie wartości, zwłaszcza wartości wiedzy, w życiu społeczeństw oraz związków społecznie akceptowanego systemu aksjologicznego z nauką i wiedzą. Jedną z inspiracji do podjęcia tej problematyki jest obserwowana od jakiegoś czasu tendencja do zbyt, moim zdaniem, swobodnego posługiwania się określeniem wiedza, traktowania go jako synonimu informacji (niekiedy nawet danych) w sytuacjach, gdy takie zrównanie semantyczne nie tylko nie jest uprawnione, ale wręcz błędne. Obserwowane zjawiska terminologiczne w pewnych przypadkach są symptomem nie tylko (lub nie tyle) niefrasobliwości, co głębszych zmian dotykających przede wszystkim naszej relacji do prawdy oraz języka i metody naukowej.

Informacja

„Pojęcie informacji to jest prapojęcie, pojęcie elementarne, wyjściowe, służące między innymi do definiowania innych pojęć, dlatego samo definiuje się źle"[2]. W rozważaniach o informacji stosuje się różne definicje i wyjaśnienia znaczenia tego wyrazu (terminu), różne podejścia i perspektywy badawcze, przyczyniające się do lepszego rozpoznania, opisu i stanu wiedzy o informacji. Można powiedzieć, że informacja jest kategorią transdyscyplinarną, której zastosowanie rozciąga się od poziomu subatomowego (informacja kwantowa)

[1] Szkic artykułu został przedstawiony na IV Konferencji Biblioteki Politechniki Łódzkiej *Biblioteka w kryzysie czy kryzys w bibliotece?* 15-17 czerwca 2010 r., gdzie spotkał się z dużym zainteresowaniem i żywą dyskusją.

[2] J. Wojciechowski: *W kręgu informacji i nieinformacji*. „Bibliotekarz" 1998, nr 4, s. 2.

i molekularnego (informacja genetyczna) po poziom społecznych procesów poznawczych (polegających na odbieraniu i przetwarzaniu informacji) i komunikacyjnych. „O informacji mówi się dziś w naukach przyrodniczych i społecznych, a także w humanistyce i filozofii. Dzięki ujęciu aspektu informacyjnego możliwe stało się względnie jednolite modelowanie procesów należących do różnych poziomów organizacji świata. Entuzjaści oczekują, a jest ich wielu, że przyszła teoria informacji stanie się źródłem kolejnej rewolucji naukowej (...). Autorzy o bardziej krytycznym nastawieniu zwracają uwagę, że wszechobecna moda na *informationese* (żargon informacyjny) nie przyniosła, jak dotychczas, oczekiwanego sukcesu eksplanacyjnego. Zdaniem K. Devlina (...) rozwój wyrafinowanych technologii informacyjnych nie idzie w parze z konstrukcją zunifikowanej teorii informacji. Pod tym względem przypominamy ludzi epoki żelaza, którzy umieli wytwarzać narzędzia z żelaza, lecz byli bezradni, jeśli chodzi o udzielenie odpowiedzi na pytanie: czym jest żelazo? Ich bezradność wynikała z braku odpowiedniej teorii na temat atomowej struktury materii. Nasza sytuacja jest pod pewnym względem podobna: mamy dostęp do olbrzymiej ilości informacji, które umiemy gromadzić, przechowywać, przetwarzać i przesyłać, lecz pytanie: Czym jest informacja? – wciąż nas przerasta"[3].

Wielość i różnorodność typów desygnatów terminu informacja oraz sposobów podejścia do nich czyni obecnie niemożliwym skonstruowanie jednej definicji i teorii informacji. Być może na tym etapie rozwoju jesteśmy w stanie stworzyć jedynie definicje i uogólnienia definicyjne dla różnych aspektów, typów, okoliczności przejawiania się i istnienia informacji. Również w nauce o informacji nadal nie ma całkowitej zgody co do tego, czym jest informacja[4], choć potrafimy wskazać pewne jej własności, atrybuty i cechy pożądane, czasami zmierzyć ilość, rzadziej jakość.

Za ojca refleksji teoretycznej na temat informacji uważa się Claude'a E. Shannona, który prawdopodobnie w 1945 r. po raz pierwszy użył tego terminu w pracy *A Mathematical Theory of Cryptography*. Natomiast w 1948 r. w kolejnej pracy *A Mathematical Theory of Communication* przedstawił najważniejsze zagadnienia teorii informacji. Shannon stworzył podstawy ilościowej teorii informacji, późniejsi badacze próbowali stworzyć teorie wyjaśniające wartość, cenność czy jakość informacji. W Polsce autorem oryginalnej teorii opisującej zarówno ilość jak i jakość informacji jest Marian Mazur (*Jakościowa teoria informacji* 1970), który wprowadził rozróżnienie między informacjami opisującymi a informacjami identyfikującymi i wykazał, że tylko ilość informacji identyfikujących jest tym samym, co ilość informacji wyrażona wzorem Shannona – wbrew panującemu wcześniej przeświadczeniu, że odnosi się on do wszelkich informacji.

Imperatyw informacyjny – przymus[5] zdobywania i obcowania z informacją – stał się jednym z wyznaczników czasów, w których żyjemy. Wyrazem tej sytuacji

[3] R. Poczobut: *Fenomen wielowymiarowości umysłu a emergencja. Z ontologii i metodologii badań inter- i transdyscyplinarnych*.W: *Między unifikacją a dezintegracją: kondycja wiedzy we współczesnym społeczeństwie*. Pod red. A. Jabłońskiego i M. Zemły. Lublin, 2008, s. 12-13.

[4] Por. T. Saracevič: *Information Science*. „Journal of the American Society for Information Science" 1999, vol. 50, no. 12, p. 1054.

[5] W patologicznym wymiarze może on przybrać postać uzależnienia od informacji.

jest między innymi wielość i częstość posługiwania się takimi określeniami jak cywilizacja informacyjna, społeczeństwo informacyjne, środowisko informacyjne czy infosfera. Imperatywowi informacyjnemu towarzyszy imperatyw wysokiego tempa. Jeśli założyć, że istnieje coś takiego jak duch czasu, to duch naszych czasów, niedoskonale widziany w trakcie trwania „epoki", bez wątpienia charakteryzuje się takimi cechami jak imperatyw informacji i imperatyw czasu, a dokładniej szybkiego tempa. I choć imperatywy te ujawniają się z różnym natężeniem w różnych miejscach i sferach życia, to nie ma wątpliwości, że wpływają na nasze obyczaje, sposób zachowania, komunikowanie się z innymi, procedury itp. Według mnie są jeszcze dwa kolejne wyznaczniki ducha naszych czasów: sieć (i będąca jej efektem sieciowość środowiska informacyjnego) oraz marginalizacja znaczenia i wspólnotowo-kulturowego charakteru systemu aksjologicznego.

Nowe media, do których ekspansji walnie przyczynia się sieć, obok wielu pozytywnych rezultatów prowadzą również do zjawiska, które można nazwać inwersją kultury, czyli odwróceniem się od człowieka i wartości, a zwróceniem ku narzędziom, technikom i środkom. Za niebezpieczne i niepokojące uważam nieuwzględnianie lub zbyt rzadkie uwzględnianie w obszarze rozważań dotyczących relacji techniki i kultury udziału i roli etyki oraz aksjologii. Użytek, jaki czynimy z nowoczesnych technik i technologii informacyjnych i komunikacyjnych, tylko w części zależy od nas samych - w części jest zdeterminowany przez naturę samej techniki. Wysoce niebezpieczne jest oddawanie tej części, na którą możemy mieć wpływ, we władanie narzędzi. „Wielu poważnych badaczy – między innymi Manuel Castells – jest zdania, że wszystkie systemy kulturowej komunikacji ulegają marginalizacji za przyczyną ekspansji przekazu medialnego. Ma to w szczególności dotyczyć układu kultury, który polska socjolog Antonina Kłoskowska nazywa pierwotnym, a który dotyczy procesu socjalizacji za pośrednictwem kontaktów *face to face*"[6].

Amerykański socjolog C. Wright Mills, autor między innymi pojęcia świadomości socjologicznej, trafnie zauważył, że ani baza materialna nie określa bezpośrednio świadomości[7] ludzi, ani świadomość ludzka nie kształtuje bezpośrednio bytu materialnego. Między świadomością a bytem sytuuje się informacja, która wpływa na uświadomienie ludziom ich własnego bytu (jego niedostatków i możliwości). Kolejnym czynnikiem jest komunikacja[8]. To dzięki komunikacji świadomość ludzka kształtuje byt. Bez komunikacji nie byłoby możliwe tworzenie kultury, w tym wiedzy, uczestnictwo w niej i przekazywanie jej kolejnym generacjom. Cywilizacja istnieje między innymi dzięki procesowi komunikowania. Pojęcia komunikowania i kultury są nierozłączne: kultura jest komunikacją, a komunikacja jest kulturą – jak stwierdził Edward T. Hall

---

[6] K. Krzysztofek: *Jaka polityka kulturalna w epoce globalizacji i mediów elektronicznych?* „Kultura współczesna" 2005, nr 1, s. 9.

[7] Ciekawym przykładem publikacji na ten temat jest książka A. Scotta: *Schody do umysłu: nowa kontrowersyjna wiedza o świadomości.* Warszawa 1999. Na gruncie bibliologii na uwagę zasługuje tekst A.Radwańskiego: *Potrzeba rewizji podstaw dyscyplin bibliotekoznawczych.* „Roczniki Biblioteczne" 2000, R. 44, s. 207-215.

[8] Bardzo ciekawe i wartościowe rozważania na temat zjawisk komunikacyjnych w relacji do bibliotek i bibliotekarstwa można znaleźć w najnowszej książce J. Wojciechowskiego: *Biblioteka w komunikacji publicznej.* Warszawa 2010.

w *Bezgłośnym języku*. Wymiana informacji, wzajemne porozumiewanie się, komunikowanie jest jedną z podstawowych funkcji człowieka i w ogóle istot żywych. Bez wymiany informacji nie mogłyby powstać wspólnoty społeczne, cywilizacje, kultury, nauka, nic, co jest atrybutem ludzkości.

Żeby żyć racjonalnie, osiągać cele życiowe trzeba posiadać informacje, odpowiednie informacje. Nie dużo informacji czy wszelkie informacje, ale właśnie odpowiednie. Tymczasem często ten aspekt ginie z pola uwagi i z rozsądku. „I chociaż tylko część informacji krążących w informacyjnym środowisku może przydać się do realizacji osobistych bądź grupowych celów, to właśnie nieokreślona bliżej perspektywa potencjalnego ich zastosowania tworzy ów społeczny fantom, który dziś wspiera postnowoczesny mit o potędze informacji w społeczeństwie wiedzy i w społeczeństwie ryzyka. Uleganie fetyszowi informacji ogranicza wolność, wywołuje negatywne emocje (strach lub przesadny zachwyt), a także skłania do nadmiernego gromadzenia, „na wszelki wypadek" informacji, które mają zapewnić sukces lub ochronę w sytuacji zagrożenia"[9].

Nauka o informacji jest tylko jedną z dyscyplin naukowych, które zajmują się informacją, jej aspektami, miarami, cechami itp. Infospecjaliści odszukują informacje, selekcjonują je, oceniają, przetwarzają, „opakowują" itd. w celu nadania informacji cechy odpowiedniości. Czynią to między innymi dlatego, że mają odpowiedni potencjał:

- „doświadczenie w „radzeniu sobie" z informacją naukową: jej filtrowaniem; oceną wiarygodności, istotności, aktualności; przystępnym i zindywidualizowanym przedstawieniem,
- znajomość użytkownika, jego potrzeb realizowane przez bezpośrednie kontakty,
- zaplecze techniczne (systemy informatyczne, sieci komputerowe, serwery, dostęp do Internetu) obecne w zdecydowanej większości dużych bibliotek, zwłaszcza akademickich,
- bazy katalogowe, biblioteki cyfrowe, specjalistyczne bazy informacyjne,
- fizyczna „przestrzeń" umożliwiająca kontakty międzyludzkie"[10].

Informację można scharakteryzować za pomocą własności, czyli stałych cech jakościowych, niestopniowalnych i niezależnych od użytkownika, oraz cech pożądanych, zależnych od interpretacji i oceny użytkownika. Najważniejszymi własnościami i cechami pożądanymi są:

- Znaczenie. Treść – rozumiane jako odniesienie informacji do jej przedmiotu.
- Relewantność. Pertynentność. Użyteczność – informacja jest interpretowana i oceniana przez pryzmat zróżnicowanych potrzeb, zainteresowań, zadań, aktualnego stanu wiedzy odbiorcy itd.
- Aktualność.
- Wiarygodność.

[9] B. Kamińska-Czubała: *Informacja jako fetysz w społeczeństwie wiedzy i w społeczeństwie ryzyka*. W: *Edukacja w społeczeństwie „ryzyka": bezpieczeństwo jako wartość*. Pod red. M. Gwoździckiej-Piotrowskiej, J. Wołejszo i A. Zduniaka. Poznań 2007, s. 92.

[10] P. Szefliński: *Społeczeństwo informacyjne – o czym biblioteka XXI w. powinna wiedzieć?* W: *II Konferencja Biblioteki Politechniki Łódzkiej: Biblioteki XXI w. Czy przetrwamy? Łódź, 19-21 czerwca 2006 r. Materiały konferencyjne*. Łódź 2006, s. 40.

- Prawdziwość (brak przekłamań, fałszu, zatajenia, manipulacji, zgodność ze stanem tego wycinka rzeczywistości pozatekstowej, którą informacja odwzorowuje).
- Obiektywność – rozumiana tu jako wolność od subiektywizmu w przekazie, odbiorze, interpretacji itp.
- Kompletność. Pełność – żadnej informacji jednostkowej lub nawet wyselekcjonowanego podzbioru nie można traktować jako wyczerpującej charakterystyki obiektu. W praktyce stworzenie pełnego obrazu informacyjnego obiektu nie jest możliwe ze względu na nieograniczoną różnorodność jego charakterystyk. W tym kontekście kompletny oznacza zatem wystarczający, by móc przetworzyć informację w wiedzę; poziom szczegółowości informacji jest zależny od potrzeb odbiorcy.
- Dokładność – rozumiana jako zgodność z poziomem szczegółowości informacji oczekiwanym przez odbiorcę.
- Dostępność – wyraża się między innymi w prostocie formalności, szybkości uzyskiwania, braku utrudnień, łatwości łączenia informacji w odpowiednim czasie, miejscu, z odpowiednimi osobami.
- Spójność – poszczególne elementy, dane współgrają ze sobą; forma odpowiada treści, aktualizacja danych jest zgodna z celami itp.
- Odpowiedniość – zwana również adekwatnością lub przyswajalnością, odnosi informację do poziomu wiedzy i kompetencji, np. językowej, odbiorcy.
- Przystawalność – informacja jest zgodna z inną informacją, interpretowana we właściwym kontekście.
- Redundantność – traktowana również jako wada informacji; zbyt często zapomina się jednak, że bez odpowiedniego poziomu redundancji informacji nie byłoby możliwe realizowanie wielu operacji przetwarzania informacji, np. streszczania.
- Przetwarzalność. Transferowalność – z perspektywy między innymi nauki o informacji ta własność informacji jest szczególnie ważna, umożliwia bowiem procesy przetwarzania informacji (streszczania, kondensowania, selekcjonowania, interpretowania itd.) oraz przenoszenia jej w czasie i przestrzeni.
- Zgodność z zasadami etycznymi.

Wartości w życiu człowieka i społeczeństw

Problematyka wartości jest stałym przedmiotem badań i refleksji filozoficznej począwszy od Sokratesa, Platona i Arystotelesa. Naukowej analizy istoty wartości w ujęciu filozoficznym jako pierwszy dokonał na początku drugiej połowy XIX w. Rudolf Lotze. Dowodził on, że normy etyczne reprezentują określone dobra, czyli coś, co posiada wartość. O tym, jak ceniona jest dana wartość, przesądza to, jaką rangę nadają jej ludzie. W psychologii kategoria wartości jest rozpatrywana ze względu na jej znaczenie w życiu psychicznym człowieka. W socjologii szczególną wagę przypisuje się regulacyjnym funkcjom wartości w życiu społecznym. W antropologii kulturowej wartość widziana jest w kontekście tworzenia, upowszechniania i trwałości dóbr kultury. Nadal najważniejszą i w pewnym sensie regulującą rolę odgrywa aksjologia filozoficzna, która podkre-

śla, że „nasz świat jest światem wartości"[11]. Do wartości najwyższej rangi, tzw. transcendentnych, należą dobro, piękno i prawda. Dziś bardzo jest potrzebna głęboka refleksja i namysł nad wartościami ze względu na coraz powszechniejsze zjawiska impasu i zagubienia człowieka w świecie i społeczeństwie.

Werner Heisenberg napisał, że wartości są kompasem, według którego mamy się orientować, dokonując wyborów w życiu, szukając drogi życiowej[12]. Człowiek żyje wartościami i wartościuje, choć nie ma powszechnej zgody ani co do tego, co to jest wartość, ani co do tego, jaka jest ontologia wartości. Pragmatyzm[13] i koncentrowanie uwagi na wartościach ekonomicznych prowadzą do zawężania horyzontu intelektualnego, do usuwania lub marginalizowania wartości, które na ogół powszechnie nazywa się wyższymi. Czasami słyszymy, a niektórzy z nas to mówią, że dominacja pragmatyzmu eliminuje wartości z naszego życia lub co najmniej odsuwa je na drugi plan. Wobec wizji pustyni aksjologicznej lub jej antytezy w postaci dżungli antywartości rodzą się liczne pytania, między innymi o dzisiejszy system wartości i przekonań oraz o to, czy można obecnie zachować utrwaloną w kulturze aksjologię.

Po Starożytności, załamaniu się wzorców Średniowiecza i Nowożytności łączącej Odrodzenie z Oświeceniem znaleźliśmy się na czwartym etapie rozwoju, kształtowanym przez konsumpcję i medializację. Współczesny człowiek, w wymiarze, którego nie przewidział Wittgenstein (z biegiem lat w znacznym stopniu odchodzący od swych wczesnych poglądów na naturę języka), prowadzi grę językową ze środowiskiem, w którym przyszło mu żyć. Podstawą tej gry jest często zanegowanie istnienia prawdy, a towarzyszy jej akcentowanie roli rozczarowań oraz pustki duchowej i aksjologicznej.

W Europie od lat co najmniej siedemdziesiątych XX w. trwa ożywiona dyskusja znana pod nazwą postmodernizmu lub ponowoczesności. W Polsce ze zrozumiałych powodów śmielej i powszechniej zaczęła się ujawniać dopiero po roku 1989. Najważniejszym znakiem postmodernizmu wydaje się być radykalny sposób myślenia i analizowania kultury po upadku wielkich ideologii (zwłaszcza oświeceniowego racjonalizmu), zwanych tutaj narracjami. Według koncepcji postmodernistycznych prawda nie istnieje, etyka zasługuje na destrukcję, nie istnieją autorytety, w zamian coraz więcej jest relatywizmu[14]. W kontekście tytułowych rozważań szczególnie istotny jest relatywizm poznawczy i aksjologiczny. Dziś stosunkowo często mamy do czynienia z pragmatyzacją prawdy. Wiedza, jak i prawda, coraz częściej są wynikiem umowy społecznej.

„Relatywizacja prawdy i obalenie tzw. mitu obiektywizmu spowodowały wprowadzenie na równych prawach do nauki, której poszczególne dyscypliny traktowane są zresztą jako odrębne dyskursy, gry językowe bez prawa wyja-

---

[11] J. Tischner: *Myślenie według wartości.* Kraków 1982, s. 483.

[12] W. Heisenberg: *Część i całość. Rozmowy o fizyce atomu.* Warszawa 1987.

[13] Pragmatyzm to system lub postawa filozoficzna, których podstawowym elementem jest pragmatyczna teoria prawdy, uzależniająca prawdziwość tez od praktycznych skutków, przyjmująca praktyczność za kryterium prawdy. Inaczej mówiąc, pragmatyzm przyjmuje wynikające z tez skutki i ich użyteczność za kryterium prawdy. Potoczne rozumienie pragmatyzmu określa go jako postawę polegającą na realistycznej ocenie rzeczywistości, liczeniu się z konkretnymi możliwościami i podejmowaniu działań, które gwarantują skuteczność.

[14] J. A. Majcherek: *Źródła relatywizmu w nauce i kulturze XX w. Od teorii względności do postmodernizmu.* Kraków 2004.

śniania ostatecznego i bez możliwości uogólnień, nowych metod i nowych procedur poznawczych traktowanych do tej pory jako nienaukowe i odrzucanych. Badania naukowe dają (...) wiedzę (...) uzyskiwaną na drodze współdziałania, współwystępowania, swoistej mozaiki różnych paradygmatów i dyskursów. (...) Nie ma też podstaw do jakiejkolwiek weryfikacji wiedzy. Zresztą zmienił się status samej wiedzy, która w wielu opracowaniach traktowana jest jako czysta informacja"[15]. Do myśli zawartej w ostatnim z przytoczonych zdań wrócę w dalszej części rozważań, utożsamianie informacji i wiedzy jest bowiem we współczesnych publikacjach z zakresu nauki o informacji i bibliotekoznawstwa stosunkowo częste, a tym niebezpieczniejsze, że w wielu przypadkach chyba czynione nieświadomie (chciałoby się rzec: z niewiedzy).

„Wiedza i teorie naukowe w postmodernizmie stanowią przejaw dążenia do kontrolowania granic danego wymiaru rzeczywistości, a poznanie różnych wymiarów rzeczywistości, w danym miejscu i czasie, jest uzasadnione tylko w danym kontekście. Stąd teorie społeczne nie odzwierciedlają jednej jedynej prawdy o świecie, ale są formami jego kształtowania, konstruowania przez specyficzne działania społeczne. (...) Teoretycy postmodernizmu odrzucają dążenie do poszukiwania ogólnych i absolutnych praw, gdyż takie skażone są zawsze narzucaniem interesów, władzy i punktu widzenia ich twórców, «terroryzują rzeczywistość», przez co nie są wiarygodne. Wola osiągnięcia prawdy jest wolą zdobycia władzy. (...) Zdaniem Zygmunta Baumana, [postmodernizmowi – przypis JWK] brakuje spójności szczególnie w aksjologii, a główną wartością stał się towar, pieniądz, pluralizm, chaos, tolerancja, konsumpcjonizm i liberalizm obyczajów. Z. Bauman podkreśla, iż w postmodernizmie dozwolone są różne style i konwencje zachowań ludzkich oraz synkrazja znaczeń różnych kultur. Pojawia się chaos kultury duchowej, a wartości stały się niestabilne stwarzając poczucie tymczasowości i ciągłej zmiany"[16].

W zacieraniu różnic między nauką, teorią naukową a anegdotą, narracją pragmatyczną tkwi między innymi źródło negowania czy wręcz wyśmiewania tych badań naukowych, które nie mają natychmiastowego i bezpośredniego przełożenia na praktykę[17]. Tymczasem, jak napisał Tadeusz Kotarbiński, „(...) rozsądny umiar każe nam zauważyć fundamentalną zależność metod projektowania nowych urządzeń[18] od zastosowania metod czysto poznawczych"[19].

Wartość zawsze łączy się z dobrem i złem (antywartość). Wartościowanie to myślenie, odczuwanie, mówienie, że coś ma jakąś wartość, pozytywną lub negatywną[20]. Inaczej mówiąc, wartościowanie jest interpretowaniem rzeczywi-

---

[15] K. Ożog: *O prymacie wolności nad prawdą – wartościowanie w dobie postmodernistycznej*. W: *Człowiek wobec wyzwań współczesności: upadek wartości czy walka o wartość?* Pod red. J. Mazura, A. Małyski, K. Sobstel. Lublin 2007, s. 52.

[16] D. Wiśniewska: *Postmodernizm w procesie edukacji akademickiej*. W: *Edukacja w społeczeństwie „ryzyka": bezpieczeństwo jako wartość*. Pod red. M. Gwoździckiej-Piotrowskiej, J. Wołejszo i A. Zduniaka. Poznań 2007, s. 382.

[17] Pomijam tu (bez względu na jego rozmiary) negatywne zjawisko tzw. pseudonauki, która nie tylko nie może, ale wręcz nie powinna mieć wpływu na praktykę.

[18] Urządzeń rozumianych w najszerszym sensie – przypis JWK.

[19] T. Kotarbiński: *Metodologia umiejętności praktycznych: pojęcia i zagadnienia*. „Projektowanie i Systemy" T. 4, 1983, s. 20.

[20] W komunikacji potocznej wartość kojarzona jest niemal zawsze z wartością pozytywną.

stości i jej elementów jako dobrych lub złych pod jakimś względem. Ta interpretacja, pełniąca bardzo ważną rolę w życiu jednostek i społeczeństw, nasuwa jednocześnie wiele problemów i pytań.

W literaturze można znaleźć wiele różnych typologii wartości. Zwykle jako wartości wyższego rzędu kwalifikuje się wartości sakralne (in. transcendentne), poznawcze, moralne i estetyczne. Na drugim planie umiejscawia się wartości witalne, obyczajowe i hedonistyczne. Nawet z tego okrojonego wyliczenia widać, że wartości mogą być ujmowane zarówno z punktu widzenia praw człowieka, jak i jego postaw wobec siebie samego, innych ludzi, otaczającej rzeczywistości, Boga (dla ludzi wierzących). Drugi podział jest nakierowany przede wszystkim na obowiązki człowieka związane z moralnością, ochroną życia, ale również ukazuje to, co wspiera człowieka w wypełnianiu jego zadań, a zarazem może stanowić cel działania, między innymi prawdę, której ludzie całe życie szukają i częściowo znajdują dzięki możliwościom poznawczym, oraz piękno. „Chwalebna jest walka o prawa człowieka, np. wolność i demokrację, ale jeśli nie towarzyszy jej usilne staranie o jakość ludzkich postaw moralnych, to i wolność, i demokracja okazują się (jak dobrze wiemy) rozczarowujące, podlegają patologiom i demoralizacji. (...) Bardzo ważne dla całej problematyki aksjologicznej są różnice w traktowaniu wartości ze względu na samą ich istotę jako absolutnych, powszechnych lub też przekonaniowych. Wartości absolutnych bronią przede wszystkim (choć nie tylko) filozofowie chrześcijańscy, dla których wiążą się one z wiarą w Boga i objawienie. Powszechności pewnych wartości broni również część uczonych niezwiązanych z chrześcijaństwem, na zasadzie aksjomatycznego uznawania lub tzw. kontraktyzmu. Współcześnie przeważa jednak relatywistyczny sposób traktowania wartości jako sprawy ludzkich przekonań, po części upowszechnionych w określonych grupach społecznych i poszczególnych kulturach"[21].

„Pod wpływem dyskursu filozofów XIX i XX w. wokół natury wartości oraz sposobu ich istnienia i poznawalności, wyłaniają się trzy charakterystyczne stanowiska: naturalizm, intuicjonizm i emotywizm aksjologiczny. Dla naturalizmu aksjologicznego rzecz jest dlatego wartościowa, bo jest pożądana, a nie zaś dlatego pożądana, że jest wartościowa. W naturalizmie pojawia się, zdaniem krytyków, błąd polegający na utożsamianiu wartości z niektórymi jej przejawami empirycznymi. Poglądy naturalizmu odrzucają przedstawiciele intuicjonizmu i emotywizmu aksjologicznego. Dla intuicjonizmu wartość jest prostą i nieempiryczną cechą przedmiotu, ujmowaną wprost aktem intuicji. Emotywizm eksponuje niedefiniowalność terminów wartościujących, ponieważ brak jest dla nich empirycznych desygnatów. Zatem wartości nie istnieją i nie ma problemu ich poznawalności. Terminy i sądy wartościujące służą jedynie do wyrażania postaw emocjonalnych oraz wzbudzania ich u innych. (...) wszyscy jednak są zgodni, że (...) istnieje głęboka potrzeba namysłu nad wartościami"[22].

---

[21] J. Puzynina: *Co znaczy „walka o wartości"?* W: *Człowiek wobec wyzwań współczesności: upadek wartości czy walka o wartość?* Pod red. J. Mazura, A. Małyski, K. Sobstel. Lublin 2007, s. 28.

[22] J. Fiedorczuk: *Filozoficzne spojrzenie na pojęcie wartości.* [online]. [dostęp: 17.07. 2010]. Dostępny w World Wide Web: <http://www.eid.edu.pl/publikacje/filozoficzne_spojrzenie_na_pojecie_wartosci,202.html>.

Wiedza jako wartość

Kryzys pewności, prawdy i wiedzy dotyka ludzi i organizacje społeczne, w tym również biblioteki i inne instytucje informacji oraz, a właściwie przede wszystkim, pracujących w nich ludzi. Jednym z łatwiej dostrzegalnych wyrazów owego kryzysu jest dewaluacja pewnych kategorii pojęciowych i aksjologicznych, w tym kategorii wiedzy. Wyraża się to między innymi w nadużywaniu pojęcia wiedzy oraz w nierzadko niewłaściwym stosowaniu takich określeń jak organizacje wiedzy, pracownicy wiedzy, zarządzanie wiedzą, systemy organizacji wiedzy. Na tę „nieznośną lekkość terminologiczną", będącą rezultatem zarówno braku odpowiedzialności za słowo, jak i niewiedzy, nakłada się jeszcze zjawisko, które można nazwać paradoksem inteligencji, wyrażające się w nieodróżnianiu zdobywania, gromadzenia wiedzy od jej wykorzystania, użycia. Nie wystarczy wiedzę mieć – trzeba jeszcze wiedzieć, po co się ją ma i umieć we właściwym momencie zastosować.

Wiedza różni się od informacji, wiedza jest zdolnością do efektywnego działania. Wiedza jest związana z ludźmi, z osobą posiadacza, jest konstruktem umysłu ludzkiego powstałym dzięki motywowanej wewnętrznie różnorodnej aktywności własnej – informacja może istnieć niezależnie (np. w postaci dokumentu). Wiedza należy do jednostki i do społeczności, przepływa przez społeczności, powstaje na styku wiedzy poprzedniej. Bez wiedzy „starej" nie można tworzyć wiedzy „nowej". „(...) transformacja informacji w wiedzę zachodzić może tylko w umyśle człowieka. Subiektywna wiedza jednostki przekazywana lub komunikowana bezpośrednio od jednej osoby do drugiej musi być najpierw konwertowana do postaci informacji (nie można jej nabyć wprost). Informacja jest w tym przypadku traktowana jako obiektywna forma wiedzy – poddająca się komunikowaniu i zapisywaniu (tzw. wiedza skodyfikowana). W trakcie tej konwersji część wiedzy niekodyfikowalnej (np. doświadczenie, intuicja, umiejętności) umyka"[23]. Zapewne i w nauce, i w zarządzaniu utrata części wiedzy, według nazewnictwa zastosowanego przez Materską, niekodyfikowalnej, jest zawsze stratą, niekiedy dotkliwą. Trudno negować (nie będąc reprezentantem postawy postmodernistycznej) istnienie wiedzy społecznie akceptowanej, publicznej. Brytyjski fizyk John Ziman zwrócił uwagę, że nauka, a zatem także w dużej części wiedza, ma wymiar społeczny. Sztukę, literaturę i wiele różnych dziedzin ludzkiej działalności można uprawiać samemu, natomiast metoda naukowa zakłada, że to, co odkryjemy, tak długo nie jest wiedzą naukową, dopóki nie zostanie przedstawione innym badaczom oraz przez nich krytycznie ocenione i przyjęte lub odrzucone. Publiczna prezentacja, ocena i akceptacja (lub odrzucenie), to nie produkt uboczny działalności naukowej, ale sama jej istota, istota wiedzy naukowej. Inaczej mówiąc, krytyka[24] stanowi jedną z podstaw działalności naukowej.

Wyraz wiedza jest nie tyle wieloznaczny, co posiada znaczenia analogiczne, mówiąc za Wittgensteinem, należące do jednej rodziny znaczeń. Wiedza jest rezultatem poznania. „(...) najczęściej oznacza układ wiadomości, które jakoś

[23] K. Materska: *Informacja w organizacjach społeczeństwa wiedzy*. Warszawa 2007, s. 52.

[24] *Krytyka i krytycyzm w nauce*. Warszawa 1998.

uznajemy i potrafimy w pewien przynajmniej sposób uzasadnić (przeważnie na podstawie doświadczenia). (...) Relacje znaczeniowe między wyrażeniami: *znać*, *umieć*, *wiedzieć*, *posiadać wiadomości i mieć pojęcie* nie bywają dokładnie i jednolicie podawane. Można je tak przedstawić: znać to niekiedy tyle, co posiadać informacje raczej powierzchowne[25]; umieć zaś to potrafić coś zdziałać przy niekoniecznej znajomości teorii tego działania; umiejętność oznacza czasem nie tylko znajomość praktyczną, lecz także wiedzę teoretyczną (np. w nazwie Polska Akademia Umiejętności)[26]; wiedzieć to znać nie tylko częściowo i biernie, lecz także kompletnie i ze zdolnością do właściwego sformułowania wiadomości oraz ich uzasadnienia[27]; a wreszcie określenia: *posiadać wiadomości* oraz *mieć pojęcie* przeciwstawiają się sobie tak, jak: *posiadać tylko oderwane informacje* oraz *powiązaną* – aczkolwiek niepełną – *wiedzę*. Często używa się terminu *wiedza* jako nadrzędnego dla zespołu wiadomości potocznych, umiejętności i nauki, stąd mówi się *wiedza naukowa*"[28]. Wiedza naukowa nie jest ani trwała, ani pewna. Ulega zmianom, rozwija się, a jednym z tego efektów jest starzenie się wiedzy i jej stopniowa dezaktualizacja, choć tempo starzenia się wiedzy jest różne w poszczególnych naukach i dyscyplinach. Natomiast wiedza potoczna, choć często niespójna i nie zawsze prawdziwa, jest znacznie bardziej trwała.

Badaniem natury wiedzy zajmują się między innymi filozofowie, psychologowie, naukoznawcy, specjaliści od sztucznej inteligencji i inżynierii wiedzy. Stąd też bierze się wielość definicji, klasyfikacji czy typologii wiedzy. Wiedzę możemy między innymi podzielić: 1) według dopuszczalnych źródeł i kryteriów poznania na: wiedzę racjonalną – zgodną z krytycznie usposobionym rozumem i doświadczeniem oraz irracjonalną (pozaracjonalną) – dopuszczającą również źródła pozaracjonalne, jak np. emocjonalne czy wolitywne; 2) według sposobu wnioskowania na: wiedzę dedukcyjną – wywnioskowaną z apriorycznie przyjętych przesłanek oraz indukcyjną – wyjaśniającą w drodze uogólnienia fakty dane w doświadczeniu; 3) według przedmiotu poznania wiedza może być specjalistyczna (jednoaspektowa) i ogólna (wieloaspektowa); 4) według kryterium zastosowania wiedza dzieli się na wiedzę teoretyczną (wyjaśniającą, dlaczego jest tak lub dlaczego tak było) i wiedzę praktyczną (uzasadniającą, dlaczego należy tak, a nie inaczej działać). W kategorii wiedzy można wprowadzić również następujący porządek: 1) wiedza deklaratywna (propozycjonalna, sądowa, przekonaniowa, „wiedza że") – dotyczy faktów, np. koty mają cztery łapy, i jest uzasadnionym prawdziwym przekonaniem; 2) wiedza proceduralna („wiedza jak") – dotyczy umiejętności czynienia czegoś, zazwyczaj jest trudniejsza do werbalizacji od „wiedzy deklaratywnej", np. jak pływać; 3) metawiedza (wie-

---

[25] „Atoli znawstwo nie jest wyłącznie posiadaniem oderwanych informacji, lecz pełną i biegłą znajomością określonej dziedziny". S. Kamiński: *Nauka i metoda: pojęcie nauki i klasyfikacja nauk*. Do druku przygotował A. Bronk. Lublin 1992, s. 24.

[26] „Szczególnym przypadkiem umiejętności jest inteligencja, czyli umiejętność trafnego wyzyskania posiadanej wiedzy w nieszablonowej sytuacji". S. Kamiński: *Nauka*... op. cit., s. 24.

[27] „(...) analizy zwrotów: *wiedzieć coś, wiedzieć że, wiedzieć jak* przeprowadzili filozofowie analityczni. Niektórzy (np. J. L. Austin) uważają, że *wiem* pełni funkcję nie deskryptywną, lecz performatywną. Przeczy temu R. M. Chisholm, przyznając *wiem* podstawową rolę opisową. S. Kamiński: *Nauka*..., op. cit., s. 24.

[28] S. Kamiński: *Nauka*..., op. cit., s. 24.

dza o wiedzy; „wiem, że wiem"). Inny podział wyróżnia wiedzę jawną i wiedzę niejawną. Dziś w kontekście sieci, a zwłaszcza zjawisk określanych mianem 2.0, należałoby koniecznie wspomnieć o wiedzy autorytatywnej, eksperckiej i rozproszonej, zbiorowej mądrości (wiedzy) użytkowników sieci.

Wiedza jest jedną z wyższych wartości poznawczych, informacja nie. „*Informacja* nie jest *wiedzą*. Może ona być źródłem wiedzy. Informacje mogą tworzyć wiedzę, wówczas kiedy są porządkowane, systematyzowane, hierarchizowane, (...) konfrontowane, oceniane i krytykowane. Bez tych niełatwych operacji intelektualnych strumień informacji, który winien być „wodą życia" dla umysłów dążących do wiedzy, nader łatwo stać się może destruktywnym potopem, przyczyną dezinformacji"[29]. Wiedza zakłada istnienie świadomego podmiotu interpretującego ją i nowe informacje, „czekające" na ewentualną inkorporację w struktury wiedzy. Wiedza jest intencjonalna (świadomościowa), podczas gdy informacja nie.

O przekształcaniu informacji w wiedzę oraz o łączeniu procesów informacyjnych z systemami wiedzy inspirująco pisał, między innymi, nieco dziś zapomniany, a przypomniany przez Annę Sitarską[30], Klemens Szaniawski. Zdaniem Szaniawskiego, celem badania naukowego jest rewizja informacji niepewnej, mało prawdopodobnej czy wręcz fałszywej oraz umiejętne włączanie zweryfikowanych informacji w spójny system wiedzy, dążenie do prawdy. Anna Sitarska dziesięć lat temu sformułowała niezwykle ważne i chyba zignorowane przez środowisko pytania. Czy „nie należałoby w dobie formowania «społeczeństwa informacji i wiedzy» włączyć do refleksji nad kryteriami ocen i systemami wartościowania informacji w ogólniejszym kontekście wartości informacji? Czy przy identyfikowaniu właściwości procesów przekształcania informacji w wiedzę wtedy, kiedy wartościujemy informację niezależnie od jej zakresu (treści) oraz poziomu (twórcy i/lub odbiorcy), nie jest uprawnione wartościowanie wg kryterium prawdziwości i poziomu hipotetyczności?"[31]. I dalej: „Swoista dla postmodernistycznego świata relatywizacja systemów wartości, zwłaszcza dystans wobec wartości *prawdy* i *prawdziwości*, tak w sferach nauki i kultury, jak i w innych obszarach ludzkiej działalności, nie powinna przesądzać o odrzucaniu tych wartości z repertuaru kryteriów oceny informacji i wiedzy. Nie można unikać tych fundamentalnych pojęć przede wszystkim w procesach kształcenia i nauczania, które należą do najważniejszych przykładów procesu przekształcania informacji w wiedzę. Tak więc, jeśli chcemy rozważać dwa zasadnicze problemy informacji naukowej – (1) przekształcanie informacji w wiedzę oraz (2) społeczne uwarunkowania i konsekwencje komunikowania się, w tym również poprzez rozległe sieci komputerowe, właśnie „konfrontacja *informacji* i *prawdy*" wydaje się nieunikniona"[32].

Pojęcie wiedzy możemy rozpatrywać w dwóch ujęciach. Ujęcie węższe przez wiedzę rozumie „ogół wiarygodnych informacji o rzeczywistości wraz z umie-

---

[29] M. Agnosiewicz: *Racjonalista – Prolog*. [online]. [dostęp: 10.07.2010]. Dostępny w World Wide Web: <http://www.racjonalista.pl/kk.php/s,1>.

[30] A. Sitarska: *Problemy przekształcania informacji w wiedzę*. W: *Informacja. Wiedza. Gospodarka*. Pod red. W. Pindlowej i D. Pietruch-Reizes. Warszawa 2001, s. 199-208.

[31] A. Sitarska: *Problemy*..., op. cit., s. 201.

[32] A. Sitarska: *Problemy*..., op. cit., s. 202.

jętnością ich wykorzystywania; w społeczeństwach współczesnych wiedza w tym znaczeniu to przede wszystkim, choć nie wyłącznie, wiedza naukowa; zajmuje się nią głównie teoria poznania i filozofia nauki"[33]. Jeżeli natomiast spróbujemy spojrzeć na wiedzę w szerszej perspektywie, wówczas uznamy za nią „wszelki zbiór informacji, poglądów, wierzeń itp., którym przypisuje się wartość poznawczą i/lub praktyczną; wiedza w tym znaczeniu może nie mieć z nauką nic wspólnego, gdyż często odnosi się do zjawisk, którymi nauka w ogóle się nie zajmuje, zawiera twierdzenia jawnie z nauką sprzeczne oraz nie zakłada konieczności uzgadniania głoszonych twierdzeń za pomocą procedur uznanych w nauce"[34].

Przyjęło się mówić, że informacja to zinterpretowane dane, a wiedza to informacje powiązane relacjami. Czy to rzeczywiście wystarcza do zrozumienia istoty wiedzy? Informacje powiązane relacjami to co najwyżej wiedza potencjalna; żeby stała się realna, trzeba być w stanie nadać jej pewną strukturę, być w stanie wniknąć w jej istotę, zinternalizować ją, włączyć w już istniejącą strukturę wiedzy. Wiedzę można opanować na różnym poziomie, scharakteryzowanym np. jednym z następujących „stopni wtajemniczenia":

- poznanie (znajomość faktów, metod, definicji),
- stosowanie (posiadanie umiejętności, znajomość algorytmów),
- zrozumienie (głęboka znajomość, intuicja, dostrzeganie powiązań),
- twórczość (odkrywanie nowych obszarów, rozwiązywanie trudnych problemów)[35].

Spośród czterech wymienionych stopni opanowania wiedzy za niedocenianiy uważam poziom zrozumienia. Samo pojęcie rozumienia „ma wiele odcieni znaczeniowych"[36]. Niekiedy oznacza poprawne stosowanie terminologii. I choć takie „terminologiczne" zrozumienie jest tylko jednym z pierwszych, płytszych warstw, to jednak wiedza nie może się obyć bez niego. Innym razem jest sumowaniem wiadomości. Ale zrozumieć to znaczy także znaleźć coś więcej, prostotę lub głębię, które nie były znane nikomu wcześniej. „Właśnie o takie zrozumienie zabiegają uczeni"[37]. Zrozumienie mocno łączy się ze świadomością (także z inteligencją). Związki informacji i wiedzy ze świadomością są bardzo mało znanym i zbadanym obszarem nauki o informacji, choć wydaje się, że zwłaszcza w badaniach użytkowników miałyby wielkie i znaczące pole do popisu. „Poznawanie istoty i mechanizmów zmian zachodzących w trakcie przekształcania informacji w wiedzę, przez pryzmat własności informacji oraz cechy kontekstu działań ludzkich podejmowanych pod wpływem informacji, może być traktowane również, jako wnikanie w społeczne tło formowania się wiedzy pragmatycznej, a więc również ważnych warstw świadomości".[38]

---

[33] *Nowa Encyklopedia Powszechna PWN.* T. 6. Warszawa 1997, s. 733.

[34] Ibidem.

[35] J. L. Cieśliński: *Zrozumienie kluczem do wiedzy.* W: *Między unifikacją a dezintegracją: kondycja wiedzy we współczesnym* społeczeństwie. Pod red. A. Jabłońskiego i M. Zemły. Lublin 2008, s. 83.

[36] Ibidem, s. 41.

[37] Ibidem.

[38] A. Sitarska: *Problemy...*, op. cit., s. 202.

Zarządzanie wiedzą

Nieuprawnione i nadmierne rozszerzanie znaczenia *wiedzy* sprawia, że bardzo często myli się organizację i/lub zarządzanie wiedzą z organizacją i/lub zarządzaniem informacją, źródło wiedzy ze źródłem informacji. Producenci systemów takich jak zarządzające dokumentami, w tym zapisami bibliograficznymi, niekiedy nazywają je systemami zarządzania wiedzą, podczas gdy w rzeczywistości wspomagają one działania jedynie na poziomie informacyjnym. Zarządzanie wiedzą nie może być kwestią tylko i wyłącznie technologiczną, gdyż wiedza z definicji wymaga udziału człowieka, również odpowiedniej kultury organizacyjnej, kanałów komunikacyjnych, stosownych reguł dostępu i przechowywania informacji. Zarządzanie informacjami ustala pewne zasady, wytyczne, ocenia źródła danych, podczas gdy zarządzanie wiedzą jest zintegrowane z człowiekiem, jego inteligencją, zdolnością rozumienia (jako kluczem do wiedzy), wyobraźnią, intuicją, miejscem pracy, składa się na nie pewna kultura współpracy w określonym środowisku, dzielone dobra intelektualne itd.

Wiedza nie jest systemem technologicznym i choć technologie informacyjne mogą tu być przydatne, to nic nie zastąpi rozumu. Autorzy wielu publikacji[39] zakładają, że wiedza jest „ważniejsza" od informacji (co oczywiście nie znaczy, że zawsze jest wystarczająca, dostatecznie aktualna itp.). „(...) «społeczeństwo wiedzy» (...) winno cechować się między innymi umiejętnością dostrzegania względności informacji i czegoś, co można nazwać «relacyjnością» informacji, ich wzajemnych powiązań, umiejętnością własnej obserwacji świata i wyciągania z tej obserwacji własnych wniosków, umiejętnością dostrzegania i definiowania problemów, umiejętnością prowadzenia dyskusji, której celem pozostaje dochodzenie do prawd alternatywnych"[40]. Technologie teleinformatyczne są odpowiednie do generowania, przetwarzania i przesyłania danych i informacji, nie wiedzy. I to niezależnie od niezwykłej zdolności komputerów do „kojarzenia" danych. „(...) społeczeństwo informacyjne dostarcza nam wielu informacji na temat tego, «jak» mamy coś czynić; milczy jednak w kwestiach tego, «dlaczego i po co?»"[41].

Masowa produkcja wiedzy przez ludzi i odpowiednio zaprogramowane komputery bardzo często jest dziś produkcją tandety, „wiedzy" bezużytecznej i bezwartościowej (lub małowartościowej). „Mam nadzieję, że masowa produkcja wiedzy i słuszne dążenie do praktycznych zastosowań nie wyeliminują tego najpiękniejszego aspektu nauki, jakim jest twórczość i sztuka. Zwłaszcza, że historia nauki uczy, iż często na pozór bezużyteczne badania okazywały się później przełomowe z punktu widzenia rozwoju nowych technologii"[42]. Społeczeństwo wiedzy to między innymi promowanie twórczości, oryginalnego rozwiązywania

[39] M.in.: J. Gocłkowski, K. M. Machowska: *Społeczeństwo wiedzy a społeczeństwo informatyczne*. W: *Społeczeństwo informatyczne: szanse czy zagrożenie*. Pod red. B. Chyrowicz. Lublin 2003 s. 141-179; L. W. Zacher: *Od społeczeństwa informacyjnego do społeczeństwa wiedzy (dylematy tranzycyjne: między informacją, wiedzą i wyobraźnią)*. W: *Społeczeństwo informacyjne. Wizja czy rzeczywistość?* Pod red. L. Haber . T.1. Kraków 2004, s. 103-112.

[40] M. Golka: *Bariery w komunikowaniu i społeczeństwo (dez)informacyjne*. Warszawa 2008, s. 161.

[41] Ibidem, s. 163.

[42] J. L. Cieśliński: *Zrozumienie...*, op. cit., s. 84.

problemów. „Istnieje tzw. wielka twórczość, która oddziaływa na znaczne grupy ludzi (idee, pomysły, hipotezy, wizje artystów, pisarzy, uczonych), oraz twórczość mała, która dotyczy niesformalizowanych obszarów kultury: pracy, obyczajów, mieszkania, gospodarstwa rolnego, czyli kultury jako regulatora życia. Chodzi o innowacje dokonywane przez „ludzi z ulicy". Ta twórczość decyduje o poziomie kulturalnym społeczeństw. Polityka kulturalna i oświatowa powinna wspierać twórczość wielką i małą oraz przenikanie się obydwu poziomów twórczości"[43].

Tom Wilson w artykule *The nonsense of «knowledge management»*[44] napisał, że kiedy chcemy wyrazić to, co wiemy, posługujemy się komunikatami (ustnymi, tekstowymi, graficznymi, gestami, mową ciała). Nie przenoszą one „wiedzy", są informacją, którą umysł drugiego człowieka może przyswoić, zrozumieć, zinterpretować i włączyć do już istniejących struktur wiedzy. Struktury te nie są takie same u osoby nadającej komunikat, co u osoby go odbierającej, ponieważ struktury wiedzy każdego z nas są zdeterminowane przez nasze życiorysy (ang. *biographically determined*). Informacjami można zarządzać, można też zarządzać zasobami czy źródłami informacji. Wiedza może być przedmiotem zarządzania tylko wtedy, gdy zarządza nią jej posiadacz, ale nawet wtedy zarządzanie to jest niedoskonałe. W rzeczywistości często nie wiemy, co (ile) naprawdę wiemy. To, że jednak posiadamy pewną wiedzę, może się okazać dopiero wtedy, kiedy jej potrzebujemy, aby coś wykonać. Większość z tego, czego się uczyliśmy, ulega zapomnieniu, ale może niespodziewanie ujawnić się, przypomnieć, kiedy jest potrzebne lub kiedy nie jest potrzebne. Innymi słowy, zdajemy się mieć bardzo małą kontrolę nad tym, „co wiemy"[45].

Z badania przeprowadzonego przez Wilsona wynika, że termin zarządzanie wiedzą jest używany w wielu różnych kontekstach znaczeniowych, między innymi na określenie zarządzania kadrami, „ludzkiego" (w sensie zasobów ludzkich) aspektu zarządzania, podnoszenia i aktualizowania kwalifikacji personelu, technologii informacyjno-komunikacyjnych oraz, wcale nie tak rzadko, jako synonim zarządzania informacją. Wyrazem pewnej bezradności i jednak niedoskonałości terminu zarządzanie wiedzą zdaje się być zauważalne w literaturze przesunięcie leksykalne z zarządzania wiedzą na dzielenie się wiedzą. Zdaniem Wilsona termin zarządzanie wiedzą stał się dziś tak popularny z kilku powodów, wśród których są następujące: (1) w wielu organizacjach termin informacja jest używany jako synonim terminu dane, co z kolei wywołuje zapotrzebowanie na „nowy" termin, oznaczający coś ważniejszego, o wyższej randze; (2) dla wielu firm z branży IT „zarządzanie wiedzą" jest atrakcyjną etykietą, która pozwala sprzedawać produkty, także te od dawna dostępne na rynku, pod nową nazwą (np. Lotus Notes nie jest już aplikacją klienta w środowisku do pracy grupowej Lotus Notes/Domino, ale KnowledgeWare); (3) jeśli organizacja nie może się pochwalić, że zarządza wiedzą, wypada z gry; (4) niektóre instytucje kształcące studentów, w tym w zakresie informacji naukowej i bibliotekoznawstwa,

---

[43] K. Krzysztofek: *Jaka polityka kulturalna w epoce globalizacji i mediów elektronicznych?* „Kultura współczesna" 2005, nr 1(43), s. 14.

[44] T. D. Wilson: *The nonsense of 'knowledge management'.* „Information Research" 2002 no. 8. [online]. [dostęp: 27.07.2010]. Dostępny w World Wide Web: <http://InformationR.net/ir/8-1/paper144.html>.

[45] Ibidem.

uznały „zarządzanie wiedzą" za dobry środek do odróżnienia się od innych, wzmocnienia własnej odrębności, „unikatowości" i pozycji na uczelni. Podobną refleksję można znaleźć w artykule Radwańskiego: „Wielu dostawców informacji sprzedaje nam swoje produkty pod hasłem sprzedawania wiedzy. Wystarczy tylko, że pełnotekstowa baza danych zostanie zaindeksowana, już pretenduje do rangi zasobu wiedzy"[46].

Dość podobnie jest z terminem system organizacji wiedzy. W literaturze[47] znajdziemy dość powszechnie akceptowane wyjaśnienia, że terminem tym objęte są wszystkie typy systemów organizacji informacji, ułatwiające zarządzanie wiedzą. „Ponieważ systemy organizacji wiedzy są narzędziami organizacji informacji (sic!-JWK), lokują się w sercu każdej biblioteki, muzeum czy archiwum"[48].

Są takie dziedziny działalności praktycznej, w których rozróżnienie między danymi, informacją i wiedzą nie jest konieczne. Także w języku potocznym wiedza i informacja często są traktowane jako synonimy. Jednak w nauce o informacji i bibliotekoznawstwie nie można się zgodzić na taką swobodę terminologiczno-ontologiczną. Nie o to chodzi, żeby nie posługiwać się terminem wiedza, ale żeby, przynajmniej w wypowiedziach naukowych czy paranaukowych, używać go zgodnie z jego znaczeniem (znaczeniami). Pewnie jest już za późno na wyeliminowanie z użycia pewnych nieudanych tworów terminologicznych (chyba że znikną ich desygnaty, jak to się nie raz działo), ale posługując się nimi, należy mieć świadomość ich sztuczności i umowności, a ucząc, uwrażliwiać studentów, uczniów, słuchaczy na pewną umowność terminologii oraz konsekwencje braku szacunku dla języka.

Odpowiadając na pytanie zawarte w tytule artykułu, z dużym prawdopodobieństwem mogę stwierdzić, że mamy dziś do czynienia z kryzysem informacji i wiedzy. Kryzys informacji jest w moim odczuciu w znacznej mierze pochodną ilości informacji, jej potopu, zalewu, co wcale nie jest wartością lub nie musi nią być. „Wartością byłoby «społeczeństwo dobrze poinformowane» (niezależnie od trudności wymierzenia owego «dobrego» czy «właściwego» stopnia poinformowania"[49]. Trudniejsza sprawa jest z wiedzą. Współczesne czasy sprzyjają niewiedzy. „(...) po co uczyć się czegokolwiek, skoro w razie potrzeby każdą wiedzę[50] można mieć na zawołanie. Przekonanie to opiera się jednak na złudzeniu, że procesy myślowe w mózgu mogą przebiegać bez zasobów informacyjnych znajdujących się w pamięci jednostki. Owszem, Internet (...) umożliwia pewne mechaniczne skojarzenia, które wobec tradycyjnych nośników (np. książek) musiał spełniać umysł odbiorcy. Ciągle jednak «nasze umysły potrafią spożytkować tylko taką ilość informacji, jaką są zdolne sobie przyswoić» – słusznie zauważył Charles Jonscher (...). A coraz częściej przyswajają sobie niewiele albo zgoła nic"[51]. Jeśli na to nałożymy jeszcze brak jakiegokolwiek

[46] A. Radwański: *Potrzeba...*, op. cit., s. 160.

[47] Np. G. Hodge: *Systems of Knowledge Organization for Digital libraries. Beyond traditional authority files*. [online]. [dostępny: 27.07.2010]. Dostępny w World Wide Web: <http://clir.org/pubs/reports/pub91/1knowledge.html>.

[48] Ibidem.

[49] M. Golka: *Bariery...*, op. cit., s. 156.

[50] W oryginale nie ma cudzysłowu, ale moim zdaniem, wyraz wiedza w tym kontekście należy rozumieć przenośnie – przypis JWK.

[51] M. Golka: *Bariery...*, op. cit., s. 158.

wyposażenia aksjologicznego[52], to w rezultacie otrzymamy to, co Lech Zacher nazwał „tandetą informacyjną", a ja rozszerzając, nazwałabym „niby-wiedzą".

Bibliografia

1. Bauman Z.: *Etyka ponowoczesności*. Warszawa 2003.
2. Cieśliński J. L.: *Zrozumienie kluczem do wiedzy*. W: *Między unifikacją a dezintegracją: kondycja wiedzy we współczesnym społeczeństwie*. Pod red. A. Jabłońskiego i M. Zemły. Lublin 2008 s. 37-84.
3. Fiedorczuk J.: *Filozoficzne spojrzenie na pojęcie wartości*. [online]. [dostęp: 17.07.2010]. Dostępny w World Wide Web: <http://www.eid.edu.pl/publikacje/filozoficzne_spojrzenie_na_pojecie_wartosci,202.html>.
4. Goćkowski J., Machowska K. M.: *Społeczeństwo wiedzy a społeczeństwo informatyczne*. W: *Społeczeństwo informatyczne: szansa czy zagrożenie*. Pod red. B. Chyrowicz. Lublin 2003 s. 141-179.
5. Golka M.: *Bariery w komunikowaniu i społeczeństwo (dez)informacyjne*. Warszawa 2008.
6. Heller M.: *Filozofia nauki: wprowadzenie*. Kraków, 2009.
7. Hodge G.: *Systems of Knowledge Organization for Digital libraries. Beyond traditional authority files*. [online]. [dostęp: 27.07.2010] <http://clir.org/pubs/reports/pub91/1knowledge.html>.
8. Kamińska-Czubała B.: *Informacja jako fetysz w społeczeństwie wiedzy i w społeczeństwie ryzyka*. W: *Edukacja w społeczeństwie „ryzyka": bezpieczeństwo jako wartość*. Pod red. M. Gwoździckiej-Piotrowskiej, J. Wołejszo i A. Zduniaka. Poznań 2007, s. 92-99.
9. Kamiński S.: *Nauka i metoda: pojęcie nauki i klasyfikacja nauk*. Pod red. A. Bronk. Lublin, 1992.
10. Kleiber M.: *Racjonalna wizja rozwoju to warunek przyszłych sukcesów*. [online]. [dostęp: 10.07.2010]. Dostępny w World Wide Web: <http://www.aktualnosci.pan.pl/index.php?option=com_content&view=article&id=254:racjonalna-wizja-rozwoju-to-warunek-przyszych-sukcesow&catid=16:opinie-i-pogldy&Itemid=21>.
11. *Krytyka i krytycyzm w nauce*. Warszawa, 1998.
12. Krzysztofek K.: *Jaka polityka kulturalna w epoce globalizacji i mediów elektronicznych?* „Kultura współczesna" 2005 nr 1(43), s. 5-18.
13. Majcherek J. A.: *Źródła relatywizmu w nauce i kulturze XX w. Od teorii względności do postmodernizmu*. Kraków, 2004.
14. Materska K.: *Informacja w organizacjach społeczeństwa wiedzy*. Warszawa, 2007.
15. Muraszkiewicz M.: *Esej: Nowy paradygmat, czyli od systemu do sieci*. W: *Od informacji naukowej do technologii społeczeństwa informacyjnego*. Praca zbior. pod red. B. Sosińskiej-Kalaty i M. Przastek-Samokowej przy współpracy A. Skrzypczaka. Warszawa 2005, s. 83-86.
16. Ożog K.: *O prymacie wolności nad prawdą – wartościowanie w dobie postmodernistycznej*. W: *Człowiek wobec wyzwań współczesności: upadek wartości czy walka o wartość?* Pod red. J. Mazura, A. Małyski, K. Sobstel. Lublin 2007, s. 46-57.
17. McQuail D.: *Prawda i jakość informacji*. [online]. [dostęp: 10.07. 2010]. Dostępny w World Wide Web: <http://wiedzaiedukacja.eu/archives/546>.

[52] Nie możemy narzekać na brak kodeksów etycznych dla bibliotekarzy i pracowników informacji (np. *Kodeksy etyki bibliotekarskiej na świecie: antologia narodowych kodeksów etycznych*. Pod red. Z. Gębołysia i J. Tomaszczyka. Warszawa 2008). Pozostaje mieć nadzieję, że publikowanie kodeksów idzie w parze ze znajomością etyki i umiejętnością dokonywania wyborów moralnych będących w zgodzie z określonym systemem wartości. „(...) nieetyczność pojawia się wszędzie tam, gdzie nieznajomość podstawowych pojęć, brak kompetencji, a dopiero na końcu zła wola. Już Sokrates twierdził przecież, że możemy postępować dobrze tylko wtedy, gdy wiemy czym jest dobro". (A. Siewierska-Chmaj: *Co z tym dziennikarstwem?* W: *Teoria, praktyka, etyka. O kształceniu dziennikarzy w Polsce i na świecie*. Pod red. A. Siewierska-Chmaj. Rzeszów 2005, s. 9).

18. Pacek J.: *Uwolnić informację!* EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2009 nr 1 (101). [online]. [dostęp: 27.07 2010]. Dostępny w World Wide Web: <http://www.ebib.info/2009/101/a.php?pacek>.
19. Poczobut R.: *Fenomen wielowymiarowości umysłu a emergencja. Z ontologii i metodologii badań inter- i transdyscyplinarnych.* W: *Między unifikacją a dezintegracją: kondycja wiedzy we współczesnym społeczeństwie.* Pod red. A. Jabłońskiego i M. Zemły. Lublin 2008, s. 11-35.
20. Puzynina J.: *Co znaczy „walka o wartości"?* W: *Człowiek wobec wyzwań współczesności: upadek wartości czy walka o wartość?* Pod red. J. Mazura, A. Małyski, K. Sobstel. Lublin 2007, s. 24-36.
21. Radwański A.: *Psychologiczne i socjologiczne aspekty prezentacji informacji i wiedzy.* W: *Informacja. Wiedza. Gospodarka.* Pod red. W. Pindlowej i D. Pietruch-Reizes. Warszawa 2001, s. 157-163.
22. Saracevič T.: *Information Science.* „Journal of the American Society for Information Science" 1999, vol. 50, no. 12, pp. 1051-1063.
23. Sitarska A.: *Problemy przekształcania informacji w wiedzę.* W: *Informacja. Wiedza. Gospodarka.* Pod red. W. Pindlowej i D. Pietruch-Reizes. Warszawa 2001, s. 199-208.
24. Stanula-Boroń M.: *Informacja, język i wiedza w ujęciu Karla R. Poppera.* „Zagadnienia Informacji Naukowej" 2000, nr 1, s. 3-16.
25. Szaniawski K.: *Nauka jako proces poszukiwania informacji.* W: Szaniawski K.: *O nauce, rozumowaniu i wartościach: pisma wybrane.* Wybrał i opracował J. Woleński. Wstępem opatrzyli S. Amsterdamski i J. Woleński. Warszawa 1994, s. 110-119.
26. Szefliński P.: *Społeczeństwo informacyjne – o czym biblioteka XXI w. powinna wiedzieć?* W: *II Konferencja Biblioteki Politechniki Łódzkiej: Biblioteki XXI w. Czy przetrwamy? Łódź, 19-21 czerwca 2006 r. Materiały konferencyjne.* Łódź 2006, s. 31-40.
27. Tischner J.: *Myślenie według wartości.* Kraków, 1982.
28. Wilson T. D.: *The nonsense of The nonsense of 'knowledge management.* „Information Research" no. 8. [online]. [dostęp: 27.07.2010]. Dostępny w World Wide Web: <http://InformationR.net/ir/8-1/paper144.html>.
29. Wiśniewska D.: *Postmodernizm w procesie edukacji akademickiej.* W: *Edukacja w społeczeństwie „ryzyka": bezpieczeństwo jako wartość.* Pod red. M. Gwoździckiej-Piotrowskiej, J. Wołejszo i A. Zduniaka. Poznań 2007, s. 382-388.
30. Wojciechowski J.: *W kręgu informacji i nieinformacji.* „Bibliotekarz" 1998, nr 4, s. 2-5.
31. Wojciechowski J.: *Biblioteka w komunikacji publicznej.* Warszawa, 2010.
32. Zacher L. W.: *Od społeczeństwa informacyjnego do społeczeństwa wiedzy (dylematy tranzycyjne: między informacją, wiedzą i wyobraźnią).* W: *Społeczeństwo informacyjne. Wizja czy rzeczywistość?* Pod red. L. Haber. T. 1. Kraków 2004, s. 103-112.
33. Ziman J.: *Public knowledge: the social dimension of science.* Cambridge, 1968.
34. Żbikowska-Migoń A: *Rola czasopism w krytyce piśmiennictwa naukowego.* W: *Dokument, książka i biblioteka w badaniach naukowych i nauczaniu uniwersyteckim.* Pod red. M. Skalskiej-Zlat i A. Żbikowskiej-Migoń. Wrocław 2008, s. 139-155.
35. Życiński J.: *Formacja do wartości humanistycznych w kulturze postmoderny.* W: *Człowiek wobec wyzwań współczesności: upadek wartości czy walka o wartość?* Pod red. J. Mazura, A. Małyski, K. Sobstel. Lublin 2007, s. 13-23.

Summary

The author's reflections on the value of information. The context is limited to the issues relevant and important for the librarians in Poland in the first decade of the 21st century. The article concerns (among others): expected attributes and features of information, information imperative, values in human and social life, knowledge as a value, mutual relations of information and knowledge, devaluation of knowledge, and misusage of the "knowledge" term. Answering the title question the author states, referring to sociological and axiological contexts, that we are witnessing the crisis of the value of knowledge nowadays.

ROZWÓJ BADAŃ NAD PRZETWARZANIEM JĘZYKA NATURALNEGO

Piotr Malak
Instytut Informacji Naukowej i Bibliologii
Uniwersytet M. Kopernika w Toruniu

Przetwarzanie języka naturalnego, metody statystyczne NLP, gramatyki generatywne, wyszukiwanie informacji

Rosnąca popularność cyfrowych form przechowywania i przekazywania informacji powoduje, że konieczne staje się opracowywanie nowych, wydajnych metod ułatwiających zarządzanie czy też wyszukiwanie takiej informacji. Z informacją cyfrową związane są częściowo lub w całości m.in. nowe formaty opisu dokumentu (np. Dublin Core), biblioteki cyfrowe czy też wyszukiwarki sieciowe.

Ogromną zaletą źródeł elektronicznych w porównaniu do ich tradycyjnych poprzedników jest możliwość swobodnego przeszukiwania tekstu. Użytkownik dokumentu elektronicznego za pomocą specjalnego interfejsu może przeszukać całą treść dokumentu dla wybranych haseł czy słów kluczowych, tworzonych w sposób dowolny, całkowicie swobodnie. Nie jest już ograniczony tylko do haseł klasyfikacji rzeczowej, których wprawne stosowanie wymaga często odpowiedniego szkolenia.

Jednymi z najbardziej zaawansowanych systemów informacyjno-wyszukiwawczych są wyszukiwarki internetowe. Są to specjalistyczne systemy indeksujące treść dokumentów dostępnych online i zwracające użytkownikowi listy dokumentów spełniających zadane przez niego kryteria wyszukiwania. Przy czym w systemach zaawansowanych technologicznie dopasowanie treści dokumentu odbywa się z uwzględnieniem zasad gramatyki danego języka. Analizowane automatycznie są wszystkie prawidłowe formy gramatyczne wyrazów tworzących zapytanie użytkownika.

Od początków stosowania komputerów rozwija się dziedzina badań nad automatycznym przetwarzaniem języka naturalnego (ang. NLP – *Natural Language Processing*). Dziedzina ta dostarcza rozwiązań zarówno teoretycznych, na poziomie lingwistyki, jak i praktycznych zastosowań wykrytych prawidłowości i zależności lingwistycznych. Jako przykłady można podać moduły korekty ortograficznej i gramatycznej, systemy automatycznej translacji czy wspomniane wcześniej wyszukiwarki sieciowe.

Artykuł ma na celu wprowadzenie czytelnika w powstanie i rozwój tej niezwykle dynamicznej i interdyscyplinarnej dziedziny.

Początki NLP

Podstawy teoretyczne dla NLP przygotował m.in. Alan Turing, opracowując w 1936 r. teorię automatu. Teorię tę rozwinął następnie w latach 50. XX w. Stephen C. Kleene, wzbogacając ją o pojęcia automatu skończonego oraz zbiorów regularnych. Kolejnym badaczem, który wniósł duży wkład w rozwój nowego kierunku badań, był Claude E. Shannon. Dostosował on modele Markova do tworzenia modeli lingwistycznych oraz opracował koncepcję entropii w teorii informacji, a także pojęcie ilości informacji. NLP zaadaptowało również teorię języków formalnych i gramatyk Noama Chomskyego z 1956 r. Na podstawie prac tych badaczy wypracowano wiele teorii opisu języka oraz analizy języków naturalnych i sztucznych[1].

Początki przetwarzania języka naturalnego jako dziedziny badań naukowych związane są w oczywisty sposób z początkiem ery komputerów, maszyn dysponujących mocą obliczeniową wystarczającą do automatycznego przeprowadzania operacji na danych. Już w latach 40. XX w. w Stanach Zjednoczonych podjęto próby automatycznego tłumaczenia tekstów. Próby te nie były zbyt skuteczne i szybko z nich zrezygnowano, wskazując jednocześnie inne obiecujące kierunki badań NLP.

W latach 50. XX w. stosowano przetwarzanie danych w postaci wyrażeń języków naturalnych dla celów wyszukiwania, klasyfikacji i selekcji informacji w dużych zbiorach. Do końca lat 80. minionego stulecia w badaniach NLP rozwijały się dwa niezależne trendy: analiza statystyczna oraz gramatyki generatywne. Metody statystyczne wykorzystywane są szeroko do wyszukiwania dokumentów w dużych zbiorach (nurt IR – *Information Retrieval*). W analizie statystycznej poszczególne słowa stanowiące treść analizowanych dokumentów tworzą zbiór wspólny, z zachowaniem informacji o częstości wystąpień danego słowa w zbiorze. W literaturze angielskiej zbiór ten nazywany jest *bag-of-words*, w piśmiennictwie polskim można spotkać powtórzenie nazwy angielskiej lub wyrażenia typu *wielozbiór.* W metodach statystycznych słowa kluczowe dla poszczególnych dokumentów wskazywane są z reguły na podstawie porównania częstości ich wystąpienia w danym dokumencie z liczbą wystąpień w całym zbiorze. Podejście takie cechuje się łatwością implementacyjną, relatywnie niskimi kosztami operacyjnymi, wysoką skutecznością wyszukiwania dokumentów oraz niezależnością od konkretnego języka naturalnego. Alternatywny nurt bazował na teorii automatów A. Turinga oraz pracach N. Chomsky'ego dotyczących gramatyk formalnych i generatywnych. Tworzone w ich wyniku gramatyki wykorzystywane były następnie do odtwarzania struktury zdania w dowolnym wyrażeniu. Do zalet takiego podejścia zalicza się głównie sformalizowanie i pogłębienie wiedzy lingwistycznej. Praktycznym zastosowaniem wyników badań nad gramatykami są np. parsery, czyli analizatory składniowe[2].

Obie metody nie są wolne od wad. Podejście statystyczne pomija znaczenie analizowanej treści, nie są rozpoznawane ani rejestrowane związki pomiędzy

---

[1] Ref. za: D. Jurafsky, J. H. Martin: *Speech and language processing. An introduction to Natural Language Processing, Computational Linguistics and Speech Recognitio.* New Jersey 1999, s. 10-11.

[2] Por. Ibidem, s. 12-15.

wyrazami, dlatego metody takie najlepiej sprawdzają się w wyszukiwaniu informacji na podstawie wystąpienia haseł wyszukiwawczych. Natomiast w przypadku metod bazujących na gramatykach i formalizmach problemem są koszty (obliczeniowe oraz czasowe) przygotowania poprawnej gramatyki oraz analizy treści dokumentu. Rezultatem zastosowania przygotowanych gramatyk jest duży zbiór możliwych interpretacji analizowanych zdań, których ujednoznacznienie często nie jest możliwe na drodze wyłącznie automatycznego przetwarzania.

Od końca lat 80. XX w. coraz większego znaczenia nabierają metody statystyczne NLP. Wiąże się to z dostępnością odpowiednio przygotowanych korpusów tekstów reprezentatywnych dla danego języka, co zwiększa wartość analiz statystycznych. Dzięki wykorzystaniu adnotowanych korpusów w opracowaniu frekwencyjnym tekstu można wykorzystać nie tylko lokalne, typowe dla danego dokumentu czy zboru wartości, ale również uwzględnić związki formalne pomiędzy poszczególnymi wyrazami[3].

Etapy rozwoju badań nad przetwarzaniem języka naturalnego

W rozwoju badań nad przetwarzaniem języka naturalnego wyodrębnić można kilka następujących etapów:

1) do 1957 r. można wskazać następujące osiągnięcia:
   - model obliczeń – na podstawie pracy A. Turinga (Turing, A., *On computable numbers, with an application to the Entscheidungsproblem,* [w:] *Proceedings of the London Mathematical Society,* 2(42), s. 230-165),
   - zastosowanie modeli Markova do analizy języka (w pracy Shannon, C. E., *A mathematical theory of communication*, [w:] *The Bell Systems Technical Journal,* 27, 279-423, 623-656),
   - wprowadzenie automatów skończonych oraz wyrażeń regularnych,
   - zastosowanie automatów do reprezentowania gramatyk (Chomsky, N., *Three models for the description of language* [w:] *IRE Transactions of Information Theory*, 2(3)),
   - wprowadzenie formalnego opisu języka oraz gramatyk bezkontekstowych (głównie N. Chomsky, op. cit., ale też Backus, J. W., *The syntax and the semantics of the proposed international algebraic of the Zürich ACM-GAMM Conference.* [w:] *Proc. of the Inf. Conf. on Information Processing,* Paris 1959; oraz Naur., P, et all., *Report on the algorithmic language algol 60.* [w:] *Communications of the ACM,* 1960),
   - zdefiniowanie entropii jako miary ilości informacji (Shannon, C. E., *Prediction and entropy of printed English* [w:] *The Bell Systems Technical Journal, 30, 1951);*
2) lata 1957-1970 – wtedy wykrystalizowały się dwa, konkurencyjne podejścia do przetwarzania języka naturalnego:
   a) metody symboliczne (formalne), do których można zaliczyć:
      - gramatykę generatywną,
      - parsery syntaktyczne,

[3] O trendach w NLP por. m.in. A. Przepiórkowski: *Powierzchniowe przetwarzanie języka polskiego*. Warszawa 2008, s. 9-11.

- metody sztucznej inteligencji;

b) metody statystyczne, a wśród nich:
- metody Bayes'a,
- optyczne rozpoznawanie liter (ang. *Optical Character Recognition*, OCR),
- identyfikacja autorów tekstów,
- tworzenie korpusów;

3) lata 1970-1983 – wypracowano cztery paradygmaty NLP:
   a) modele statystyczne: rozpoznawanie mowy, synteza mowy; Ukryte Modele Markowa (ang. *Hidden Markov Models*, HMM),
   b) logika formalna (język Prolog; Definite Clasue Grammar, DCG; teoria Lexical-Functional Grammar, LFG),
   c) rozumienie języków naturalnych,
   d) modelowanie dyskursu;
4) lata 1983-1993 – odrodzenie modeli skończonych stanów oraz empiryzmu:
   a) morfologia i fonologia za pomocą modeli skończonych,
   b) modele skończonych stanów składni,
   c) metody stochastyczne wykraczające poza rozpoznawanie mowy (IBM);
5) lata 1993 – obecnie – integracja dotychczasowych osiągnięć i metod:
   a) metody statystyczne w symbolicznych metodach analizy języka na wszystkich poziomach,
   b) systemy ekstrakcji informacji,
   c) komercyjne zastosowanie wyników badań (na komputerach domowych): rozpoznawanie mowy, korektory ortograficzne i gramatyczne[4].

Wybrane kierunki działań przetwarzania języka naturalnego

Zestawienie historyczne zaprezentowane powyżej prezentuje bogaty zasób wypracowanych na potrzeby przetwarzania języka naturalnego metod i narzędzi badawczych oraz ukazuje, że ta młoda dziedzina rozwijała się od samych początków bardzo prężnie. Poniżej zostaną zaprezentowane wybrane, popularne i sprawdzone metody NLP związane z tematyką niniejszej pracy. Metody te należą do nurtu statystycznego, który (przypomnijmy) cechuje się stosunkowo niskimi kosztami operacyjnymi analizy. Jednym z najstarszych zastosowań automatycznego przetwarzania danych językowych jest wyszukiwanie informacji w dokumentach tekstowych, kolejnym, wnoszącym przydatne rozwiązania, jest grupowanie dokumentów.

Wyszukiwanie informacji w dokumentach

W pracy *An introduction to information retrieval* jej autorzy w następujący sposób definiują pojęcie w y s z u k i w a n i e i n f o r m a c j i :

[4] Dane do zestawienia za: A. Mykowiecka: *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*. Warszawa 2007, s. 17-18, 329-343; por. również D. Jurafsky, J. H. Maritn: *Speech…*, op. cit., s. 10-15.

Wyszukiwanie informacji (IR) jest znajdowaniem materiału (najczęściej dokumentów) w postaci niestrukturalnej (przeważnie tekstu) w dużych zbiorach (zazwyczaj przechowywanych komputerowo), które zaspokajają potrzeby informacyjne[5].

Ch. Manning i pozostali autorzy zdefiniowane tak wyszukiwanie informacji przeciwstawiają modelowi wyszukiwania strukturalnego, stosowanego najczęściej w bazach danych (m.in. relacyjnych) lub w zautomatyzowanych katalogach bibliotecznych. Wyszukiwanie w zbiorach informacji strukturalnej wymaga znajomości struktury wykorzystanej do przechowywania danych, przeznaczenia poszczególnych pól oraz powiązań zachodzących pomiędzy różnymi elementami rekordu. Proces wyszukiwania polega między innymi na wskazaniu pola, którego zawartość ma zostać porównana do zapytania oraz sposobu lub metody porównawczej, jest więc dostępny dla osób przeszkolonych w wyszukiwaniach tego typu. Metodologia IR zakłada w pełni swobodne przeszukiwanie pełnego tekstu oraz ewentualnych pól metainformacyjnych dokumentu. Zapytania budowane są zazwyczaj w postaci listy słów bądź wyrażeń kluczowych opisujących informacje, na których zależy użytkownikowi. Wynikiem takiego wyszukiwania jest lista dokumentów zawierających wskazane w zapytaniu wyrażenia. W przypadku wyszukiwania swobodnego zbiór dokumentów do przeszukania nie musi być wstępnie opracowany, a przeglądanie i porównywanie zawartości dokumentów tekstowych jest procesem w pełni zautomatyzowanym. Pozwala to na obniżenie kosztów samego procesu przetwarzania dokumentów poprzez pominięcie etapu opracowania rzeczowego i formalnego dokumentu. Założenia wyszukiwania informacji (IR) zostały w praktyce zaimplementowane w wyszukiwarkach sieciowych. Intuicyjne interfejsy użytkownika wyszukiwarek oraz możliwość wskazania dokumentów jedynie na podstawie słów kluczowych występujących w treści pozwoliły na swobodne prowadzenie wyszukiwania informacji przez miliony użytkowników Internetu[6].

Kolejna oszczędność kosztów operacyjnych, a jednocześnie racjonalizacja procesu wyszukiwania pełnotekstowego została osiągnięta po wprowadzeniu plików indeksów odwróconych. W plikach tych przechowywana jest informacja o lokalizacji każdego wystąpienia każdego tokenu lub słowa we wszystkich dokumentach kolekcji. Proces odwzorowania lokalizacji poszczególnych słów i tokenów nazywany jest procesem indeksowania. Przeprowadzany jest w momencie akwizycji dokumentu do zbioru. System wyszukiwawczy może odwołać się bezpośrednio do wskazanego przez użytkownika słowa lub wyrażenia kluczowego i w krótkim czasie wskazać wszystkie dokumenty zawierające dane wyrażenie, bez konieczności każdorazowego analizowania treści dokumentów dla poszczególnych zapytań od użytkowników.

Łatwość dostępu oraz zastosowania tej formy wyszukiwania informacji nie jest jedyną zaletą IR. Autorzy pracy *An introduction...* wskazują, że kolejnym

[5] Tłumaczenie własne na podstawie definicji: *Information retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers).* Za Ch. D. Manning, P. Raghavan, H. Schütze: *An introduction to Information Retrieval*. Cambridge 2009 s. 1. [on-line]. [dostęp: 17. 08.2009]. Dostępny w World Wide Web: <http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>.

[6] Ibidem, s. 2-3.

udogodnieniem jest wyszukiwanie pełnotekstowe, które uniezależnia systemy informacyjno-wyszukiwawcze od danych przechowywanych w postaci strukturalnej. Pozwala to na przechowywanie dokumentów w postaci tekstu, bez tworzenia i wypełniania treścią specjalnych pól, jak się to dzieje w systemach bazodanowych. Innym wskazywanym zastosowaniem jest możliwość wyszukiwania łącznego informacji w różnych elementach dokumentu, na przykład w tytule oraz w treści. IR można również stosować do filtrowania i grupowania dokumentów w zbiorze w zależności od ich zawartości[7].

Systemy wyszukiwawcze

W tej samej pracy został zaproponowany podział systemów wyszukiwawczych ze względu na skalę operacyjną, sprowadzającą się w tym przypadku do ilości dokumentów w kolekcji. Jedną z wymienionych kategorii są wyszukiwarki internetowe, indeksujące miliony dokumentów. Problemy występujące w wyszukiwaniu na tak ogromną skalę to między innymi pozyskiwanie dokumentów do indeksowania zawartości, przechowywanie zaindeksowanej treści oraz zapewnienie efektywnego przetwarzania uzyskanych w ten sposób danych. Ponadto należy rozważyć specyfikę dokumentów webowych, m.in. hiperłącza, czy też próby zafałszowania treści witryn w celu zwiększenia pozycji na listach rankingowych wyszukiwarek. Kolejną wskazaną kategorią są wyszukiwarki indywidualne, instalowane na komputerach osobistych. Programy tego typu domyślnie w sposób ciągły indeksują zawartość plików na komputerach użytkowników i odwzorowują ją również w plikach indeksów odwróconych. Oprócz aplikacji systemowych do kategorii tej można zaliczyć również programy klienckie poczty elektronicznej. Aplikacje te oprócz wyszukiwania tekstu w treściach wiadomości oferują również funkcje klasyfikacji tekstu (ang. *clustering*) wskazując spam, który dociera na konta poczty elektronicznej. W tym przypadku klasyfikacja odbywa się na podstawie indeksowanej treści wiadomości i wykrycia typowych dla spamu słów kluczowych. Ponadto użytkownik programów pocztowych ma możliwość klasyfikowania pozostałych wiadomości według różnych kryteriów jak nazwa nadawcy, wątki tematyczne, itp. Problemy tego typu wyszukiwania to duża liczba formatów plików, w jakich użytkownicy przechowują swoje dane, zapewnienie ciągłej pracy systemu wyszukiwawczego bez zbytniego obciążania systemu operacyjnego, który powinien cały czas pracować w sposób wydajny i wygodny dla użytkownika oraz efektywne zarządzanie zajmowanym przez indeksy miejscem na dysku twardym. Ostatnią z proponowanych kategorii, mieszczącą się pomiędzy poprzednimi dwiema, są wyszukiwarki korporacyjne. Wyszukiwania korporacyjne zasięgiem wykraczają poza komputery indywidualne, ale nie obejmują całej sieci Internet. Ograniczone są do sieci korporacyjnych, intranetów, bądź domen webowych. W systemach tego typu zbiorem danych są zazwyczaj dokumenty wewnętrzne organizacji, przechowywane w scentralizowanym systemie plików, np. na wydzielonym serwerze plików lub baz danych, ponieważ ten typ wyszukiwania obejmuje również indeksowanie zasobów wewnętrznych baz danych. W przypadku

[7] Ibidem, s. 2.

wyszukiwania korporacyjnego również można spotkać wiele formatów plików i różne źródła danych. Konieczne jest wdrożenie przetwarzania rozproszonego oraz implementacja takiego sposobu indeksowania treści zasobów, który nie będzie blokował dostępu do zasobów innym użytkownikom sieci korporacyjnej[8].

Modele wyszukiwania informacji

Można wyróżnić dwa podstawowe podejścia do wyszukiwania informacji: model logiki Boola (ang. *Boolean Logic Model*, BLM) oraz model rankingowy (ang. *ranked-output model*). W przypadku modelu Boolowskiego zapytanie buduje się ze słów lub fraz połączonych operatorami logicznymi. Metoda ta pozwala wyłonić ze zbioru dokumentów te, których treść spełnia zadany warunek. Model rankingowy pozwala ocenić podobieństwo treści dokumentów z treścią zapytania i utworzyć na tej podstawie listę rankingową dokumentów trafnych. Przy tworzeniu rankingów wykorzystywane są najczęściej następujące modele oceny podobieństwa[9]:

1) model wektorowy (ang. *Vector Space Model*, VSM),
2) model probabilistyczny (ang. *Probabilistic Model*, PM).

Obie metody można połączyć, wyszukując za pomocą algebry Boola dokumenty zgodne z zapytaniem, a następnie, oceniając stopień zgodności, przedstawić je użytkownikowi w postaci listy rankingowej.

Grupowanie dokumentów (*Clustering*)

Celem klasteryzacji dokumentów jest podział analizowanego zbioru na grupy (zwane też klastrami) jednorodne tematycznie, gdzie podstawą podziału jest treść dokumentu. W wyniku tego procesu otrzymuje się podzbiory dokumentów podobnych do siebie, a różniących się od dokumentów w pozostałych podzbiorach.

Autorzy pracy *An introduction to information retrieval* opisują grupowanie jako najpowszechniejszą formę uczenia się nienadzorowanego – zdobywanie wiedzy przez systemy komputerowe bez udziału człowieka. Przypisanie dokumentu do wybranej klasy odbywa się automatycznie, podobnie zresztą jak wskazanie klas tematycznych dla analizowanego zboru dokumentów. Grupowanie przeciwstawiane jest klasyfikacji, gdzie dokumenty przypisywane są do ustalonych *a priori* klas, w związku z czym klasyfikacja nazywana jest uczeniem się nadzorowanym[10].

Podstawą wskazywania klas i przypisywania do nich dokumentów jest wyznaczenie odległości między dokumentami w przestrzeni dwuwymiarowej. Badacze zagadnienia wskazują dwa poziomy grupowania: g r u p o w a n i e

[8] Ibidem, s. 2.

[9] Por. A. Kempa: *Zastosowanie rozszerzonej metodologii wnioskowania na podstawie przypadków – Textual CBR w pracy z dokumentami tekstowymi*. [online]. [dostęp: 10.06. 2009]. Dostępny w World Wide Web: <http://www.swo.ae.katowice.pl/_pdf/221.pdf>.

[10] Ch. D. Manning, P. Raghavan, H. Schütze: *An introduction...*, op. cit., s. 349.

płaskie oraz grupowanie hierarchiczne. Pierwszy rodzaj tworzy klasy niepowiązane ze sobą żadnymi relacjami, natomiast drugi dostarcza układ hierarchiczny klas, ze wskazanymi relacjami zależności między poszczególnymi klasami. Metody grupowania hierarchicznego można podzielić na nieostre, gromadzące (ang. *hard*) oraz ostre, wyróżniające (ang. *soft*). W przypadku ostrych algorytmów grupowania każdy dokument przypisywany jest tylko i wyłącznie do jednej klasy, natomiast algorytmy nieostre mogą przypisać dokument do kilku klas. Klasy powstałe w wyniku grupowania hierarchicznego można przedstawić w postaci dendrogramu (drzewa zależności)[11].

Ze względu na różne dostępne metody przyporządkowywania dokumentów do klas wyróżnia się kilka rodzajów klasyfikacji. Charakterystyki wyszukiwawcze dokumentów przedstawione w postaci wektorów pozwalają przyporządkować konkretne dokumenty do zdefiniowanych, stałych klas. Możliwe jest również automatyczne generowanie klas tematycznych na podstawie dodatkowej analizy podobieństwa pomiędzy dokumentami relewantnymi do zapytania.

Stałe, zdefiniowane wcześniej klasy wykorzystywane są z kolei w procesie klasyfikacji, który z poziomu mechanizmów porównujących ma wiele wspólnego z grupowaniem. W procesie klasyfikacji można wskazać dwa główne nurty: klasyfikację opartą o wzorce oraz bezwzorcową.

Klasyfikacja oparta o wzorce

Jest to najprostszy sposób klasyfikowania dokumentów, polegający na przypisaniu do jednej z ustalonych klas tematycznych. Dla każdej określonej klasy dokumentów należy wskazać wzorce pozwalające przypisać treść do klasy. Metoda ta świetnie sprawdza się w katalogach internetowych, które operują właśnie na zbiorach ustalonych klas. Klasy tematyczne należy określać na tyle elastycznie, żeby można było przypisać do nich dokumenty nie w pełni klasyfikujące się do danej klasy. Ewentualnie można utworzyć klasę INNE, ale nie jest to rozwiązanie eleganckie i w pełni profesjonalne.

Klasyfikacja bezwzorcowa

Pewną alternatywą jest automatyczne tworzenie klas dostosowanych do posiadanej kolekcji dokumentów. Metoda ta nadaje się dobrze do zastosowania w wyszukiwarkach internetowych, ponieważ opiera się na zbiorze niesklasyfikowanych dokumentów. Dopiero na podstawie charakterystyk treściowych oraz rozkładzie częstotliwości podobnych reprezentacji system generuje klasy i przypisuje do nich poszczególne dokumenty.

W przypadku zamkniętych, kontrolowanych systemów informacyjno-wyszukiwawczych dostępny zestaw słów kluczowych jednoznacznie lokalizuje zakres treściowy dokumentu, ułatwiając użytkownikowi wybór najbardziej relewantnego. Wyszukiwarki internetowe pracują w środowisku otwartym, bez obowiązu-

[11] O grupowaniu dokumentów por. Ibidem, s. 349 oraz D. Jurafsky, J. H. Martin: *Speech...*, op. cit., s. 679.

jących powszechnie reguł tworzenia charakterystyk wyszukiwawczych, dlatego w związku z potrzebą standaryzowania i ujednolicenia sposobu komunikacji wyszukiwarki z użytkownikiem (w zakresie prezentacji listy wyników) stosowano różne metody prezentacji treści dokumentów zgodnych z zapytaniem. Ze względów praktycznych (poziomu akceptacji przez użytkowników) najpopularniejszą metodą prezentowania tematyki dokumentu użytkownikowi jest wyświetlenie kilku pierwszych zdań lub kilku zdań sąsiadujących z miejscem zlokalizowania w treści słowa kluczowego z pozycji zwróconych w odpowiedzi na kwerendę.

Podsumowanie

Artykuł miał na celu zapoznanie czytelnika z rozwojem badań i możliwościami zastosowania badań nad przetwarzaniem języka naturalnego. Wyniki tych badań znajdują szerokie zastosowanie we współczesnym przetwarzaniu i zarządzaniu informacją. Aplikacje stosujące wypracowane przez badaczy NLP prawidłowości lingwistyczne są wykorzystywane w świecie biznesu, ale również w coraz większym zakresie w domenie publicznej, na przykład w bibliotekach cyfrowych. Znajomość praw i zależności lingwistycznych może przyczynić się do bardziej precyzyjnego dostarczania użytkownikowi potrzebnej mu informacji, ale też do obniżenia kosztów technicznych, finansowych i czasowych przetwarzania i zarządzania informacją.

Bibliografia

1. Jackson P., Moulinier I.: *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization.* Amsterdam/Philadelphia 2002.
2. Jurafsky D., Martin J. H.: *Speech and language processing. An introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.* New Jersey 1999.
3. Manning Ch. D., Schütze H.: *Foundations of Statistical Natural Language Processin.* Cambridge 1999.
4. Mykowiecka A.: *Przegląd systemów automatycznej generacji tekstów w języku naturalnym.* Warszawa 1987. Prace IPI PAN nr 614.
5. Mykowiecka A.: *Generacja zdań w języku polskim na podstawie reprezentacji ich semantyki.* Warszawa 1988. Prace IPI PAN nr 644.
6. Mykowiecka A.: *Text planning.* Warszawa 1989. Prace IPI PAN nr 665.
7. Mykowiecka A.: *Planowanie struktury tekstu przy wykorzystaniu RTS.* Warszawa 1994. Prace IPI PAN nr 756.
8. Mykowiecka A.: *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym.* Warszawa 2007.
9. Piasecki M.: *Cele i zadania lingwistyki informatycznej.* W: *Metodologie językoznawstwa. Współczesne tendencje i kontrowersje.* Pod red. P. Stalmaszczyka. Kraków 2007.
10. Przepiórkowski A.: *Powierzchniowe przetwarzanie języka polskiego.* Warszawa 2008.
11. *Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych.* Oprac. B. Bojar. Warszawa 2002.

Summary

The article presents the development of researches in natural languages' processing. It discusses beginnings of these studies, as well as changes in either the research methods, or their range and scope during last 60 years. The author describes two the most popular research methods: statistical analysis and generative grammars. He also evaluates briefly advantages and disadvantages of them both. Additionally, the article presents selected modern trends of NLP activities. Among the most popular current research trends in natural language processing, one can mention information retrieval or documents' grouping.

ZAPOŻYCZENIA JĘZYKOWE W SYSTEMIE JĘZYKÓW HASEŁ PRZEDMIOTOWYCH

Anna Stanis
Biblioteka Uniwersytecka w Warszawie

Język haseł przedmiotowych, globalizacja, zapożyczenia językowe, internacjonalizmy, jhp KABA, jhp BN, Słownik haseł przedmiotowych z dziedzin policyjno-prawnych

Koniec XX i początek XXI w. jest czasem wielkich przemian. Zmiany polityczne, ekonomiczne, społeczne spowodowały przemiany kulturowe, pojawienie się nowych zachowań, stylu bycia i idącą za tym zmianę mentalności, odcisnęły też wyraźne piętno na poszczególnych językach. Wśród wielu zmian, jakie zachodzą w tych językach, najbardziej widoczny jest przyrost słownictwa obcego, a szczególnie zapożyczeń angielskich i internacjonalizmów.

Według Encyklopedii Językoznawstwa Ogólnego „zapożyczenie językowe to element przejęty z obcego języka. Najczęściej jest nim wyraz rzadziej prefiks lub sufiks" natomiast „internacjonalizm [to] wyraz albo wyrażenie frazeologiczne występujące (w postaci adaptowanej do ortografii, struktury fonologicznej i gramatycznej danego języka) w wielu językach świata[1].

Wzmożony napływ zapożyczeń i internacjonalizmów, jaki obserwujemy w ostatnim dwudziestoleciu, jest procesem, który wpływa na system słowotwórczy współczesnego języka polskiego. Wiele z nich jest często używanych w ogólnie dostępnych mediach (telewizji, prasie).

Pierwszym z powodów, który wpłynął na intensywne przejmowanie zapożyczeń do słownictwa dzisiejszej polszczyzny, była chęć jej wzbogacenia (nowe zjawiska – nowe wyrażenia), drugim powodem tak masowej skali pojawiania się elementów obcych jest uleganie modzie językowej. Dotyczy to zarówno anglicyzmów (amerykanizmów), jak i wyrazów pochodzących z innych języków: z francuskiego, np. *balejaż*, z włoskiego np. *paparazzi*, z hiszpańskiego np. *macho,* z japońskiego np. *jacuzzi, sushi*, z arabskiego np. *dżihad.*

Dominacja języka angielskiego

Dziś status angielszczyzny jest dominujący: korzysta z niej trzecia część ludności świata, mimo że jest macierzystym językiem dla 380 mln ludzi, dominuje w sferze książek, czasopism, filmów; ok. 80% treści zamieszczanych w Internecie ma angielską formę językową, chociaż 44% użytkowników Internetu mówi innymi językami. Angielski jest używany przede wszystkim przez elity techniczne, gospo-

[1] *Encyklopedia językoznawstwa ogólnego*. Wrocław, Warszawa, Kraków 1993.

darcze i polityczne[2]. Język ten rozpowszechniony jest tak bardzo, że został nazwany „łaciną Internetu" („Łacina Internetu" – uniwersalny kod, przez który rozumie się język angielski)[3].

Piotr Wróblewski przeprowadził anonimową ankietę w środowisku młodego pokolenia Polaków na temat zmian zachodzących we współczesnej polszczyźnie, badaniom poddano 100 osób, studentów I roku polonistyki Uniwersytetu w Białymstoku oraz słuchaczy II i III roku Kolegium Nauczycielskiego w Suwałkach. W odpowiedzi na pytanie dotyczące zmian zachodzących w języku 98% ankietowanych odpowiedziało, że zauważa zmiany we współczesnej polszczyźnie. Większość (66%) z tych osób do takich zmian zaliczyło zapożyczenia z języków obcych (w tym 34% mówiło o zapożyczeniach z języków obcych, 32% – o zapożyczeniach z języka angielskiego)[4].

Według Krystyny Waszakowej różnorodne dziedziny życia, w których nastąpiła ekspansja wyrazów obcych i tworzonych neologizmów, można podzielić na cztery ogólne kategorie zjawisk:

„1. Przeobrażenia ustrojowe, polityka w tym m.in. uwolnienie Polski spod wpływu ZSRR, przełamanie monopolu partii, zniesienie cenzury, tworzenie podstaw systemowych demokratycznego państwa, jego instytucji, form działania, dążenie Polski do uczestniczenia w strukturach europejskich, zwłaszcza w Unii Europejskiej – wyrażenia typu antytotalitaryzm, brukselizacja, euro deputowany, euro integracja, lobbista, postkomunista, uniokracja.

2. Ekonomia, finanse, nowe formy bankowości, tworzenie gospodarki rynkowej [...] – wyrażenia typu developer, fiskalizacja, leasing, monetaryzacja, refinansować.

3. Nowe zjawiska techniczne, a więc wszystko to co wiąże się z rozwojem cywilizacji i techniki, szczególnie informacji i komputeryzacji, telefonizacji, Internetu – wyrażenia typu billing, blog chip, cybernauta, e-biznes, e-mail, e-podpis, faks, hacker, infolinia, internauta, laptop, modem, monitoring, notebook, skaner, telebim.

4. Zdrowie, sport, styl życia, formy wypoczynku [...] – wyrażenia typu aerobik, bungee, cheesburger, dogoterapia, hipermarket, hydromasaż, jogging, lifting, peeling, skateboarding, survival"[5].

Do wyżej wymienionych kategorii można dodać jeszcze terminy z dziedziny muzyki, gdzie występuje stosunkowo dużo internacjonalizmów pochodzenia francuskiego, włoskiego, angielskiego. Dotyczy to zarówno nazw form muzycznych, nazw instrumentów, nazw tańców, jak i nazw stylów w muzyce rozrywkowej.

Ta sama autorka omówiła dwie zasadnicze grupy zapożyczeń: zapożyczenia leksykalne oraz zapożyczenia semantyczne (pominięto inne bardziej szczegółowe typy zapożyczeń, jako mniej istotne z punktu widzenia przedstawionych rozważań).

Zapożyczenia leksykalne (lub też właściwe) to wyrazy obce o różnym stopniu adaptacji w języku polskim, natomiast termin zapożyczenia semantyczne odnosi się do wyrazów, które pod obcym wpływem zmieniły swe dotychczasowe znaczenie.

---

[2] A. Fishman: *The New English Order*. „Foreign Policy" 1998/1999, no. 113 s. 26-40.

[3] M. Czarnecka: *Łacina Internetu* „Wprost" 1997, nr 20.

[4] P. Wróblewski: *O możliwości przewidywania kierunku ewolucji języka polskiego*. W: *Przyszłość języka*. Białystok 2001, s. 75-84.

[5] K. Waszakowa: *Przejawy internacjonalizacji w słowotwórstwie współczesnej polszczyzny*. Warszawa 2005, s.12, s. 19.

Włączanie do poszczególnych języków zapożyczeń o angielskim rodowodzie, zaadaptowanych już przez kilka języków, prowadzi do ich integracji, a raczej globalizacji językowej. Problemowi polskiej terminologii języków informacyjno-wyszukiwawczych, w tym problemowi zapożyczeń i internacjonalizmów w dobie globalizacji, obszerny artykuł poświęcił Wiesław Babik[6], podkreślając, jak szczególne miejsce zajmują internacjonalizmy w terminologii naukowej, technicznej i zawodowej.

„Międzynarodowa Organizacja Neologizmów Terminologicznych (obecnie Międzynarodowa Organizacja Terminologiczna – MOT), działająca na rzecz umiędzynarodowienia terminologii naukowej, technicznej i zawodowej, opublikowała w 1996 r. *Multilingual Dictionary of International Terms* zawierający internacjonalizmy z dziewięciu dziedzin w tym lingwistyki i informatyki".

Języki haseł przedmiotowych wobec przemian w językach naturalnych

Zjawisko wzmożonego napływu zapożyczeń i internacjonalizmów w polskim języku naturalnym, jakie obserwujemy w ostatnim dwudziestoleciu, nie pozostaje bez wpływu na kształtowanie systemów słowotwórczych tworzonych w Polsce języków haseł przedmiotowych. Każdy język haseł przedmiotowych jest ściśle powiązany z konkretnym językiem naturalnym, ponieważ budowany jest w oparciu o słownictwo języka naturalnego, czerpane z różnego rodzaju źródeł: norm terminologicznych, słowników językowych, encyklopedii, opracowywanych dokumentów, a przy braku innych specjalistycznych źródeł, z Internetu. Tworzony w ten sposób zasób leksykalny o słownictwie paranaturalnym, (quasinaturalnym) – równokształtnym z wyrażeniami języka naturalnego – odzwierciedla wszystkie przemiany zachodzące w tym języku.

Problem zapożyczeń i internacjonalizmów, a także ich wpływ na wybór terminu preferowanego na hasło wzorcowe w językach haseł przedmiotowych, postanowiłam prześledzić na przykładzie uniwersalnych języków haseł przedmiotowych: jhp BN, jhp KABA oraz specjalistycznego *Słownika haseł przedmiotowych z dziedzin policyjno-prawnych* (dalej: SHP) stosowanego w bibliotece Wyższej Szkoły Policji w Szczytnie. Hasła do ostatniego ze słowników tworzono w oparciu o język haseł przedmiotowych Biblioteki Narodowej. Założeniem autorów słownika było uściślenie haseł przedmiotowych z dziedzin policyjno-prawnych, a w szczególności policji, przestępczości, kryminalistyki, kryminologii, medycyny sądowej i prawa karnego. Słownik obejmuje również, w wybranym zakresie, hasła z dziedzin takich jak: psychologia, pedagogika, polityka, zarządzanie, filozofia, historia oraz pozostałe dziedziny prawa. Wybór tego słownika podyktowany był faktem, iż opracowanie zbiorów specjalistycznych wymaga znacznego uszczegóławiania haseł, co może stwarzać różne problemy terminologiczne. Hasła utrzymywane są w bazie kartoteki haseł wzorcowych w systemie ALEPH, obecnie baza liczy 26,5 tys. haseł.

Przy wyborze terminów na jednostki leksykalne języka haseł przedmiotowych istotne znaczenie ma wypracowanie ogólnych zasad/reguł, jakimi należy się kierować. Wspólną dla wszystkich tych języków zasadą jest zgodność z regułami języka polskiego, co potwierdzają cytowane materiały metodyczne.

[6] W. Babik: *Polska terminologia języków informacyjno-wyszukiwawczych w dobie globalizacji*. „Zagadnienia Informacji Naukowej" 2006, nr 1(87), s. 3-13.

„W każdej sytuacji dąży się do tego, by redakcja językowa haseł wzorcowych zapewniała zachowanie podstawowych zasad języka polskiego.[...] Jest to szczególnie trudne gdy brakuje odpowiednich słowników specjalistycznych w języku polskim, zdarza się także, że brakuje odpowiednich terminów w słownikach. [...] Proces tworzenia leksyki KABA polega na znajdowaniu najwłaściwszej dla danego znaczenia formy językowej poprawnej w języku polskim"[7].

„Współczesny język haseł przedmiotowych Biblioteki Narodowej stanowi kontynuację języka i metodyki opracowanej przez Adama Łysakowskiego" napisała Jadwiga Sadowska[8]. Na stronie internetowej Biblioteki Narodowej poświęconej jhp BN w rozdziale „Zasady doboru i redakcji merytorycznej haseł", potwierdzono tę zasadę: „Autorki *Słownika*, Ewa Stępniakowa i Janina Trzcińska, podtrzymując podstawową zasadę metodyki Adama Łysakowskiego, tj. zasadę wyszczególniającego formułowania tematów i uogólniającego formułowania określników, określiły zasady tworzenia JHP BN, które w większości są stosowane do dziś: słownictwo ma charakter uniwersalny, terminy służące jako tematy powinny odpowiadać polskiej terminologii z danej dziedziny i pochodzić z wiarygodnych polskich źródeł informacji. Początkowo jako źródła tematów preferowano najbardziej popularne, jednotomowe encyklopedie, wychodząc z założenia, że jest to źródło najbardziej dostępne dla przeciętnego użytkownika. Obecnie podstawowym zasobem terminów są wielkie encyklopedie uniwersalne oraz encyklopedie dziedzinowe i słowniki terminologiczne, w mniejszym stopniu monografie i artykuły naukowe, wydawnictwa urzędowe, akty prawne. Terminy są ponadto weryfikowane za pomocą aktualnego słownika języka polskiego"[9].

Oprócz zasady zgodności z regułami języka polskiego, oba jhp przy redagowaniu haseł wzorcowych preferują: rozpowszechnienie nazwy w piśmiennictwie, bazach danych, zasobach internetowych, poprawność językową, aktualne stosowanie, zwięzłość (pierwszeństwo ma forma krótsza pod warunkiem, że jest dostatecznie zrozumiała – jhp KABA), jednoznaczność, pochodzenie językowe (pierwszeństwo mają formy o polskim źródłosłowie).

Oba języki dopuszczają możliwość stosowania zapożyczeń w ściśle określonych warunkach. W przypadku jhp BN argumentacja jest następująca „hasła do opisu nowych zjawisk społecznych, nowych technologii, aktualnych wydarzeń politycznych zwykle nie są jeszcze ujmowane i definiowane w publikacjach typu encyklopedycznego i słownikowego, stąd problem ze znalezieniem odpowiedniej, a nie całkiem potocznej nazwy. Hasła często bywają więc formułowane na podstawie innych dostępnych źródeł – zagranicznych baz danych, publicystyki, zasobów internetowych; czasem po utrwaleniu się bardziej właściwego nazewnictwa bywają zmieniane. Pewna grupa terminów, np. z zakresu nauk społecznych, ekonomicznych czy technicznych, które nie mają jeszcze polskiego odpowiednika jest wprowadzana w brzmieniu obcym, najczęściej angielskim (np.:Timesharing, Public relation, Hosting, Outsourcing, Streetworking)"[10].

[7] *Język haseł przedmiotowych KABA. Zasady tworzenia słownictwa*. Pod red. Teresy Głowackiej. Warszawa, 2000.

[8] J. Sadowska: *Język haseł przedmiotowych Bibłioteki Narodowej. Poradnik*. Warszawa 2001, s. 7.

[9] Biblioteka Narodowa. Język Haseł Przedmiotowych. [online]. [Dostęp 20.08.2010] <http://www.bn.org.pl/download/document/1280922227.pdf>.

[10] Ibidem.

W przypadku języka KABA dopuszcza się, a nawet nakłada obowiązek, stosowania wyrażeń zapożyczonych z języków obcych w przypadkach, gdy nie mają one odpowiedników w języku polskim, np. Art Brut, Public relations, Streamer, lub gdy odpowiedniki te są mniej popularne i/lub używane w piśmiennictwie rzadziej od wyrażeń zapożyczonych np. Alergia zamiast Uczulenie (medycyna), Kosaćce zamiast Irysy.

Zwykle status hasła wzorcowego uzyskuje forma powszechnie używana. Wyrażenia nieprzyjęte za wzorcowe, ale brane pod uwagę przy redagowaniu formy hasła wzorcowego (w tym nazwy łacińskie), są wiązane z formą przyjętą (wzorcową) relacją ekwiwalencji. W obu językach – jhp BN i jhp KABA – przyjęto zasadę zachowania naturalnego szyku wyrażenia preferowanego jako hasło wzorcowe tzn. szyku porzeczownikowego. Na pierwszym miejscu występuje rzeczownik, a po nim inne wyrażenia (przymiotniki, imiesłowy, liczebniki) określające ten rzeczownik. W obu językach istnieją od tej zasady wyjątki dla stałych związków frazeologicznych np. Czarne dziury.

W przypadku SHP ogólne zasady wyboru formy hasła wzorcowego są takie same jak w jhp BN.

Na wybór terminu (w tym zapożyczenia) jako hasła wzorcowego ma również wpływ analiza zagranicznych słowników języków haseł przedmiotowych.

Wybór zapożyczenia jako hasła wzorcowego następuje w przypadkach, gdy:

a) polski odpowiednik zapożyczenia nie przyjął się w piśmiennictwie lub jest zbyt wieloznaczny, np.:

Reenginering – w języku KABA przyjęto jako hasło wzorcowe, umieszczając w relacji ekwiwalencji termin Reinżynieria;

Design – stosowany też w pisowni dizajn. W języku KABA przyjęty jako hasło wzorcowe, podobnie jak w językach LCSH i RAMEAU. Wprowadzono także termin Wzornictwo przemysłowe i w relacji ekwiwalencji do niego termin Design (przemysł). W słownictwie jhp BN przyjęto Wzornictwo przemysłowe, ale dla kategorii osób parających się tym zajęciem przyjęto hasło Designerzy;

Stalking – jhp BN, jhp KABA, SHP, przyjęły ten sam termin dla określenia pewnego typu zachowań przejawiających się nagabywaniem, osaczaniem, natrętnym uwielbianiem innych osób;

Dżihad – święta wojna islamska, w obu jhp BN i KABA przyjęte jako hasło wzorcowe;

World Wide Web – anglojęzyczne rozwinięcie skrótu WWW, tzw. Ogólnoświatowa Pajęczyna. To dosłowne polskie tłumaczenie nie przyjęło się: jhp KABA przyjął World Wide Web, jhp BN hasło Portal internetowy, SHP – Strony WWW. Jak wynika z tego przykładu, nie ma jednoznacznie przyjętego odpowiednika w języku polskim;

Talk show – internacjonalizm – przyjęty jako hasło wzorcowe w językach: LCSH, RAMEAU, jhp BN, jhp KABA;

Animaloterapia – termin przyjęty jako hasło wzorcowe w jhp BN – w leksyce jhp KABA nie ma takiego terminu;

Jogging – internacjonalizm, przyjęty w LCSH, RAMEAU i jhp KABA, w leksyce jhp BN nie ma takiego terminu;

Alibi – termin przyjęty w jhp BN, nie występuje w słownictwie KABA;

Curriculum vitae – termin przyjęty w jhp KABA i RAMEAU jako hasło wzorcowe (w jhp KABA na podstawie *Nowy Leksykon* PWN Warszawa, 1998); w jhp BN występuje jako termin ekwiwalentny do hasła Biografie, w SHP jako termin ekwiwalentny do hasła Życiorysy urzędowe;

Biznesplan – termin przyjęty jako hasło wzorcowe w leksyce KABA (na podstawie *Słownika Współczesnego Języka Polskiego)*; w jhp BN i SHP jest terminem ekwiwalentnym do hasła przedmiotowego rozwiniętego Przedsiębiorstwo – planowanie;

Casting – termin przyjęty w jhp BN jako hasło wzorcowe, w jhp KABA termin ekwiwalentny do hasła przedmiotowego rozwiniętego Kino – obsada;

Leasing – termin przyjęty w jhp KABA, jhp BN i SHP;

Outsourcing – termin przyjęty w jhp KABA, jhp BN i SHP.

b) w polskiej rzeczywistości nie występuje zjawisko, dla którego wprowadzono termin, np.:

Fabliaux – przyjęty w tej samej formie co w RAMEAU również w językach LCSH i KABA, gatunek w literaturze francuskiej XII-XIV. W słownictwie BN i SHP nie ma takiego terminu;

Kasyda – gatunek literatury arabskiej, termin przyjęty zarówno w jhp BN jak i w jhp KABA;

Ramadan – miesiąc z kalendarza muzułmańskiego, przyjęty zarówno w jhp BN jak i w jhp KABA;

Commentaire compose (dysertacja) – rodzaj komentarza w pracy naukowej francuskiej, przyjęty w jhp KABA w formie zaproponowanej przez RAMEAU. W słownictwie BN i SHP nie ma takiego terminu;

Sumo – zapasy japońskie, termin przyjęty zarówno w jhp BN jak i w jhp KABA;

Sushi – potrawa japońska, termin przyjęty zarówno w jhp BN jak i w jhp KABA;

Judo – japońska sztuka walki, termin przyjęty w jhp KABA, w jhp BN i SHP również termin przyjęty ale w wersji spolszczonej Dżudo.

Jednym z najważniejszych problemów przy opracowywaniu słownictwa języków haseł przedmiotowych jest ustalanie ekwiwalentów terminów wybranych na hasło wzorcowe.

Jeśli termin zapożyczony nie został wybrany jako hasło wzorcowe pełni funkcję terminu ekwiwalentnego np.:

Souvenirs – przyjęty w RAMEAU i LCSH, w języku KABA jest ekwiwalentem hasła Pamiątki (przedmioty). W słownictwie jhp BN nie ma takiego terminu;

Product placement – w języku KABA termin ekwiwalentny do hasła wzorcowego Lokowanie produktu (polski termin został uznany za przyjęty w piśmiennictwie na podstawie dostępnych źródeł), w jhp BN Product placement jest hasłem wzorcowym a Lokowanie produktu terminem ekwiwalentnym;

Bodyguard – w języku KABA termin ekwiwalentny do hasła wzorcowego Agenci ochrony osobistej, w jhp BN i w SHP przyjęto jako hasło wzorcowe Pracownicy ochrony mienia i osób;

Bilingwizm – w obu językach jhp KABA i jhp BN termin ekwiwalentny do hasła wzorcowego Dwujęzyczność;

Puzzle – w jhp BN jest terminem ekwiwalentnym do hasła wzorcowego Układanki, w jhp KABA termin przyjęty tak jak w LCSH i RAMEAU;

Gender roles – w jhp KABA termin ekwiwalentny do hasła Rola wynikająca z płci (*Encyklopedia psychologii*, Warszawa, 1998). W jhp BN nie ma takiego terminu.

e-[przedmiot] w jhp BN, jhp KABA, SHP

e-[przedmiot]	JHP KABA	JHP BN	SHP
e-bank		Bank internetowy	
e-biznes	Handel elektroniczny	Gospodarka elektroniczna	
e-book	Książki elektroniczne	Książki elektroniczne	
e-czasopisma		Czasopisma elektroniczne	
e-duszpasterstwo	Internet w duszpasterstwie		
e-edukacja	Internet w edukacji		
e-firma		Przedsiębiorstwo internetowe	
e-gospodarka		Gospodarka elektroniczna	
e-government	Elektroniczna administracja publiczna		
e-hazard	Gry hazardowe w Internecie		
e-kontrakty		Umowy elektroniczne	Umowy elektroniczne
e-księgarnie		Księgarstwo internetowe	
e-learning	Internet w edukacji, Kształcenie online	Nauczanie na odległość	Nauczanie na odległość
e-mail	Poczta elektroniczna	Poczta elektroniczna	Poczta elektroniczna
e-mail art	Sztuka Internetu		
e-marketing	Marketing w Internecie	Marketing elektroniczny	
e-muzea		Muzea wirtualne	
e-papierosy	Papierosy elektroniczne	Papieros elektroniczny	
e-pieniądze		Pieniądz elektroniczny	
e-podpis	Podpis elektroniczny	Podpis elektroniczny	Podpis elektroniczny
e-pornografia	Pornografia internetowa		
e-przedsiębiorstwo		Przedsiębiorstwo internetowe	
e-śmieci	Odpady elektroniczne		
e-terapia	Internet w psychoterapii		
e-umowy	Umowy elektroniczne	Umowy elektroniczne	Umowy elektroniczne
e-urząd		Administracja elektroniczna	
e-usługi	Usługi sieciowe	Usługi elektroniczne	

Brak jakiegoś terminu z podanych wyżej przykładów w słownictwie omawianych języków nie oznacza jego eliminacji, najczęściej nie zostały jeszcze opracowane z powodu braku dokumentów na dany temat w bibliotece lub tak jak w

przypadku SHP dobór terminów ogranicza zakres tematyczny. Szczególnym problemem przy opracowywaniu słownictwa jhp jest ekspansja w języku naturalnym wyrazów z pierwszym członem -e, wzorowanych na międzynarodowym skrócie e-mail, w którym przymiotnik *electronic* zredukowano do elementu -e. Element -e występuje zarówno w zapożyczeniach, jak i strukturach tworzonych na gruncie rodzimym, a jego znaczenie rozszerzono na 'internetowy, wirtualny, związany z siecią elektroniczną'. Wszechobecny w słownictwie naturalnym, stanowi coraz większy problem również w językach haseł przedmiotowych. Generalnie w językach haseł przedmiotowych wyrazom z e- elementem wyznaczono funkcję terminu odrzuconego, zarówno w jhp BN jak i w językach KABA i SHP. Poniżej w tabeli podano przykłady złożeń z członem -e często spotykane w języku naturalnym wraz z ich odpowiednikami w wyżej wymienionych językach haseł przedmiotowych.

Poniżej złożenia z członem e- spotykane w źródłach, dla których nie znaleziono ekwiwalentów w żadnym z wymienionych jhp:

e-archiwum, e-autor, e-banking, e-bankowość, e-bay, e-biblioteka, e-bilet, e-biuletyn, e-biznesmen, e-czytanie, e-demokracja, e-dokument, e-donos, e-dyktando, e-ekonomia, e-film, e-gazeta, e-giełda, e-handel, e-kasyno, e-klient, e-konsumpcja, e-konto, e-literatura, e-nauka, e-notatki, e-państwo, e-papier, e-PIT-37, e-powieść, e-praca, e-przyszłość, e-publikacje, e-rada, e-rynek, e-sklep, e-spółka, e-sprawozdanie, e-student, e-studia, e-superkonto, e-szkoła, e-świat, e-teksty, e-wybory, e-wykluczenie, e-zakupy, e-zbiory, e-złodzieje.

Wyznaczenie e-złożeniom funkcji terminu ekwiwalentnego łączy się z akceptacją pojawienia się w jednym miejscu w indeksie wciąż rosnącego zbioru terminów z przedrostkiem e-.

Podobne problemy sprawia napływ do języka polskiego konstrukcji z euro- (szczególnie w dziedzinach ekonomii, bankowości, finansów, gospodarki, polityki, handlu, sportu). Konstrukcje zawierające euro- to głównie zapożyczenia (najczęściej z języka angielskiego) np. *euroczek* lub „hybrydy" zawierające internacjonalne euro- i element rodzimy np. *eurourzędnik*.

Słowniki języków haseł przedmiotowych są systemami otwartymi i muszą być stale modyfikowane. W przypadku wprowadzania nowych haseł wzorcowych (szczególnie zapożyczeń) istnieje konieczność dokonania zmian w relacjach z hasłami już istniejącymi. Poprawne „ulokowanie" zapożyczenia w leksyce jhp sprawia więcej trudności niż termin od dawna istniejący w języku naturalnym.

W praktyce większość jiw pełni dwie funkcje równocześnie: języka informacyjnego, służącego do charakteryzowania dokumentów oraz języka wyszukiwawczego, stosowanego do określania potrzeb informacyjnych użytkowników systemu. Realizacja funkcji języka informacyjnego ma wpływ na jego funkcję wyszukiwawczą i chociaż funkcja wyszukiwawcza jest brana pod uwagę przy tworzeniu języka, to w przypadku zapożyczeń dążenia twórców języka i oczekiwania jego użytkowników mogą być rozbieżne, a przyjęta w jhp zasada zgodności z regułami języka polskiego i źródłosłowu polskiego może być przez użytkowników ignorowana na rzecz preferowanych modnych zapożyczeń z innych języków. Większość użytkowników chciałaby znaleźć informację, posługując się pytaniem sformułowanym intuicyjnie, zaś ich ocena jhp często dokonywana jest przez pryzmat jednoznaczności wszystkich wyrażeń, łatwości nauczenia się i posługiwania się tym językiem.

Przenikanie się kultur i języków jest zjawiskiem naturalnym. We Francji rząd wspiera walkę z anglicyzmami, tworząc np. francuską nomenklaturę komputerową. Nie obawia się znacznego napływu zapożyczeń język rosyjski, przyjmując tylko terminologię związaną z funkcjonowaniem sprzętu i oprogramowania oraz pisząc cyrylicą obce wyrazy. Skoro „uchwalono ustawę dotyczącą ochrony języka ojczystego (rosyjskiego) nie trzeba niepokoić się napływem angielskiej terminologii"[11]. W Polsce w 1999 roku Sejm uznał język polski za dobro narodowe i uchwalił *Ustawę o języku polskim*[12]. O czystość polszczyzny ma dbać jako instytucja opiniodawczo-doradcza Rada Języka Polskiego, ale jej rola jest ograniczona.

Na temat spolszczania terminów z zakresu informatyki wypowiadał się Stanisław Lem, traktując humorystycznie próbę spolszczenia wyrazu interfejs na „międzymordzie" i pisząc: „ skoro Internet niezbyt przeze mnie lubiany tak się zadomawia, to angielskiego trzeba się uczyć, albowiem języki etnicznie lokalne tworzą silnie erodowane agresją angielską wysepki. Ja więc spolszczać na siłę nie zamierzam"[13].

Intensywne przejmowanie zapożyczeń, „umiędzynaradawianie" słownictwa dzisiejszej polszczyzny, a zatem również słownictwa jhp wynika z potrzeby jej wzbogacania w związku ze zmianami rzeczywistości. Chociaż wśród najnowszych zapożyczeń leksykalnych dominują anglicyzmy, nie znaczy to bynajmniej, że zupełnie nie ma wyrazów pochodzących z innych języków: z francuskiego, włoskiego, hiszpańskiego, japońskiego, rosyjskiego, arabskiego. W procesie globalizacji politycznej, gospodarczej, ekonomicznej dochodzi do zacieśniania kontaktów między państwami, narodami, w tym także kontaktów językowych, co przejawia się powiększeniem obszaru internacjonalizmów w poszczególnych językach i procesem unifikacji języków. Zjawisko przenikania wyrazów obcych do języka polskiego budzi mieszane uczucia – od tolerancji, po obawy o zachowanie tożsamości narodowej, kulturowej i językowej. Twórcy języków haseł przedmiotowych są wnikliwymi obserwatorami zmian w języku naturalnym podejmując decyzje mające wpływ na wybór terminów użytych jako hasło wzorcowe.

Bibliografia

1. Babik W.: *Termin i jego status w systemie leksykalnym języka informacyjno-wyszukiwawczego*. „Zagadnienia Informacji Naukowej" 1999, nr 1(73), s. 13-14.
2. Babik W.: *Polska terminologia języków informacyjno-wyszukiwawczych w dobie globalizacji*. „Zagadnienia Informacji Naukowej" 2006, nr 1(87), s. 3-13.
3. Chmiołek M.: *Słownik haseł przedmiotowych z dziedzin policyjno-prawnych stosowanych w bibliotece Wyższej Szkoły Policji w Szczytnie*. Szczytno, 2008.
4. *Encyklopedia językoznawstwa ogólnego*. Wrocław, Warszawa, Kraków, 1993.
5. *Język haseł przedmiotowych KABA. Zasady tworzenia słownictwa*. Pod red. T. Głowackiej. Warszawa, 2000.
6. Kurek-Kokocińska S.: *Związki języka naturalnego i języków informacyjno-wyszukiwawczych – na przykładzie języka haseł przedmiotowych*. „Zagadnienia Informacji Naukowej" 2004(83), s. 46-65.
7. *Przyszłość języka*. Pod red. S. Krzemień-Ojak, B. Nowowiejskiego. Białystok, 2001.

[11] G. Trifomova: *Anglojazyčnye zaimstvovanija russkij jazyk pererabatyvaet, ustaivajet i kodificiruet*. „Ex Libris" 2004, nr 28, s. 5-6.

[12] Dz. U. z 1999 r. Nr 90, poz. 999.

[13] St. Lem: *Bomba megabitowa*. Kraków 1999, s. 55.

8. Sadowska J.: *Język haseł przedmiotowych Biblioteki Narodowej. Poradnik.* Warszawa, 2001.
9. Waszakowa K.: *Przejawy internacjonalizacji w słowotwórstwie współczesnej polszczyzny.* Warszawa, 2005.

Summary

The article discusses how the phenomenon of increased inflow of borrowings in natural language being observed within last twenty years influences formative system of subject headings languages. Internationalisation and effect of globalization are mentioned as a trend indicating the direction of changes in natural and subject heading languages.

ROZWÓJ METOD MAPOWANIA DOMEN NAUKOWYCH I POTENCJAŁ ANALITYCZNY W NIM ZAWARTY

Veslava Osińska
Uniwersytet M. Kopernika w Toruniu

Wizualizacja informacji, mapowanie domen naukowych, analiza współcytowań, sieci społeczne

Historia wizualizacji domen naukowych

Mapowanie i wizualizacja domen naukowych jako metodyka badań zdobywa coraz większe zainteresowanie wśród specjalistów informacji naukowej i naukoznawstwa. Doprowadziły do tego co najmniej dwa czynniki. Pierwszy to cyfryzacja zasobów piśmiennictwa naukowego i innych źródeł wiedzy i co za tym idzie, przyrost i rozwój bibliograficznych baz danych dostępnych online, takich jak np.: ISI Web of Knowledge, Scopus, Google Scholar. Drugim jest nieliniowy wzrost mocy obliczeniowej komputerów, wykorzystywanie superkomputerów i technologii rozproszonych [6, 21].

Aby lepiej zrozumieć strukturę i dynamikę rozwoju poszczególnych działów nauki oraz znaleźć sposób identyfikacji w nich trendów tematycznych, naukowcy analizują literaturę naukową i ścieżki cytowań poszczególnych publikacji. Analiza cytowań jest znana od dawna z bibliometrii. Termin **wizualizacja dyscyplin wiedzy** lub **nauki** (*KDViz – Knowledge Domain Visualization)* wymyślony został przez Ch. Chena w 2001 r., redaktora czołowego czasopisma w zakresie wizualizacji informacji *Information Vizualization*[1] (powszechny skrót: *InfoViz*). Ponieważ metody te prowadzą do generowania map graficznych, równolegle używana jest inna nazwa: **mapowanie nauk** (*Mapping Science*) lub rzadziej naukografia (*scientography*) [8]. Tę ostatnią wprowadził w 1960 roku E. Garfield - założyciel Instytutu Informacji Naukowej w Filadelfii (*ISI*). Na przestrzeni ostatniego 10-lecia badania wizualizacji domen wiedzy zostały rozszerzone do zadań wyszukiwania informacji [10].

Nad rozwojem technik i metod wizualizacji pracują interdyscyplinarne zespoły informatyków, programistów od projektowania interakcji człowiek – komputer (*Human Computer Interface – HCI*), analityków, naukoznawców, grafików i plastyków. Warto tu wspomnieć o projekcie Wirtualne Studio Wiedzy (*Virtual*

[1] http://www.palgrave-journals.com/ivs/index.html.

Knowledge Studio – VKS)[2], który niedawno powstał w Holenderskiej Akademii Nauk i wspiera naukowców nauk humanistycznych i społecznych w Holandii w opracowywaniu nowych rozwiązań naukowych. Celem jest tworzenie metaforycznych modeli struktur nauk humanistycznych i społecznych. Główną cechą takiego centrum jest integracja wizualizacji i analizy na tle współpracy badaczy nauk społecznych, humanistycznych oraz informatyków.

O atrakcyjności takich graficznych zasobów wiedzy dla odbiorców przesądzają nie tylko walory estetyczne, lecz także aktualność, potencjał analityczno-merytoryczny oraz dostępność online. Dlatego nie bez powodu wśród praktyków *KDViz* istnieje tendencja do szybkiej publikacji wyników badań w sieci. Na przykład, wirtualna wystawa *Places@Spaces*[3] jest ogólnoświatowym projektem, ukierunkowanym na studiowanie infrastruktury nauki w skali globalnej w celu zademonstrowania osiągnięć w dziedzinie mapowania informacji o roku 1998. Portal redagowany jest przez K. Börner, *School of Library and Information Science, Indiana University*, której zespół prowadzi zaawansowane badania nad metodami wizualizacji dziedzin nauki, obszarów współpracy naukowców oraz bibliotek cyfrowych. Zamieszczane tam na bieżąco mapy domenowe, kartograficzne i pojęciowe mają na celu inspirowanie interdyscyplinarnych dyskusji, umożliwienie komunikacji pomiędzy działalnością człowieka i postępem naukowym. Coraz więcej w sieci udostępnia się aplikacji do wizualizacji własnych, niekoniecznie akademickiego przeznaczenia, danych w standardowych formatach (np. ASCII, CSV), zachęcając dociekliwych użytkowników do testów rozmaitych konfiguracji graficznych, od wykresów bąbelkowych i rozproszonych (*skater chart*) począwszy, na chmurkach tekstowych kończąc (*text cloud*). Takim narzędziem jest eksplorator autorstwa IBM *ManyEyes*[4], gdzie użytkownicy mają możliwość nie tylko wizualizacji własnych danych przygotowanych w odpowiednich formatach, lecz i wytypowania najlepszych projektów.

Proces wizualizacji

A. Elementy składowe

W ponad 10-letniej historii wizualizacji informacji na czoło wysunęło się kilka zespołów badawczych specjalizujących się w metodach wizualizacji wiedzy i nauki. Najbardziej znaczące osoby w *Infoviz*, kierujące tymi zespołami i występujące w większości źródeł cytowań, zostały wymienione w poprzednim rozdziale.

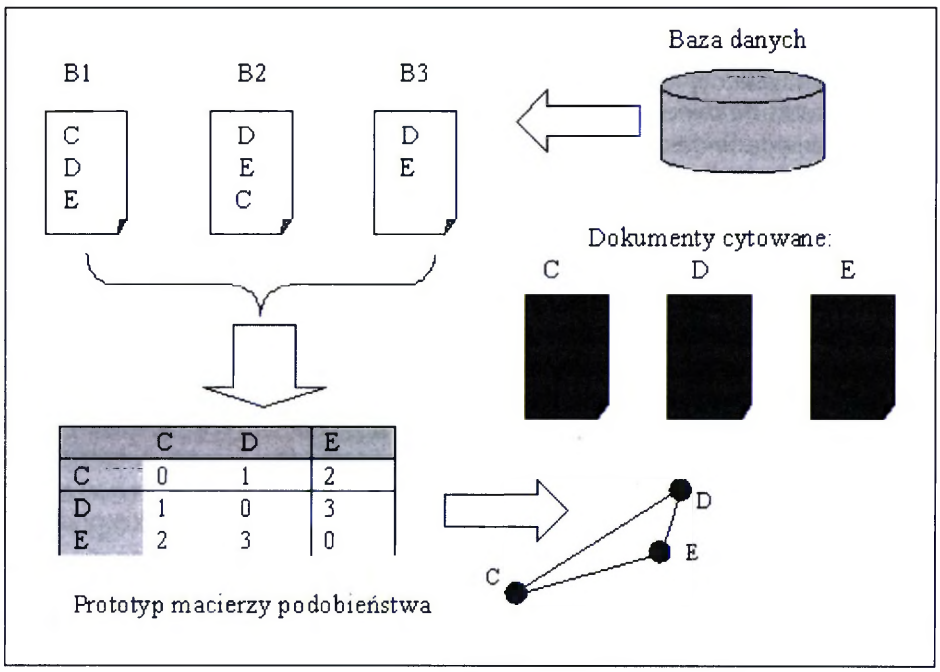
W każdej analizie wykorzystującej mapy wizualizacji zawarte są trzy podstawowe składniki: obiekty analizy, relacje pomiędzy nimi oraz metoda mapowania. Wcześniejsze metody wykorzystywały najprostsze grafy, gdzie węzły (*nodes*) reprezentowały obiekty badań, a krawędzie (*edges, links*) – związki pomiędzy nimi. Im bliżej usytuowane były obiekty, tym bardziej były do siebie podobne pod względem wybranej miary (kierunki badań, źródła cytowań, wspólni auto-

[2] http://virtualknowledgestudio.nl/aboutvks/.
[3] http://scimaps.org.
[4] http://manyeyes.alphaworks.ibm.com/manyeyes/.

rzy itp.). Później zaczęto opierać się na algorytmach reprezentujących wyniki w postaci bardziej kognitywnej, bo dwu- lub trójwymiarowej sieci, z wyeksponowaniem różnych poziomów abstrakcji. Krótka historia *KDViz* i interdyscyplinarny charakter tej dyscypliny badawczej utrudniają systematykę metodologii badań.

Ponieważ tradycyjne podejście do wyboru danych analizy nie zmienia się od lat, można spróbować podzielić sposoby mapowania ze względu na obiekty analizy i miary podobieństwa.

Wspólne cytowanie autorów (ACA – A*uthor* Co-***citation analysis***) lub artykułów (DCA – *document co-citation analysis*) to metoda używana w naukometrii od lat 70., pierwsza, na której opierały się technologie wizualizacyjne [16]. Dwa dokumenty (autorzy) są wspólnie cytowane, jeśli równocześnie występują w wykazie bibliograficznym trzeciego dokumentu. Podobieństwo pomiędzy dokumentami jest obliczane na podstawie częstości ich wspólnych cytowań. Uproszczony schemat wizualizacji na podstawie źródeł cytowań przedstawia rys. 1. Wizualizacja za pomocą DCA charakteryzuje dynamikę domeny wiedzy, natomiast ACA pozwala na wnioskowanie o strukturze intelektualnej obszaru badawczego [2, 5]. O wiele rzadziej za miarę podobieństwa przyjmuje się wspólne pozycje bibliograficzne w dwóch różnych dokumentach – tzw. bibliografie sprzężone (*bibliographic coupling*).



Rys. 1. Schemat analizy i mapowania wspólnego cytowania dokumentów. Dokumenty DE mają największe podobieństwo, dokumenty CD – najmniejsze.

Wspólne występowanie pojedynczych słów lub sekwencji w tytułach, abstraktach lub treści publikacji. Mogą to być słowa kluczowe, deskryptory tema-

tyczne, terminy główne, hasła przedmiotowe lub najważniejsze wyrażenia i frazy w tekście. Nie ma tu potrzeby używania programów do analizy częstości występowania i współwystępowania słów, można to bowiem wykonać za pomocą albo własnych aplikacji, albo ogólnie dostępnych narzędzi, chociażby arkusza kalkulacyjnego EXCEL. Takim graficznym reprezentacjom opartym na wspólnej terminologii trochę na wyrost nadano nazwę mapy semantycznej, ponieważ w założeniu ma ona pomagać w zrozumieniu kognitywnej struktury danej domeny naukowej. Typowym przykładem autentycznej mapy semantycznej jest wizualna reprezentacja tezaurusa (Visual Thesaurus[5]). W wizualizacji słów kluczowych istotny jest algorytm wyszukujący semantycznie podobne terminy, w przypadku różnotematycznej bazy danych występuje wysoka przypadkowość.

W mapowaniu nauk używa się także kolekcji poklasyfikowanych dokumentów. Liczba wspólnych artykułów naukowych w przypisanych kategoriach tematycznych przekłada się na podobieństwo semantyczne ich tematów. Mapowanie dokumentów na podstawie wspólnych klas lub kategorii tematycznych daje najbardziej prawidłowy obraz organizacji przedmiotowej danej domeny naukowej. Jednak prac wykorzystujących tę metodę jest niewiele [12, 17], zapewne z powodu braku danych o klasyfikacji w bazach bibliograficznych.

Jak się można domyślić, większość jednostek analizy pochodzi z bibliograficznych baz danych, dlatego przeważająca część publikacji na temat wizualizacji domen naukowych to zmapowane kolekcje danych baz przedmiotowych ISI Web of Knowledge. W pracach [16, 17] podano przykład zmapowania uniwersum klasyfikacyjnego w dziedzinie nauk komputerowych oraz zaproponowano metodę modernizacji pierwotnego schematu klasyfikacji ACM CCS. Bibliograficzne dane dokumentów naukowych nie wystarczą, jeśli obiekty analizy wyłaniają się z treści artykułów. Oprócz dostępu do pełnotekstowych zasobów naukowych niezbędne jest również zastosowanie odpowiednich algorytmów przetwarzania języka naturalnego.

Na etapie rozwoju Web 2.0 popularne jest mapowanie zasobów wiedzy pochodzących z ogólnodostępnych serwisów sieciowych. Na przykład mapa artykułów angielskiej Wikipedii wygenerowana została przy użyciu miary rejestrującej wspólne kategorie artykułów [11]. Aktywność użytkowników serwisów specjalistycznych również inspiruje do analizowania i mapowania takich danych jak blogi i fora internetowe, logi użytkowników [1] itd. Dotychczasowe metody wizualizacji informacji zostały z powodzeniem zaadaptowane w środowisku sieciowym, gdzie strony internetowe z pewnym przybliżeniem można traktować jako artykuły naukowe, a hiperłącza jako powiązania cytowań. Takie podejście wspomogło szybki rozwój webometrii [21].

B. Etapy procesu wizualizacji

Z praktycznego punktu widzenia model mapowania informacji tekstowej – a z taką mamy do czynienia w analizie treści publikacji naukowych – obejmuje następujące etapy [2,4]:

[5] www.visualthesaurus.com

Wybór odpowiedniej bazy danych: bibliograficzne bazy danych oparte na modelu meta, a więc zawierające metadane, np. tytuł, autora, datę publikacji, słowa kluczowe. Dodatkowym atutem są informacje bibliometryczne, np. liczba cytowań, współpracownicy autora, miara oddziaływania czasopisma itp., ułatwiające ustalenie związków pomiędzy badanymi obiektami.

Wytypowanie i obróbka jednostek analizy, którymi mogą być artykuły, czasopisma, autorzy, słowa kluczowe, określniki tematyczne.

Wybór stosownej miary podobieństwa, np. wspólne pary cytowań, wspólne kategorie, współautorzy itp. oraz obliczenie częstości współwystępowania wraz z wyznaczeniem progu zliczeń przypadków, poniżej którego analiza staje się niemiarodajna.

Stworzenie czytelnego rozkładu danych, dającego się zmapować na przestrzeń 2D lub 3D. Posługując się skalarnym modelem obliczania (liczony iloczyn skalarny jako podstawowa miara podobieństwa) wartości podobieństwa dla każdej pary obiektów, uzyskuje się macierz podobieństwa. Wartości macierzy odwzorowują odległości topologiczne pomiędzy danymi w wybranej przestrzeni; za pomocą odpowiednich algorytmów grupujących analizuje się powstające klastry obiektów. http://www.ebib.info/publikacje/matkonf/mat19/osiewalska.Php – 10. Można też zastosować wektorowy model przestrzeni informacyjnej, np. analizę ukrytych grup semantycznych (*Latent Semanics Analysis – LSA*) [15]. Ponieważ na tym etapie wielowymiarowy charakter danych nie pozwala na ich zobrazowanie, następnym krokiem musi być redukcja wymiarów do dwóch lub trzech. Używa się tu technik statystycznych, takich jak analiza czynnikowa, analiza głównych składowych, skalowanie wielowymiarowe (*Multidimensional Scaling - MDS*). Do inteligentnego rozmieszczania wielowymiarowych danych wykorzystuje się również sztuczne sieci neuronowe, mapy samoorganizujące się (*Self Organizing Map – SOM*) lub algorytmy klasteryzacji [2, 5, 16].

Zastosowanie aplikacji do wizualizacji i wyświetlenie informacji w sposób umożliwiający użytkownikowi wygodną obserwację. Współczesne interfejsy wizualizacyjne zapewniają przejrzystą i interaktywną eksplorację danych, umożliwiając szerokokątne przeglądanie, powiększanie, filtrowanie, wyszukiwanie i wyszczegółowienie wybranych fragmentów danych.

Walidacja statystyczna w celu ewaluacji i doboru odpowiednich parametrów początkowej miary podobieństwa. Nieznaczna zmiana warunków początkowych może całkowicie zmodyfikować końcowy rozkład danych. Przy całej różnorodności metod i technik wizualizacji informacji problemem pozostaje ewaluacja końcowego rozkładu wizualizacji. Dlatego do oceny wyników wizualizacji podchodzi się empirycznie: sprzężenie zwrotne „parametry początkowe – wyniki" pozwala na wygenerowanie najbardziej dokładnej mapy wizualizacji.

Oprogramowanie do wizualizacji

Proces wizualizacji wymaga rozwiązań programistycznych na co najmniej dwóch poziomach. Dysponując bazą danych, analityk staje przed problemem ekstrakcji relewantnych informacji, takich jak cytowania, słowa kluczowe, abstrakty artykułów w badanej kolekcji. Następnie należy umiejętnie obliczyć przy-

padki współwystępowania jednostek analizy i wypełnić macierz podobieństwa tymi wartościami, po uprzednim przeskalowaniu i normalizacji (konieczna jest automatyzacja tej procedury). Odpowiedniego programu wymaga mapowanie danych na przestrzeń obserwacji, a przedtem redukcja wymiarów. Tworzenie takiego oprogramowania i jego wdrożenie jest czasochłonne, dlatego większość badaczy *KDViz* korzysta z gotowych programów lub modułów softwarowych typu *Open Source* [3,6,20].

Do obróbki danych bibliograficznych i bibliometrycznych korzysta się z oprogramowania:

– Bibexcel[6] – narzędzie mające wiele zastosowań bibliometrycznych autorstwa szwedzkiego naukowca O. Perrsona, założyciela dziedziny BIN w Szwecji. Akceptowalny format danych – *Dialog*,

– CiteSpace[7] – aplikacja stworzona przez Ch. Chena do rozmaitych analiz bibliometrycznych oraz wizualizacji wyników. Współpracuje z formatami danych pobieranych ze strony Web of Science; aktualnie opracowywane formaty baz Scopus i Google Scolar.

Oprogramowanie do mapowania i wizualizacji:

– CiteSpace,

– Pajek[8] – słoweński darmowy, ceniony przez profesjonalistów program do analizy i wizualizacji obszernych sieci danych,

– UCINET[9] – program dostępny w wersji trial do konstruowania i analizy społecznościowej sieci świata nauki; preferowane są dane macierzowe,

– XLSTAT[10] – komercyjne rozszerzenie programu MS Excel o funkcje analizy statystycznej i wizualizacji danych,

– Permap[11] – darmowy, nieduży program do mapowania metodą skalowania wielowymiarowego; szczególnie polecany w obszarach zastosowań psychometrii i socjometrii,

– Aplikacje do typowej analizy statystycznej dużych zbiorów danych: Statistica, Origin, Matlab, Mathematica.

Z powyższej listy wyróżnić należy ciągle rozwijany CiteSpace – kompletne narzędzie obsługujące wszystkie etapy procesu wizualizacji: od ekstrakcji danych do walidacji wynikowych konfiguracji graficznych.

Przykład wizualizacji domeny naukowej

Spektrum mapowania nauk jest bardzo szerokie: nauki ścisłe, nauki medyczne, społeczne i humanistyczne. Te ostatnie stanowią prawdziwe wyzwanie dla ekspertów *KDViz*, bazy Web of Science bowiem indeksują nieznaczną część wszystkich światowych zasobów w zakresie nauk humanistycznych [18], również prace, gdzie się podejmuje mapowania nauk komputerowych i inżynierii

[6] http://www8.umu.se/inforsk/Bibexcel/.
[7] http://cluster.cis.drexel.edu/&u0007E;cchen/citespace/.
[8] http://vlado.fmf.uni-lj.si/pub/networks/pajek/.
[9] http://www.analytictech.com/ucinet6/ucinet.htm.
[10] http://www.xlstat.com/.
[11] http://www.ucs.louisiana.edu/~rbh8900/old%20web%20page.html.

są bardzo nieliczne [17]. Nie jest tajemnicą, że w bazach ISI ciągle przeważają źródła amerykańskie, co nie odwzorowuje aktualnej struktury świata nauki, gdzie ostatnio znaczący jest wkład europejski, a szczególnie chiński. Skala materiału badanego, tj. liczba czasopism wpływa na rozpiętość dziedzinową mapowanej wiedzy. Na przykład, na podstawie 2 tys. czasopism zwizualizowano 32 dyscypliny i 140 specjalności naukowe [4], natomiast zobrazowanie całościowej nauki wymagało użycia 16 tys. publikacji i materiałów konferencyjnych, co dało ponad 7 mln publikacji za okres 2001-2005[12].

Interdyscyplinarny charakter dziedziny Bibliotekoznawstwo i Informacja Naukowa (BIN) powoduje, iż jest ona najczęściej wybierana do wizualizacji. Badania struktury BIN przeprowadza się od lat, w podobny sposób używając źródeł; na podstawie analizy najważniejszych 12-tu czasopism anglojęzycznych z tego zakresu:

- Annual Review of Information Science and Technology,
- Electronic Library,
- Information Processing & Management,
- Information Technology and Libraries,
- Journal of Documentation,
- Journal of Information Science,
- Journal of the American Society for Information Science,
- Journal of the American Society for Information Science and Technology – JASIST,
- Library & Information Science Research,
- Library Resources & Technical Services,
- PROGRAM-Automated Library and Information Systems,
- Scientometrics.

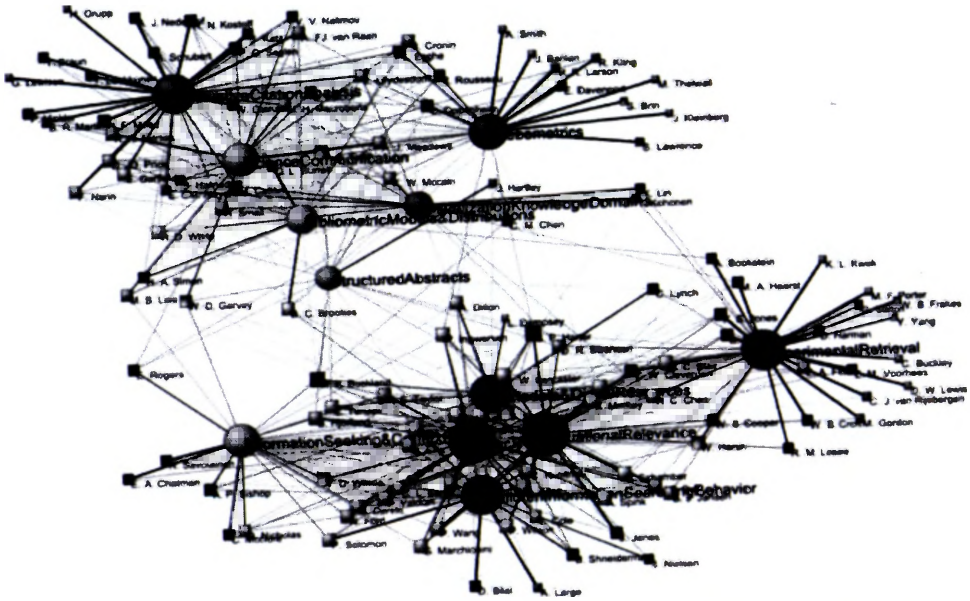
Spośród wielu wizualizacji BIN wybraliśmy przykład najbardziej czytelny i kompletny, a jednocześnie aktualny opisany w pracy [21]. Jest to wynik zmapowania literatury naukowej z lat 1996-2005 przy zastosowaniu analizy współcytowań autora (*ACA*), wzbogaconej o algorytmy rotacji ortogonalnej (rys. 2).

W wyniku analizy współcytowań zidentyfikowano główne specjalności badawcze. Największe na mapie węzły to:

- wyszukiwanie informacji,
- metadane i zasoby cyfrowe,
- Webometria,
- analiza cytowań (naukometria, bibliometria),
- badanie zachowań użytkownika,
- wizualizacja domen naukowych (ACA, analiza współcytowań),
- wyszukiwanie doświadczalne (algorytmy, modele, systemy, ewaluacja),
- komunikacja naukowa,
- użyteczność sieciowa, projektowanie interfejsów,
- abstrakty strukturalne (piśmiennictwo naukowe).

Autorzy podkreślają, iż w porównaniu z wynikami z końca lat 90., czyli sprzed ekspansji technologii sieciowych i używania Google'a („googlowania"),

[12] R. Klavans, K. Boyack: *Maps of Science: Forecasting Large Trends in Science. Places@Spaces: Mapping Science.* [online]. [dostęp: 19.05.2009]. Dostępny w World Wide Web: <http://www.scimaps.org/dev/big_thumb.php?map_id=164>.



Rys. 2. Wizualizacja dziedziny Bibliotekoznawstwa i Informacji Naukowej metodą ACA na podstawie zmapowania artykułów z 12 czasopism za okres 1996-2005 [21]

powstał klaster obejmujący zagadnienia Webometrii oraz klaster odnoszący się do *KDViz*, a reprezentacje przestarzałych obszarów badań, takie jak OPAC i Online Retrival uległy rozmyciu. Wyniki mapowania wskazują też na rosnące znaczenie studiów użyteczności sieciowej i zachowań użytkowników. Widać też nowy temat badawczy: analiza strategii wyszukiwawczej dla dzieci (*Children's information searching behaviour*). Na mapie dystrybucji ognisk tematów badawczych zauważalne jest również szerzące się kognitywne podejście do rozwiązywania problemów w naukach o informacji.

Podsumowanie: mapy wiedzy a tradycyjne analizy bibliometryczne

Jako że metodologia wizualizacji informacji, zaadaptowana do nowej w ostatnim 10-leciu dyscypliny badawczej jaką jest mapowanie nauki/wiedzy, wyrosła z bibliometrii i naukometrii, nie dziwi fakt korzystania z cytowań jako z podstawowego materiału badawczego. Oceny bibliometryczne bazują na cytowalności, uznawanej za obiektywną miarę jakości źródeł piśmienniczych. Potencjał informacji zawartej w indeksach cytowań od dawna wykorzystuje się do badań naukoznawczych, co spowodowało, iż analiza cytowań stała się samodzielnym działem bibliometrii [14].

Miary wpływu czasopism (np. Impact Factor – IF, index Hirscha, PageRank), czy autora podawane są w postaci liczbowej, analiza sprowadza się więc do porównania w tabelach liczb, które uważane są za parametry świadczące o wydajności naukowej danego czasopisma/autora. Uzyskane w ten sposób wskaźniki efektywności są podstawą do tworzenia rankingów i często

decydują o skali dotacji budżetowych, stąd kontrowersje w świecie nauki dotyczące powiązań komercyjności przedsięwzięć naukowych z nadinterpretacją wskaźnika IF [7]. Tabela bibliometryczna pozwala na wyodrębnienie grupy najbardziej lub najmniej wydajnych czasopism/autorów, jednak z tradycyjnej analizy bibliometrycznej nie da się wywnioskować o relacjach pomiędzy badanymi jednostkami i ich podobieństwem, np. tematycznym. Problem ten jest rozstrzygany na poziomie analizy wielowymiarowej, badań właściwości dużych zbiorów danych i ich dynamiki przy zastosowaniu technik datamining. W artykule [14] zademonstrowano za pomocą dendrogramów możliwości analizy skupień, jednej z metod bibliometrii strukturalnej. Na podstawie danych współcytowań pogrupowane zostały polskie czasopisma ekonomiczne, w grupy podobne pod względem opisywanych tematów.

Kolejny etap analiz wielowymiarowych – prezentacja wyników – daje nie tylko wgląd w związki, hierarchie i podobieństwo pomiędzy obiektami, lecz i obraz globalnej struktury danych, czasem też tendencje zmian, a co za tym idzie, możliwości prognozowania. Naukowcy wykorzystują materiał badawczy pochodzący nie tylko z indeksów cytowań, lecz coraz częściej z abstraktów i tekstu dokumentów. Eksploracja tekstu (*text mining*) w dużych zbiorach danych jest modnym przedmiotem badań w naukach komputerowych. Informatycy wypracowują metody, które mogłyby weryfikować tematyczną klasyfikacje/kategoryzację badanych dokumentów. Niebagatelne znaczenie ma tu stały postęp w rozwoju technik prezentacji informacji. Cechą większości współczesnych aplikacji do wizualizacji, gdzie znajdują zastosowanie najnowsze odkrycia nauk kognitywnych, jest interakcja z użytkownikiem. To powoduje, iż analityk ma możliwość modernizacji konfiguracji danych wynikowych, posługując się empirycznie doborem parametrów wejściowych. Ważnym aspektem badawczym jest to, że mapy wizualizacyjne mogą służyć użytkownikowi jako interfejs systemu wyszukiwawczego [16].

Liczne prace w ostatnich latach w zakresie wizualizacji nauki/wiedzy udowadniają zasadność takich badań i przydatność wynikowych map nie tylko w naukoznawstwie i opracowaniach bibliometrycznych, lecz również w biologii, obrazowaniu medycznym i genetyce [3]. Warto tu wymienić możliwości analityczne, które niesie wizualizacja informacji i danych masowych. Mapowanie nauki umożliwia:

– określenie aktualnych trendów tematycznych i dominujących obszarów badawczych;

– poznanie bazy naukowej danej dziedziny oraz jej frontów badawczych, kierunki i zakresy integracji multidyscyplinarnych, identyfikację subdyscyplin;

– wykrycie społecznościowej (intelektualnej) struktury pola badań, włączającej zarówno aktywność autorów i wydawców, jak i osób zaangażowanych w tworzenie baz wiedzy; nakreślenie wzoru sieci współpracy pomiędzy naukowcami;

– analizę dorobku naukowego badaczy, instytucji naukowej, bazy naukowej danego kraju;

– wykrycie ewolucji za pomocą mapowania ciągłego (*longitudinal mapping*) [9] oraz prognozowanie przyszłych zachowań domeny naukowej poprzez wizualizację i modulowanie struktury;

– lepsze zrozumienie struktury, dynamiki dziedziny, wiedzy oraz kierunków rozprzestrzeniania się (dyfuzji) wiedzy. Naukowcy *Infoviz* są w stanie wykryć zmiany krytyczne w nauce, posługując się teorią paradygmatów [2]; może to wspomóc dokonanie odkryć naukowych;

– wyszukiwanie semantyczne informacji (za pomocą map semantycznych);

– weryfikację i modernizację klasyfikacji przedmiotowej na podstawie semantycznej organizacji dokumentów na mapie wizualizacji.

Wizualizacja domen wiedzy jako odnowione i wzbogacone o techniki datamining podejście metodologii biblio- i naukometrycznych zawiera spory potencjał badawczy. Jak widać, wizualizacja zasobów naukowych decyduje o właściwym postrzeganiu i rozumieniu danych na wszystkich poziomach organizacji wiedzy. Twórcy KDViz podkreślają wieloperspektywiczność zastosowań metod wizualizacji: w wyszukiwaniu informacji, monitorowaniu trendów w nauce i technologiach, polityce finansowania ośrodków badań i konkretnych naukowców.

Bibliografia

1. *Bollen J. [et al.]: Clickstream Data Yields High-Resolution Maps of Science. PLoS ONE. 2009 vol. 4 no. 3.* [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <http://www.plosone.org/ article/info:doi/10.1371/journal.pone.0004803>.
2. Börner K., Chen Ch., Boyack K. W.: *Visualizing Knowledge Domains*. W: B. Cronin (red.): *Annual Review of Information Science & Technology*. 2003 vol. 37, pp. 179-255.
3. Boyack K. W. [et al.]: *Domain visualization using VxInsight for science and technology management*. „Journal of the American Society for Information Science and Technology" 2002, no. 53(9), pp. 764-774.
4. Boyack K. W. [et al.]: *Mapping the backbone of Science*. „Scientometrics" 2005 vol. 64, no. 3, pp. 351-374.
5. Chen Ch.: *Information Visualization. Beyond the Horizon*. 2nd ed. London, 2006.
6. Chen Ch. [et al.]: *The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis*. „Journal of the American Society for Information Science and Technology" 2010 vol. 61(7), pp. 1386-1409.
7. *European Association of Science Editors statement on impact factors*. [online]. [dostęp: 31.08. 2010]. Dostępny w World Wide Web: <http://www.ease.org.uk/statements/EASE_statement_on_impact_factors.shtml. Retrieved, 2009-03-25>.
8. Garfield E.: *Scientography: Mapping the tracks of science*. „Current Contents: Social & Behavioural Sciences" 1994, no. 7(45), pp. 5-10.
9. Garfield E.: *Essays/Papers on Mapping the World of Science*. Eugene Garfield, Ph. D. Home Page. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <http://garfield.library.upenn.edu/ mapping/mapping.html>.
10. Hjørland B.: *Domain analysis in information science: eleven approaches – traditional as innovative*. „Journal of documentation" 2002, no. 58, pp. 422-462.
11. Holloway T. [et al.]: *Analyzing and Visualizing the Semantic Coverage of Wikipedia and Its Authors*. Wyd. specjalne: *Understanding Complex Systems. Complexity*. 2007, vol. 12, no. 3, pp. 30-40.
12. Moya-Anegón F. [et al.]: *A new technique for building maps of large scientific domains based on the cocitation of classes and categories*. „Scientometrics" 2004, vol 61, no. 1, pp. 129-145.
13. Osiewalska A.: *Mierniki oceny czasopism i naukowców*. EBIB Biuletyn Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2008, nr 11(99). [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <http://www.ebib.info/2008/99/a.php?osiewalska>.
14. Osiewalska A.: *Bibliografie czasopism naukowych Biblioteki Głównej UEK jako źródło danych dla analiz bibliometrycznych*. W: *Bibliograficzne bazy danych: kierunki rozwoju*

i możliwości współpracy. Ogólnopolska konferencja naukowa z okazji 10-lecia bazy danych BazTech, Bydgoszcz 27-29 maja 2009 r. [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <http://www.ebib.info/publikacje/matkonf/mat19/osiewalska.php>.
15. Osińska V.: *Przybliżenie semantyczne w wizualizacji informacji w Internecie i bibliotekach cyfrowych.* EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2006, nr 7(77). [online]. [dostęp: 31.08.2010]. Dostępny w World Wide Web: <http://www.ebib.info/2006/77/osinska.php>.
16. Osińska V.: *Wizualizacja i wyszukiwanie dokumentów.* Warszawa 2010.
17. Osińska V., Bala P.: *New Methods for Visualization and Improvement of Classification Schemes: The Case of Computer Science.* „Knowledge Organization" 2010, no. 37, pp. 157-172.
18. Persson O., Danell R.: *How to use Bibexcel for various types of bibliometric analysis.* W: *Celebrating scholarly communication studies: A festschrift for Olle Persson.* Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. *2009 vol. 05-S* pp. 9-24. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>.
19. Sivertsen G.: *Publication patterns in all fields 2009.* W: *Celebrating scholarly communication studies: A festschrift for Olle Persson.* Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. *2009 vol. 05-S* pp. 55-60. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>.
20. Thelwall M.: *A Webometric Analysis of Olle Persson.* W: *Celebrating scholarly communication studies: A festschrift for Olle Persson.* Special volume of the e-zine of the ISSI. Pod red. F. Åström, R. Danell, B. Larsen and J. W. Schneider. 2009 vol. 05-S, pp. 61-71. [online]. [dostęp: 12.11.2010]. Dostępny w World Wide Web: <http://www8.umu.se/inforsk/Bibexcel/ollepersson60.pdf>.
21. Zhao D., Strotmann A.: *Information Science during the first Decade of the Web: An enriched author cocitation analysis.* „Journal of the American Society for Information Science and Technology" 2008, vol. 59(6), pp. 916-937.

Summary

The article describes a new research method, popular among bibliometric and sciencemetric specialists, i.e. science visualization. Genesis, technologies, and stages of the visualization process are presented, with an example of matrix map for library and information science for the years 1996-2005. Research potential of this method were discuseed, in comparison with bibliometric methods. The article's aim is to encourage Polish scientists, interested in studies regarding science visualization, hardly popular in our country. The author describes mapping tools popular among visualization specialists, like Open Source, and their usability and technological features.

LINKED DATA – MODEL DANYCH POWIĄZANYCH W SEMANTIC WEB

Marcin Roszkowski
Instytut Informacji Naukowej
i Studiów Bibliologicznych
Uniwersytetu Warszawskiego

Semantic Web, Linked Data, reprezentacja wiedzy

Tim Berners-Lee w 2007 r., wypowiadając się na temat przyszłości World Wide Web [5], stwierdził, że sukces tego otwartego środowiska informacyjnego jest wynikiem trzech czynników:

– nieograniczonych możliwości łączenia informacji, dokumentów za pomocą np. hiperłączy;

– zastosowania otwartych standardów jako podstawy rozwoju i innowacji aplikacji (m.in. działalność W3C i standardy opisu informacji, np. HTML, XML, RDF, Dublin Core, czy przesyłania danych, np. HTTP, FTP,);

– rozdzielenia warstw sieciowych odpowiedzialnych za komunikację, przesyłanie i przetwarzanie informacji i jej wizualizację.

Te trzy cechy charakteryzują dzisiejszą sieć, która jest siecią:

– stworzoną przez ludzi i dla ludzi,

– zbudowaną w większości z dokumentów, które zawierają dane zamknięte w ich strukturze,

– zamkniętych i niepowiązanych ze sobą baz danych, które często porównywane są z silosami danych [5].

Dzisiejsze World Wide Web to sieć dokumentów – statycznych lub generowanych przez aplikacje bazodanowe, ale to wciąż sieć niezrozumiała dla aplikacji je przetwarzających, które „widzą" w nich ciągi znaków oraz zestawy komend związane ze sposobem ich wyświetlania. Hipertekst, dzięki węzłom i hiperłączom, dał możliwość wizualizacji i eksploracji przestrzeni informacyjnej. Powiązania hipertekstowe, zarówno inter- jak i intratekstualne, odsyłają jednak do dokumentów lub konkretnych miejsc w ich strukturze (np. nagłówek, paragraf). Tym samym centralnym punktem w takiej koncepcji sieci jest dokument, jego metadane oraz powiązania z innymi dokumentami w sieci. Dane zawarte w dokumencie, jego struktura oraz sposoby wyświetlania stanowią jeden ciąg kodu, co utrudnia automatyczną ekstrakcję informacji treściowej i związków znaczeniowych z innymi dokumentami.

W ciągu ostatnich dwudziestu lat dokonała się ewolucja zarówno koncepcji WWW, jak i technologii w niej wykorzystywanych. Obecnie jesteśmy świadkami funkcjonowania trzech generacji sieci. Web 1.0, czyli pierwsza generacja sieci obecna od początku lat 90. ubiegłego stulecia, charakteryzowała się opisem

struktury dokumentu za pomocą języków znacznikowych, przede wszystkim HTML. Mieliśmy do czynienia w większości z dokumentami statycznymi, gdzie metadane nadawane był praktycznie wyłącznie przez twórców dokumentów, projektantów serwisów, rzadko automatycznie przez oprogramowanie i przez użytkowników. Druga generacja sieci – Web 2.0, to sieć partycypacji, interakcji i integracji. Jest to sieć współuczestnictwa użytkowników w tworzeniu jej zawartości, ale także masowe tworzenie metadanych przez użytkowników (np. folksonomie i tagowanie). Integracja w wymiarze społecznym polega na tworzeniu grup użytkowników wokół serwisów społecznościowych, a w wymiarze technologicznym na tworzeniu systemów pobierających informacje z innych serwisów, na przykład na zagnieżdżaniu aplikacji Google, Youtube. Trzecia generacja sieci, czyli Web 3.0 czy też Semantic Web to sieć danych, które dzięki wspólnemu językowi opisu oraz tworzonym przez użytkowników powiązaniom mogą być automatycznie przetwarzane przez programy komputerowe. Semantic Web to sieć, w której najważniejsze są metadane i struktura [22].

Proces tworzenia i rozpowszechniania metadanych w przypadku sieci Web 3.0 zostaje wsparty przez mechanizmy automatycznie je przetwarzające, które generują rezultaty wyszukiwania. Społecznościowy charakter metadanych jest tutaj zachowany, nie jest to sieć maszyn. Zmienia się tylko sposób opisu rzeczywistości poprzez jego sformalizowanie. Mamy więc do czynienia z przejściem od sieci dokumentów zawierających dane dostępne za pomocą wyszukiwarek i przeglądarek, przez sieć aplikacji prezentujących w jednym miejscu zawartość wielu serwisów i aplikacji, do sieci danych, które wewnętrznie są uporządkowane i powiązane dzięki wspólnemu modelowi reprezentacji wiedzy.

Semantic Web to wizja sieci, która dąży do utrzymania postulatu „każdy może wypowiedzieć się na dowolny temat" (tzw. postulat 3xA – Anyone can say Anything about Any topic) [1], która ma się „przyczynić do utworzenia i rozpowszechnienia standardów opisywania treści w Internecie w sposób, który umożliwi maszynom i programom (tzw. agentom) przetwarzanie informacji w sposób odpowiedni do ich znaczenia." [34] Jest to wizja Tima Berners'a-Lee, którego zdaniem zrozumienie przez aplikację danych polegałoby na zastosowaniu takiego modelu ich opisu, który zakładałby możliwość powiązania ich znaczenia w ramach wspólnego kontekstu.

Linked Data – założenia

Semantic Web nie polega tylko na umieszczaniu danych w sieci, lecz także na tworzeniu powiązań między nimi. Ma to umożliwić manualną lub automatyczną eksplorację zbiorów i odkrywanie nowych danych. Za pomocą danych powiązanych ze sobą tego typu przeszukiwanie staje się możliwe [6]. Koncepcja danych powiązanych Tima Berners'a-Lee, znana jako *Linked Data,* polega na wykorzystaniu World Wide Web i jego technologii do tworzenia formalnych połączeń między danymi pochodzącymi z różnych zbiorów. Umożliwia tworzenie relacji pomiędzy zbiorami danych operującymi różnymi formatami opisu, które różnią się zarówno pod względem treściowym, jak i formalnym. Jest to możliwe dzięki zastosowaniu wspólnego modelu reprezentacji wiedzy oraz rozwiązań

technologicznych uznawanych obecnie za standardy w WWW. Linked Data odwołuje się do metod ekspresji, reprezentacji, łączenia i współdzielenia danych w Semantic Web, które wykorzystują istniejące standardy i narzędzia sieciowe [36]. Nie jest to nowy schemat metadanych, system reprezentacji wiedzy, lecz zestaw wytycznych dotyczących publikowania i udostępniania danych w sieci semantycznej.

Koncepcja Linked Data jest oparta na języku RDF (Resource Description Format), formalnym modelu reprezentacji wiedzy, rekomendowanym przez Konsorcjum WWW. Polega na tworzeniu deklaracji opisujących otaczającą nas rzeczywistość i ustanawianiu związków znaczeniowych pomiędzy jej elementami. Linked Data wychodzi więc poza organizację rzeczywistości wirtualnej i jest gotowe do opisu świata realnego.

Linked Data jako model publikowania, łączenia i udostępniania danych w Semantic Web zakłada cztery ogólne reguły postępowania [6]:

- używaj URI (Uniform Resource Identifier) jako nazw i sposobu odróżniania obiektów;
- stosuj protokół HTTP, aby uzyskać informacje o opisywanych obiektach;
- udostępniaj użyteczne informacje o obiekcie identyfikowanym przez jego URI za pomocą standardów RDF/XML;
- wykorzystuj powiązania z innymi obiektami za pomocą URI, aby zapewnić możliwość eksploracji i odkrywania informacji o innych obiektach.

Pierwsza zasada odnosi się do tworzenia unikalnych nazw dla obiektów i zasobów za pomocą standardu URI. Druga i trzecia zasada odwołują się do wykorzystania sieciowego modelu organizacji World Wide Web do pozyskiwania informacji o zidentyfikowanych obiektach. Czwarta kładzie nacisk na tworzenie powiązań między obiektami i zasobami [11].

Technologie Linked Data

Koncepcja Linked Data jest oparta na dwóch, fundamentalnych dla WWW, technologiach: identyfikatorach URI [4] (Uniform Resource Identifier) oraz protokole przesyłania danych HTTP. Identyfikatory URI nie muszą w swojej budowie wskazywać na umiejscowienie obiektu, który oznaczają (tak jak w przypadku URL). Za ich pomocą można zidentyfikować dowolną jednostkę zarówno w świecie wirtualnym, jak i rzeczywistym, konkretną i abstrakcyjną. Celem standardu URI jest identyfikacja zasobów i obiektów za pomocą unikalnych ciągów znaków [32].

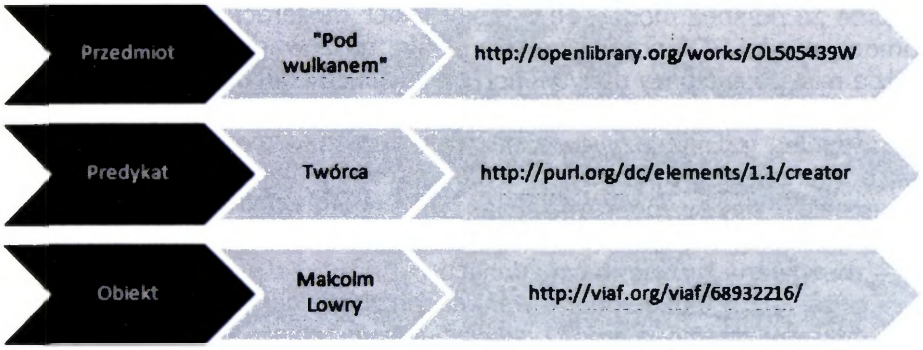
Protokół komunikacyjny HTTP (Hypertext Transfer Protocol) zapewnia znormalizowany sposób komunikacji między komputerami. Umożliwia przesłanie informacji o obiekcie lub zasobie zidentyfikowanym za pomocą URI do maszyny, w której wywołano identyfikator za pomocą poleceń protokołu HTTP. Jest on uniwersalnym mechanizmem pozyskiwania zasobów lub informacji o obiektach nieposiadających postaci cyfrowych, które nie mogą zostać przesłane bezpośrednio za jego pomocą.

O ile identyfikatory URI i protokół komunikacyjny HTTP stanowią podstawę technologii Linked Data, to ideą tej koncepcji jest zastosowanie jednego, wspól-

nego dla wszystkich źródeł danych modelu reprezentacji wiedzy o obiektach i zasobach. Tę funkcję pełni RDF jako język reprezentacji wiedzy stosowany do wyrażania metainformacji o obiektach i relacjach, w jakie wchodzą z innymi. „Jest to specyfikacja modelu metadanych, ustanowiona przez World Wide Web Consortium (W3C) na podstawie reguł języka XML. Podstawowym założeniem wyjściowym jest tutaj rozdzielenie warstwy wizualizacji informacji od warstwy ich przetwarzania, czyli oddzielenie struktury dokumentu od jego treści. RDF udostępnia przede wszystkim standaryzowaną syntaktykę wykorzystywaną do reprezentacji metainformacji dla obiektu informacyjnego. Oznacza to, że określa reguły gramatyczne tworzenia interpretowalnych przez programy „wypowiedzi" o obiektach informacyjnych i ich właściwościach. RDF zakłada trójelementową budowę takich wyrażeń, które w jej dokumentacji noszą nazwę *trójek RDF* (ang. *RDF triples*). Składają się one z:

– podmiotu, który jest opisywanym zasobem,
– predykatu określającego opisywaną własność,
– obiektu, który określa wartość tej własności" [31].

Przedmiot, który jest identyfikowany za pomocą jego unikalnego URI w danym zbiorze, jest charakteryzowany poprzez wskazanie na jego własność (predykat) oraz wartość tej cechy (obiekt). Najbardziej pożądaną sytuacją jest taka, w której wszystkie elementy takiej deklaracji są przywoływane za pomocą ich unikalnych URI (rys. 1).



Rys. 1. Budowa deklaracji RDF

Odwzorowanie informacji o przedmiocie odbywa się poprzez odwołania bezpośrednio do identyfikatorów wskazujących na własność oraz jej wartość, a pośrednio do sformalizowanych przestrzeni nazw. Są to zbiory słownictwa, na przykład w postaci schematów metadanych, słowników systemów organizacji wiedzy, ontologii internetowych, które opisuje się za pomocą standaryzowanych języków reprezentacji. Należą do nich *RDF* Schema (RDFS) czy Web Ontology Language (OWL). W przytoczonym przykładzie deklaracja reprezentująca stwierdzenie, że autorem dzieła „Pod wulkanem" jest Malcolm Lowry, stosuje odwołania do:

– URI dzieła „Pod wulkanem" z bibliograficznego zbioru danych projektu Open Library (http://openlibrary.org/),

– URI atrybutu dc:Creator ze schematu metadanych Dublin Core,

– URI rekordu kartoteki wzorcowej nazw osobowych VIAF (The Virtual International Authority File, http://www.viaf.org).

Taki sposób ekspresji powoduje, że aplikacja, która ją odczytuje, za każdym razem odwołuje się do zadeklarowanych przestrzeni nazw, w których zidentyfikowano obiekty, ich własności czy relacje, jakie mogą tworzyć z innymi. Tym samym mamy do czynienia nie z ciągami znaków „niezrozumiałymi" dla programów je przetwarzających, lecz elementami, których kontekst znaczeniowy został zdefiniowany w przestrzeniach nazw. Stąd obecność przymiotnika „semantyczna" w odniesieniu do Semantic Web. U jej podstaw leży stosowanie sformalizowanego systemu odniesienia do opisu jej zasobów. Pozwala to na przetwarzanie zasobów przez aplikacje, które każdorazowo, aby „zrozumieć" znaczenie terminów w deklaracjach, odwołują się do przestrzeni nazw, z których one pochodzą. Termin ma znaczenie dla aplikacji tylko w takim stopniu, w jakim zostało ono wcześniej zdefiniowane w deklaracjach oraz wywoływanych przestrzeniach nazw. Koncepcja Linked Data zakłada jedynie łączenie i ekspresję danych wykorzystujących istniejące standardy. W tym przypadku w jednej deklaracji wykorzystano elementy pochodzące z trzech różnych zbiorów słownictwa: dwóch kartotek wzorcowych oraz schematu metadanych. Pomimo heterogenicznego charakteru tych zasobów istnieje możliwość odwoływania się do nich w jednej deklaracji, zachowując jednocześnie warunek przetwarzalności przez aplikacje. Zarówno deklaracja, jak i struktura tych zbiorów są bowiem opisane za pomocą modelu RDF. Zbiór takich deklaracji, skonstruowany za pomocą składni RDF oraz odwołania do sformalizowanych przestrzeni nazw, będzie miał postać bazy danych reprezentującej model Linked Data.

Tak skonstruowana sieć zbiorów danych i powiązań między nimi, operująca wspólną rodziną języków reprezentacji wiedzy, może być postrzegana jako odrębna warstwa World Wide Web. Jest ona umiejscowiona pod warstwą dokumentów sieciowych, rozdzielając tym samym warstwę opisu danych od warstwy ich prezentacji. Jest to sieć:

– zbudowana z różnego typu danych,

– gdzie każdy może opublikować własne zbiory danych,

– gdzie twórcy mogą wykorzystywać dowolne zbiory słownictwa do reprezentacji danych,

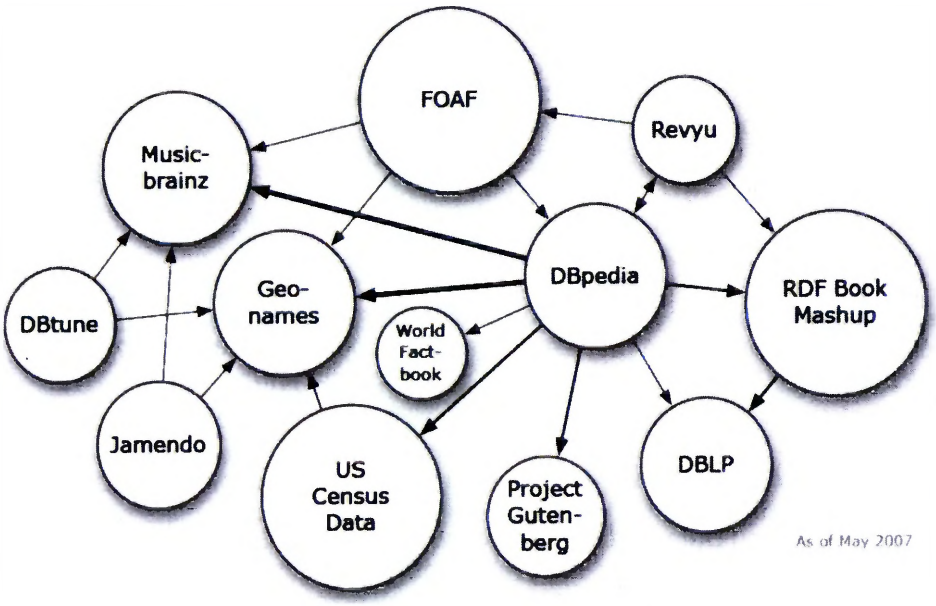
– której elementy są połączone ze sobą za pomocą powiązań w składni RDF i tworzą globalną przestrzeń danych obejmującą także całe zbiory danych i słownictwa wykorzystywanego do ich opisu.

Z punktu widzenia przetwarzania informacji, sieć danych cechuje:

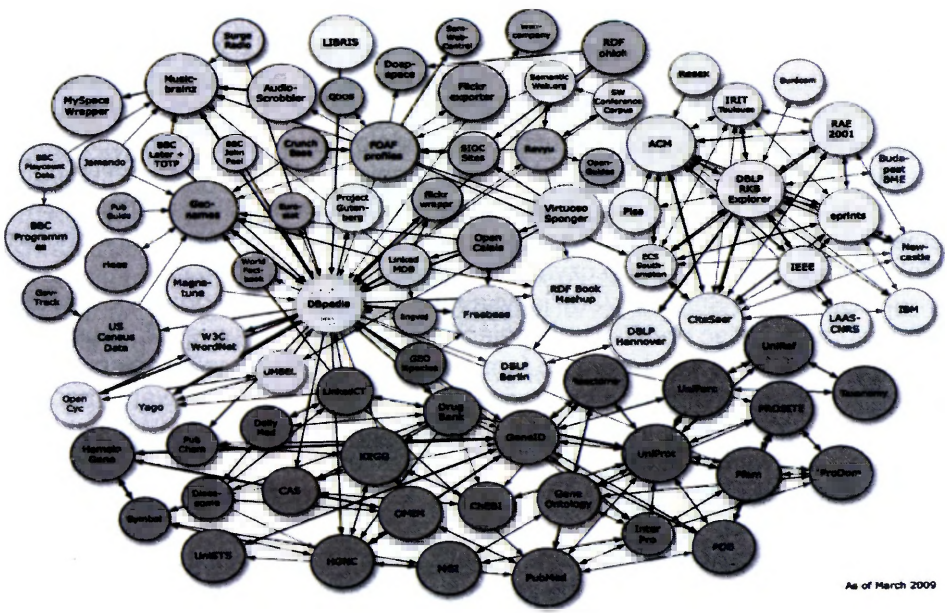
– oddzielenie danych od warstw formatowania i prezentowania;

– samoopisujący się charakter danych (*self-describing*). Deklaracje RDF zawierają odwołania do identyfikatorów własności oraz ich wartości, które są zdefiniowane w przestrzeniach nazw. Aplikacje przetwarzające takie deklaracje są przekierowywane do takich zbiorów w celu identyfikacji kontekstu znaczeniowego dla zastosowanego terminu;

– dostęp do zbiorów danych poprzez interfejsy programowania aplikacji (API wykorzystujące standaryzowany język zapytań SPARQL) oraz protokół przesyłania danych HTTP;



Rys. 2. Chmura Linked Data w maju 2007 r. [35]



Rys. 3. Chmura Linked Data w marcu 2009 r. [10]

– otwarty charakter (mechanizmy pozyskiwania danych nie są programowane dla konkretnego zbioru danych, lecz dzięki wspólnemu modelowi reprezentacji oraz powiązaniom RDF mogą eksplorować inne zasoby) [10].

Chmura Linked Data

Zbiory danych opublikowane zgodnie z modelem Linked Data tworzą tzw. chmurę danych powiązanych (*linked data Cloud*), czyli globalną sieć zbiorów danych stosujących wzajemne odwołania. W maju 2007 r. liczyła ona zaledwie kilka zbiorów danych realizujących ten model. Obecnie jest ich ponad 200 i szacuje się, że zawarta w nich łączna liczba deklaracji przekracza 25 miliardów trójek RDF [16].

Diagram chmury danych powiązanych pokazuje wielkość zbiorów danych i występowanie relacji pomiędzy nimi. Zmiany w rozmiarze chmury oraz zakres i kierunki odwołań można prześledzić na stronie internetowej Richarda Cyganiaka [15], który od 2007 r. opracowuje tego typu diagramy.

Zasoby chmury Linked Data

Zasoby Linked Data można interpretować w kategoriach gotowych danych opublikowanych za pomocą technologii Semantic Web oraz założeń tego modelu publikowania lub w postaci sformalizowanych przestrzeni nazw (schematów metadanych, ontologii), czyli zbiorów słownictwa stosowanych do ich opisu.

Złożoność otaczającego nas świata oraz jego projekcji zawartej w zasobach sieciowych wymusza stosowanie wielu narzędzi do jego efektywnego opisu i uchwycenia istotnych cech oraz relacji, jakie w nim zachodzą. W kontekście Semantic Web mamy do czynienia z dużą liczbą narzędzi wykorzystywanych do tego celu, zaczynając od schematów metadanych, przez zbiory słownictwa stosowane do wyrażania wartości określonych własności obiektów, po skomplikowane ontologie internetowe służące do modelowania fragmentu rzeczywistości, który za ich pomocą ma zostać formalnie opisany. Formalny charakter opisu sporządzany za ich pomocą, czyli zgodny ze standardami sieci semantycznej, powoduje, że są one gotowymi narzędziami opisu danych włączanych w chmurę Linked Data. Mamy więc do czynienia z wykorzystaniem np.:

– RDFS (RDF Schema) [30], języka reprezentacji wiedzy opartego na RDF, który wprowadzając podstawowe pojęcia (np. pojęcie klasy, jej własności, zakres), służy przede wszystkim do formalnego uporządkowania zapisów;

– OWL (Web Ontology Language) [26], narzędzia do budowania bardziej szczegółowych opisów fragmentów rzeczywistości niż RDFS. OWL jest standardem sieciowym przyjętym przez W3C. Rozszerza on w dużym stopniu możliwości modelowania rzeczywistości RDFS i jest narzędziem o wyjątkowo dużej sile semantycznej;

– Dublin Core Metadata Element Set [20], najczęściej stosowanego schematu metadanych dla obiektów cyfrowych, którego zapisy są wyrażane za pomocą RDF;

– FOAF (Friend of a Friend) [23], schematu formatowania metadanych służących do budowania wizytówek sieciowych i tworzenia sieci społecznych;

– SKOS (Simple Knowledge Organization System) [38], który „jest rozszerzeniem RDF wykorzystywanym do reprezentacji wiedzy w sieciowych systemach organizacji wiedzy. Jest to język opisu prostych systemów organizacji

wiedzy dla zasobów sieciowych" [31]. SKOS jest formalnym językiem opisu słowników systemów leksykalnych zarówno tradycyjnych, jak i sieciowych;

– SIOC (Semantically-Interlinked Online Communities) [37], ontologii stworzonej na bazie RDFS I OWL służącej do modelowania i opisu obiegu informacji w społecznościach sieciowych. Jest to standard służący do opisu treści generowanych przez użytkowników takich społeczności (np. fora dyskusyjne, blogi, galerie zdjęć), który wraz z FOAF jest wykorzystywany do tworzenia zbiorów danych przetwarzalnych przez aplikacje sieciowe;

– DOAP (Description of a Project) [21], ontologii wykorzystującej RDFS oraz m.in. zestaw elementów Dublin Core i FOAF do tworzenia opisów projektów oprogramowania, w szczególności Open Source;

– GeoNames Ontology [42], ontologii dedykowanej, opracowanej na potrzeby projektu GeoNames (http://www.geonames.org/), służącej do formalnej ekspresji metadanych geoprzestrzennych;

– Music Ontology [27], ontologii identyfikującej obiekty, własności oraz relacje dotyczące dokumentów dźwiękowych (np. wykonawca, album, utwór, wykonanie, aranżacja, itd.).

W efekcie prac prowadzonych nad technologiami dla Semantic Web opracowano wiele narzędzi tego typu w postaci schematów metadanych, zbiorów słownictwa czy gotowych ontologii internetowych je wykorzystujących.

Zagadnienia związane z Linked Data można rozpatrywać z punktu widzenia zbiorów danych opublikowanych dla sieci semantycznej. Rozmiar chmury Linked Data pokazuje, że mamy do czynienia z ciągle rozrastającą się globalną siecią zbiorów powiązanych danych. Warto wymienić kilka z nich. Są to:

– DBpedia (http://dbpedia.org/), czyli „semantyczna" wersja Wikipedii. Jest to projekt Freie Universitat w Berlinie oraz Uniwersytetu w Lipsku. W 2007 r. opublikowano pierwszą wersję Wikipedii w formie danych powiązanych. Najnowsza wersja pochodzi z kwietnia 2010 r. i zawiera opisy 3,4 miliona obiektów [9]. Konwersja danych z Wikipedii do jej semantycznej wersji polega na zakodowaniu abstraktu artykułu oraz metadanych zawartych w bocznej części opisu do postaci deklaracji RDF za pomocą ontologii stosowanej w tym projekcie, np. „semantyczna" forma artykułu w DBpedii dla hasła Malcolm Lowry – http://dbpedia.org/page/Michael_Lowry;

– The DBLP Computer Science Bibliography [40], bibliograficzna baza danych z zakresu informatyki. Zawiera opisy ponad 800 tys. jednostek bibliograficznych. Dostępna jest w postaci hipertekstowej oraz „surowej", czyli deklaracji RDF;

– GeoNames (http://www.geonames.org), baza danych zawierająca dane geograficzne o ponad 2,6 miliona jednostek geograficznych na całym świecie. Została opublikowana w formie graficznej oraz w modelu Linked Data;

– BBC Music (http://www.bbc.co.uk/music/artists), baza danych o artystach prezentowanych w rozgłośni BBC. Również dostępna w formie graficznej oraz w modelu Linked Data, np.:

- informacje o grupie Nirvana w formie graficznej: http://www.bbc.co.uk/music/artists/5b11f4ce-a62d-471e-81fc-a69a8278c7da,
- informacje o grupie Nirvana w formie deklaracji RDF: http://www.bbc.co.uk/music/artists/5b11f4ce-a62d-471e-81fc-a69a8278c7da.rdf;

– CIA World Factbook (http://www4.wiwiss.fu-berlin.de/factbook/), publikacją Centralnej Agencji Wywiadowczej USA, która zawiera podstawowe dane na temat wszystkich krajów świata. Wersję bazy danych dla przetwarzania przez aplikacje przygotowano w Freie Universitat w Berlinie;

– Dane rządowe, raporty, analizy, zestawienia statystyczne:

- rządu USA (http://www.data.gov/catalog/raw),
- rządu Wielkiej Brytanii (http://data.gov.uk/).

Przytoczone przykłady to tylko wybrane inicjatywy związane z publikowaniem danych w chmurze Linked Data. Obszerny zbiór baz danych dla Semantic Web dostępny jest w serwisie Comprehensive Knowledge Archive Network (CKAN) [16], projektu realizowanego przez Open Knowledge Foundation.

Zasada tożsamości

Kluczem do eksploracji chmury danych powiązanych i pozyskiwania danych z wielu źródeł jest możliwość sformalizowania zasady tożsamości. Polega to na zastosowaniu relacji wykładnika języka OWL – owl:sameAs jako związku tożsamości znaczeniowej ustanawianego pomiędzy identyfikatorami obiektów z różnych zbiorów. W warstwie technicznej pozwala to na agregację rozproszonych metadanych odnoszących się do wielu obiektów, dla których ustanowiono ten typ zależności na przykład we wspomnianej DBpedii, URI obiektu odnoszącego się do osoby Woodego Allena powiązano wykładnikiem owl:sameAs z URI obiektów z baz New York Times, Open Cyc i Zitgist. Aplikacja napotykająca na taki zapis będzie mogła uzupełnić dane o Woodym Allenie z Dbpedii o zapisy pozostałych baz dzięki formalnemu wykładnikowi tożsamości. Formalizacja prawa identyczności ma istotne znaczenie dla pozyskiwania danych w chmurze Linked Data. Ważne są także własności tego typu relacji ustalone w ramach języka OWL, np. symetryczność (zob. [18]).

Dostęp do danych i publikowanie Linked Data

Dostęp do zbiorów danych polega na tzw. negocjacji zawartości pomiędzy serwerem udostępniającym dane a aplikacją, która wymusza ich pobieranie. Odbywa się to za pomocą standaryzowanego języka SPARQL (Simple Protocol And RDF Query Language), który jest językiem zapytań i protokołem dla plików RDF. Forma i zakres zapytań oraz zasady przetwarzania pobranych danych są opracowywane dla aplikacji je przetwarzających i tworzą ich interfejsy programowania (API – Application Programming Interface). Inną formę dostępu realizują narzędzia wyszukiwania informacji, które wykorzystują interfejsy graficzne do prezentacji rezultatów wyszukiwania i prezentowania powiązań między danymi. Umożliwiają one również eksplorację tych danych dzięki wizualizacji powiązań za pomocą hipertekstu. Są to różnego typu przeglądarki (np. Marbles – http://www5.wiwiss.fu-berlin.de/marbles, Tabulator – http://www.w3.org/2005/ajar/tab) oraz wyszukiwarki (np. Falcons – http://iws.seu.edu.cn/services/falcons/).

Problematyka publikowania danych w modelu Linked Data została przedstawiona w licznych opracowaniach i poradnikach (np. [8], [19]). Powstało także wiele aplikacji, w większości darmowych, które umożliwiają konwersję danych w istniejących formatach do postaci RDF (tzw. RDFizer) oraz oprogramowania służącego do generowania relacyjnych baz danych na podstawie zbiorów deklaracji (np. D2R Server). [10]

Linked Data a dane bibliograficzne

W styczniu 2010 r. ukazał się raport przygotowany przez American Library Assiocation pt. „Understanding the Semantic Web: Bibliographic Data and Metadata" [40]. W całości poświęcono go zagadnieniom uczestnictwa bibliotek w Semantic Web. Stwierdzono tam, że zmiany, które muszą nastąpić w bibliotekach, mają polegać na przekształceniu katalogu bibliotecznego z formy zamkniętej do postaci zbioru danych powiązanych, gdzie odwołuje się do innych zasobów sieciowych i do którego prowadzą powiązania z innych źródeł danych [13]. Uczestniczenie bibliotek w tworzeniu zasobów Semantic Web oznacza konieczność stosowania modelu Linked Data. Dane bibliograficzne są sporządzane w bibliotekach zgodnie z przyjętymi standardami opisu (m.in. normy krajowe, ISBD, AACR2) oraz na podstawie przyjętych formatów zapisu (np. MARC 21). Zapowiadana w tym dokumencie zmiana paradygmatu ma wymusić formalizację sposobu zapisu metadanych bibliograficznych z wykorzystaniem technologii Semantic Web. Celem zmiany ma być współtworzenie zasobów sieci trzeciej generacji, a nie wyłącznie obecność katalogów w środowisku sieciowym. Wymaga to takiego zapisu i ekspresji danych, który będzie „czytelny" dla aplikacji Semantic Web.

21 maja 2010 r. w ramach konsorcjum World Wide Web powstał projekt o nazwie Library Linked Data Incubator Group [42]. Celem projektu jest pomoc w zwiększeniu interoperacyjności danych bibliotecznych w sieci poprzez stworzenie platformy wymiany poglądów dla osób zainteresowanych tematyką wdrażania koncepcji Linked Data w środowisku bibliotekarskim. Grupę tworzy 43 członków, 20 organizacji współpracujących oraz 10 zaproszonych ekspertów. Do zadań grupy należy:

– zbieranie informacji o projektach wdrażania technologii Semantic Web w bibliotekach,

– wspieranie współpracy pomiędzy podmiotami (biblioteki, muzea, archiwa, wydawcy) zainteresowanymi wprowadzeniem danych o dobrach kultury do przestrzeni informacyjnej Semantic Web,

– identyfikacja modeli danych, zbiorów słownictwa, ontologii oraz zasad i tworzenia lub poprawy interoperacyjności pomiędzy nimi,

– wskazywanie na zapotrzebowanie na nowe standardy, wytyczne i metodykę działań [2].

Efektem prac prowadzonych w ramach projektu są tzw. wzorce rozwiązań (ang. *case templates*), które zawierają rozwiązania i algorytmy postępowania dla konkretnych projektów dotyczących uczestnictwa bibliotek w Semantic Web. Aktualnie opracowano 13 takich miniprojektów, które dotyczą m.in. adaptacji

istniejących modeli konceptualnych i systemów organizacji wiedzy, wykorzystania technologii Semantic Web do przetwarzania danych bibliograficznych oraz zarządzania i dystrybucji metadanych (http://www.w3.org/2005/Incubator/lld/wiki/Use_Cases).

Problematyka danych bibliograficznych jako Linked Data jest obszerna i obejmuje zagadnienia związane zarówno z konwersją danych, jak i metodami i licencjami na jakich są udostępniane. Pierwszym krokiem jest jednak przygotowanie nowych narzędzi i adaptacja już istniejących do procesu konwersji danych. Wymaga to formalizacji opisu zarówno standardów opisu bibliograficznego, jak i kartotek wzorcowych oraz słowników języków informacyjno-wyszukiwawczych stosowanych do opisu zawartości dokumentów.

Functional Requirements for Bibliographic Records (FRBR) [28] jest modelem konceptualnym, który na najwyższym poziomie abstrakcji porządkuje uniwersum bibliograficzne. Wyznacza on grupy jednostek oraz relacje, jakie mogą między nimi zachodzić w szerokim kontekście opisu bibliograficznego. Konsekwencją takiej perspektywy interpretacji uniwersum bibliograficznego są dwa kolejne modele opracowane przez IFLA: Functional Requirements for Authority Data (FRAD) [24] i Functional Requirements for Subject Authority Data (FRSAD) [25], wykorzystywane do porządkowania struktury kartotek wzorcowych. Pomimo trudności z ich zastosowaniem w prezentacji informacji w katalogach i kartotekach bibliotecznych (np. [3], [12]) modele te pełnią ważną funkcję w aspekcie konceptualnym i terminologicznym. Dlatego też zostały wykorzystane do modelowania uniwersum bibliograficznego w postaci ontologii internetowych z przeznaczeniem do wykorzystania w opisie danych dla Semantic Web. Obecnie trwają prace nad przełożeniem zapisu struktury tych modeli z wykorzystaniem składni RDF i opracowaniem gotowych ontologii internetowych, do których będzie można się odwoływać w deklaracjach [17]. W serwisie Open Metadata Registry (http://metadataregistry.org) można na bieżąco śledzić postępy tych prac. Zamieszczono tam wyniki prac nad „semantyzacją" FRBR, ISBD, FRSAD, a także Resource Description and Access (RDA) [29], czyli nowych anglo-amerykańskich zasad katalogowania, które w założeniu mają być wykorzystywane do opisu zasobów sieciowych.

Obok projektów zmierzających do formalizacji zapisu modeli danych bibliograficznych i standardów katalogowania prowadzi się również badania nad zastosowaniem nowych narzędzi do zapisu metadanych dla jednostek bibliograficznych. Jednym z takich projektów jest ontologia Bibliographic Ontology (BIBO – http://bibliontology.com). Jest to prosta ontologia definiująca własności jednostek bibliograficznych, która oprócz własnej specyfikacji wykorzystuje także schemat Dublin Core i wspominany FOAF. Innym przykładem jest działalność Biblioteki Kongresu na rzecz „semantyzacji" zbiorów słownictwa i formatów opisu. Serwis internetowy (http://www.loc.gov/standards/marcxml/) tej biblioteki udostępnia darmowe narzędzia i dokumentację, która pozwala na konwersję formatu MARC 21 do formatu RDF z wykorzystaniem m.in. elementów schematu Dublin Core. Ciekawy przykład inicjatywy spoza środowiska bibliotekarskiego to projekt MarcOnt Bibliographic Ontology [33]. Ontologia MarcOnt (http://www.marcont.org/) jest przeznaczona dla bibliotek cyfrowych i wykorzystuje schemat Dublin Core rozszerzony o jednostki, klasy i własności zaprojektowane za

pomocą OWL. MarcOnt ma niewiele wspólnego z formatem MARC pomimo nazwy, która może nasuwać takie skojarzenie.

Podstawę semantycznego charakteru sieci trzeciej generacji stanowią sformalizowane przestrzenie nazw, do których odwołują się deklaracje identyfikujące i opisujące obiekty. Są to zbiory słownictwa wyznaczające kontekst znaczeniowy dla terminów w nich występujących. W licznych pracach analizujących szanse udziału danych bibliograficznych w chmurze Linked Data szczególnie podkreśla się rolę zbiorów słownictwa opracowanego w społeczności bibliotekarskiej. Kartoteki wzorcowe, słowniki języków informacyjno-wyszukiwawczych to gotowe narzędzia leksykalne, na których „semantyzację" czekają twórcy aplikacji dla Semantic Web. Do najważniejszych prac prowadzonych w środowisku bibliotekarskim, które zmierzają do konwersji tego typu źródeł, należą projekt VIAF oraz publikowanie zasobów leksykalnych za pomocą języka SKOS.

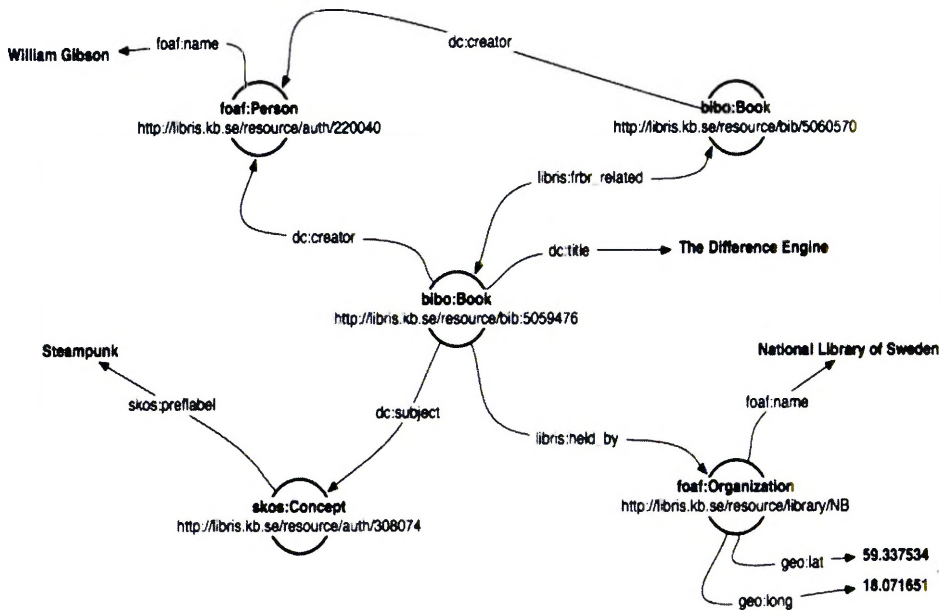
VIAF (The Virtual International Authority File, http://viaf.org/) to projekt realizowany przez Bibliotekę Kongresu oraz Bibliotekę Narodową Niemiec. Jest to międzynarodowa kartoteka nazw osobowych, której zadaniem jest ułatwienie wyszukiwania informacji poprzez zniesienie barier językowych. W projekcie bierze udział 18 instytucji z całego świata, w tym także z Polski (NUKAT). Obok podstawowego założenia projektu, czyli udostępnienia międzynarodowej kartoteki nazw osobowych, celem ma być publikacja tych danych w chmurze Linked Data (zob. rys 1).

Dzięki opracowaniu formalnego języka opisu słowników prostych systemów organizacji wiedzy, czyli SKOS [33, 38], udostępniono narzędzie służące do przekładu struktury słowników tradycyjnych języków informacyjno-wyszukiwawczych do postaci przetwarzalnej przez aplikacje sieciowe. SKOS oprócz umożliwiania opisu struktury paradygmatycznej systemów organizacji wiedzy działa także jako język przejścia. Pozwala na odwzorowanie zależności semantycznych pomiędzy różnymi systemami organizacji wiedzy. W strukturze SKOS wyodrębniono wykładniki typów zgodności zakresowej pomiędzy wiązanymi w ten sposób jednostkami z różnych źródeł. Cały czas trwają prace nad konwersją zapisu słowników wybranych języków informacyjnych oraz źródeł słownictwa wykorzystywanych do opisu zasobów bibliotek. Do zakończonych projektów należą m.in. :

- AGROVOC Agricultural Thesaurus,
- Australian Public Affairs Information Service (APAIS) Thesaurus,
- DDC Dewey Decimal Classification (projekt eksperymentalny, wybrane klasy najwyższego stopnia podziału),
- GEMET General Multilingual Environmental Thesaurus,
- GeoNames, kartoteka wzorcowa nazw geograficznych,
- HPMulti Thesaurus for Health Promotion,
- ISO 3166-1 Country Codes,
- ISO639 Codes for the representation of names of languages,
- IVOA astronomy vocabularies,
- Kartoteki wzorcowe Biblioteki Narodowej Niemiec,
- Library of Congress Subject Headings,
- Medical Subject Headings,
- NASA taxonomy,

– New York Times subjects,
– RAMEAU,
– STW Thesaurus for Economics,
– The General Multilingual Environmental Thesaurus (GEMET),
– Thesaurus for Graphic Materials,
– Thesaurus for the Social Sciences,
– Thesaurus of territorial units of Spain and France,
– UK Archival Thesaurus (UKAT),
– UK Government Category List (GCL),
– UNESCO Thesaurus [39].

Dotychczas niewiele bibliotek podjęło decyzję o opublikowaniu swoich zasobów w chmurze Linked Data. Są to Niemiecka Biblioteka Narodowa, Biblioteka Narodowa Węgier oraz LIBRIS – katalog zbiorczy szwedzkich bibliotek akademickich. Pierwsza inicjatywa bibliotek w tym kontekście to opublikowanie pod koniec 2008 roku katalogu LIBRIS. Dane z tego zbioru opisano za pomocą wielu narzędzi ekspresji metadanych (rys. 4), m.in.:
– Schematu metadanych Dublin Core,
– Schematu FOAF,
– Ontologii BIBO,
– SKOS jako sposobu ekspresji opisu rzeczowego,
– Opracowanej na potrzeby systemu ontologii LIBRIS.



Rys. 4. Schemat modelowania uniwersum bibliograficznego w projekcie LIBRIS [http://blog.libris.kb.se/semweb/wp-content/uploads/2008/12/linked-libris1.jpg]

Rysunek 4 przedstawia w formie schematu zestaw deklaracji dotyczących dokumentu pt. *The Difference Engine*, którego autorem jest William Gibson. Reprezentacja wiedzy dla tego fragmentu uniwersum bibliograficznego została

skonstruowana za pomocą wspomnianych ontologii oraz własności i relacji, które można wyrazić za ich pomocą. W centrum grafu znajduje się URI dla dokumentu, który zidentyfikowano jako *książka* (bibo:Book), której tytuł (dc:title) ma formę *The Difference Engine,* jej autorem (dc:creator) jest obiekt zidentyfikowany jako osoba (foaf:Person), która nazywa się (foaf:name) *William Gibson.* Treść dokumentu jest opisana poprzez wskazanie na jego przedmiot (dc:subject), który jest zidentyfikowany jako *pojęcie* (skos:Concept) pochodzące z określonego systemu organizacji wiedzy i ma formę (skos:preflabel) *Steampunk.* Książka ta jest w posiadaniu (libris:held_by) organizacji (foaf:Organization) o nazwie (foaf:name) *Biblioteka Narodowa Szwecji (National Library of Sweden)*, która zlokalizowana jest na 59 stopniu szerokości (geo:lat) i 18 stopniu długości geograficznej (geo:long), itd.

Siła ekspresji tego modelu reprezentacji wiedzy (RDF) jest olbrzymia. Wymaga jednak stosowania adekwatnych narzędzi modelowania fragmentu rzeczywistości (ontologii, schematów metadanych, zasobów leksykalnych). W przypadku uniwersum bibliograficznego ten wymóg jest spełniony poprzez funkcjonowanie ugruntowanych standardów i schematów opisu oraz systemów porządkowania pojęć, które pozwalają na identyfikację dzieła na odpowiednim poziomie szczegółowości, wskazanie na istotne cechy jego treści oraz odwzorowanie relacji bibliograficznych, w jakie wchodzi z innymi dziełami. Charakterystyczną cechą modelu reprezentacji wiedzy opartego na RDF jest jego elastyczność. Z punktu widzenia konstruowania metadanych bibliograficznych mamy do czynienia z otwartą formą rekordu bibliograficznego, który tworzą deklaracje składające się na konkretny komunikat o charakterze metainformacyjnym. Wykorzystanie różnych narzędzi opisu jednostek bibliograficznych pozwala na tworzenie kompleksowych narzędzi opisu, dostosowanych do potrzeb organizacji lub opisywanej dziedziny czy domeny. Otwarty model reprezentacji wiedzy oparty na współdzielonej składni (RDF) pozwala na budowanie profilów aplikacyjnych dostosowanych do wymogów reprezentacji metadanych w poszczególnych dziedzinach wiedzy, systemach, aplikacjach, usługach (zob. np. zalecenia ustaleń singapurskich dla Dublin Core) [14]. Korzyści płynące dla bibliotek z tego otwartego modelu reprezentacji są dwojakie: z jednej strony mamy możliwość uzupełniania opisów jednostek bibliograficznych poprzez odwoływanie się do źródeł zewnętrznych, z drugiej, to dane bibliograficzne mogą zostać wykorzystane poza kontekstem wyznaczanym przez miejsce i usługi bibliotek w systemach i serwisach informacyjnych.

Problemy

Tim Berners-Lee, mówiąc w 2009 r. o Linked Data [6] na prestiżowej konferencji „TED – Ideas worth spreading", operował sloganem „raw data now!". Odniósł się do konieczności przede wszystkim udostępniania zbiorów danych zgodnie z założeniami koncepcji Linked Data. Problem tkwi w tym, że do końca nie jesteśmy w stanie określić wszystkich możliwych sposobów wykorzystania takiego modelu danych w odniesieniu do naszego środowiska. Duże zainteresowanie chmurą danych powiązanych z różnych sfer rejestrowania dorobku

intelektualnego i dziedzictwa kulturowego spowodowało pierwszą falę krytyki oraz wskazanie jej niedostatków. Pierwszy problem to brak skodyfikowanej wiedzy typu know-how, a więc wszelkiego rodzaju poradników, instrukcji i metodyki tworzenia zasobów w ramach koncepcji Linked Data. Cały czas poruszamy się na poziomie przygotowania i publikacji danych, nie odnosząc się do konieczności zaprojektowania i wdrożenia mechanizmów ich przetwarzania i pozyskiwania informacji. Nie ma ostatecznych rozstrzygnięć dotyczących długoterminowej dostępności do zbiorów danych oraz stabilności mechanizmów odsyłania (URI). Równie istotny problem to wiarygodność deklaracji oraz mechanizmów jej weryfikacji, czyli problem fałszywych metadanych i tego konsekwencji. Z punktu widzenia ekspresji i reprezentacji danych pojawił się także problem opisu całych kolekcji danych, czyli metadanych na poziomie kolekcji (ang. *collection level description*) oraz braku mechanizmów weryfikacji nieścisłości w opisie tego samego atrybutu dla dokumentu za pomocą różnych narzędzi (przy jednoczesnym wykorzystaniu zasady tożsamości – owl:sameAs).

Należy pamiętać, że Semantic Web to wizja, której jedną z materializacji jest chmura danych powiązanych. Koncepcja Linked Data dotyczy metod ekspresji, reprezentacji, łączenia i współdzielenia danych w semantycznej sieci, które wykorzystują istniejące standardy i narzędzia sieciowe. Stabilność modeli opisu uniwersum bibliograficznego i narzędzi wykorzystywanych w bibliotekach do tego celu pozwala mieć nadzieję, że będziemy aktywnymi uczestnikami tworzenia zawartości Semantic Web, o ile znajdą się chętni i pieniądze na sfinansowanie tego typu inicjatyw.

Bibliografia

1. Allemang D., Hendler J.: *Semantic web for the working ontologist modeling in RDF, RDFS and OWL*. Burlington 2007, s. 13.
2. Baker T., Bermes E.: *Library Linked Data Incubator Group Charter*. [online]. [dostęp: 05.11. 2010]. Dostępny w World Wide Web: <http://www.w3.org/2005/Incubator/lld/charter>.
3. Bennett R., Lavoie B., O'Neill E.: *The Concept of a Work in WorldCat: An Application of FRBR*. „Library Collections, Acquisitions, and Technical Services" 2003, vol. 27, no. 1, pp. 45-59.
4. Berners-Lee T.: *Al. Uniform Resource Identifier (URI): Generic Syntax*. [online]. [dostęp: 05. 11.2010]. Dostępny w World Wide Web: <http://tools.ietf.org/html/rfc3986>.
5. Berners-Lee T.: *Hearing on the Digital Future of the United States: Part I – The Future of the World Wide Web*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html>.
6. Berners-Lee T.: *Linked Data – Design Issues*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/ DesignIssues/LinkedData.html>.
7. Berners-Lee T.: *On the next Web*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html>.
8. Bizer C., Cyganiak R., Heath T.: *How to Publish Linked Data on the Web*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/>.
9. Bizer Ch.: *DBpedia 3.5 released*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://blog.dbpedia.org/2010/04/12/dbpedia-35-released/>.
10. Bizer Ch., Heath T., Berners-Lee T.: *Linked Data – The Story So Far*. „International Journal on Semantic Web and Information Systems" 2009, no. 3, pp. 1-22.
11. Campbell L., MacNeill S.: *The Semantic Web, Linked and Open Data*. [online]. [dostęp: 05. 11.2010]. Dostępny w World Wide Web: <http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf>.

12. Clinker H.: *FRBR: Functional Requirements for Bibliographic Records Application of the Entity-Relationship Model.* „Library Resources & Technical Services" 2002, vol. 46, no. 4. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.oclc.org/research/publications/archive/2002/oneill_frbr22.pdf>.
13. Coyle K.: *Library Data in a Modern Context.* „Library Technology Reports" 2010, no. 1, pp. 5-13.
14. Coyle K.: *Metadata models for the World Wide Web.* „Library Technology Report" 2010, no. 46, pp. 12-19.
15. Cyganiak R.: *About the Linking Open Data dataset cloud.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://richard.cyganiak.de/2007/10/lod/>.
16. *Datasets in the next LOD Cloud.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www4.wiwiss.fu-berlin.de/lodcloud/>.
17. Davis I., Newman R.: *Expression of Core FRBR Concepts in RDF.* [online]. [dostęp: 05.11. 2010]. Dostępny w World Wide Web: <http://vocab.org/frbr/core.html>.
18. Ding L., Finin T., Shinavier J., McGuinness D.: *owl:sameAs and Linked Data: An Empirical Study.* W: *Proceedings of the WebSci10: Extending the Frontiers of Society Online*, Raleigh. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://journal.webscience.org/403/>.
19. Dodds L., Davis I.: *Linked Data Patterns. A pattern catalogue for modelling, publishing, and consuming Linked Data.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://patterns.dataincubator.org/book/linked-data-patterns.pdf>.
20. *Dublin Core Metadata Element Set, Version 1.1.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://dublincore.org/documents/dces/>.
21. Dumbill E.: *Description of a Project.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://trac.usefulinc.com/doap>.
22. Fay R.: *Linked: Semantic Web & metadata.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.slideshare.net/robinfay/semantic-web-cataloging-metadata>.
23. *FOAF Vocabulary Specification 0.98.* Red. D. Brickley, L. Miller. [online]. [dostęp: 05.11. 2010]. Dostępny w World Wide Web: <http://xmlns.com/foaf/spec/>.
24. *Functional Requirements for Authority Data.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.ifla.org/publications/functional-requirements-for-authority-data>.
25. *Functional Requirements for Subject Authority Data (FRSAD) A Conceptual Model.* Red. M. Zeng, M. Žumer, A. Salaba. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://nkos.slis.kent.edu/FRSAR/FRSAD-Report.pdf>.
26. Herman I.: *Web Ontology Language (OWL).* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/2004/OWL/>.
27. *Music Ontology Specification.* Red. Y. Raimond, F. Giasson. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://musicontology.com/>.
28. Padziński A.: *Wymagania funkcjonalne dotyczące rekordów bibliograficznych – FRBR: możliwości zastosowania w katalogach bibliotecznych.* „Przegląd Biblioteczny" 2004, z. 3/4, s. 173-194.
29. *RDA: Resource Description and Access.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.rda-jsc.org/rda.html>.
30. *RDF Vocabulary Description Language 1.0: RDF Schema.* Pod red. D. Brickley, V. N. Guha. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/TR/rdf-schema/>.
31. Roszkowski M.: *Simple Knowledge Organization System (SKOS) – reprezentacja wiedzy w sieciowych systemach organizacji wiedzy.* „Zagadnienia Informacji Naukowej" 2009, nr 1(93), s. 89-102.
32. Segaran T., Evans C., Taylor J.: *Programming the Semantic Web.* Sebastopol 2009.
33. *Semantic Digital Libraries.* Pod red. S. Kruk, B. McDaniel. Berlin, Heidelberg 2009, s. 114-119.
34. Sequeda J.: *Consuming Linked Data.* [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.slideshare.net/juansequeda/consuming-linked-data-semtech2010>.

35. Shakya A., Takeda H.: *A Report on Linked Data*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://sigswo.org/papers/SIG-SWO-A801/SIG-SWO-A801-07.pdf>.
36. *SIOC Core Ontology Specification*. Pod red. U. Bojārs, J. Breslin. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://rdfs.org/sioc/spec/>.
37. *SKOS Simple Knowledge Organization System Primer*. Pod red. A. Isaac, E. Summers. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/TR/2009/NOTE-skos-primer-20090818/>.
38. *SKOS/Datasets*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/2001/sw/wiki/SKOS/Datasets>.
39. *The DBLP Computer Science Bibliography*. Pod red. M. Ley. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.informatik.uni-trier.de/~ley/db/>.
40. *Understanding the Semantic Web: Bibliographic Data and Metadata*. „Library Technology Reports". [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.ala-techsource.org/library-technology-reports/understanding-the-semantic-web-bibliographic-data-and-metadata>.
41. Vatant B.: *GeoNames Ontology*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.geonames.org/ontology/>.
42. *W3C Launches Library Linked Data Incubator Group*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://www.w3.org/News/2010#entry-8803>.
43. Wikipedia. *Semantic Web*. [online]. [dostęp: 05.11.2010]. Dostępny w World Wide Web: <http://pl.wikipedia.org/wiki/Semantic_Web>.

Summary

The paper presents foundations of Linked Data – a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web. Practical applications of this model were presented. The paper includes issues on bibliographic data on Semantic Web, major problems and possible advantages for library community.

„TERMINY METADANYCH DCMI" I MOŻLIWOŚCI ICH WYKORZYSTANIA W OPISIE RZECZOWYM[1]

Agnieszka Brachfogel
Biblioteka Narodowa

Metadane, Dublin Core, DCMI, opis rzeczowy, MARC21

Metadane Dublin Core to jedne z najpopularniejszych zestawów elementów opisu dokumentów w repozytoriach cyfrowych lub internetowych multiwyszukiwarkach. Elementy Dublin Core (*DCMES v 1.1*[2]) sformułowano w połowie lat 90. w postaci zalecenia Dublin Core Metadata Initiative (DCMI) i aż do 2008 r. nie podlegały właściwie większym zmianom[3]. Upowszechnienie zaleceń międzynarodowych organizacji (m.in. Konsorcjum WWW) dotyczące struktur i schematów danych (m.in. sieci semantycznej, modelu RDF) spowodowało, że w 2008 r. DCMI opublikowała nowy dokument – *Terminy metadanych DCMI*, w którym uwzględniła te wytyczne. Podstawowy zestaw elementów *DCMES v 1.1* nie został wycofany, a jego zawartość, przejęta do nowego opracowania, stanowi w nim jeden z rozdziałów.

Terminy metadanych DCMI zawierają najnowszy wykaz wszystkich elementów metadanych utrzymywanych przez DCMI. Lista uporządkowana jest według następujących kategorii danych:

– własności (*properties*),
– słownikowe schematy kodowania (*vocabulary encoding schemes*),
– syntaktyczne schematy kodowania (*syntax encoding schemes*),
– klasy (*classes*).

---

[1] Tekst jest zapisem wystąpienia przygotowanego na konferencję *Cyfrowość bibliotek i archiwów* zorganizowaną w Warszawie przez Stowarzyszenie Bibliotekarzy Polskich i Bibliotekę Narodową, 26-27 listopada 2009 r. Prezentacja dostępna jest na: http://www.bn.org.pl/dla-bibliotekarzy/jhp-bn/aktualnosci/konferencja-cyfrowosc-bibliotek-i-archiwow.

[2] *Dublin Core Metadata Element Set, Version 1.1.* [online]. [dostęp: 5.11.2010]. Dostępny w World Wide Web: http://dublincore.org/documents/dces/; wersja polska *Dublin Core Metadata Element Set, Version 1.1: Reference Description.* Oprac. M. Nahotko. [online]. [dostęp: 27.10.2010]. Dostępny w World Wide Web: http://ebib.oss.wroc.pl/standard/dc.html; patrz też: PN-ISO 15836 *Informacja i dokumentacja – Zestaw elementów metadanych Dublin Core.*

[3] Uzupełnieniem podstawowego 15-elementowego zestawu Dublin Core (*DCMES v.1.1*) był *Dublin Core Qualifiers*, wykaz zawierający dodatkowe elementy (tzw. kwalifikatory), które uszczegóławiały wybrane elementy zestawu *DCMES*. *Dublin Core Qualifiers* został zastąpiony przez *Terminy metadanych DCMI*, a kwalifikatory stały się integralnymi elementami nowego zestawu metadanych. Więcej patrz: http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/.

Dodatkowo, w dwóch osobnych rozdziałach, przedstawiono terminy *Słownika typów DCMI*, którymi można wskazać formę opisywanego dokumentu, oraz dwa terminy związane abstrakcyjnym modelem danych DCMI[4].

Każdy termin prezentowany jest za pomocą następującego podstawowego zestawu atrybutów:

NAZWA – jednoznaczne oznaczenie terminu, niepowtarzalne wśród terminów DCMI,

ETYKIETA – nazwa terminu zrozumiała dla człowieka,

URI – jednoznaczny identyfikator terminu,

DEFINICJA – objaśnienie terminu, jego znaczenie i cechy,

RODZAJ TERMINU – rodzaj terminu zgodnie z definicją określoną w abstrakcyjnym modelu DCMI.

Większość terminów uzupełniono o dodatkowe informacje (np. komentarz będący rozszerzeniem definicji, odnośniki do innych terminów). Dzięki temu można zorientować się w strukturze metadanych, poznać zakres ich stosowania i ewentualne zależności występujące pomiędzy elementami.

Własności, czyli elementy opisu

Na początku wykazu wymienione są własności (*properties*). Są to kluczowe metadane Dublin Core, bo to za ich pomocą opisuje się dany zasób. Rozdzielono je na dwie kategorie:

– własności w przestrzeni nazw (*terms*)[5],

– własności w dawnej przestrzeni nazw (*elements/1.1*)[6].

Własności w dawnej przestrzeni nazw */elements/1.1/* to piętnaście elementów Dublin Core, którymi dotychczas operowano. Natomiast nowością są własności w przestrzeni nazw */terms/* , które są tamtych aktualizacją i uzupełnieniem. Wykaz własności w przestrzeni nazw */terms/* składa się z pięćdziesięciu pięciu terminów[7], z których:

– czterdzieści dwa to elementy znane z podstawowego i rozszerzonego zestawu Dublin Core[8] – zachowano nazwy i definicje, a przydano nowe URI i odpowiednio powiązano je z innymi terminami.

---

[4] Abstrakcyjny model danych DCMI (*DCMI Abstract Model* http://dublincore.org/documents/2007/06/04/abstract-model/) jest to opracowanie, w którym przedstawiono – w postaci diagramów i według założeń schematu RDF – zależności pomiędzy poszczególnymi elementami metadanych Dublin Core.

[5] Patrz: *Terminy metadanych DCMI*. Rozdział 2: Własności w przestrzeni nazw */terms/*. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <http://www.bn.org.pl/download/document/1253606451.pdf>.

[6] Patrz: *Terminy metadanych DCMI*. Rozdział 3: Własności w przestrzeni nazw */elements/1.1/*. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <http://www.bn.org.pl/download/document/1253606451.pdf>.

[7] Własności DCMI jest w sumie siedemdziesiąt jeden, z czego pięćdziesiąt pięć występuje w przestrzeni nazw */terms/*, piętnaście – w dawnej przestrzeni nazw */elements/1.1/*, a jedna – w przestrzeni nazw */dcam/* (chodzi o własność *Składnik [Member of]*).

[8] Wśród tych 39 elementów: 15 to elementy znane z podstawowego zestawu Dublin Core (*DCMES v.1.1*), a pozostałe 24 z wykazu *Dublin Core Qualifiers*, na który składały się elementy równorzędne wobec *DCMES* (np. *Odbiorcy*) oraz je uszczegóławiające (np. kwalifikatorem elementu *Tytuł* był *Wariant tytułu*).

Trzynaście to nowe terminy, są to:
– prawa dostępu (*Access Rights*),
– licencja (*License*),
– metoda uzupełnienia zbioru (*Accrual Method*),
– częstotliwość uzupełniania zbioru (*Accrual Periodicity*),
– reguły uzupełniania zbioru (*Accrual Policy*),
– cytata bibliograficzna (*Bibliographic Citation*),
– zgodny z (*Conforms to*),
– data przyjęcia (*Date Accepted*),
– data copyright (*Date Copyrighted*),
– data złożenia (*Date Submitted*),
– poziom edukacyjny odbiorcy (*Education Level*),
– metoda szkolenia (*Instructional Metod*),
– pośrednik (*Mediator*).

Ze wszystkich własności przestrzeni nazw */terms/* do opisu rzeczowego wykorzystać można następujące elementy:

TEMAT (*Subject*) – tu można podawać wszelkie informacje dotyczące treści (zawartości) opisywanego obiektu; zaleca się przy tym korzystanie ze słownictwa kontrolowanego (np. hasła przedmiotowe, symbole klasyfikacyjne).

OPIS (*Description*) – może zawierać „m.in.: streszczenie, spis treści, graficzne przedstawienie treści lub opis zasobu w formie swobodnego tekstu". Jest to kolejny element dotyczący treści obiektu, ale w przeciwieństwie do *Tematu* – tu ma się większą swobodę w doborze i formie podawanych informacji.

ABSTRAKT (*Abstract*) – jest szczególnym przypadkiem *Opisu*, więc jego treść może równie dobrze znaleźć się w elemencie *Opisie*. Od przyjętych zasad zależy, który z tych elementów będzie wykorzystywany.

SPIS TREŚCI (*Table of Contents*) – podobnie jak *Abstrakt* jest szczególnym przypadkiem *Opisu* i od przyjętych zasad zależy, czy spis treści podany zostanie w *Opisie* czy *Spisie treści*.

RODZAJ (*Type*) – ogólna definicja mówi, że chodzi tu o „typ lub rodzaj zasobu". Komentarz uzupełnia ją o wskazanie, że „najlepiej jest korzystać ze słowników kontrolowanych, jak np. *Słownik typów DCMI* (DCMI Type Vocabulary). Format pliku, obiekt fizyczny czy wymiary zasobu podajemy w elemencie *Format*". Przykładem są określenia typu: kolekcja, oprogramowanie, obraz ruchomy lub terminy określające formę edytorsko-piśmienniczą dokumentów tekstowych np.: powieść, podręcznik, czasopismo.

FORMAT (*Format*) – tu podaje się cechy fizyczne opisywanego dokumentu, tzn. format pliku, nośnik lub wymiary. Informacje o rodzaju nośnika możemy podawać osobno w przeznaczonym do tego *Nośniku*.

NOŚNIK (*Medium*) – w tym miejscu podaje się informacje o fizycznym nośniku opisywanego obiektu, np.: płótno, papier, płyta DVD.

ODBIORCY (*Audience*) – wskazanie „kategorii jednostek, do których zasób jest kierowany lub dla których jest użyteczny". Mogą to być określenia typu: kobiety, dzieci, biznesmeni.

POZIOM EDUKACYJNY ODBIORCY (*Education Level*) – określenie grup odbiorców poprzez wskazanie poziomu ich wykształcenia, np.: uczniowie, studenci, wykładowcy etc.

METODA SZKOLENIA (*Instructional Method*) – tu można podawać informacje o metodzie kształcenia wykorzystanej w opisywanym zasobie.

POŚREDNIK (*Mediator*) – wskazanie uczestników procesu nauczania, do których – m.in. – skierowany jest opisywany zasób (np.: nauczyciele, rodzice, trenerzy).

ZASIĘG (*Coverage*) – „umiejscawianie treści zasobu w czasie i przestrzeni, wskazywanie zakresu zastosowania lub jurysdykcji, której zasób podlega". Preferowane są określenia słowne (nie numeryczne, a więc nie współrzędne geograficzne, ale nazwa miejsca, nie daty, ale nazwa epoki).

ZASIĘG CZASOWY (*Temporal Coverage*) – zawężenie *Zasięgu* zasobu do charakterystyki zasięgu chronologicznego.

ZASIĘG PRZESTRZENNY (*Spatial Coverage*) – zawężenie *Zasięgu* zasobu do charakterystyki zasięgu przestrzennego.

Własności DCMI a opis przedmiotowy w formacie MARC 21

Wymienione wyżej własności DCMI przekładają się na następujące pola opisu rzeczowego w formacie MARC21:

Zestawienie własności DCMI i ich odpowiedników pól formatu MARC 21 w opisie rzeczowym

Własność DCMI	Rodzaj pola/podpola	MARC 21
TEMAT	hasło rzeczowe hasło osobowe hasło korporatywne nazwa imprezy tytuł	650 600 610 611 630
RODZAJ FORMAT NOŚNIK	forma/rodzaj określnik formalny	655 6XX $v
ODBIORCY POŚREDNIK POZIOM EDUKACYJNY ODBIORCY METODA SZKOLENIA	określnik formalny	6XX $v
ZASIĘG ZASIĘG PRZESTRZENNY	hasło/określnik geograficzny	651 6XX $z
ZASIĘG ZASIĘG CZASOWY	data z hasła określnik chronologiczny	6XX $d 6XX $y
OPIS ABSTRAKT SPIS TREŚCI	streszczenie	520

W praktyce konwersja haseł utworzonych w formacie MARC21 na strukturę Dublin Core polega na fasetowaniu, tzn. rozdzieleniu poszczególnych elementów hasła na równorzędne jednostki. Np.:

650 09 $a Przemysł tytoniowy **$x** prawo **$z** Polska **$y** stan na 1925 r. **$v** źródła
(*hasło wg JHP BN*)

Temat: *Przemysł tytoniowy – prawo*
Zasięg przestrzenny: *Polska*
Zasięg czasowy: *1925 r.*
Rodzaj: *źródła (ustawa)*

650 0 $a Rośliny aromatyczne **$x** uprawy **$z** Polska **$v** czasopisma
(hasło wg KABA)

Temat: *Rośliny aromatyczne – uprawy*
Zasięg przestrzenny: *Polska*
Rodzaj: *czasopismo*

Problemy przy konwersji haseł na strukturę Dublin Core dotyczą przede wszystkim odpowiedniego wyboru i przygotowania haseł, prawidłowego zestawienia poszczególnych formatów (tzw. mapowania), a także ujednolicenia form niektórych wartości (np. liczbę mnogą *czasopisma*, *źródła* należałoby zastąpić liczbą pojedynczą *czasopismo, źródło*; zamiast ogólnego terminu *źródło* należałoby podawać formę opisywanego obiektu, np. *ustawa*, *rozporządzenie*, *statut*).

Przykłady zastosowania własności z przestrzeni nazw */terms/*

W istniejących bazach najczęściej stosowane są struktury mieszane, tzn. podstawę stanowią elementy Dublin Core w dawnej przestrzeni nazw */elements/1.1/*, ale niemal zawsze uzupełniane są elementami własnymi, stworzonymi dla potrzeb danej bazy. Marginalnie można spotkać elementy z nowej przestrzeni nazw */terms/*.

Elementy Dublin Core z nowej przestrzeni nazw */terms/* wykorzystano na przykład w następujących projektach:
– CONTENTdm
– MICHAEL Project
– XMETADIS.

CONTENTdm to oprogramowanie służące do przesyłania, opisywania i udostępniania kolekcji cyfrowych umożliwiające przeszukiwanie zawartości kolekcji cyfrowych (jest to rodzaj multiwyszukiwarki i repozytorium w jednym). Stworzone zostało przez specjalistów z University of Washington, a obecnie jest produktem oferowanym przez OCLC. Znalazły się tu następujące nowe własności Dublin Core (wyróżniony element należy do opisu rzeczowego):

Date Created <dcterms:created>
Date Published <dcterms:issued>
Collection Identifier <dcterms:isPartOf>

Subject (Place) <dcterms:spatial>
Title (Alternative) <dcterms:alternative>
Metacollection Identifier <dcterms:isPartOf>

MICHAEL Project to europejski wielojęzyczny serwis umożliwiający przeszukiwanie cyfrowych kolekcji muzeów, archiwów, bibliotek i innych instytucji kultury. W strukturze tej bazy znajdują się następujące nowe własności Dublin Core (wyróżnione elementy należą do opisu rzeczowego):

Size [dcterms:extent]
Standard [dcterms:conformsTo]
Spatial Coverage [dcterms:spatial]
Period [dcterms:temporal]
Super-Collection [dcterms:isPartOf]
Sub-Collection [dcterms:hasPart]

XMETADIS to format zestawu metadanych Niemieckiej Biblioteki Narodowej sporządzony dla dysertacji i prac postdoktorskich udostępnianych online. W jego strukturze znalazły się następujące nowe własności Dublin Core (wyróżniono elementy należące do opisu rzeczowego):

<dcterms:alternative>
<dcterms:abstract>
<dcterms:issued> lub <dcterms:dateAccepted>
<dcterms:medium>

Kontrola słownictwa

Opracowanie rzeczowe udostępnianych dokumentów powinno opierać się na słownictwie kontrolowanym. Dlatego wśród terminów metadanych DCMI wymienia się słowniki oraz wykazy, które stanowią źródło formalnego słownictwa. Podzielono je na dwie kategorie:
– słownikowe schematy kodowania (*vocabulary encoding schemes*),
– syntaktyczne schematy kodowania (*syntax encoding schemes*).

Do **słownikowych schematów kodowania**[9] należą:
– słowniki haseł i terminów przedmiotowych (LSCH, MeSH),
– klasyfikacje (DDC, LCC, NLM, UKD),
– inne (IMT, Słownik typów DCMI, TGN).

Każde z nich ma przypisane URI w przestrzeni nazwowej */terms/*, co jednoznacznie pozwala zidentyfikować źródło, z którego zaczerpnięte zostało hasło określające treść lub formę zasobu.

[9] Patrz: *Terminy metadanych DCMI*. Rozdział 4: Słownikowe schematy kodowania. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <http://www.bn.org.pl/download/document/1253606451.pdf>.

Do **syntaktycznych schematów kodowania**[10] należą standardy (m.in. ISO, RFC), które są zestawieniami różnego typu kodów (np. nazw języków, nazw krajów, schematy podawania dat czy współrzędnych miejsca), i – podobnie jak w przypadku słownikowych schematów kodowania – służą przede wszystkim do jednoznacznej identyfikacji źródła kodu podanego w opisie danego zasobu.

Poza schematami kodowania danych osobną kategorię stanowią **klasy**[11], za pomocą których porządkuje się metadane w sieci semantycznej. Klasy definiują poszczególne elementy, m.in. własności (np. własność *Zasięg czasowy* należy do klasy *Okres*, stąd wiadomo że chodzi o – zgodnie z definicją tej klasy – „nazwę przedziału czasu lub daty jego początku i końca"). Klasy określają, czym są elementy, z których buduje się bazę danych. Klasa nie jest zatem elementem opisu jakiegoś obiektu, ale pozwala na odpowiednie grupowanie informacji (kwestia odpowiedniego przygotowania i ustawienia systemu).

Dzięki temu możliwa jest kontrola poprawności przypisywania wartości danej własności (np. hasło *Adam Mickiewicz* do elementu *Twórca*). Taka kontrola jest jednak możliwa tylko wówczas, gdy wartości czerpiemy ze słowników kontrolowanych. Dodatkową zaletą zastosowania klas jest to, że można się nimi posłużyć przy kształtowaniu procesu wyszukiwania, dostosowując go do indywidualnych potrzeb (np. znalezienia wszystkich danych powiązanych bezpośrednio i pośrednio z daną klasą).

Podsumowanie

Podstawowe zmiany, które wprowadzono do formatu Dublin Core, polegają na:

– przemianowaniu i pogrupowaniu poszczególnych elementów zgodnie ze schematem RDF,

– osadzeniu elementów opisu w nowej przestrzeni nazw */terms/* i określeniu relacji pomiędzy własnościami z nowej przestrzeni nazw */terms/* i dawnej */elements/1.1/*,

– uzupełnieniu zestawu o nowe elementy.

Z pięćdziesięciu pięciu własności nowej przestrzeni nazw */terms/* czternastoma można posłużyć się do opisu treści (zawartości) lub formy zasobu, trzy, tj.: Poziom edukacyjny odbiorcy (*Education Level*), Metoda szkolenia (*Instructional Metod*) oraz Pośrednik (*Mediator*) są całkowicie nowymi elementami.

Zwiększenie liczby elementów Dublin Core pozwala tworzyć precyzyjniejsze opisy obiektów udostępnianych w repozytoriach. W istniejących bazach już

---

[10] Patrz: *Terminy metadanych DCMI*. Rozdział 5: Syntaktyczne schematy kodowania. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <http://www.bn.org.pl/download/document/1253606451.pdf>.

[11] Patrz: *Terminy metadanych DCMI*. Rozdział 6: Klasy. [online]. [dostęp: 8.11.2010]. Dostępny w World Wide Web: <http://www.bn.org.pl/download/document/1253606451.pdf>. W przestrzeni nazw */terms/* wymieniono dwadzieścia dwie klasy. Wszystkich klas utrzymywanych przez DCMI jest więcej, bo aż trzydzieści pięć: oprócz wspomnianych dwudziestu dwóch, kolejnych dwanaście należy do przestrzeni nazw */dcmitype/* (to są terminy wchodzące w skład *Słownika typów DCMI*), a jedna – do przestrzeni nazw */dcam/* (chodzi o klasę *Słownikowy schemat kodowania* [*VocabularyEncodingScheme*]).

korzysta się z tej możliwości, zastępując ogólne terminy (np. *Zasięg*) elementami o węższym zakresie (np. *Zasięg przestrzenny*, *Zasięg czasowy*). Wydaje się jednak, że w większości repozytoriów zmiany w strukturach metadanych będą marginalne, i jak to widać na przedstawionych wyżej przykładach, będą mieszanką elementów starych i nowych oraz własnych, wykorzystywanych lokalnie w danej bazie.

Summary

DCMI Metadata Terms is an up-to-date specification of all metadata terms maintained by the Dublin Core Metadata Initiative. Included are the fifteen terms of the Dublin Core Metadata Element Set (v. 1.1). The hierarchy and appropriate linking between the terms has been created. Number of the subject elements has increased. Partly it is the refinements of the basic elements but some of them are completely new. Although the conversion into and from MARC fields is already possible, some difficulties are encountered and need to be solved. Usage of the DCMI Metadata Terms has not become widespread yet, however the elements enter the structures of the catalogues.

FINANSOWANIE OPEN ACCESS

Anna Wałek
Biblioteka Główna i OINT
Politechniki Wrocławskiej

Open Access, czasopisma elektroniczne, modele finansowe Open Access, The Public Library of Science (PLoS)

Kwestia finansowania otwartego dostępu do zasobów wiedzy jest sprawą niezwykle istotną, a jednak najczęściej pomijaną w polskich opracowaniach na ten temat. W Polsce Open Access pojmowany jest nadal bardziej jako pewna idea, łącząca zmagania entuzjastów, często oparta na zasadzie wolontariatu. Niektórzy uważają, że debata na temat Open Access w istocie nie dotyczy spraw ekonomicznych, ale nieograniczonego dostępu do publikowanych wyników badań naukowych, również za pomocą modelu OA[1], który z założenia ma być bezpłatny. Niestety, nie jest on pozbawiony kosztów. Mimo iż autor nie pobiera honorarium za swoją pracę, są koszty administracyjne, systemowe, koszty recenzji i inne, które muszą być pokryte.

Obecnie funkcjonuje kilka modeli finansowania OA:

– publikację finansuje autor – aby opublikować swój artykuł w czasopiśmie OA, musi uiścić opłatę autorską (*publication fee*);

– publikowanie prac swoich pracowników finansuje instytut naukowy lub uczelnia – często w ramach prowadzonego repozytorium (archiwum e-printów), czasopisma OA lub na stronie domowej;

– dofinansowanie pochodzi z zewnętrznych instytucji, dotacji rządowych, grantów;

– model podwójny – dochody pochodzące ze sprzedaży wersji drukowanej przeznaczone są na sfinansowanie wolnego dostępu w wersji online;

– model współdzielony – instytucje członkowskie, takie jak biblioteki czy towarzystwa naukowe wspólnie utrzymują czasopismo OA i całe jego zaplecze techniczne;

– płatne członkostwo w wydawnictwie OA – wydawnictwo OA oferuje instytucjom naukowym udział wiążący się ze składką członkowską w zależności od wybranej formy członkostwa, instytucja ma prawo do opublikowania określonej liczby artykułów w ciągu roku;

– model hybrydowy – wydawnictwo oferuje autorom możliwość wpłacenia pewnej kwoty w celu „uwolnienia" artykułu publikowanego w czasopiśmie nie mającym statusu OA. Po dokonaniu opłaty artykuł dostępny jest bez ograniczeń.

[1] A. Gass: *Paying to Free Science: Costs of Publication as Costs of Research*. „Serials Review" 2005, vol. 31, no. 2, p. 103.

Pytanie o to, jak finansować koszty recenzowanych artykułów naukowych w wolnym dostępie jest niezwykle ważne. Odpowiedzią wydawców OA, takich jak The Public Library of Science (PLoS) jest, aby przy wyeliminowaniu opłat za dostęp wydawcy byli opłacani za usługi, których dostarczają autorom, instytucjom i środowisku naukowemu (przynajmniej w zakresie biologii i medycyny), jako dostawcy tak usług jak i systemu. W tym wypadku wydawcy pobierają opłaty za udostępnianie i prowadzenie serwisu. Obejmują one recenzje, przygotowanie zaakceptowanego tekstu do publikacji, umieszczenie go w sieci itd. W zamian za jednorazową zapłatę wydawcy OA, tacy jak PLoS czy BioMed Central, nie nakładają na publikację żadnych ograniczeń i jest ona w pełni dostępna w ramach czasopisma czy repozytorium[2].

Jednym z najliczniej reprezentowanych stanowisk w sprawie finansowania publikowania w wolnym dostępie jest opinia, że powinni to robić sami autorzy ze środków otrzymywanych w ramach grantu od instytucji sponsorującej badania naukowe lub od pracodawcy.

Jedną z propozycji pokrycia opłat autorskich był też pomysł, aby były one realizowane z budżetów bibliotek instytucji, z którymi związani są autorzy[3]. Pomysłodawcy tłumaczyli koncepcję tym, że to właściwie biblioteki mają szansę zaoszczędzić najwięcej na wprowadzeniu OA. Rozwiązanie to miałoby bowiem wpływ na zmniejszenie nakładów ponoszonych przez biblioteki na prenumeratę czasopism. Biblioteki Uniwersytetu Cornell w Stanach Zjednoczonych (Cornell University Libraries) w 2004 r. przeprowadziły analizę nakładów, jakie ponoszą uczelnie na prenumeratę czasopism naukowych w porównaniu z tym, ile publikują ich pracownicy. Wynikiem był *Report of the Cornell University Library's Task Force on Open Access Publishing*[4]. Jego autorzy jednak są sceptycznie nastawieni do modelu OA, starają się bowiem udowodnić, że model OA nie przyniesie uczelni rzeczywistych oszczędności. Wręcz przeciwnie, biorąc pod uwagę liczbę artykułów publikowanych rocznie przez pracowników Cornell University i średnie koszty związane z ich udostępnieniem w recenzowanym czasopiśmie OA, obliczyli, że nakłady na czasopisma wzrosną, jeśli biblioteka przeznaczy swoje dotychczasowe fundusze na pokrycie składek autorskich, nawet w przypadku gdyby większość wydawców przeszła na model OA[5]. Autorzy raportu podkreślają, że model OA nie może być traktowany jako uniwersalne rozwiązanie kwestii czasopism naukowych, ale bez wątpienia może być pragmatycznym rozwiązaniem w konkretnych przypadkach[6].

Jedna z konkluzji raportu Cornell University, która zakłada, że koszty wielu instytucji naukowych i ich bibliotek wzrosną, jeśli OA stanie się powszechny, jest zdaniem wielu komentatorów nietrafna, z założenia bowiem koszty publikowania w modelu OA powinny być traktowane jako integralna część kosztów badań naukowych, a dotyczą one zarówno instytucji, w której badania te są prowadzone, jak i zewnętrznych agencji sponsorujących naukę poprzez przy-

[2] Ibidem.

[3] Ibidem.

[4] *Report of the Cornell University Library's Task Force on Open Access Publishing.* [online]. [dostęp: 14.05.2010]. Dostępny w World Wide Web: <http://ecommons.cornell.edu/bitstream/1813/193/3/OATF_Report_8-9.pdf>.

[5] Ibidem.

[6] Ibidem.

znawanie grantów. Zatem koszty publikacji powinny być ponoszone nie tylko przez instytucje naukowe (czy, jak niektórzy twierdzą, biblioteki), ale również, a może przede wszystkim, przez grantodawcę finansującego badania, których wynikiem jest publikacja[7].

W Stanach Zjednoczonych wiele instytucji finansujących badania naukowe pośrednio lub bezpośrednio finansuje również publikację artykułów naukowych. Największa na świecie rządowa instytucja finansująca badania z zakresu tzw. life sciences – NIH (The US National Institutes of Heath) pozwala autorom na wykorzystywanie funduszy z grantów do finansowania publikacji, także OA. W 2005 r. na ten cel przeznaczano ok. 1% wartości grantów, co dawało ok. 2500 dolarów na artykuł[8]. NIH deklarowało już w 2005 r. wydawanie ok. 30 milionów dolarów na bezpośrednie koszty publikacji. Wynikiem sponsorowanych przez NIH badań jest średnio 60-65 tys. publikacji rocznie. Zatem na każdy artykuł przeznaczono ok. 450-500 dolarów. W porównaniu z prawie 14 mld dolarów, które w 2003 r. NIH przeznaczyła na granty badawcze, kwota na finansowanie publikacji to jedynie 0,22% ogólnego budżetu[9].

Jednym z problemów autorskiego modelu finansowania OA jest rozbieżność, jaka może zaistnieć pomiędzy naukami ścisłymi, medycznymi czy biologicznymi a naukami humanistycznymi i społecznymi, gdzie system finansowania i uzyskiwania grantów różni się zasadniczo, zwłaszcza jeśli chodzi o same kwoty dotacji[10].

Innym modelem finansowania jest praktykowany przez niektórych wydawców model oparty w całości lub częściowo na opłatach członkowskich. Dotyczy to głównie stowarzyszeń i towarzystw naukowych będących jednocześnie wydawcami czasopism naukowych. W przypadku PLoS Institutional Membership członkowie zwykle pokrywają do 50% autorskiej opłaty, resztę ponosi instytucja, która finansuje prowadzone przez autora badania. W przypadku autorów z instytucji, które nie mają członkowskiej umowy z wydawcą OA (większość autorów piszących dla czasopism PLoS), opłaty autorskie pochodzą zwykle z grantów lub innych zewnętrznych źródeł finansowania, rzadziej natomiast z funduszy instytucji macierzystych[11].

Niektóre opłaty ponoszone przez instytucje macierzyste są rekompensowane przez odpowiednie agencje finansujące badania. Dwie największe światowe fundacje finansujące badania z nauk biomedycznych: Welcome Trust i Howard Hughes Medical Institute w całości pokrywają koszty publikacji sponsorowanych przez siebie badaczy w czasopismach Open Access[12].

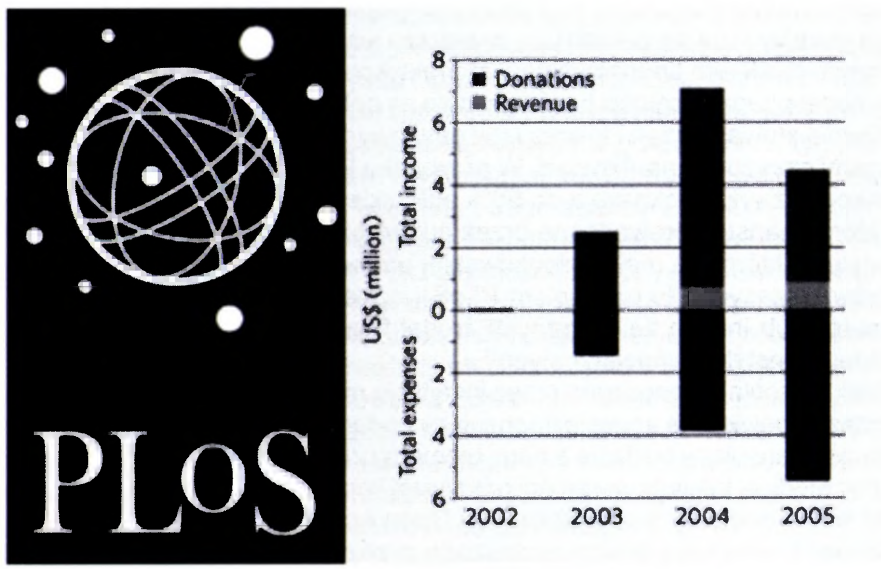
Model finansowy i koszty ponoszone przez wydawców OA można przeanalizować na przykładzie PLoS. W 2003 r., kiedy na rynku ukazał się tytuł „PLoS Biology", przewidywano, że koszt opublikowania jednego artykułu w czasopiśmie wyniesie ok. 1500 dolarów[13]. Ponieważ PLoS zapowiedział, że na łamach swoich czasopism będzie publikował tylko najbardziej wartościowe i najlepsze artykuły, oznacza to, że około 90% nadesłanych do redakcji

[7] A. Gass: *Paying to Free Science...*, op. cit., s. 104.
[8] Ibidem, s. 105.
[9] Tamże, s. 104.
[10] Ibidem.
[11] Ibidem.
[12] Ibidem.
[13] D. Butler: *Who will pay for open access?* „Nature" 2003, vol. 425, pp. 554-555.

i zrecenzowanych prac zostanie odrzucona. Sam proces recenzowania tysięcy artykułów jest jednak niezwykle kosztowny, dlatego też specjaliści zakładali, że koszty publikacji będą wyższe, niż zapowiadane 1500 dolarów[14]. Dla wielu wydawców polityka finansowa PLoS była niezrozumiała. Publikowanie w wersji elektronicznej powinno zredukować koszty produkcji o około 20% w porównaniu z wielonakładowym czasopismem. Pozostaje jednak wciąż wiele kosztów o charakterze personalnym i administracyjnym[15].

W 2006 roku PLoS musiał zmierzyć się z kryzysem finansowym. Z obliczeń specjalistów wynikało bowiem, że organizacja zaczęła ponosić straty sięgające w 2005 r. mln dolarów. Dochody PLoS pochodzące z reklam i opłat autorskich pokrywały zaledwie 35% kosztów. Chociaż dochody z roku na rok rosły (od 0,75 mln dolarów w latach 2003-2004 do 0,9 mln w 2005 r.), to mimo wszystko nie pokrywały wydatków szacowanych na 1,5 do 5,5 mln w ciągu trzech lat[16].

Polityka firmy w początkowych latach działalności bazowała w dużej mierze na dotacjach. 9 mln dolarów PLoS otrzymał od Fundacji Gordona i Betty Moore'ów, 4 mln dolarów od Sandler Family Supporting Foundation. Pieniądze z tych źródeł pokryły 65% kosztów funkcjonowania PLoS, szybko jednak się wyczerpywały[17].



Rys. 1. Przychody i wydatki PLoS w latach 2002-2005 (na wykresie w kolorze jasnoszarym zaznaczono dochody z tytułu reklam i opłat autorskich, natomiast kolorem ciemnoszarym otrzymane dotacje zewnętrzne).
Źródło: D. Butler: *Open Access journal hits rocky times*. „Nature" 2006 vol. 441 [online]. [dostęp: 3.09.2010]. Dostępny w World Wide Web: <http://www.nature.com/nature/journal/v441/n7096/full/441914a.html>

[14] Ibidem, s. 555.
[15] Ibidem.
[16] D. Butler: *Open-access journal hits rocky times*. „Nature" 2006 vol. 441 [online]. [dostęp: 16.07.2010]. Dostępny w World Wide Web: <http://www.nature.com/nature/journal/v441/n7096/full/441914a.html>.
[17] Ibidem.

Obecnie sytuacja PLoS uległa polepszeniu. Częściowo korzysta wciąż z różnego typu grantów, dotacji i darowizn, jednak według raportu na rok 2009 organizacja zakłada, że w 2009 r. jej przychody pokryją ok. 85% wydatków, natomiast w 2010 r. uzyska już pełną niezależność finansową[18]. Chociaż PLoS jest organizacją non profit, jej działalność można określić jako model hybrydowy, w pewnych aspektach bowiem działa ona jako typowe przedsiębiorstwo rynkowe[19]. OA stał się jednym z modeli biznesowych, odkąd komercyjne wydawnictwa dostrzegły korzyści finansowe w jego publikowaniu. Wystarczy spojrzeć na przejęcie przez Springer Verlag wydawnictwa BioMed Central w 2008 roku[20].

Wyjście PLoS z kryzysu, odbyło się jednak poprzez podniesienie kosztów obciążających autorów. Tak, jak się spodziewano, w ciągu ostatnich kilku lat opłaty ponoszone przez autorów chcących publikować na łamach czasopism PLoS wzrosły prawie dwukrotnie. Według aktualnego cennika dostępnego na stronie PLoS artykuły w tytułach wydawanych przez wydawnictwo wynoszą: 2250 dolarów w przypadku „PLoS Computational Biology", „PLoS Genetics", „PLoS Pathogens" i „PLoS Neglected Tropical Diseases", 2900 w przypadku „PLoS Medicine" i „PLoS Biology", jedynie publikowanie w interdyscyplinarnym czasopiśmie „PLoS ONE" kosztuje autora obecnie 1350 dolarów[21]. Na preferencyjne warunki finansowe (zniżka 10%) mogą liczyć autorzy, których instytucje legitymują się członkowstwem w PLoS. Wydawnictwo uruchomiło również specjalny program pomocowy dla tych autorów, których nie stać na samodzielne pokrycie opłat, polegający na całkowitym lub częściowym umorzeniu należności[22].

BioMed Central również pobiera opłaty autorskie za publikowanie na łamach większości swoich 206 czasopism. Wynoszą one od 760 dolarów w przypadku takich czasopism jak: „Journal of Cheminformatics" i „Chemistry Central Journal" do 2440 dolarów w przypadku „Genome Medicine" czy „Genome Biology". Tytuły: „Sports Medicine, Arthroscopy, Rehabilitation, Therapy & Technology", „Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine", „Journal of the International AIDS Society", „Journal of Orthopaedic Surgery and Research", „Journal of Biomedical Science", „Italian Journal of Pediatrics", „Chiropractic & Osteopathy" i „Chinese Medicine" są całkowicie zwolnione z opłat autorskich, natomiast nowy tytuł „Journal of Systems Chemistry" jest objęty promocyjnym okresem zwolnienia[23].

Coraz więcej komercyjnych wydawców sięga po model Open Access również dlatego, że jest to dobra forma promocji. Niektórzy wydawcy udostępniają tytuły w całkowicie otwartym dostępie lub zapewniają dostęp do artykułów po upływie określonego okresu karencji. Takie tytuły finansowane są przez ich

---

[18] *PLoS Progress Report June 2009.* [online]. [dostęp: 10.08.2010]. Dostępny w World Wide Web: <http://www.plos.org/downloads/progress_report.pdf>.

[19] Ibidem.

[20] Ibidem.

[21] *Publication Fees for PLoS Journals.* [online]. [dostęp: 15.08.2010]. Dostępny w World Wide Web: <http://www.plos.org/journals/pubfees.php>.

[22] Ibidem.

23 *Frequently asked questions about BioMed Central's article-processing charges.* [online]. [dostęp: 17.07.2010]. Dostępny w World Wide Web: <http://www.biomedcentral.com/info/about/apcfaq>.

komercyjne wersje lub przeznacza się na nie środki z wpływów pochodzących z innych czasopism danego wydawcy, udostępnianych komercyjnie. Istnieje również model hybrydowy, w przypadku którego część artykułów w obrębie jednego tytułu czasopisma może być udostępniona w wolnym dostępie, natomiast dostęp do pozostałych jest odpłatny. W przypadku niektórych tego typu czasopism, gdzie zwykle stosowany jest model polegający na subskrypcji, autor, po dokonaniu opłaty, może „uwolnić" swój artykuł, tak aby był on dostępny dla wszystkich w wolnym dostępie. Pierwszym czasopismem, które zastosowało model hybrydowy był „Florida Journal of Entomology". Obecnie praktycznie każde wydawnictwo naukowe daje autorom możliwość publikowania w tym modelu, np. Cambridge University Press: „Cambridge Open Option", Elsevier: „Sponsored Article", Nature Publishing Group („British Journal of Cancer"): „BJC OPEN, („British Journal of Pharmacology"): „BJP Open", Oxford University Press: „Oxford Open", Royal Society (UK): „EXiS Open Choice", Springer: „Open Choice".

Różnica pomiędzy OA a hybrydowym OA polega również na tym, że w modelu hybrydowym publikacja często opatrzona jest pewnymi restrykcjami dotyczącymi redystrybucji lub wykorzystania[24].

Poniższa tabela przedstawia wysokość pobieranych przez wydawnictwa opłat autorskich:

WYDAWCY OPEN ACCESS		
Nazwa wydawcy	**Wysokość opłaty autorskiej**	**Przykładowe tytuły**
BioMed Central	625-2365 $, średnio 1535 $ (15% na podstawie członkostwa UCB membership)	BMC Evolutionary Biology, BMC Informatics, Genome Biology, Nutrition Journal
Co-Action Publishing	od 65 € za stronę do 712 € za artykuł	Food and Nutrition Research, Ethics and Global Politics
Hindawi Publishing	600 $-1500 $	International J. of Plant Genomics, J. of Applied Mathematics, J. of Nanomaterials, Modelling and Simulation in Engineering
Oxford University Press	1800 $ (jeśli mamy drukowaną subskrypcję) (2250 $ od 12.01.09)	Pełna oferta czasopism OA
PhysMath Central	1515 $ (15% zniżki na podstawie członkostwa UCB membership)	PMC Physics A, PMC Physics B
Public Library of Science (PLoS)	1350 $-2800 $ (10% zniżki na podstawie UC membership)	PLoS Biology, PLoS Computational Biology, PLoS Medicine, PLoS One
WYDAWCY HYBRYDOWEGO MODELU OA		
Nazwa wydawcy (nazwa programu OA)	**Wysokość opłaty autorskiej**	**Tytuły czasopism**
American Chemical Society (ACS Author Choice)	1500 $-3000 $	ACS Chemical Biology, Analytical Chemistry, J. of the American Chemical Society, J. of Organic Chemistry

[24] *Selective list of Open Access and Paid Access Fees.* [online]. [dostęp: 15.08.2010]. Dostępny w World Wide Web: <http://www.lib.berkeley.edu/scholarlycommunication/oa_fees.html>.

American Institute of Physics (AIP) (Author Select)	1500 $-1800 $, 2500 $	Applied Physics Letters, Chaos, J. of Mathematical Physics, J. of Rheology
American Physical Society (APS) (Free to Read)	975 $, 1300 $	Physical Review A-E, Physical Review Letters, Reviews of Modern Physics
Blackwell Publishing (Online Open)	2600 $	International Review of Finance, International Statistical Review, Renaissance Studies, J. of Applied Crystallography
Cambridge Journals (Cambridge Open Option)	1700 $	American Political Science Review, Geological Magazine, J. of Social Policy, Systematics and Biodiversity
Cold Spring Harbor Laboratory Press (Open Access Option)	2000 $	tylko Genome Research
Elsevier (Sponsored Articles dostępny dla ok. 40 z 1800 tytułów)	3000 $	Discrete Applied Mathematics, J. of Molecular Biology, Physics Letters B, Vision Research
National Academy of Sciences (Open Access Option)	950 $ (w ramach UCB membership)	Proceedings of the National Academy of Sciences (PNAS)
Oxford Journals (Oxford Open — optional open access model)	1500 $	Biostatistics, Health Policy and Planning, J. of Design History, Quarterly J. of Mathematics
Royal Society of Chemistry (RSC) (RSC Open Science)	1800 $-4500 $ (15% zniżki w przypadku subskrypcji)	Chemical Biology, Geochemical Transactions, Methods in Organic Synthesis
Springer (Open Choice)	2000 €/3000 $	Acoustical Physics, Computational Economics, Current Psychology, J. of Religion and Health
Taylor and Francis (iOpenAccess)	3100 $	Annals of Human Biology, Food Biotechnology, International J. of Computer Mathematics, Philosophical Psychology
Wiley InterScience (Funded Access for biomedical titles)	3000 $	J. of Basic Microbiology, Proteomics, Statistics in Medicine
World Scientific (WorldSciNet OPEN ACCESS)	2500 $	International J. of Mathematics, Modern Physics Letters B

Obok wydawnictw, które pobierają opłaty od autorów nadsyłających teksty, są również takie, które opłat nie pobierają. W zasadzie ponad dwie trzecie czasopism z listy Directory of Open Access Journals (DOAJ) i ponad 80% czasopism OA wydawanych przez towarzystwa naukowe nie pobiera opłat od autorów[25]. Czasopisma te utrzymywane są z dotacji rządowych, grantów lub z pieniędzy towarzystw naukowych i instytucji, pod których szyldem funkcjonują. Wiele instytucji i organizacji finansujących badania naukowe, zwłaszcza tych,

[25] B. Hooker: *If it won't sink in, maybe we can pound it in…* [online]. [dostęp: 19.07.2010]. Dostępny w World Wide Web: <http://www.sennoma.net/main/archives/2007/12/if_it_wont_sink_in_maybe_we_ca.php>.

które dysponują środkami publicznymi, nakłada na grantobiorców obowiązek publikowania wyników badań w czasopiśmie lub repozytorium OA. Niektóre czasopisma i repozytoria powstają właśnie w celu publikowania wyników badań przez grantobiorców konkretnej organizacji.

Po przeanalizowaniu różnych przypadków nasuwa się stwierdzenie, że najlepszym rozwiązaniem byłoby finansowanie OA za pomocą modelu mieszanego. Częściami składowymi opłat ponoszonych przez autorów powinny być fundusze pochodzące z macierzystych placówek i instytucji, z organizacji finansujących badania naukowe oraz funduszy rządowych. Fora do publikowania wyników badań naukowych powinny również tworzyć instytuty naukowe oraz wyższe uczelnie, tworząc repozytoria, elektroniczne archiwa publikacji oraz recenzowane czasopisma naukowe, za publikowanie w których ani autor, ani jednostka, z której się wywodzi, nie musiałaby ponosić opłat.

Wielu zwolenników Open Access, podkreśla, że podobnie jak w przypadku badań nad AIDS czy innymi chorobami ponad kwestie finansowe powinien być ważniejszy interes i dobro ludzkości. Tym bardziej w przypadku, kiedy badania te prowadzone są za pieniądze publiczne. W modelu OA chodzi bowiem zasadniczo o to, aby podatnik finansujący badania naukowe nie musiał płacić za dostęp do ich wyników. W większości wypadków komercyjne wydawnictwa utworzyły na tej podstawie nowy, zyskowny model biznesowy, w którym opłaty z czytelnika zostały przeniesione na autora.

Summary

Funding Open Access is an extremely important question, however hardly mentioned in Polish literature. The Open Access model, assumed to assure free access to resources, generates system, organizational, and administrative costs, which we shall be aware of. The author discusses current methods of Open Access publications' funding, like: author, institutional, double, shared, and hybrid models, as well as governmental funding, grants, and payable membership in an OA publishing house. Particular attention is put on an author model of financing OA publications, assuming that costs of publishing texts are paid by their authors. The levels of authorship fees in most of commercial OA and hybrid journals' publishers are described.

KOMPUTERYZACJA GROMADZENIA ZBIORÓW A ZINTEGROWANY SYSTEM BIBLIOTECZNY. DOŚWIADCZENIA BIBLIOTEKI JAGIELLOŃSKIEJ

Ewa Dąbrowska
Biblioteka Jagiellońska

Gromadzenie zbiorów, komputeryzacja, VTLS, AFAS, moduł gromadzenia VIRTUA, systemy zintegrowane, ewidencja zbiorów

W ciągu ostatnich dwudziestu lat rozwój technik informatycznych sprawił, że w polskich bibliotekach zaszły rewolucyjne zmiany, przekształcające je w biblioteki elektroniczne i hybrydowe. Systemy biblioteczne, obejmujące całość lub większość procesów bibliotecznych, wpłynęły znacząco na organizację i charakter pracy bibliotekarza, dla którego jeszcze niedawno podstawowym narzędziem pracy był ołówek i fiszka. Biblioteki mogą wybierać między systemami wyspecjalizowanymi, spełniającymi tylko jedną funkcję, systemami złożonymi z wielu wyspecjalizowanych podsystemów oraz systemami zintegrowanymi. Za system zintegrowany uważa się system wielofunkcyjny, obejmujący wszystkie lub większość funkcji bibliotecznych, w którym pracuje się na jednej wspólnej bazie danych[1]. Jego główną ideą jest wykorzystywanie informacji jednego podsystemu przy obsłudze drugiego[2], a zwłaszcza stopniowe rozbudowywanie i wzbogacanie opisu bibliograficznego zapoczątkowanego lub skopiowanego z dostępnego źródła już na etapie gromadzenia zbiorów[3]. Przystępując do komputeryzacji Biblioteka Jagiellońska (dalej BJ), wzięła pod uwagę taki właśnie typ systemu i organizacji pracy. W artykule chciałabym zatem przedstawić przebieg komputeryzacji procesów gromadzenia zbiorów w BJ w odniesieniu do wdrażanego tu systemu komputerowego i pokazać, jak wyglądało w praktyce zetknięcie idei systemu zintegrowanego z rzeczywistością biblioteczną.

Przebieg komputeryzacji w polskich bibliotekach jest często, niestety, słabo udokumentowany, przez co staje się trudny do zrekonstruowania[4]. Tym bardziej warto przeanalizować historię tego, z różnych powodów skomplikowanego i jak okazało się, w praktyce wciąż jeszcze – w zakresie gromadzenia zbiorów w BJ – niezakończonego procesu. Podjęłam się tego jako nie tylko obserwator, ale aktywny uczestnik wydarzeń, które znacząco zmieniły oblicze biblioteki.

---

[1] A. Jacquesson: *Automatyzacja bibliotek: zarys historyczny, strategia, perspektywy*. Warszawa 1999, s. 288.

[2] Ibidem, s. 81. Autor zauważa, że filozofia ZSB przeważała do lat 80. XX w., a ponieważ nie zaspokajały one wszystkich potrzeb danej biblioteki, rosnąć zaczęło zainteresowanie systemami złożonymi.

[3] A. Radwański: *Jak komputeryzować bibliotekę: poradnik*. Warszawa 2000, s. 60.

[4] Ibidem, s. 104.

Pierwsze kroki w świat komputerów

Wprawdzie pierwsze polskie próby komputeryzacji czy też mechanizacji sięgają lat 60. XX w., dotyczyły jednak głównie zastosowań informacyjnych oraz bibliograficznych. W latach 70. zaczęto komputeryzować czynności biblioteczne, przerwał to jednak kryzys polityczno-gospodarczy lat 80.[5]. Tymczasem w Wielkiej Brytanii już w połowie lat 80. ubiegłego wieku w 85% ogłoszeń dotyczących pracy w bibliotece koniecznym warunkiem było doświadczenie w pracy z komputerem[6].

Sytuację w Polsce zmieniło upowszechnienie się mikrokomputerów, na przełomie lat 80. i 90. pojawiły się na biurkach bibliotekarzy, co pozwoliło na oswojenie się z nową techniką przed przystąpieniem do planowej komputeryzacji biblioteki. Komputery zaczęto stosować w bieżącej pracy, od pisania korespondencji poczynając, do zakładania różnych baz danych. Bazy te były prowadzone równolegle do kartotek fiszkowych, do których drukowano karty z programu, nie potrafiono bowiem rozstać się od razu z tradycyjnym warsztatem pracy. Należy przyznać też, że początkowo były traktowane jako ciekawostka, coś, co można pokazać gościom, by się pochwalić nowoczesnością.

W Oddziale Gromadzenia i Uzupełniania Zbiorów (dalej OGR) powstała, w programie Micro CDS/ISIS, kartoteka książek zagranicznych, pochodzących z kupna, wymiany i darów. Sekcja Wymiany zaczęła wykorzystywać program, napisany pierwotnie dla księgarni, do obsługi wysyłki publikacji uczelnianych do kontrahentów w kraju i za granicą. Sekcję Kupna w sprawach związanych z rozliczeniami wydatków wspomagał arkusz kalkulacyjny Lotus 123. Dla wszystkich ułatwieniem w prowadzeniu korespondencji stały się edytory tekstu, a zakupiona baza danych dostarczała nie tylko Sekcji Egzemplarza Obowiązkowego informacji o nowo powstałych firmach wydawniczych. Wkrótce też powszechnie wykorzystywanym narzędziem stał się internet i poczta e-mail.

Gromadzenie po podjęciu decyzji o komputeryzacji BJ

Gdy na początku lat dziewięćdziesiątych zapadła decyzja o zakupie i wdrożeniu zintegrowanego systemu bibliotecznego VTLS, dla dyrekcji Biblioteki jasne było, że system musi objąć również gromadzenie zbiorów, gdzie miał powstawać zaczątek opisu bibliograficznego pozyskiwanych w różny sposób publikacji. Ówczesna wersja tego systemu nie zawierała jednak w pełni zintegrowanego modułu wspomagającego proces gromadzenia, oferowany był natomiast program o nazwie AFAS [Acquisitions ans Fund Accounting System][7], którego najważniejsze funkcje to kontrola realizacji zamówień oraz prowadzenie księgowości. BJ jako jedyna biblioteka w Polsce zdecydowała się na jego na-

[5] E. Ścibor: *Zarys historii mechanizacji i automatyzacji bibliotek w Polsce*. W: A. Jacquesson: *Automatyzacja bibliotek...*, op. cit., s. 12-13.

[6] P. F. Burton: *Technologia informacyjna i jej wpływ na służby biblioteczno-informacyjne*. W: *Automatyzacja bibliotek: wybór materiałów z konferencji „Automatyzacja bibliotek"* Wrocław, 11-13 grudnia 1992. Wrocław 1993, s. 17.

[7] Dokładne omówienie programu zob. E. Dąbrowska: *AFAS – system obsługi zamówień i księgowania wydatków w ocenie użytkownika*. „Bibliotekarz" 1998, nr 3, s. 11-14.

bycie, wkrótce przystąpiono więc do przygotowań niezbędnych do wdrożenia. Oprócz zapoznania się z samym systemem przez pracowników i przetłumaczenia dokumentacji oraz komend w programie, należało opracować procedury pracy w nim i przygotować niezbędne na starcie wzory typowej korespondencji, system kont księgowych, kartotekę walut oraz bibliotek instytutowych. Zasilono także bazę dostawców danymi wgranymi z zakupionej bazy wydawców polskich oraz adresami kontrahentów wymiany z programu wymiany.

Zdecydowano, że AFAS obejmie całość procedur związanych z zakupami wydawnictw zwartych, zbiorów specjalnych oraz niektórych wydawnictw ciągłych[8], scali wszystkie kartoteki zamówień i przybytków oraz kartoteki kontrahentów wymiany, dostawców sekcji kupna, darczyńców i wydawców. Już samo to znacznie ułatwiło i uprościło pracę, gdyż dotąd w wielu wypadkach, sprawdzając jeden tytuł, należało przeszukać kilka kartotek. Zaprzestano oczywiście prowadzenia ich tradycyjnych odpowiedników. Czasopisma prowadzono dalej na kartach kontynuacyjnych, gdyż AFAS nie oferował tu optymalnego rozwiązania. Powoli natomiast, w miarę opracowywania kolejnych tytułów ciągłych przez Oddział Opracowania Czasopism, przechodzono do rejestrowania wpływu czasopism w rekordzie zasobu w katalogu online.

Niestety, AFAS nie zawierał funkcji rejestru przybytku w rozumieniu obowiązujących w BJ przepisów o wstępnej ewidencji wpływów, stąd był on prowadzony dalej w sposób tradycyjny. Brakowało też w tym programie możliwości obsługi wysyłki wydawnictw w ramach wymiany i gospodarki dubletami i drukami zbędnymi. Był zatem rozwiązaniem cząstkowym, choć już w tym zakresie, który oferował, znacząco poprawiał jakość pracy, a producent obiecywał w przyszłości doskonalsze rozwiązania.

Komputeryzacja biblioteki – a więc zmiana narzędzi i systemu pracy – to dobry moment do rozważenia, co można lub należy zmienić w organizacji pracy. Na zebraniach toczono dyskusje nad ewentualnymi usprawnieniami, nad tym, z czego można zrezygnować, a z czego nie. Omawiano znane z literatury fachowej i przykładów innych bibliotek przypadki łączenia działów i funkcji, rozważano, czy i na ile może to poprawić jakość pracy. Zdarzały się postawy niechętne wobec wprowadzanych zmian, dla niektórych problemem była konieczność przełamania utartych schematów myślenia, wytrącenie z wygodnego toru dawno wypracowanych metod pracy. Przeważała jednak postawa jeśli nie entuzjastyczna, to przynajmniej aprobująca. Nie brakowało pomysłów dobrych, ale przedwczesnych, nie do zrealizowania w ówczesnych warunkach, co pokazały następne lata. Ostatecznie przystąpiono do pracy w systemie bez większych zmian w organizacji pracy, modyfikacji uległ natomiast jej charakter. Zmiany następowały stopniowo, jako logiczna konsekwencja przemian następujących w bibliotekarstwie pod wpływem postępującej informatyzacji, a posunięcia, które wydawały się na początku nie do przyjęcia, z czasem okazywały się najlepszym wyjściem.

[8] Dotyczyło to wówczas przede wszystkim czasopism zakupywanych poza główną prenumeratą, m.in. w Ośrodku Rozpowszechniania Wydawnictw Naukowych PAN, figurujących na jednym rachunku razem z wydawnictwami zwartymi.

Pół kroku w kierunku systemu zintegrowanego

AFAS nie był integralną częścią systemu VTLS, ale odrębnym programem. Dawał do wyboru dwie możliwości pracy: zupełnie samodzielnej, w oderwaniu od całości systemu, albo we współpracy z nim. Biblioteka Jagiellońska zdecydowała się na ten drugi model. Współpraca ta polegała przede wszystkim na kopiowaniu z katalogu online do AFAS-u danych bibliograficznych, wykorzystywanych następnie przy zamawianiu lub rejestrowaniu wpływających książek. Siłą rzeczy pracownicy OGR musieli najpierw te dane bibliograficzne samodzielnie do katalogu wprowadzić, co sprowadzało się do zapisu w kilku polach formatu USMARC tytułu i odpowiedzialności, adresu wydawniczego, wydania i ISBN. Były to informacje wystarczające do zidentyfikowania książki na poziomie gromadzenia. Zarejestrowana w ten sposób książka funkcjonowała w katalogu online ze statusem „zamówiona", a jej opis był korygowany i uzupełniany przez Oddział Opracowania, kiedy już tam dotarła. Problem stwarzały opisy tytułów, które bardzo długo nie wpływały, a niektóre zamówienia nigdy nie zostały zrealizowane. W opinii wielu „zaśmiecało" to bazę i obniżało wartość naszego katalogu.

Dane skopiowane do AFAS-u wykorzystywano do tworzenia zamówień, wiążąc je z rekordami dostawców, kont księgowych, walut, realizacji, itp. W ten sposób rejestrowano nie tylko rzeczywiste zamówienia, ale również wpływające bez wcześniejszego zamówienia publikacje ze wszystkich źródeł. Miało to zapewnić szybką informację o zasobie bibliotecznym w celu jego bieżącego uzupełniania oraz uniknięcia dubletów. Było to zajęcie czasochłonne, tym bardziej, że wpływy do Biblioteki wzrastały, ale wówczas nie wyobrażano sobie, że można by przekazywać książki do opracowania formalnego bez zarejestrowania danych bibliograficznych na etapie gromadzenia zbiorów, jak to wówczas sugerowali konsultanci z firmy VTLS.

Istniała także możliwość przekazywania z AFAS-u do VTLS-u informacji o stanie realizacji zamówienia oraz danych adresowych z kartoteki dostawców AFAS do modułu obsługi wpływu czasopism VTLS w celu urgowania czasopism. Występowały przy tym jednak trudności techniczne, które ostatecznie wpłynęły na zaniechanie tych czynności.

Taki był tryb pracy z książką. Wpływ czasopism do czasu ich opracowania w katalogu online rejestrowano na tradycyjnych kartach kontynuacyjnych, a wpływające zbiory specjalne rejestrowano tylko w AFAS-ie.

Sekcja Egzemplarza Obowiązkowego wykorzystywała elektroniczną wersję Przewodnika Bibliograficznego, z której kopiowano do AFAS-u opisy książek, wykorzystywane potem do upominania się o nie u wydawców. Przez jakiś czas praca EO w tym zakresie była jeszcze bardziej uproszczona, gdyż zautomatyzowano monitowanie poprzez mechaniczne porównywanie zawartości Przewodnika Bibliograficznego z zawartością katalogu online, a upomnienia wysyłano na wyselekcjonowane w ten sposób opisy. Jednak nie była to optymalna metoda, gdyż w ten sposób przy urgowaniu pomijano książki będące w posiadaniu bibliotek instytutowych Uniwersytetu Jagiellońskiego, a których nie miała Biblioteka Jagiellońska. Komputerowy katalog jest bowiem katalogiem centralnym Uniwersytetu Jagiellońskiego, obejmującym również zbiory biblio-

tek instytutowych. Po jakimś czasie zrezygnowano więc z tej metody w trosce o kompletność zbiorów BJ.

Istotnym problemem był brak możliwości nadawania lub choćby notowania numeru akcesji w AFAS-ie. W Bibliotece Jagiellońskiej prowadzi się zbiorczą ewidencję wpływu, obejmującą wszystkie otrzymywane materiały, niezależnie od tego, czy wejdą następnie do zbiorów. W dalszym ciągu prowadzono więc rejestr przybytków w sposób tradycyjny, pisany ręcznie, a nadany w ten sposób numer akcesji wpisywano w polu notatek przy rekordzie zamówienia. Prowadzono nawet dyskusje, czy numer akcesji jest w ogóle potrzebny, co się odzwierciedliło tym, że przez kilka pierwszych lat w katalogu online nie umieszczano go w żadnym polu opisu bibliograficznego czy rekordu egzemplarza. Praktyka wykazała, że w bibliotece takiej jak BJ akcesja, świadcząca o proweniencji książki czy innego materiału bibliotecznego, jest czymś ważnym.

AFAS nie zaspokajał także wymagań Sekcji Wymiany. Do jej zadań należy nie tylko zamawianie i przyjmowanie książek, ale i wysyłanie wydawnictw uczelnianych i innych do kontrahentów, czego AFAS nie obsługiwał. W dalszym ciągu więc posługiwano się do tego celu programem, stworzonym w zasadzie na potrzeby księgarni, który był w stanie tworzyć listy wysyłkowe do kontrahentów mających u nas stałe zamówienia na wydawnictwa ciągłe. Nie jest to tylko problem spotykany w produktach firmy VTLS, moduł gromadzenia systemu PROLIB również nie oferuje takiej funkcji[9]. Idealna byłaby sytuacja, gdyby całość prac wymiany obejmował jeden system, potrafiący jeszcze dokonać bilansu publikacji wysyłanych w stosunku do otrzymywanych.

Tak więc AFAS tylko częściowo zaspokajał potrzeby OGR, pozostawiając poza swoim zakresem wiele działań, przy których musiano posiłkować się innymi programami, nie był też w stanie dostarczyć wymaganych od OGR danych statystycznych. Dane bibliograficzne, kopiowane z katalogu online, mogły być w AFAS-ie modyfikowane lub usuwane, co nie wpływało na stan danych w katalogu.

I krok wstecz

Po dwóch latach zdecydowano się na zaprzestanie wstępnego katalogowania książek przez pracowników OGR w katalogu online, ponieważ było to zbyt czasochłonne, a kłopotliwe dla oddziału opracowania, który argumentował, że łatwiej założyć nowy rekord bibliograficzny, niż poprawiać już istniejący[10].

[9] J. Dziak: *Zastosowanie nowoczesnych technologii informacyjnych w procesie gromadzenia zbiorów na przykładzie biblioteki wyższej uczelni technicznej*. W: *Usługi – aplikacje – treści w gospodarce opartej na wiedzy*. Pod red. D. Pietruch-Reizes i W. Babika. Warszawa 2004, s. 168-173; K. Żmigrodzka: *Nowe formy pracy Oddziału Gromadzenia Zbiorów*. W: *Ewolucja procesów bibliotecznych na tle dziejów Biblioteki Głównej Akademii Ekonomicznej we Wrocławiu*. Wrocław 2007, s. 62-63.

[10] Nie jest to tylko nasze doświadczenie – jak pisze Alain Jacquiesson w cytowanej wyżej książce, coraz rzadziej stosowanym rozwiązaniem jest ponowne wykorzystanie w procesie katalogowania opisów dokumentów opracowanych wstępnie przy okazji zamówienia, gdyż korekta takiego opisu trwa dłużej niż sporządzenie nowego. A. Jacquiesson: *Automatyzacja*..., op. cit., s. 85. Tym niemniej wiele bibliotek polskich przyjęło taką organizację pracy, łącząc przy tym często działy gromadzenia i opracowania formalnego druków zwartych.

Zaczęto więc już ok. 1996 r. rejestrować książki tylko w AFAS-ie. Tymczasem gwałtownie wzrastająca liczba wpływających do biblioteki publikacji, zwłaszcza pochodzących z egzemplarza obowiązkowego, spowodowała duże kłopoty z bieżącym opracowaniem wpływających materiałów. Gdy w 1994 r. ze wszystkich źródeł wpływu razem otrzymaliśmy ok. 45 tys. woluminów zwartych i ciągłych, to w 1998 r. było tych woluminów prawie 63 tys., przy czym nie wzrosła liczba pracowników OGR[11].

Należy pamiętać, że BJ jest ustawowo zobowiązana do wieczystego archiwizowania jednego z dwu otrzymywanych egzemplarzy obowiązkowych, nie może więc, jak inne biblioteki, prowadzić selekcji otrzymywanych bezpłatnie od wydawców publikacji i wprowadzać do zbiorów tylko interesujące pozycje. Choć nie wszyscy wydawcy przysyłają egzemplarz obowiązkowy, to i tak jego ilość wzrasta z roku na rok. Ale nie tylko w tym odzwierciedla się uznaniowa rola BJ jako drugiej biblioteki narodowej w kraju: specjalizuje się ona w gromadzeniu rękopisów, starodruków, grafik, zbiorów kartograficznych, gdzie wprawdzie decyzja o pozyskaniu do zbiorów należy już do właściwych oddziałów zbiorów specjalnych, ale samo nabycie musi znaleźć odzwierciedlenie w dokumentacji OGR, a jako biblioteka uniwersytecka ma na celu zaspokajanie potrzeb czytelniczych i informacyjnych środowiska akademickiego.

Wskutek znaczącego wzrostu wpływów w 1998 r. podjęto decyzję o całkowitym zaprzestaniu wstępnego katalogowania na etapie gromadzenia zbiorów. Odtąd AFAS był wykorzystywany przez Sekcję Kupna, która za jego pomocą prowadziła zamówienia i rejestrowała rachunki, przez Sekcję Egzemplarza Obowiązkowego do monitowania wydawców o niedostarczone zgodnie z przepisami książki, a przez Sekcję Wymiany i Darów głównie do prowadzenia kartoteki ofiarodawców i kontrahentów.

Od tej pory książki, objęte tylko wstępną, zbiorczą ewidencją wpływu w OGR, kierowane są do Oddziału Opracowania, a katalog online stał się głównym źródłem informacji o nabytkach. W polu 910 rekordu bibliograficznego umieszczany jest nadany przez OGR numer akcesji, dzięki czemu jeszcze przed utworzeniem rekordów egzemplarza wiadomo, że dana książka jest własnością BJ, a nie jednej z bibliotek instytutowych. W wypadkach, które tego wymagają, można wspierać się dokumentacją związaną z rejestrem przybytków, m.in. wykazami otrzymywanych publikacji.

Producent oprogramowania nie przykładał się zbytnio do rozwijania AFAS-u, w którym przez wszystkie lata użytkowania były funkcje „niedostępne w tej wersji oprogramowania", sposób szeregowania znaków pozostawiał wiele do życzenia (np. polskie znaki szeregowane po alfabecie łacińskim), a opcja archiwizacji danych obejmowała tylko zakupy i dary. Dla egzemplarza obowiązkowego nie przewidziano nawet odpowiedniego typu zamówienia, użytkowano więc „federal depository", co wymagało podawania fikcyjnego w tym wypadku konta, z którego zamówienie miałoby być opłacone.

Pracowano już nad systemem Virtua, która miała być doskonałym produktem docelowym. Jednak testowanie modułu dla gromadzenia przeprowadzone w kilku polskich bibliotekach ok. 2004 r. dowiodło, że w dalszym ciągu nie

[11] Rok 2009 przyniósł ponad 80 tys. woluminów zwartych i ciągłych z samego tylko egzemplarza obowiązkowego przy tej samej obsadzie oddziału.

zaspokaja on ich potrzeb, skutkiem czego nie został w Polsce wdrożony[12]. W BJ rozpoczęto zatem w 2007 r. prace nad własnym programem dla OGR, na razie obsługującym tylko jeden, ale najpilniejszy wycinek pracy – upominanie się o książki niedostarczone w ramach egzemplarza obowiązkowego. Powstał program intuicyjny w obsłudze, przyjazny pracownikowi. Obejmuje kartotekę wydawców, do której można przenieść dane adresowe zebrane w AFAS-ie, kartotekę opisów bibliograficznych, zamówień, upomnień i ich wydruków, oraz dziennik korespondencji rejestrujący wychodzące pisma, daje dostęp do stron www i poczty e-mail z rekordu dostawcy oraz szersze niż w AFAS-ie możliwości przeszukiwania rekordów[13]. Pracę usprawnia system podpowiedzi w trakcie wypełniania pól i łatwy mechanizm wprowadzania danych kolejnych tomów serii monograficznych. Od razu dało się zauważyć, że praca w tym programie jest nie tylko bardzo wygodna, ale i szybka. Jest to program „szyty na miarę", który może być na bieżąco uzupełniany i rozbudowywany zgodnie z potrzebami. Natomiast programy czy też moduły większych systemów oferowane na rynku wykazują zwykle w sprawach związanych z gromadzeniem zbiorów braki, z powodu których biblioteki muszą się wspierać dodatkowymi oprogramowaniami w procesie gromadzenia zbiorów.

Czy gromadzenie zbiorów musi być objęte zintegrowanym systemem ogólnobibliotecznym?

Mówiąc o zintegrowanym systemie bibliotecznym, mamy na myśli system obejmujący całość procesów bibliotecznych, poczynając od gromadzenia zbiorów. System, w którym opis publikacji rejestrowanej przez oddział gromadzenia jest na następnych etapach opracowania rozbudowywany, a zarówno bibliotekarz, jak i czytelnik ma dostęp do informacji nie tylko o nabytkach od momentu ich wpływu, ale i o pozycjach dopiero zamawianych do zbiorów.

W kontekście uwarunkowań wewnętrznych, jakimi są cele i zadania Oddziału Gromadzenia Zbiorów, a także możliwości oferowanych przez funkcjonujący w Bibliotece system komputerowy, oraz zewnętrznych, czyli w naszym przypadku – współpracy z NUKAT, powstaje pytanie, czy objęcie procesów gromadzenia zbiorów BJ w ramach jednego, ogólnobibliotecznego systemu zintegrowanego jest konieczne i w ogóle możliwe.

---

[12] Zob. m.in: J. Ratkowska: *Trójmiejski Zespół Biblioteczny jako przykład współpracy środowisk naukowych w zakresie komputeryzacji bibliotek: wady, zalety, propozycje rozwiązań normalizacyjnych*. W: *Elektroniczny wizerunek biblioteki*. Pod red. Mai Wojciechowskiej. Gdańsk 2008, s. 201; P. Korobczak, P. Domino: *Baza danych Oddziału Gromadzenia Biblioteki Uniwersyteckiej we Wrocławiu*. W: *Komputeryzacja Biblioteki Uniwersyteckiej we Wrocławiu – dziesięć lat doświadczeń*. [dokument elektroniczny]. Praca zbior. pod red. Grażyny Piotrowicz. Wrocław 2006, s. 95-114; K. Mojejko-Kotlińska: *Bazy danych w automatyzacji funkcji bibliotecznych w BG UMCS w Lublinie*. W: *Infobazy '2002 – bazy danych dla nauki: materiały z konferencji, Gdańsk, 24 czerwca – 26 czerwca 2002 r.* [pod red. Antoniego Nowakowskiego] Gdańsk 2002, s. 183-188.

[13] W AFAS-ie jedynym elementem opisu bibliograficznego dostępnym do wyszukiwania był tytuł.

Moduł gromadzenia Virtua, będący częścią zintegrowanego systemu, dalej jest przede wszystkim narzędziem do zamawiania publikacji i rozliczania wydatków, a tymczasem główne źródło wpływu w BJ to egzemplarz obowiązkowy, wielokrotnie przewyższający wszystkie inne źródła wpływu razem wzięte. Znamienne jest, że żadna z bibliotek w Polsce użytkujących system firmy VTLS nie wprowadziła, poza Biblioteką Jagiellońską, produktów tej firmy obsługujących gromadzenie zbiorów, ponieważ w ocenie tych bibliotek nie były zadowalające. Potrzebne były inne wspomagające programy, a pewne czynności, które mogły być skomputeryzowane, wykonywano dalej tradycyjnie.

AFAS nie przewidywał nadawania numeru akcesji. Producent oprogramowania obiecywał wprowadzenie akcesji w module Virtua, jak jednak mogliśmy się przekonać na pokazie podczas spotkania użytkowników VTLS w Gdańsku w czerwcu 2010 r., polega to na dodaniu jednego pola w rekordzie egzemplarza, w którym można wpisać nadaną już wcześniej na daną pozycję akcesję[14]!

Kolejnym problemem jest to, że system zintegrowany zakłada gromadzenie metadanych już na etapie gromadzenia zbiorów, tymczasem w warunkach BJ byłoby to ich tworzenie, a więc katalogowanie. W literaturze często można napotkać koncepcje zakładające możliwość ściągania opisów bibliograficznych z baz danych dostawców[15] czy z ogólnokrajowych serwisów księgarskich[16] już na etapie gromadzenia zbiorów w momencie zamawiania publikacji, co w efekcie miało doprowadzić do zaniku oddziałów opracowania formalnego[17]. Jak do tej pory polscy wydawcy nie potrafili porozumieć się w sprawie centralnego katalogu składowego ani nie widzą takiej potrzeby, o czym można przeczytać w relacjach z zebrań Polskiej Izby Książki[18]. Bardziej realne jest pobieranie danych z bibliotecznego katalogu centralnego takiego jak NUKAT lub z Biblioteki Narodowej. Jest to jednak rozwiązanie dla mniejszych bibliotek, nie o tak rozbudowanych i różnorodnych zbiorach jak BJ, która stoi na czele bibliotek intensywnie katalogujących w NUKAT[19]. W ogóle trudno by było znaleźć kompletne źródło danych bibliograficznych, z którego pracownicy OGR mogliby czerpać opisy publikacji, wpływają tu bowiem książki z całego świata, zarówno całkiem nowe, jak i bardzo stare. Lata temu zrezygnowano z katalogowania wszystkich publikacji na poziomie OGR w katalogu online w trosce o zapewnienie szybkiej informacji – nie tylko czytelnikowi, ale i bibliotekarzowi, co jest możliwe dzięki naciskowi położonemu na bieżące katalogowanie wpływów przez oddział opracowania. Także dla innych bibliotek – instytutowych,

[14] Wynika to zapewne z faktu, że jest to system amerykański, utworzony pierwotnie dla tamtejszej biblioteki, a biblioteki amerykańskie uważają księgi i rejestry akcesyjne za zbędne. Zob. *Encyklopedia Wiedzy o Książce*. Wrocław 1971, s. 19.

[15] P. F. Burton: *Wdrażanie technologii informacyjnej: uwagi ogólne*. W: *Automatyzacja bibliotek: wybór materiałów z konferencji „Automatyzacja bibliotek" Wrocław, 11-13 grudnia 1992*. Wrocław 1993, s.11.

[16] A. Radwański: *Jak komputeryzować...*, op. cit., s. 60.

[17] J. Maj, M. Nahotko, W. Szczęch: *Zastosowanie komputera w bibliotece*. Warszawa 1996, s. 11.

[18] *Odpowiedź IKP na decyzję PIK w sprawie Polskiego Katalogu Składowego Książki*. [online]. [dostęp: 9.11.2010]. Dostępny w World Wide Web: <http://www.ksiazka.net.pl/?id=archiwum09&uid=13202>

[19] E. Chrzan: *Komputeryzacja bibliotek naukowych w Polsce: ewolucja świadomości*. W: *Elektroniczny wizerunek biblioteki*. Pod red. M. Wojciechowskiej. Gdańsk 2008, s. 190.

czy działających w tym samym mieście – sygnał, że dana książka znajduje się w BJ, pozwala zaoszczędzić na zakupie, zwłaszcza gdy chodzi o drogą książkę zagraniczną.

Natomiast podjęcie katalogowania na etapie zamawiania budzi poważne wątpliwości w kontekście współpracy z katalogiem centralnym NUKAT. Dane bibliograficzne wykorzystywane przy zamawianiu są czasami niepełne lub błędne, brane z folderów, katalogów i internetowych stron wydawców, z reklam w czasopismach, z dezyderatów czytelników. Problemem są też pozostające w katalogu opisy bibliograficzne niezrealizowanych zamówień, których usuwanie nie byłoby właściwe, gdyż stanowią informację dla OGR, że podejmowano działania w celu sprowadzenia danej pozycji, co okazało się niemożliwe na przykład z powodu wyczerpania nakładu.

Należy pamiętać o tym, że zadaniem OGR jest ciągłe dbanie o pozyskiwanie materiałów do zbiorów, uzupełnianie posiadanego zasobu i zdobywanie nowości z różnych źródeł. Pracownik OGR powinien koncentrować się na publikacjach, których nie ma w bibliotece, a więc potrafić dotrzeć do informacji o nich oraz postarać się je sprowadzić (darmowo lub odpłatnie), przeprowadzić selekcję wpływających darów czy publikacji otrzymanych w ramach wymiany zgodnie z profilem biblioteki i oczekiwaniami czytelników. Rejestrowanie otrzymywanych materiałów to efekt końcowy tej pracy i nie powinno zajmować zbyt dużo czasu. Rozwiązaniem może być łączenie działów gromadzenia i opracowania, jak to miało miejsce w niektórych bibliotekach, pytanie jednak, na ile było to faktyczne połączenie struktur i ich funkcji, a na ile zabieg formalny, zachowujący w obrębie działu podział na odrębne jednostki gromadzenia i opracowania. Nie jest to jednak jedyne możliwe wyjście. Naszym zdaniem, wypracowany przez nas w trakcie lat obserwacji i doświadczeń system pracy, w którym OGR koncentruje swoje wysiłki na uzupełnianiu zbiorów, a otrzymywane publikacje, po nadaniu numeru akcesji, kierowane są do oddziału opracowania i tam katalogowane, jest zadowalający, zarówno ze względu na potrzeby czytelników, jak i pracowników biblioteki.

Daleko jeszcze do pełnej komputeryzacji procesów gromadzenia zbiorów w BJ. Stosowany przez jakiś czas AFAS, choć nie w pełni satysfakcjonujący, znacząco zmienił komfort pracy, trudno jednak na trwałe zadowalać się taką sytuacją, ponosząc jeszcze dodatkowo koszty finansowe. Także obecnie oferowany moduł gromadzenia firmy VTLS nie jest w stanie spełnić naszych oczekiwań, bo choć jest częścią zintegrowanego systemu bibliotecznego, to w zakresie oferowanych funkcji nie odbiega wiele od AFAS-u. Mimo to są biblioteki w Polsce rozważające obecnie jego wdrożenie. Aby uniknąć problemów związanych ze współkatalogowaniem w NUKAcie, Biblioteka Uniwersytetu Gdańskiego proponuje ukryć przed czytelnikiem rekordy bibliograficzne utworzone przez pracowników gromadzenia zbiorów na poziomie katalogu lokalnego, a po dotarciu książki do oddziału opracowania zastępować je w całości na podstawie numeru systemowego przez właściwe rekordy skopiowane z NUKAT-u. Nie będzie więc mowy w takim przypadku o dostarczaniu czytelnikowi informacji o książce już na etapie zamawiania, co uważane jest za jedną z zalet systemu zintegrowanego, choć często nie jest zbyt dobrze

widziane przez samych bibliotekarzy[20]. Braknie w tym wypadku także drugiej cechy systemu zintegrowanego - stopniowego rozbudowywania i wzbogacania rekordu bibliograficznego zapoczątkowanego na etapie gromadzenia.

* * *

Przystępując do komputeryzacji procesów gromadzenia zbiorów w BJ funkcjonował pewien teoretyczny obraz tego, jak powinna ona przebiegać. Wychodzono z założenia, że już na etapie gromadzenia do katalogu online należy wprowadzać opisy bibliograficzne dla zamawianych i otrzymywanych publikacji, uzupełniane na następnych etapach drogi książki, zgodnie z preferowaną ogólnie ideą systemów zintegrowanych. Miało to zapobiec dublowaniu pracy przez oddział gromadzenia i oddział opracowania oraz dostarczyć czytelnikowi szybkiej informacji na temat już otrzymanych lub dopiero oczekiwanych przez Bibliotekę publikacji. Założenie to zostało jednak zweryfikowane przez praktykę. Z czasem zarzucono całkowicie wstępne katalogowanie otrzymywanych w darze, drogą wymiany czy w ramach egzemplarza obowiązkowego książek, nawet tylko na poziomie wewnętrznej bazy OGR, czyniąc to jedynie dla publikacji zamawianych, czy to w drodze kupna, czy w którejś z darmowych form pozyskiwania zbiorów.

Komputeryzacja gromadzenia zbiorów w przypadku BJ nie ograniczyła się do jednorazowego zaimplementowania zakupionego oprogramowania, ale jest to żywy, wciąż trwający proces, w którym zmianom podlegają zarówno wykorzystywane narzędzia informatyczne, jak i sama praca, jej tryb i organizacja. Zmiany te wynikają z obserwacji i wniosków wynikających z bieżącej pracy oraz z uwarunkowań zewnętrznych, głównie z rozwijającej się w kraju elektronicznej infrastruktury informacyjnej. Nie przeprowadzano gruntownych zmian organizacyjnych w strukturze działów, rozwijano natomiast współpracę pomiędzy nimi w celu usprawnienia pracy i zapewnienia czytelnikowi jak najszybszej możliwości dotarcia do informacji. Wyeliminowano zbędne, a czasami powtarzające się w różnych działach czynności. Osoby zajmujące się gromadzeniem zbiorów koncentrują się na pozyskiwaniu nowości i uzupełnień, katalogowanie pozostawiając przeznaczonym do tego oddziałom opracowania formalnego. Dzięki temu liczba zatrudnionych przy gromadzeniu osób nie wzrosła mimo wielokrotnego wzrostu wpływu i zmiany jego charakteru, a wykorzystywane narzędzia informatyczne, choć niejednorodne i rozproszone, znacznie ułatwiają pracę. Prowadzone będą działania zmierzające do rozbudowy zapoczątkowanego tu własnego oprogramowania obsługującego procesy gromadzenia zbiorów, co spowodowane zostało niesatysfakcjonującymi propozycjami firmy VTLS. Na ten stan rzeczy wpływ miał szereg czynników, przede wszystkim specyficzna

[20] B. Grocholska: *Status rekordu „odkryty" – czy warto pokazywać opis bibliograficzny na etapie zamawiania książki.* W: *II Seminarium: Gromadzenie zbiorów – sztuka wyboru: Wrocław, 23-24 czerwca 2005 r.* EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy 2005. Materiały konferencyjne nr 11. [online]. [dostęp: 9.11.2010]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/matkonf/grom2/grocholska.php>. Na podstawie przeprowadzonej ankiety autorka stwierdza, że zaledwie 20% z przebadanych przez nią polskich bibliotek uwidacznia swoje plany zakupowe przed czytelnikami.

sytuacja Biblioteki Jagiellońskiej, będącej biblioteką uniwersytecką, a jednocześnie pełniącej rolę biblioteki narodowej, co determinuje charakter i wielkość gromadzonych zbiorów. Obowiązek wieczystego archiwizowania egzemplarza obowiązkowego sprawia, że polityka gromadzenia skierowana jest nie tylko na potrzeby współczesnego, ale i przyszłego czytelnika, dla którego staramy się zachować całość polskiej produkcji wydawniczej jako świadectwo dorobku umysłowego i kulturalnego oraz dokument życia społecznego Polski.

Summary

The article describes the computerization process in the Library Collection and Completion Department (LCCD) of the Jagiellonian Library, regarding its integrated library system. The AFAS software for collection processes, offered by the VTLS Company at the beginning of the 90ties, and being used by the Jagiellonian Library as the only one in Poland, has not been satisfying, regardless its significant work improvement. However, a narrow range of functions resulted in the decision on creating an own software, instead of buying current collection module of Virtua. As one can observe, the LCCD does not have to register bibliographical data of incoming unordered volumes, but only those which are going to be ordered. Books incoming for free, after collective registration and accession, are passed at once to the Indexing Department.

CZY ISTNIEJE JĘZYK SŁÓW KLUCZOWYCH?

WIESŁAW BABIK: SŁOWA KLUCZOWE. KRAKÓW: WYDAW. UNIWERSYTETU JAGIELLOŃSKIEGO. 2010, 242 S.

O słowach kluczowych mówi się od ponad pół wieku, zwłaszcza w kontekście wyszukiwania informacji w dokumentacyjnych systemach komputerowych, a od co najmniej kilkunastu lat także w związku z przeszukiwaniem zasobów sieciowych. Nie jest to termin jednoznaczny, co wynika stąd, że funkcjonuje nie tylko w dziedzinie informacji, ale także w informatyce, językoznawstwie, literaturoznawstwie, neurobiologii. Można wręcz powiedzieć, że współcześnie poszerza swój zakres i zasięg, zwłaszcza w związku z zastosowaniami sieciowymi.

Każdy komputerowy system informacyjno-wyszukiwawczy daje możliwość wyszukiwania przez słowa kluczowe, a dokładniej indeks słów kluczowych (*key-words*), ale nawet pobieżna obserwacja tego sposobu wyszukiwania pozwala stwierdzić, że czasem chodzi o pojedyncze słowa języka naturalnego (przejęte z opisów bibliograficznych i innych tekstów) z zachowaniem ich form gramatycznych, czasem natomiast są to wyrażenia wielowyrazowe (frazy), a czasem mamy do czynienia z wykorzystaniem słowników terminologicznych. W pierwszych dwóch przypadkach mówi się o języku swobodnych słów kluczowych, w drugim – o kontrolowanym (znormalizowanym) języku słów kluczowych. To rozróżnienie i niejednoznaczność terminu *słowo kluczowe* prowadzą do pytania o to, czy możemy mówić o jednym języku i jego dwóch odmianach, a nawet dalej, czy możemy mówić o języku sztucznym, czy też tylko o słowach języka naturalnego, które nazwaliśmy słowami kluczowymi, ponieważ w pewnych warunkach korzystamy z nich jako „kluczy", które pozwalają dotrzeć do jakiegoś tekstu/informacji. Takie wątpliwości pojawiają się, zwłaszcza gdy korzystamy z zupełnie dowolnych słów w przeglądarkach sieciowych.

Podstawowy więc problem to ustalenie zakresu terminu *słowo kluczowe*, z jednej strony w kontekście języka naturalnego, z drugiej zaś jako jednostki leksykalnej sztucznego języka paranaturalnego, wykorzystywanego do indeksowania i wyszukiwania w systemach informacyjno-wyszukiwawczych i zasobach sieciowych. Jest to punkt wyjścia do dalszych rozważań:

– jaki jest związek słów kluczowych z językiem naturalnym, jaki z językami sztucznymi paranaturalnymi,
– jakie są cechy charakterystyczne języka słów kluczowych,
– jakie ma wersje,
– jak funkcjonuje,
– jak się mają słowa kluczowe do słownika mentalnego człowieka.

Na te pytania próbuje odpowiedzieć Wiesław Babik w swojej książce, która jest pierwszą monografią tego zagadnienia na gruncie polskim. Celem głównym, który sobie postawił autor, była próba określenia tożsamości języka słów kluczowych, jego charakterystyki (natury) poprzez określenie granic pomiędzy nim a językiem naturalnym oraz innymi paranaturalnymi językami informacyjno-wyszukiwawczymi. Inaczej mówiąc, zbadanie, czy język słów kluczowych jest językiem sztucznym, czy naturalnym.

Książka składa się z trzech części, z których dwie (druga i trzecia) poświęcone są językowi słów kluczowych jako językowi informacyjno-wyszukiwawczemu, a dokładniej słowom kluczowym jako jednostkom leksykalnym tego języka, natomiast pierwsza stanowi wprowadzenie do problematyki szeroko rozumianych słów kluczowych.

W części pierwszej (*Słowa kluczowe w różnych kontekstach*, s. 23-78) autor przedstawia definicje słów kluczowych spotykane w piśmiennictwie różnych dziedzin, skupiając się bardziej na rozumieniu słów kluczowych w wyszukiwarkach internetowych (przeglądarkach sieciowych) i systemach hipertekstowych, rozpatruje słowa kluczowe w kontekście języka naturalnego – m.in. związek słów kluczowych ze słownikami mentalnymi, kojarzenie terminów i ustalanie ich powiązań semantycznych (gniazda semantyczne). Są to pytania o to, jak myślimy, jakie znaczenie ma język naturalny w wyszukiwaniu informacji, jaka jest językowa reprezentacja świata w umyśle, wreszcie jak poszukujemy informacji? Ta różnorodność aspektów badawczych słów kluczowych sprawia, że stają się one przedmiotem zainteresowania interdyscyplinarnego (językoznawców, psychologów, pracowników informacji, informatyków).

Część druga dotyczy słów kluczowych jako jednostek leksykalnych języka informacyjno-wyszukiwawczego (*Słowo kluczowe jako jednostka leksykalna języka słów kluczowych*, s. 79-120). Autor daje przegląd historii i genezy języka słów kluczowych, sięgając do czasów bardzo odległych (słowa z tytułu), a we współczesnym znaczeniu odwołuje się do systemu unitermów Mortimera Taubego (połowa XX w.), podaje definicje, wymienia cechy istotne, wskazuje typy języków słów kluczowych. Mamy tu charakterystykę funkcjonalną i strukturalną języka słów kluczowych rozpatrywanego jako pewnego rodzaju model, omówienie systemu leksykalnego i gramatycznego, porównanie z językiem naturalnym i językami sztucznymi paranaturalnymi – językiem deskryptorowym i językiem haseł przedmiotowych. Autor przyjmuje definicję zawartą w *Słowniku encyklopedycznym informacji, systemów i języków informacyjno-wyszukiwawczych* (B. Bojar, 2002), mianowicie: „język słów kluczowych – język informacyjno-wyszukiwawczy o notacji paranaturalnej, bez określonej explicite w słowniku paradygmatyki, w którym jednostki słownika nazywane są słowami kluczowymi, a na gramatykę składają się reguły indeksowania współrzędnego". Definicję tę nieco poszerza, nie uwzględniając cechy paradygmatyki (s. 85).

Część trzecia poświęcona jest wykorzystaniu słów kluczowych (nie języka słów kluczowych) w wyszukiwaniu informacji (*Słowa kluczowe w systemach wyszukiwania informacji*, s. 121-216). Są tu uwzględnione słowniki słów kluczowych, polskie i obce, z różnych dziedzin, których wykaz (37) autor zamieścił i poddał analizie w swojej publikacji (s. 235-237). Są to wykazy słów kluczowych m.in. z zakresu medycyny, farmakologii, zdrowia publicznego, teologii,

religioznawstwa, pedagogiki, psychologii, logopedii, językoznawstwa, literaturoznawstwa, marksizmu-leninizmu, nauk politycznych, socjologii, ekonomii politycznej, filozofii, etnografii, botaniki. W przypadku słowników mamy do czynienia ze słowami kluczowymi odpowiednio dobranymi i uporządkowanymi, a więc terminologią danej dziedziny. Ten aspekt jest czasem wyraźnie podkreślany w podtytułach słowników (wykaz podstawowych terminów z...). Jak pisze autor, celem tego rozdziału była weryfikacja tezy o końcu języka słów kluczowych.

Zacznijmy od tytułu i części wstępnej. Autor nadał swojej książce tytuł *Słowa kluczowe*, nie: *Język słów kluczowych*. Jest to pewnego rodzaju dysonans, ponieważ już w pierwszym akapicie wstępu metodologicznego pisze: „Przedmiotem rozprawy są słowa kluczowe używane w indeksowaniu i wyszukiwaniu informacji. Jak wszystkie języki informacyjno-wyszukiwawcze stanowią one lingwistyczny instrument służący do organizacji wiedzy o świecie, utrwalonej w postaci dokumentów i udostępnianej za pomocą dokumentacyjnych systemów informacyjno-wyszukiwawczych, zarówno w tradycyjnych dokumentacyjnych systemach wyszukiwania informacji online, jak i zasobach Internetu. Zostały przedstawione na szerokim tle ich wykorzystania w różnych dziedzinach, nazwanych tu kontekstami. Rozważania dotyczą przede wszystkim ich statusu jako jednostek leksykalnych języka informacyjno-wyszukiwawczego" (s. 9, podkr. JS).

Z tekstu tego możemy wnioskować, że:

a) autor stawia znak równości między terminami *słowa kluczowe* i *język słów kluczowych* w odniesieniu do zastosowania w systemach informacyjno-wyszukiwawczych; pytanie, czy jest to tylko zmiana terminologiczna, czy także zakresowa?

b) język słów kluczowych traktuje jako język informacyjno-wyszukiwawczy, w którym słowa kluczowe są jednostkami leksykalnymi; pytanie, czy we wszystkich przypadkach i zastosowaniach?

Można dalej zapytać, dlaczego autor unika w tytule terminu *język słów kluczowych:*

– Czy dlatego, że omawia słowa kluczowe, wykraczając poza zastosowania w systemach informacyjno-wyszukiwawczych?

– Czy dlatego, że ma wątpliwości, czy język słów kluczowych istnieje, choć przecież opisuje go, podając jego cechy konstytutywne (słownictwo, gramatykę), zajmuje się jego semantyką, typami, wykorzystaniem?

– Czy dlatego, że termin 'język słów kluczowych' jest niejednoznaczny (język swobodnych słów kluczowych lub kontrolowanych słów kluczowych), co jest wynikiem niejednolitej natury słów kluczowych?

Zatrzymajmy się przy języku swobodnych słów kluczowych, który jest prototypem właściwego (kontrolowanego) języka słów kluczowych. Język swobodnych słów kluczowych znany jest nam z wyszukiwania w Internecie. Są to słowa wybierane z tekstów języka naturalnego i, jak spostrzega słusznie autor, „Cały aparat indeksowania i wyszukiwania nie należy do języka, lecz do systemu i został tu rozbudowany o wiele dodatkowych opcji, zależnych od aktualnych możliwości technologii komputerowej, a nie od samego języka" (s. 157). Te obserwacje prowadzą do stwierdzenia: „Języki słów kluczowych w odmianie języków swobodnych słów kluczowych stanowią więc zredukowaną

do potrzeb optymalizacji procesów informacyjnych odmianę języka naturalnego" (s. 119). W zakończeniu książki autor zajmuje zdecydowane stanowisko wobec języka swobodnych słów kluczowych, stwierdzając, że „nigdy nie był i nie jest językiem sztucznym, podobnie jak swobodne słowa kluczowe w Internecie" (s. 204). Jeśli nie był i nie jest językiem sztucznym, to znaczy, że jest językiem naturalnym. I to jest bardzo ważna konkluzja opracowania Wiesława Babika, z którą należy się zgodzić.

Co zatem z językiem słów kluczowych w wersji kontrolowanej, czyli nieco znormalizowanej? Autor uznaje, że słownictwo tej wersji języka jest słownictwem naturalnym, tylko w niewielkim stopniu poddającym się normalizacji. Zwraca też uwagę na brak explicite relacji paradygmatycznych w słownikach języków słów kluczowych (inaczej niż w językach deskryptorowych i językach haseł przedmiotowych), co oznacza właśnie brak normalizacji. (Co do relacji paradygmatycznych, to może dopuszczalne jest uznanie, że istnieją one w sposób niejawny, potencjalny, podobnie jak słownik tego języka, a ujawniają się dzięki możliwościom kojarzeniowym umysłu). Wszystkie jednostki leksykalne są samodzielne, podobnie jak w językach deskryptorowych. Autor stwierdza, że pierwotną wersją tego języka był język swobodnych słów kluczowych ze słownictwem niekontrolowanym. „Później próbowano z tego zrobić język kontrolowanych (znormalizowanych) słów kluczowych" (s. 203). Czy zrobiono? Można wątpić, skoro autor pisze: „Nazwy 'język (swobodnych, kontrolowanych) słów kluczowych' utrzymują się przede wszystkim ze względów marketingowych. Po prostu przyzwyczajono się do stosowania tych nazw na określenie słów kluczowych używanych w wyszukiwaniu" (s. 204). To potwierdzałoby zmianę zakresową terminów 'język słów kluczowych' i 'słowa kluczowe' a jednocześnie uzasadniało tytuł monografii. Ostatecznie, po opisaniu tego języka w części drugiej i trzeciej, autor dochodzi do wniosku, że język słów kluczowych nie istnieje jako sztuczny język informacyjno-wyszukiwawczy. Jest to tylko termin, usankcjonowany tradycją, który nie ma uzasadnienia w rzeczywistości, bowiem właśnie w rzeczywistości są tylko słowa języka naturalnego, które bywają używane w funkcji wyszukiwawczej, metainformacyjnej. „Obecnie słowa kluczowe *są formą wykorzystywania języka naturalnego w procesach zautomatyzowa*nego wyszukiwania informacji" (s. 204). W zakończeniu autor pisze, że nie udało mu się udzielić pełnej odpowiedzi na postawione w rozprawie pytanie o tożsamość i granice tego języka, bo wszystko tu jest „rozmyte", a więc jest to trochę „dziwny język", stwierdza (s. 203). Stąd już tylko krok, aby uznać, że język taki nie istnieje obiektywnie, istnieją tylko słowa języka naturalnego, używane w funkcji wyszukiwawczej, nazywane słowami kluczowymi.

Niewątpliwie wnioski autora są wynikiem nie tylko analizy piśmiennictwa na ten temat, ale także wynikiem obserwacji indeksowania tekstów i wyszukiwania w zasobach sieciowych oraz dokumentacyjnych bazach danych, a także wpływu technologii na sposoby wyszukiwania. Obserwacje te rzeczywiście prowadzą do wniosku, że albo musimy uznać, że język słów kluczowych jest „dziwny" zarówno w odniesieniu do słownictwa, jak i gramatyki (zerowej), albo jest to język naturalny. Słownik języka słów kluczowych jest otwarty i tożsamy ze słownictwem języka naturalnego (w zasadzie nie podlega normalizacji, każde wyrażenie jest dozwolone z wyjątkiem tzw. stop-listy, czyli pewnej grupy

wyrażeń wykluczonych, nieznaczących, np. przyimki, rodzajniki), znaczenia słów kluczowych pokrywają się ze znaczeniami w języku naturalnym, a nawet gdyby było inaczej, to o tym nie wiemy, bo nie ma objaśnień explicite. Natomiast w przypadku zasobów internetowych i niektórych systemów dokumentacyjnych słowa kluczowe są wybierane z tekstów dokładnie w takiej formie, w jakiej zostały zapisane w tekście. Jeśli odwołamy się do kompetencji użytkownika w korzystaniu z tego języka, to musimy stwierdzić, że jest to wyłącznie kompetencja wynikająca z kompetencji języka naturalnego. Znaczeń słów kluczowych nie uczymy się, ponieważ znamy je skądinąd. A zatem:

– Czy nie popełniamy nadużycia, nazywając coś językiem sztucznym, co nim nie jest?

– Czy nie przypisaliśmy językowi naturalnemu funkcji języka sztucznego tylko dlatego, że odnosi się do rzeczywistości tekstowej?

– Czy mechaniczne wybieranie słów z tekstu może być uznane za proces indeksowania?

Są to pytania, które nasuwają się podczas lektury książki Babika.

W tym kontekście pojawia się pytanie o przyszłość języków informacyjno-wyszukiwawczych[1], zwłaszcza wobec rozwoju technologicznego sprawiającego, że coraz rzadziej w wyszukiwaniu informacji pośredniczy jakiś „bezpośredni pomocnik". Użytkownik jest raczej skazany na siebie, a to oznacza, że najchętniej będzie korzystał z języka naturalnego, który jest mu najbliższy mentalnie. Czy potrzebne są nam języki typu klasyfikacji, skoro, aby je zrozumieć, potrzebny jest język przekładowy paranaturalny (czy może naturalny?) w postaci indeksu przedmiotowego, prowadzącego do odpowiednich miejsc (klas) struktury klasyfikacyjnej, oznaczonych symbolicznie? Wreszcie, trzeba pamiętać o najnowszym wpływie technologii na systemy informacyjne, mianowicie tzw. tagowaniu, czyli nadawaniu przez czytelników/użytkowników całkowicie dowolnych określeń (nazywanych jednak słowami kluczowymi) tekstom dostępnym przez przeglądarki sieciowe. Wszystko to na pewno powinno skłaniać do spojrzenia na języki informacyjno-wyszukiwawcze z jednej strony w kontekście rozwoju technologicznego, z drugiej zaś z punktu widzenia roli języka naturalnego w wyszukiwaniu informacji. Wydaje się, że takiego właśnie spojrzenia na język słów kluczowych dokonał Wiesław Babik w swojej książce, dowodząc, że coś, co zwyczajowo określaliśmy jako język słów kluczowych, jest co najwyżej zbiorem zwyczajowo nazywanych słów kluczowych, a te z kolei w rzeczywistości są tylko słowami języka naturalnego. Z punktu widzenia teorii języków informacyjno-wyszukiwawczych jest to spojrzenie rewolucyjne, ale może należy uznać je tylko za racjonalne.

Jadwiga Sadowska

[1] Zob. B. Bojar: *Języki informacyjno-wyszukiwawcze wczoraj, dziś... czy jutro?* „Zagadnienia Informacji Naukowej" 2009, nr 1, s. 3-24.

SEMANTIC WEB W BIBLIOTEKACH

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY „OPEN ACCESS TO KNOWLEDGE – PROMOTING SUSTAINABLE PROGRESS" 10-15 SIERPNIA 2010, GÖTEBORG

Ostatniego dnia tegorocznego 76. Kongresu IFLA, który odbył się w Göteborgu w dniach 10-15 sierpnia 2010 r., zorganizowano sesję poświęconą nowym technologiom w bibliotekarstwie. Sesja numer 149 była wspólną sesją czterech sekcji IFLA: Technologii Informacyjnej, Katalogowania, Klasyfikacji i Indeksowania oraz Zarządzania Wiedzą. Została zatytułowana „Biblioteki i Semantic Web" (Libraries and Semantic Web).

Celem sesji była charakterystyka wizji Semantic Web, ruchu na rzecz Linked Data i przekonanie bibliotekarzy do współuczestniczenia w tworzeniu i współdzieleniu zasobów wiedzy w środowisku sieciowym. W założeniu, sesja nie miała mieć charakteru technicznego, lecz prezentować rozwiązania o charakterze koncepcyjnym. Pomimo wczesnej pory (8.30) i ostatniego dnia obrad, spotkanie przyciągnęło wielu zainteresowanych (ponad 150 osób).

Sesję rozpoczął Richard Wallis (z firmy Talis – dostawcy oprogramowania i rozwiązań technologicznych dla bibliotek w Wielkiej Brytanii) prezentując referat pt. „Semantic Web i biblioteki" (Semantic Web & Libraries). W ciągu ponad godzinnego wystąpienia przedstawił koncepcję Semantic Web, prezentując jej założenia, technologie oraz nakreślając rolę i miejsce bibliotek w tworzeniu globalnej sieci semantycznej. Wprowadził słuchaczy w koncepcję Linked Data, czyli jednego z urzeczywistnień wizji Tima Bernersa Lee, w postaci sieci rozproszonych zbiorów danych operujących wspólnym modelem reprezentacji wiedzy – RDF (Resource Description Framework). Było to jedyne wystąpienie, które miało charakter wprowadzający do tematyki Semantic Web. Pozostałe referaty dotyczyły szczegółowych zagadnień. Były to: standardy katalogowania w kontekście Semantic Web, sprawozdania z projektów bibliotecznych związanych z tą tematyką, zagadnienia prawne oraz problematyka wykorzystania języków informacyjnych i kartotek wzorcowych w środowisku sieciowym.

Standardy katalogowania i Semantic Web

Gordon Dunshire z Uniwersytetu Strathclyde w Glasgow w wystąpieniu pt. „Inicjatywy na rzecz udostępnienia bibliotecznych modeli i struktur metadanych

w Semantic Web" (Initiatives to make standard library metadata models and structures available to the Semantic Web) podjął problematykę dostosowania bibliotekarskich standardów katalogowania i modeli metadanych do wymogów formalnych Semantic Web. Punktem wyjścia rozważań autora były standardy FRBR (Functional Requirements for Bibliographic Records) i jego pochodne oraz ISBD (International Standard Bibliographic Description). Służą one do modelowania i porządkowania uniwersum bibliograficznego zarówno w wymiarze abstrakcyjnym (FRBR) jak i konkretnym (ISBD). Zastosowanie tych standardów w Semantic Web polegałoby na pełnieniu przez nie funkcji ontologii internetowych czy przestrzeni nazw, wykorzystywanych do opisu obiektów sieciowych. Wiąże się to z koniecznością zapisu ich struktury w postaci sformalizowanej, którą w kontekście Semantic Web wyznacza model RDF. W 2007 r. podczas spotkania specjalistów z zakresu bibliograficznych metadanych (Data Model Meeting) w British Library (http://www.bl.uk/bibliographic/meeting.html) opracowano zalecenia dotyczące przyszłych prac nad istniejącymi standardami opisu i modelami metadanych. Znalazły się tam m.in. zapisy o konieczności opracowania rozwiązań, które pozwolą na integrację istniejących modeli danych bibliograficznych. Ma to polegać na tworzeniu tzw. profilów aplikacyjnych w ramach RDA, czyli nowych anglo-amerykańskich zasad katalogowania, których zadaniem będzie ekspresja modeli FRBR, FRAD (Functional Requirements for Authority Data) oraz FRSAD (Functional Requirements for Subject Authority Data). Kolejnym krokiem będzie zapis takich struktur za pomocą rekomendowanych przez konsorcjum WWW standardów reprezentacji w postaci RDF, RDF schema czy SKOS (Simple Knowledge Organization system). Pozwoli to środowisku bibliotekarskiemu na:

– operowanie dotychczasowymi standardami metadanych, ale w pełni dostosowanymi do architektury Semantic Web,

– opracowanie profilu aplikacyjnego Dublin Core dla środowiska bibliotekarskiego, który wykorzystywałby model FRBR,

– współtworzenie zasobów Semantic Web poprzez udostępnianie wysokiej jakości danych bibliograficznych.

G. Dunshire stwierdził, że jednym z problemów związanych z publikowaniem danych w Semantic Web jest znalezienie najbardziej efektywnego sposobu i narzędzia odwzorowania skomplikowanych relacji, jakie zachodzą pomiędzy elementami określonego zbioru, a więc zastosowanie odpowiedniej ontologii internetowej. Ten problem nie dotyczy środowiska bibliotekarskiego ponieważ uniwersum bibliograficzne jest bardzo dobrze zdefiniowane i opisane w aktualnych standardach bibliotecznych. Nasze zadanie polega na formalizacji standardów katalogowania i metadanych do postaci „zrozumiałej" dla aplikacji sieciowych, czyli zgodnej z modelem RDF. G. Dunshire poświęcił temu zagadnieniu wiele uwagi. Jedną z ostatnich inicjatyw, jakie podjęła IFLA na tym polu jest powołana pod koniec 2009 roku grupa zadaniowa IFLA Namespaces Task Group. Jednym z jej celów jest koordynowanie prac nad ekspresją standardów IFLA dotyczących katalogowania do postaci zgodnych z modelem RDF. Zostaną one opublikowane z wykorzystaniem zbioru identyfikatorów URI odwołujących się do zaprojektowanego na potrzeby projektu zasobu identyfikatorów – iflastandards.info.

W szerszym kontekście rozważań, G. Dunshire stwierdził, że obecność danych bibliograficznych w Semantic Web spowoduje zmianę „paradygmatu jednostki bibliograficznej", z rekordu bibliograficznego zawierającego zbiór metadanych określonych przez format i zasady sporządzania, do postaci zestawu pojedynczych deklaracji reprezentowanych w składni tzw. trójek RDF (RDF tripples). Konkludując, autor podkreślił zadanie, jakie stoi przed IFLA w rozwoju Semantic Web w postaci promowania wykorzystania istniejących standardów i rozwiązań praktycznych w publikowaniu danych w środowisku chmury danych powiązanych (Linked Data Cloud) oraz korzyści jakie mogą płynąć stąd dla środowiska bibliotekarskiego we współuczestniczeniu w tworzeniu zasobów globalnej sieci semantycznej.

Biblioteki i Semantic Web – projekty

Duża część wystąpień miała charakter sprawozdań z projektów, które dotyczyły szeroko rozumianej obecności danych bibliograficznych w Semantic Web. Tę tematykę poruszyli przedstawiciele Europeany – Europejskiej biblioteki cyfrowej, Biblioteki Narodowej Niemiec oraz projektu Theseus. Problematyka poruszana w tych wystąpieniach była zróżnicowana i dotyczyła zarówno rozwiązań konkretnych problemów jak i analizy przebiegu całego projektu.

Steffen Hennicke z Uniwersytetu Humboldta w Berlinie zaprezentował sieciowy model danych stosowany w Europeanie. Europeana Data Model, to nowy sposób strukturyzacji i reprezentacji danych dostarczanych do Europeany. Jest bardziej elastycznym, o większej sile ekspresji następcą wcześniejszego modelu danych stosowanych w tej zintegrowanej platformie wyszukiwawczej. Zbudowano go na wytycznych modeli danych zgodnych z wizją Semantic Web oraz Linked Data (RDF(S), OAI-ORE, SKOS oraz Dublin Core). Działa jako ontoiogia najwyższego rzędu, która umożliwia zachowywanie obecnych modeli metadanych, ale pozwala jednocześnie na ich wspólne przetwarzanie i interoperacyjność. Autor przedstawił charakterystykę Europeana Data Model w kontekście zgodności ze standardami Semantycznej Sieci.

Jan Hanneman w referacie pt. „Linked Data dla bibliotek" (Linked Data for Libraries) przedstawił doświadczenia Biblioteki Narodowej Niemiec w publikowaniu i udostępnianiu zbiorów danych bibliograficznych zgodnych z koncepcją Linked Data. Celem projektu DNB Linked Data Service jest udostępnienie w pierwszym etapie wybranych kartotek wzorcowych opracowywanych w Bibliotece Narodowej Niemiec. Są to:

– Personennamendatei (PND) – kartoteka wzorcowa haseł osobowych, 1,8 mln rekordów,

– Schlagwortnormdatei (SWD) – kartoteka wzorcowa haseł przedmiotowych, 160 tys. rekordów,

– Gemeinsame Körperschaftsdatei (GKD) – kartoteka wzorcowa haseł korporatywnych, 1,3 mln rekordów.

Autor streścił przebieg projektu, napotkane problemy oraz plany związane z publikowaniem zasobów Deutsche Nationalbibliothek jako danych powiązanych (Linked Data). Pomimo panującego przekonania o prostocie koncepcji

Linked Data, sam proces konwersji i udostępniania danych może przysparzać jednak wiele problemów. J. Hanneman zidentyfikował cztery obszary, w których mogą występować niejasności i trudności:

– techniczny – wymagana jest infrastruktura informatyczna i sprzętowa oraz wiedza z zakresu programowania,

– konceptualny – problematyka modelowania danych, poszukiwanie odpowiedniej lub odpowiednich ontologii, brak praktycznych wskazówek opartych na wcześniejszych doświadczeniach,

– prawny – rozwiązanie kwestii prawa do publikacji i udostępniania danych, określanie licencji,

– ogólny – brak dokumentacji dla publikowania danych bibliograficznych jako Linked Data, małe doświadczenie bibliotekarzy na tym polu.

Wnioski z wdrożenia projektu przedstawiają się następująco:

– konfiguracja tego typu usługi udostępniania danych nie jest prosta,

– problematyka modelowania danych bibliograficznych jest złożona,

– mentalność otwartej wymiany danych nie jest powszechna,

– doświadczenia z innych wdrożeń postrzegane są błędnie jako reguły.

Taki model udostępniania danych sprawia, że użytkownik pozostaje nierozpoznany. Odpowiednio opracowany model danych to klucz do użyteczności całej usługi.

Andreas Heß (Deutsche Nationalbibliothek) w referacie pt. „CONTENTUS – w kierunku semantycznych bibliotek multimedialnych" (CONTENTUS – towards semantic multi-media libraries) przedstawił założenia i cele projektu Contentus, wśród których najważniejszym jest dostarczanie instytucjom kultury rozwiązań informatycznych umożliwiających efektywną konwersję zdygitalizowanych zasobów do postaci multimedialnego i bogatego semantycznie środowiska wyszukiwawczego. A. Heß skupił się na problematyce integracji metadanych oraz zagadnieniach wyszukiwania semantycznego. Przedstawił gotowe rozwiązania technologiczne dla wybranych baz danych w kontekście semantycznego wyszukiwania oraz nawiązał do zagadnień projektowania interfejsów wyszukiwawczych.

Aspekty prawne

Zagadnienia prawne związane z publikowaniem i udostępnianiem danych bibliograficznych w Semantic Web zostały poruszone w jednym wystąpieniu. Patrick Danowski (Scientific Information Service, CERN) w referacie pt. „Krok pierwszy – wysadzić silos! – Otwarte dane bibliograficzne, pierwszy krok w kierunku Linked Open Data" (Step one: blow up the silo! - Open bibliographic data, the first step towards Linked Open Data) skupił się na problematyce ustalania i wyboru licencji udostępniania danych bibliograficznych w ramach zasobów Semantic Web. W tym kontekście poddał analizie tradycyjną koncepcję praw autorskich oraz licencję Creative Commons i domeny publicznej. Podstawowym wnioskiem wystąpienia P. Danowskiego było stwierdzenie, że udostępnianie danych bibliograficznych w Semantic Web powinno odbywać się wyłącznie w zakresie domeny publicznej. Spełnia ona wymogi Semantic Web oraz gwa-

rantuje wielokrotne wykorzystywanie (reusability) danych bibliograficznych oraz stabilność tego rodzaju projektów.

Języki informacyjne, kartoteki wzorcowe a Semantic Web

Ważne zagadnienia dotyczące dostosowania tradycyjnych języków informacyjno-wyszukiwawczych oraz kartotek wzorcowych do wymogów formalnych Semantic Web zostały przedstawione przez Bernarda Vatanta z francuskiej firmy informatycznej Mondeca. Przedstawił on ogólną metodykę przekształcania zapisu struktury słowników języków informacyjno-wyszukiwawczych oraz kartotek wzorcowych do postaci zgodnej z modelem RDF. Scharakteryzował projekty konwersji m.in. słowników RAMEAU, LCSH oraz tezaurusów AGROVOC i EUROVOC do formatu SKOS (Simple Knowledge Organization System). Format SKOS to formalny model reprezentacji struktury systemów organizacji wiedzy, który w 2009 r. został uznany za rekomendację Konsorcjum WWW.

Dyskusja

Po wystąpieniach miała miejsce kilkunastominutowa dyskusja, w której uczestniczyli słuchacze oraz prelegenci. Zwrócono w niej uwagę na wysoką jakość danych bibliograficznych i spełnianie przez nie wymogów tzw. core data. Oznacza to, że metadane tworzone w bibliotekach spełniają wymogi stabilności i wiarygodności, na co kładzie się szczególny nacisk w kontekście publikowania danych w Semantic Web. Zwracano także uwagę na problemy technologiczne i bariery informatyczne, które utrudniają bibliotekarzom podejmowanie inicjatyw związanych z udostępnianiem ich zasobów w modelu Linked Data. Podkreślano udział tradycyjnych języków informacyjno-wyszukiwawczych w opisie zasobów sieciowych dzięki ich konwersji do formatu SKOS. Pojawiły się także pewne zastrzeżenia dotyczące efektywności tego modelu. Pozwala on na efektywny opis słownika wybranego systemu organizacji wiedzy, lecz nie dysponuje wystarczającymi środkami do ekspresji zdań oraz reguł syntaktycznych.

Podsumowanie

Sesja nr 149 „Semantic Web w bibliotekach" należała do najciekawszych posiedzeń podczas tegorocznego kongresu IFLA. Pokazała, że z jednej strony w środowisku bibliotekarskim panują duże różnice w zaawansowaniu technologicznym i nadążaniu za zmianami w środowisku udostępniania informacji o charakterze globalnym, a z drugiej, że w tym samym środowisku prowadzi się zaawansowane koncepcyjnie i technologicznie projekty badawcze, które stawiają nas na równi z innymi sektorami współtworzącymi zasoby globalnej sieci semantycznej. Cały czas brakuje nam jednak jasnych reguł publikowania danych bibliograficznych w Semantic Web, które poparte byłyby doświadczeniem i poddane zostały krytyce. Potrzebna jest również, co zauważył Jan

Hanneman, zmiana mentalności związana z otwarciem się na udostępnianie własnych danych oraz narzędzi użytkownikom i aplikacjom spoza środowiska bibliotekarskiego. Nie bez powodu bowiem otwartość dostępu, współdzielenie danych, narzędzi i doświadczeń były mottem tegorocznego kongresu IFLA w Goeteborgu.

Marcin Roszkowski

Informacja dla Autorów

Redakcja „Zagadnień Informacji Naukowej" przyjmuje wyłącznie teksty wcześniej nieopublikowane: oryginalne prace badawcze, materiały źródłowe do działu Rozprawy, Badania, Materiały, recenzje do działu Recenzje i Omówienia, sprawozdania i materiały z wydarzeń do działu Kronika. Teksty prosimy przesyłać napisane w programie Word w formatach DOC lub RTF oraz w postaci wydruku (podwójny odstęp między wierszami) na jeden z niżej podanych adresów:

bbojar@gmail.com – Bożenna Bojar – redaktor naczelny
a.stanis@uw.edu.pl – Anna Stanis – sekretarz redakcji
wydawnictwo@sbp.pl – Wydawnictwo SBP, 00-335 Warszawa, ul. Konopczyńskiego 5/7

Każdy artykuł powinien zawierać streszczenie autorskie w języku polskim o objętości nie więcej niż ½ strony formatu A4 (ok. 1000 znaków) i słowa kluczowe. Streszczenia w języku angielskim (wykonane w wydawnictwie) wraz ze słowami kluczowymi są umieszczane w „Library and Information Science Abstracts" oraz „Knowledge Organization".

Tytuły różnych typów publikacji należy wyróżnić kursywą, tytuły czasopism powinny być umieszczone w cudzysłowie. Teksty wygłaszane wcześniej na konferencji należy uzupełnić o szczegółowe dane tej konferencji.

Wszystkie materiały ilustracyjne umieszczone w tekście powinny mieć własną numerację i tytuły. Przy dużej ilości materiałów ilustracyjnych prosimy o przygotowanie ich w powyższy sposób na osobnych stronach z zaznaczeniem ich miejsca w tekście.

Przy sporządzaniu przypisów przyjmuje się zalecenia normy PN-ISO 690: 2002 Dokumentacja. Przypisy bibliograficzne. Zawartość, forma i struktura (dla dokumentów drukowanych) z pominięciem w opisie numerów ISBN i ISSN.

Przypisy bibliograficzne ponumerowane liczbami arabskimi powinny być umieszczone na dole strony.

Pierwszy przypis do danego dokumentu powinien zawierać wszystkie konieczne elementy opisu bibliograficznego tego dokumentu. W przypadku ponownego odwołania się do dokumentu już opisanego, jeżeli następujące po sobie, kolejne przypisy dotyczą tego samego dokumentu, należy zamiast pełnego opisu stosować oznaczenie: Ibidem (po którym mogą występować numery stron, np.: Ibidem, s.10).

Gdy powołujemy się na dokument wymieniony w jednym z przypisów wcześniejszych, powtarzamy początkowe elementy opisu tego dokumentu, np. autora i początek tytułu lub tylko początek tytułu danej książki (w przypadku prac zbiorowych), dodając numer odpowiedniej strony, np.: K. Wolff: *Książka wśród młodzieży...* op. cit., s. 3.

Jeśli dzieło ma 1-3 autorów wymieniamy wszystkich. W przypadku, gdy autorów jest więcej, wymieniamy tylko pierwszego autora dodając [et al.].

Przykłady wybranych przypisów bibliograficznych

M. Dembowska: *Dokumentacja i informacja naukowa: zarys problematyki i kierunki rozwoju*. Warszawa 1965.

Słownik encyklopedyczny informacji, języków i systemów informacyjno-wyszukiwawczych. Oprac. B. Bojar. Warszawa 2002.

M. Banacka: *Wybrane problemy działalności informacyjnej bibliotek do końca XIX w.* W: *Informacja naukowa w Polsce*. Pod red. E.Ścibora. Olsztyn 1998.

E. Ścibor: *Co nam zostało z tych lat. Projekt SINTO po 15 latach*. „Praktyka i Teoria Informacji Naukowej i Technicznej" 2002, nr 3-4, s. 11.

Bibliografię załącznikową należy umieścić na końcu tekstu przed streszczeniem w języku polskim w układzie alfabetycznym autorów, opisy prac tego samego autora powinny być uporządkowane chronologicznie. Dla prac tego samego autora opublikowanych w tym samym roku, należy dodać do roku wydania literę (a, b, c itd.). Jeśli wymaga tego treść

artykułu bibliografia załącznikowa może być w układzie chronologicznym lub według formy wydawniczej. W opisach publikacji zagranicznych należy uwzględnić pisownie skrótów stron i numerów w języku tekstu (np. w języku angielskim W: = In:, s. = p.).

Redakcja przykładowych opisów bibliograficznych

Wydawnictwo zwarte

Żmigrodzki Z.: *Wybrane zagadnienia bibliotekarstwa – działalność informacyjna bibliotek.* Warszawa 1983.

Nahotko M.: *Metadane sposób na uporządkowanie Internetu.* Kraków 2004. Zeszyty Naukowe Uniwersytetu Jagiellońskiego. Prace Bibliotekoznawstwa i Informacji Naukowej z. 6 [8].

Auger Ch. P.: *Information sources in grey literature.* 3 wyd. London 1998.

Artykuł w pracy zbiorowej

Frączek R.: *Infobroker – wyszukiwanie informacji na zamówienie.* W: *Informacja naukowa: rozwój – metody – organizacja.* Pod red. Z. Żmigrodzkiego, W. Babika i D. Pietruch-Reizes. Warszawa 2006, s. 148-149.

Babik W.: *Inżynieria języka naturalnego na potrzeby języka dla systemów informacyjno-wyszukiwawczych.* W: Z. Vetulani, W. Abramowicz, G. Vetulani: *Język i technologia.* Warszawa 1996, s. 66-69.

Artykuł w czasopiśmie

Bojar B.: *Języki informacyjno-wyszukiwawcze – wczoraj, dziś... czy jutro?* „Zagadnienia Informacji Naukowej" 2009, nr 1(93), s. 3-24.

Safahieh H., Asemi A.: *Computer literacy skills of librarians: a case study of Isfahan University librariens, Iran.* „The Electronic Library" 2010, vol. 28, no. 1, pp. 89-99.

Artykuł w czasopiśmie elektronicznym

Grzecznowska A.: *Użytkowanie informacji biznesowej w sektorze małych i średnich przedsiębiorstw w warunkach zmieniającego się rynku usług informacyjnych.* EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2002, nr 11(40). [online]. [dostęp: 17.04.2004]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2002/40/grzecznowska.php>.

Szczepańska B.: *Broker informacji – zawód z przyszłością czy zawód z przyszłości?* EBIB Elektroniczny Biuletyn Informacyjny Bibliotekarzy. 2002 nr 11(40). [online]. [dostęp: 10.09. 2007]. Dostępny w World Wide Web: <http://ebib.oss.wroc.pl/2002/40/szczepanska.php>.

Normy

ISO 11620:1998, *Information and documentation. Library performance indicators;* wersja polska PN-ISO 11620:2006 *Informacja i dokumentacja. Wskaźniki funkcjonalności bibliotek.*

Dokumenty elektroniczne

Strona główna Dziennika.pl. [online]. [dostęp: 26.05.2009]. Dostępny w World Wide Web: <http://www.dziennik.pl/niezamówionych>.

Moje Miasto Kraków. [online]. [dostęp: 26.02.2010]. Dostępny w World Wide Web: <http://www.mmkrakow.pl/7502/2009/11/27/mmkowy-konkurs-z-mikolajkiem? category=interwencje>.

The Cochrane Library. W: *Biblioteka Główna Uniwersytetu Medycznego w Białymstoku*. [online]. [dostęp: 14.06.2009]. Dostępny w World Wide Web: <http://biblioteka.gumed.edu.pl/index.php?strona=195#co>.

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. [online]. [dostęp: 11.09.2009]. Dostępny w World Wide Web: <http://oa.mpg.de/openaccess-berlin/berlin_declaration.pdf>.

Autorzy proszeni są o podanie następujących danych: tytuł i stopień naukowy, miejsce pracy (instytucja, adres) i e-mail. Redakcja zastrzega sobie prawo skracania tekstów oraz wprowadzania zmian w uzgodnieniu z autorem. Materiałów niezamówionych redakcja nie zwraca.

Spis treści

I. ROZPRAWY, BADANIA, MATERIAŁY

Jadwiga Woźniak-Kasperek
KRYZYS WARTOŚCI WIEDZY? 3

Piotr Malak
ROZWÓJ BADAŃ NAD PRZETWARZANIEM JĘZYKA NATURALNEGO 21

Anna Stanis
ZAPOŻYCZENIA JĘZYKOWE W SYSTEMIE JĘZYKÓW HASEŁ PRZEDMIOTOWYCH 31

Veslava Osińska
ROZWÓJ METOD MAPOWANIA DOMEN NAUKOWYCH I POTENCJAŁ ANALITYCZNY W NIM ZAWARTY 41

Marcin Roszkowski
LINKED DATA – MODEL DANYCH POWIĄZANYCH W SEMANTIC WEB 52

Agnieszka Brachfogel
„TERMINY METADANYCH DCMI" I MOŻLIWOŚCI ICH WYKORZYSTANIA W OPISIE RZECZOWYM 69

Anna Wałek
FINANSOWANIE OPEN ACCESS 77

Ewa Dąbrowska
KOMPUTERYZACJA GROMADZENIA ZBIORÓW A ZINTEGROWANY SYSTEM BIBLIOTECZNY. DOŚWIADCZENIA BIBLIOTEKI JAGIELLOŃSKIEJ 85

II. RECENZJE I OMÓWIENIA

CZY ISTNIEJE JĘZYK SŁÓW KLUCZOWYCH?
WIESŁAW BABIK: SŁOWA KLUCZOWE. KRAKÓW: WYDAWNICTWO UNIWERSYTETU JAGIELLOŃSKIEGO. 2010, 242 S.
Jadwiga Sadowska 96

III. KRONIKA

SEMANTIC WEB W BIBLIOTEKACH.
WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY. „OPEN ACCESS TO KNOWLEDGE – PROMOTING SUSTAINABLE PROGRESS" 10-15 SIERPNIA 2010, GÖTEBORG
Marcin Roszkowski 101

Informacja dla Autorów 107

Contents

I. THESIS, RESEARCH, MATERIALS

Jadwiga Woźniak-Kasperek
CRISIS OF THE VALUE OF KNOWLEGDE? 3

Piotr Malak
DEVELOPMENT TRENDS IN NATURAL LANGUAGES' PROCESSING 21

Anna Stanis
LANGUAGE BORROWINGS IN THE SUBJECT HEADINGS SYSTEM 31

Veslava Osińska
DEVELOPMENT OF THE SCIENCE DOMAINS MAPPING METHODS AND ITS ANALITICAL POTENTIAL 41

Marcin Roszkowski
LINKED DATA –THE MODEL OF DATA CONNECTED INTO THE SEMANTIC WEB 52

Agnieszka Brachfogel
"DCMI METADATA TERMS" AND THEIR POTENTIAL APPLICATION IN INDEXING 69

Anna Wałek
OPEN ACCESS FUNDING 77

Ewa Dąbrowska
COMPUTERISATION OF LIBRARY COLLECTION PROCESS VS. INTEGRATED LIBRARY SYSTEM. EXPERIENCES OF THE JAGIELLONIAN LIBRARY 85

II. REVIEWS

CAN WE TALK ABOUT KEYWORD LANUAGE? WIESŁAW BABIK: SŁOWA KLUCZOWE. KRAKÓW: WYDAWNICTWO UNIWERSYTETU JAGIELLOŃSKIEGO. 2010, 242 S.
Jadwiga Sadowska 96

III. CHRONICLE

SEMANTIC WEB IN LIBRARIES. WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY. "OPEN ACCESS TO KNOWLEDGE – PROMOTING SUSTAINABLE PROGRESS" AUGUST 10-15TH, 2010, GÖTEBORG
Marcin Roszkowski 101

Information for Authors 107

NASZ PARTNER STRATEGICZNY

ALEPH Polska Sp. z o.o.
ul. Kossaka 7
01-576 Warszawa
tel./fax: (022) 839 83 18
www.aleph.pl
www.exlibrisgroup.com

Maciej Dziubecki
m.dziubecki@aleph.pl

Aleph Polska
automatyzacja bibliotek

ExLibris The bridge to knowledge

ALEPH Polska Sp. z o.o. powstała w marcu 2001 roku. Wcześniej (od roku 1992) była Działem Wdrożeń systemu bibliotecznego Aleph® w firmie TCH Systems S.A. Obecnie jest jedynym dystrybutorem systemu Aleph oraz innych produktów firmy Ex Libris™ (MetaLib®, SFX®, DigiTool®, Verde®, Primo®) w Polsce i działa jako biuro serwisowe dla polskich użytkowników tych systemów.

Ex Libris Group jest wiodącym światowym dostawcą wysokiej jakości systemów dla bibliotek, centrów naukowych i muzeów. Zintegrowane systemy biblioteczne Aleph® oraz Voyager®, będące sztandarowymi produktami firmy znalazły użytkowników w ponad 3000 instytucji w 62 krajach. Aleph jest wiodącym na rynku oprogramowaniem do automatyzacji bibliotek uniwersyteckich, publicznych, narodowych, naukowych jak również konsorcjów, sieci krajowych i wielkich korporacji.

W ostatnich latach Ex Libris wprowadził na rynek cztery nowe produkty, wyznaczające kierunek rozwoju nowoczesnego oprogramowania dla bibliotek: MetaLib, portal zapewniający dostęp do szeroko rozumianych zasobów informacji naukowej, SFX, system powiązań do źródeł informacji, DigiTool, system zarządzania obiektami cyfrowymi oraz Verde system do skutecznego zarządzania wyborem, oceną, zakupem i obsługą (wznowieniem lub rezygnacją z prenumeraty) elektronicznych baz danych i czasopism.

Najnowszy produkt, Primo, to system skupiający w jednym miejscu dostęp do wielu aplikacji, w sposób niewidoczny dla końcowego użytkownika. Jeden interfejs, rozszerzone możliwości wyszukiwania, jedna lista wyników, dostęp do materiałów z różnych źródeł i systemów w jednym miejscu.

ExLibris Primo

Primo® firmy Ex Libris™ to zaawansowane rozwiązanie do odkrywania i dostarczania informacji. Zapewnia ono czytelnikom nowoczesny i wygodny interfejs do wszelkich lokalnych i zdalnych źródeł informacji. Primo zostało stworzone po to, by biblioteki mogły oferować swoim użytkownikom usługi dopasowane do ich potrzeb i dostarczać relewantne informacje szybko i wydajnie, niezależnie od czasu i miejsca, w których są potrzebne.

Sprawdź jak działa Primo w Bibliotece Królewskiej Danii: http://search.kb.dk/beta/

Sprawdź jak działa Primo w The University of Iowa: http://smartsearch.uiowa.edu/

PL ISSN 0324-8194