

*Klasyfikacje biblioteczne, klasyfikacje dokumentacyjne,
katalogi internetowe, ontologie internetowe,
Klasyfikacja Dziesiętna Deweya,
Klasyfikacja Biblioteki Kongresu,
Uniwersalna Klasyfikacja Dziesiętna*

Barbara SOSIŃSKA-KALATA *

STRUKTURY KLASYFIKACYJNE W ORGANIZACJI ZASOBÓW INFORMACYJ- NYCH INTERNETU

W artykule omówiono narzędzia tematycznej organizacji informacyjnych zasobów Internetu w hierarchicznych strukturach klasyfikacyjnych lub kategoryzacyjnych. Wskazano charakterystyczne cechy ontologii stosowanych przez znane ogólne serwisy wyszukiwawcze (Yahoo!, AltaVista, Onet), a także przez serwisy kontrolowanej jakości (BUBL LINK, RSCI Resources Channel). Szczególną uwagę zwrócono na coraz szersze wykorzystywanie w Internecie tradycyjnych klasyfikacji piśmiennictwa – uniwersalnych systemów takich jak Klasyfikacja Dziesiętna Deweya, Uniwersalna Klasyfikacja Dziesiętna i Klasyfikacja Biblioteki Kongresu oraz klasyfikacji specjalistycznych przeznaczonych do organizacji dokumentów z zakresu poszczególnych dziedzin wiedzy.

1. WPROWADZENIE

W latach sześćdziesiątych dość powszechny był pogląd, że tradycyjne klasyfikacje piśmiennictwa są nieprzydatne dla zautomatyzowanych systemów informacyjnych. Mimo iż badania eksperymentalne już w latach sześćdziesiątych dowiodły, że nie ma istotnych przeszkód dla wykorzystania w systemach komputerowych tego typu narzędzi reprezentowania, organizowania i wyszukiwania informacji, mit o nieprzydatności klasyfikacji piśmiennictwa w konfrontacji z technologią komputerową utrzymywany był przez kolejne niemal czterdzieści lat. Jego uwiarygodnieniu sprzyjała dominacja specjalistycznych baz danych wśród zasobów in-

* Instytut Informacji Naukowej i Studiów Bibliologicznych, Uniwersytet Warszawski, Nowy Świat 69, 00-046 Warszawa (e-mail: Barbara.Sosinska@mercury.ci.uw.edu.pl).

formacyjnych udostępnianych przez systemy skomputeryzowane – baz źródłowych i informatorowych, operujących specyficznymi językami opisu i wyszukiwania danych oraz baz bibliograficznych, wywodzących się z reguły z tradycyjnych serwisów abstraktowych, operujących najczęściej językami deskryptorowymi lub słów kluczowych, rzadziej – specjalistycznymi klasyfikacjami (por. [9]). W latach osiemdziesiątych sytuację tę w pewnym stopniu zmienił rozwój bibliotecznych katalogów online, w których dość szybko znalazły zastosowanie najpierw dwie największe klasyfikacje amerykańskie – Klasyfikacja Dziesiąta Deweya (KDD) i Klasyfikacja Biblioteki Kongresu (KBK) – a wkrótce także Uniwersalna Klasyfikacja Dziesiąta (UKD).[†] W wielu jednak krajach, w tym także w Polsce, zastosowanie klasyfikacji w katalogu online ciągle należy do rzadkości.

Mimo dość powszechnego głównie w krajach anglosaskich wykorzystywania tradycyjnych systemów klasyfikacyjnych w katalogach OPAC, generalnie niemal do końca stulecia znaczna część społeczności bibliotekarskiej preferowała stosowanie w nich alfabetycznych systemów indeksowania, uznając je za wygodniejsze dla użytkowników, bo łatwiej poddające się dostosowaniu do potrzeb opisu współczesnego piśmiennictwa i automatycznego wyszukiwania informacji. Podobne poglądy zdominowały również myślenie o narzędziach organizacji zasobów Internetu (por. [12]). W pierwszym okresie rozwoju internetowych serwisów informacyjnych także wielu specjalistów wywodzących się ze środowiska informatyków poddawało w wątpliwość użyteczność organizowania rzeczowego dostępu do zasobów sieci zarówno za pomocą klasyfikacji, jak i jakichkolwiek innych biblioteczno-bibliograficznych narzędzi opisu tematycznego. Dominowało przekonanie, że najbardziej efektywną organizację zapewnią programy indeksowania automatycznego i techniki wyszukiwania wzorowane na doświadczeniach systemów pełnotekstowych. W ciągu ostatnich kilku lat, z jednej strony pod wpływem wzrostu zasobów sieciowych i skali ich wykorzystania, z drugiej zaś pod wpływem wzrostu świadomości zróżnicowania ich jakości, następuje stopniowe przewartościowanie tego do niedawna powszechnego poglądu (por. [6]).

2. GŁÓWNE METODY ORGANIZACJI ZASOBÓW INFORMACYJNYCH INTERNETU

W toku dotychczasowej ewolucji internetowych serwisów informacyjnych ukształtowały się dwie podstawowe metody organizacji i przeszukiwania jego zasobów, stanowiące swoistą odmianę dwóch klasycznych podejść do rzeczowego opracowania i organizacji piśmiennic-

[†] Wykorzystanie KDD i KBK w katalogach OPAC ułatwiło umieszczanie ich symboli w opisach dokumentów dystrybuowanych w formie MARC przez Bibliotekę Kongresu i OCLC; w 1985 r. przeprowadzono eksperyment DDC Online Project, którego celem było przedstawienie możliwości wyszukiwawczych tej klasyfikacji. W połowie lat osiemdziesiątych podobnym testom poddano też KBK. W 1988 roku, na seminarium FID w Espoo w Finlandii po raz pierwszy zademonstrowany został system ETICS (katalog online biblioteki Wyższej Szkoły Technicznej w Zurychu), w którym jako narzędzie wyszukiwawcze zastosowano UKD.

twa – opartego na indeksowaniu i porządkowaniu alfabetycznym oraz opartego na klasyfikowaniu i porządkowaniu logicznym, dziedzinowo-tematycznym.

Pierwsza metoda, stosowana przez większość internetowych wyszukiwarek, opiera się na architekturze *crawler-indexer*, a więc polega na wykorzystaniu robotów automatycznie identyfikujących nowe lub zmodyfikowane strony WWW, przesłaniu ich na lokalny serwer i zaindeksowaniu na podstawie słów i fraz kluczowych występujących w nagłówku lub w całym tekście dokumentu z określoną częstotliwością, najczęściej relatywizowaną do długości indeksowanego tekstu (por. [4]). Obecnie istnieje ponad tysiąc tego typu serwisów wyszukiwawczych, działających w skali globalnej lub lokalnej.[‡] Są one efektywnymi narzędziami wyszukiwania według możliwie precyzyjnie wskazanych nazw zagadnień, których dotyczy poszukiwana informacja, z reguły mają jednak skłonność do generowania odpowiedzi zawierających znaczną liczbę informacji nierelevantnej, albo o niskiej relewancji. Niewątpliwie największą zaletą wyszukiwarek jest względnie szybkie indeksowanie dużej liczby stron WWW. Niektóre spośród nich deklarują rejestrację w swych bazach danych około 300 milionów zaindeksowanych dokumentów, co według współczesnych szacunków odpowiada 1/3 liczby wszystkich dokumentów dostępnych w Internecie. Ocenia się jednak, iż obecnie serwisy te przeciętnie rejestrują niecałe 20% zasobów sieci. W ostatnich latach obserwuje się stopniowy spadek tego odsetka, co wyraźnie sygnalizuje nienadążanie współczesnych metod indeksowania za rosnącą dynamiką rozwoju Internetu.

Druga metoda, stosowana w internetowych katalogach (ang. *directories, web channels*), opiera się na tworzeniu tzw. *hotlists* – wykazów adresów URL uporządkowanych hierarchicznie według kategorii tematycznych, które mają ułatwić przeglądanie zasobów sieci. W przeciwieństwie do automatycznie indeksujących wyszukiwarek, katalogi najczęściej tworzone są na podstawie klasyfikowania dokumentów przez zespoły profesjonalnych klasyfikatorów, rzadziej przez grupy wolontariuszy; niektóre spośród serwisów prowadzących katalogi zachęcają autorów dokumentów do wskazywania wybranych kategorii tematycznych wśród metadanych dokumentów. Obecnie internetowe katalogi są zwykle zintegrowane z wyszukiwarkami, dzięki czemu sklasyfikowane w nich dokumenty podlegają systematycznej weryfikacji i aktualizacji.

3. INTERNETOWE ONTOLOGIE

Większość popularnych serwisów wyszukiwawczych wykorzystuje katalogi zorganizowane za pomocą pewnego rodzaju klasyfikacji (niekiedy określanej jako amatorska), zwanej *ontologią*. Zwykle stanowi ona hierarchiczny system kategorii tematycznych, często generowanych empirycznie na podstawie wstępnego podziału zasobów Internetu na kilkanaście ogólnych klas odpowiadających najczęstszym typom informacji poszukiwanej i utrzymywanej w sieci. Pionierem w tej dziedzinie był portal Yahoo! (<http://www.yahoo.com>), którego katalog

[‡] Stale aktualizowana charakterystyka i ocena obszernego zestawu wyszukiwarek internetowych umieszczona jest m.in. na stronach: <http://www.searchengine.com>, <http://searchenginewatch.com>, <http://www.allsearchengines.com>, <http://www.searchengines.pl>, <http://bluepages.atelier.pl>.

stał się wzorcem dla kolejnych serwisów. W większości ontologii stosowanych przez popularne serwisy, np. AltaVista (<http://www.altavista.com>), America Online (AOL; <http://www.aol.com/>), Open Directory (<http://www.dmoz.org/>) wykorzystywany m.in. przez serwisy HotBot (<http://www.hotbot.lycos.com>), Google(<http://www.google.com>), Lycos (<http://www.lycos.com>) czy polskie portale Onet (<http://www.onet.pl>) i Wirtualna Polska (<http://www.wp.pl>), wśród kategorii wyróżnionych na pierwszym poziomie podziału powtarzają się takie jak: *Biznes i ekonomia; Edukacja; Gry; Komputery i Internet; Nauka; Podróże; Praca, kariera i pieniądze; Rozrywka; Samochody i motoryzacja; Sport i rekreacja; Sztuka i kultura; Zdrowie; Życie prywatne lub rodzinne; Życie publiczne*. Często powtarzają się też swoiste klasy formalne: *Aktualności* (prasa, pogoda, programy, repertuary, wiadomości, etc.), *Informatory* (ang. *Reference*; mapy, plany, słowniki, książki telefoniczne, etc.), *Ludzie i rozmowy* (ang. *People & Chat*), *Strony prywatne*, *Zakupy* (z dalszym podziałem według branż i typów oferowanych produktów). Kategorie uporządkowane są według alfabetu i nie mają z zasady nadanej żadnej sztucznej notacji. Naturalna nazwa kategorii jest równocześnie jej etykietą. W metodzie wyodrębniania tych kategorii dostrzec można pewne podobieństwo do sposobu, w jaki wyróżniane są kategorie tematyczne, za pomocą których w wielu krajach zachodnioeuropejskich organizuje się zbiory bibliotek publicznych.[§] Charakterystyczne też jest, że z reguły kategoria czy kategorie obejmujące naukę lub naukę i edukację poddane są dość głębokiemu dalszemu podziałowi, bardzo podobnemu do podziału nauk i tematów badawczych stosowanego w klasyfikacjach bibliotecznych i dokumentacyjnych lub instytucjonalnych klasyfikacjach nauki.

Hierarchiczna rozbudowa ontologii na ogół zależy od wielkości sklasyfikowanych zasobów Internetu, zwykle przyjmuje się bowiem zasadę segmentacji kategorii, do których zakwalifikowano więcej niż 20 – 25 adresów. W wielu ontologiach obejmuje ona średnio od kilku do kilkudziesięciu tysięcy podkategorii, zorganizowanych na 5 – 10 poziomach hierarchii. Np., w ontologii należącego do największych serwisu Yahoo! wyróżniono ponad 20000 kategorii szczegółowych, otrzymanych z podziału 14 kategorii podstawowych [10]; w największym polskim katalogu OKO wyróżniono 2500 kategorii szczegółowych, w których uporządkowano ponad 300 tysięcy polskich adresów WWW. W wielu serwisach przeglądanie rozbudowanej struktury katalogu ułatwiają mechanizmy tzw. skracania ścieżki dostępu do odpowiedniej kategorii szczegółowej, takie jak wyświetlanie obok kategorii podstawowych hiperłączy do podkategorii z kolejnych poziomów, prezentacja mapy katalogu i jego wybranych części, czy wyposażenie katalogu w wyszukiwarkę umożliwiającą natychmiastową lokalizację określonej kategorii. Zwykle, poczynając od kategorii drugiego poziomu, podawane są informacje o łącznej liczbie dokumentów zakwalifikowanych do niej i wszystkich jej podkategorii. Często też najczęściej wykorzystywane kategorie szczegółowe z określonego działu lub całego katalogu bywają wyświetlane w formie odrębnej listy tzw. „skrótów” (ang. *shortcuts*).

Większość internetowych ontologii ma strukturę polihierarchiczno-polirelacyjną, która umożliwia zarówno odwzorowanie złożonych zależności hierarchicznych tematów inter- i mul-

[§] Dobrym przykładem biblioteki o tak zorganizowanym księgozborze jest lokalna biblioteka w Gütersloh prowadzona jako modelowa biblioteka publiczna przez Fundację Bertelsmanna i władze miasta [5].

tidyscyplinarnych, jak i wielokrotne kwalifikowanie adresów poszczególnych dokumentów o złożonej, wielod dziedzinowej tematyce. Nie są to jednak klasyfikacje syntetyczne, zwykle bowiem wyliczają wszystkie klasy tematyczne, od stosunkowo prostych klas ogólnych po złożone, reprezentowane przez rozbudowane frazy nominalne. Wielokrotność podporządkowania kategorii szczegółowych klasom ogólniejszym często jest konsekwencją przyjmowania za podstawę kwalifikowania różnych elementów złożonego tematu. Według tej zasady kategoria tematyczna *Historia prawa w Niemczech* podporządkowana jest trzem różnym kategoriom szerszym: *Historia*, *Prawo* i *Niemcy*. Te wielokrotne podporządkowania pojawiać się mogą na różnych poziomach podziału. Np., w katalogu serwisu AltaVista dwukrotnie powtórzona jest kategoria *Neuroscience* – raz wśród kategorii trzeciego poziomu podporządkowanych kategorii *Sciences* i drugi raz wśród kategorii poziomu piątego, podporządkowanych podkategorii *Disciplines*, uzyskanej z podziału kategorii *Cognitive Science*. Z kolei w Katalogu OKO kategoria trzeciego stopnia *Historia sztuki@* jest równocześnie podkategorią w klasie tematycznej drugiego stopnia podziału *Nauki społeczne i humanistyczne*, jak i podkategorią czwartego stopnia podporządkowaną kategorii trzeciego stopnia *Historia*; kategorie *Czytelnia online@*, *Encyklopedie@*, *Jednostki miar i przeliczenia@*, *Korepetycje@*, *Materiały dydaktyczne dla nauczycieli*, *Słowniki@* rozsiane są w wielu miejscach struktury katalogu na różnych poziomach hierarchii, np.: jako podkategorie kategorii *Pomoce naukowe*, jako podkategorie nadrzędnej wobec niej kategorii *Zdobywanie wiedzy*, jako podkategorie kategorii *Pomoce naukowe* podporządkowanej kategorii *Uczniowie*, jako podkategorie ogólnej kategorii *Nauka i edukacja*. Podstawą wielokrotnej lokalizacji dużej liczby kategorii szczegółowych z pełną ich dalszą rozbudową zdaje się być zasada zbliżona do tej, którą w klasyfikacjach bibliotecznych i katalogach systematycznych stosuje się formułując odsyłacze uzupełniające. Umieszczane są w każdej klasie tematycznej, dla której istnieje relatywnie duże prawdopodobieństwo, że zainteresowanych nią użytkowników interesować będą także tematy należące do tych kategorii. Niekiedy mechaniczne przeniesienie całej rozbudowy pewnej kategorii do rozbudowy pokrewnej wobec niej klasy tematycznej prowadzi do naruszenia zasad logiki, zwykle jednak ma uzasadnienie pragmatyczne, jak na przykład umieszczenie kategorii *Materiały dydaktyczne dla nauczycieli* wśród podkategorii klasy *Pomoce naukowe* w kategorii *Uczniowie*.

Jakkolwiek w ostatnich latach równoległe wyposażenie serwisu w wyszukiwarke współpracującą z bazą adresów automatycznie zaindeksowanych dokumentów i w hierarchicznie uporządkowany katalog tematyczny stało się standardem realizowanym przez wszystkie duże serwisy wyszukiwawcze, semantyczna organizacja zasobów Internetu za pomocą pewnej struktury hierarchicznej szczególnie popularna jest w tzw. serwisach informacyjnych kontrolowanej jakości, z reguły zapewniających dostęp do pełnych tekstów dokumentów elektronicznych i najczęściej adresowanych do środowiska naukowego lub wybranych grup specjalistycznych. Część z nich wykorzystuje ontologie, czyli hierarchiczne systemy kategorii tematycznych opracowane na własny użytek. Zwykle jednak starają się one odwzorować strukturę współczesnej nauki i zachować logiczną poprawność podporządkowań w znacznie większym stopniu niż ontologie omówione wcześniej.

Wśród internetowych klasyfikacji lub kategoryzacji nauk na szczególną uwagę zasługuje *RealSci Classification Index* serwisu Academic Resources Channel i zastosowana w nim *Digital Science Classification* (<http://www.realsci.com>). Operuje obecnie 6777 klasami tematycznymi, a więc liczbą podobną do liczby klas w Broad System of Ordering – uniwersalnej klasyfikacji szczytowej opracowanej jeszcze w latach siedemdziesiątych dla potrzeb planowanego wówczas ogólnościwiatowego systemu informacji naukowej i technicznej UNISIST, propagowanej dziś jako doskonałe narzędzie tematycznej organizacji zasobów globalnej sieci informacyjnej [1], [3]. Są one zorganizowane w dziesięciu kategoriach ogólnych znakowanych notacją setną i grupujących nauki według oryginalnie zdefiniowanego przedmiotu badawczego: 00 Knowledge, 10 Language, 20 Wisdom, 30 Religion, 40 Society, 50 Universe, 60 Technology, 70 Medicine, 80 Creativity, 90 Civilization. W organizacji kategorii 50 Universe łatwo dostrzec dobrze znany ewolucyjny porządek nauk o świecie i przyrodzie; ciekawym pomysłem jest wyodrębnienie i sekwencyjne uporządkowanie nauk o wiedzy, czyli poznaniu, informacji, komputerach, bibliotekach, mediach, etc.

4. TRADYCYJNE KLASYFIKACJE PIŚMIENICTWA W INTERNECIE

W wielu przypadkach podstawą konstrukcji hierarchicznej struktury katalogów internetowych serwisów kontrolowanej jakości są schematy znanych klasyfikacji piśmiennictwa. Badania przeprowadzone w 1997 roku w ramach projektu DESIRE – *Development of an European Service for Information on Research and Education* wykazały, że w serwisach internetowych wykorzystywane są zarówno uniwersalne systemy klasyfikacyjne, takie jak KDD, UKD czy KBK, jak i ogólne klasyfikacje narodowe, takie jak Holenderska Klasyfikacja Podstawowa (hol. *Nederlandse Basisclassificatie*)^{**} opracowana dla potrzeb systemu Pica i bibliotecznych katalogów OPAC, również opracowana dla bibliotecznych katalogów online niemiecka Göttinger Online Klassifikation (GOK),^{††} czy szwedzki System Klasyfikacyjny SAB (*Sveriges Allmänna Biblioteksforening*),^{‡‡} jak wreszcie liczne klasyfikacje specjalistyczne, na przykład system Iconclass przeznaczony do organizacji informacji z zakresu ikonografii,^{§§} Klasyfikacja

^{**} Holenderska Klasyfikacja Podstawowa zastosowana została w serwisie NBW (*Nederlandese Basisclassificatie Web*) prowadzonym przez Koninklijke Bibliotheek we współpracy z holenderskimi bibliotekami akademickimi i w DutchESS: Dutch Electronic Subject Service (<http://www.konbib.nl/basisclas/basisclas.html>).

^{††} GOK od 1993 r. stosowana jest w katalogu online Dolnosaksońskiej Biblioteki Państwowej i Uniwersyteckiej w Getyndze (<http://db1-www.sub.uni-goettingen.de/gok/user/a/index.htm>) oraz w opcji wyszukiwania zaawansowanego w serwisach tematycznych: AnglistikGuide (<http://www.anglistikguide.de/>), Geo-Guide (<http://www.geo-guide.de/>) i HistoryGuide (<http://www.historyguide.de/>).

^{‡‡} System Klasyfikacyjny SAB stosują ją m.in. szwedzkie selektywne serwisy Länkskafferiet (http://www.ub2.lu.se/skolverket/sab_top.html), Internetkontakt (<http://www.btj.se/btj/saburl/saburl.html>), Systematic internetkatalog (<http://www.molndal.se/bibl/subject.htm>), Informationskällor ordnade enligt Klassifikationssystem för svenska bibliotek (SAB) (<http://www.chaplin.bibl.liu.se/sab/huvtswe.htm>).

^{§§} Informacje o klasyfikacji Iconclass i wykorzystujących ją instytucjach i projektach znajdują się na stronach: <http://iconclass.let.uu.nl/>, <http://iconclass.let.uu.nl/texts.institut.htm>.

NLM (ang. *National Library of Medicine Classification*),^{***} Engineering Information (Ei) Classification Codes,^{†††} Mathematics Subject Classification (MSC),^{†††} czy Association for Computing Machinery Computing Classification System (ACM CCS)^{§§§} [6]. Systemy te wykorzystywane bywają w różny sposób. Najczęściej służą do tematycznej segmentacji i przeglądania zarejestrowanych zasobów, są też używane w wyszukiwaniu.

W niektórych serwisach przeprowadzenie wyszukiwania za pomocą określonej klasyfikacji wymaga znajomości odpowiednich symboli, które czasami – aczkolwiek nie zawsze – można ustalić przeglądając internetową kopię tablic klasyfikacyjnych. W innych symbole klasyfikacyjne stanowią nazwy hiperłączy, obok których prezentowane są odpowiedniki słowne; użytkownik nawiguje po strukturze internetowej wersji klasyfikacji podobnie, jak czyniłby to przeglądając tradycyjne tablice klasyfikacji wyliczającej, jednak wybór któregoś z symboli w tym przypadku umożliwia mu natychmiastowe wyszukanie dokumentów, którym jest on przypisany. W jeszcze innych serwisach i wirtualnych bibliotekach operuje się po prostu frazami naturalnymi, którym przyporządkowany jest symbol niewidoczny dla użytkowników; w tym przypadku frazy wyświetlane w porządku systematycznym są nazwami tematów, które często klasyfikuje się za pomocą symboli rozwiniętych lub złożonych o wysoce skomplikowanej strukturze.

Według rejestru schematów organizacji zasobów Internetu, prowadzonego przez G. McKiernana, spośród trzech wielkich uniwersalnych klasyfikacji piśmiennictwa najczęściej wykorzystywana jest KDD (20 serwisów), następnie KBK (7 serwisów) i UKD (5 serwisów) [8]. Podane liczby są dość skromne, trzeba jednak pamiętać, że autor rejestru nie deklaruje jego kompletności. Np., KDD w ciekawy sposób wykorzystywana jest przez nieuwzględniony w nim europejski serwis o nazwie Renardus (<http://www.renardus.com>), skonstruowany w ramach 5. Programu Ramowego Komisji Europejskiej w rezultacie projektu, w którym uczestniczyły biblioteki narodowe i akademickie, ośrodki badawcze oraz serwisy typu *subject gateways* z Danii, Finlandii, Niemiec, Holandii, Szwecji i Wielkiej Brytanii. Innym interesującym serwi-

^{***} Klasyfikacja NLM umożliwia tematyczne przeglądanie zasobów medycznej informacji w Internecie w systemie OMNI (Organising Medical Networked Information; <http://www.omni.ac.uk/>) i w fińskiej bibliotece wirtualnej prowadzonej przez uniwersytet w Kuopio.

^{†††} Ei Classification Codes wykorzystywana jest w Ei Compendex – największej i najpopularniejszej bazie danych z zakresu nauk technicznych. W 1993 r. system poddano modernizacji i znacznemu pogłębieniu. Wykorzystywany jest w wielu internetowych serwisach informacji technicznej, m.in. w brytyjskim EEVL (Edinburgh Engineering Virtual Library; <http://eevl.icbl.hw.ac.uk>) i szwedzkim EELS (Engineering Electronic Library, Sweden; <http://eels.lub.lu.se>).

^{†††} Powstała w 1991 r. MSC opracowana została przez redakcje *Mathematical Reviews* i *Zentralblatt für Mathematik / Mathematics Abstracts*. Do organizacji matematycznych zasobów Internetu wykorzystywana jest przez American Mathematical Society, które na swej stronie WWW udostępnia serwis *Materials Organized by Mathematical Subject Classification* (<http://www.ams.org/mathweb/mimathbyclass.html>), a także przez niemiecki serwis *MathGuide* (<http://www.mathguide.de/>).

^{§§§} ACM CCS (dawniej: ACM CRCS – ACM Computing Reviews Classification System) jest podstawowym narzędziem identyfikacji i kategoryzacji zarówno piśmiennictwa informatycznego, jak i działalności badawczej i praktycznej w zakresie informatyki, jej zastosowań i przemysłu komputerowego. Wykorzystywana jest w bazach danych ACM, a w Internecie – przez system Ariadne, opracowany w ramach niemieckiego projektu Medoc (<http://ariadne.inf.fu-berlin.de:8000>). Elektroniczna wersja klasyfikacji z 2000 roku wraz z konwersją na edycję z 1991 roku dostępna jest na serwerze ACM: <http://www.acm.org/class/>.

sem pominiętym przez McKiernana jest amerykański CDRS (Collaborative Digital Reference Service), tzw. sieciowa „biblioteka bibliotek” prowadzona przez Bibliotekę Kongresu, OCLC, American Library Association, Library Systems and Services oraz Virtual Reference Desk, wykorzystująca równolegle KBK i KDD, a ostatnio także Klasyfikację NLM (<http://www.loc.gov/rr/digiref/>). KDD zastosowano do kategoryzacji rezultatów wyszukiwania w również nieuwzględnionej przez McKiernana bazie NetSearch, prowadzonej przez OCLC i rejestrującej zasoby dostępne w serwisie FirstSearch (<http://www.oclc.org/firstsearch>). Przykładów takich pominięć można podać wiele, a więc *de facto* zakres zastosowania najpopularniejszych uniwersalnych klasyfikacji piśmiennictwa w globalnej sieci jest znacznie większy niż wynikałoby to z cytowanego opracowania.

Każda z trzech wielkich klasyfikacji dostępna jest w Internecie w wersji elektronicznej: KDD udostępniona została w tej formie przez OCLC w serwisie WebDewey (<http://www.oclc.org/dewey/products/index.htm#webdewey/>) [2]; aktualna wersja tablic KBK znajduje się na serwerze Biblioteki Kongresu (<http://www.loc.gov/catdir/cpsol/lcco/lcco.html>) [7]; w ostatnich miesiącach 2001 roku British Standards Institution w porozumieniu z Konsorcjum UKD udostępnił serwis UDC Online (<http://www.udc-online.com>), a schemat podstawowy UKD umieszczony jest na serwerze UDC Website (<http://www.udcc.org>) [11].

Najbardziej znanym europejskim serwisem wykorzystującym KDD jest brytyjski BUBL LINK (<http://www.bubl.ac.uk/link>) – uniwersalny i selektywny przewodnik tematyczny kontrolowanej jakości, w którego zasobach znajdują się opisy stron WWW organizacji i instytucji, czasopism elektronicznych, wirtualnych bibliotek, innych serwisów tematycznych typu *gateway*, etc. Według danych z końca 2001 r., organizuje dostęp tematyczny do około 11000 adresów URL, co stanowi liczbę raczej skromną.^{****} Trzeba jednak pamiętać, że rejestrowane są tu tylko zasoby Internetu zweryfikowanej jakości, adresowane do społeczności akademickiej. Innymi przykładami stosowania KDD w sieci są amerykańskie serwisy Suite101.com (<http://suite101.com>) i CyberDewey (<http://www.anthus.com/CyberDewey/CyberDewey.html>).

Najbardziej znanymi serwisami korzystającymi z UKD są obecnie brytyjski NISS Information Gateway (<http://www.niss.ac.uk/>) i szwajcarsko-niemiecki GERHARD – GERman Harvest Retrieval and Directory (<http://www.gerhard.de/>). W NISS *Directory of Networked Resources* użytkownik może przeglądać zawartość działów za pomocą symboli UKD wyświetlanych wraz z odpowiednikami słownymi na kolejnych ekranach i pełniących funkcję hiperłączy. W opcji wyszukiwania zaawansowanego symbole UKD stanowią jeden z dostępnych kluczy wyszukiwawczych. Podobnie jak w BUBL LINK, wszystkie dokumenty i zbiory dokumentów zarejestrowane w NISS Information Gateway wyposażone są w obszerne charakterystyki. W serwisie GERHARD specjalizującym się w rejestrowaniu i organizacji dostępu tematycznego do zasobów informacji naukowej, UKD wykorzystywana jest do wyznaczenia systematycznego porządku prezentacji działów tematycznych i podporządkowanych im zagadnień i problemów badawczych oraz automatycznego klasyfikowania i wyszukiwania na podstawie słów kluczowych, którym system automatycznie przypisuje odpowiednie symbole klasyfikacyjne.

^{****} Dane liczbowe na podstawie informacji zamieszczonej na stronie <http://bubl.ac.uk/link/about.html>.

Przykładami innych znanych serwisów internetowych operujących UKD są brytyjski serwis SOSIG (Social Science Information Gateway, <http://www.sosig.ac.uk/>), w którym prowadzi się intensywne badania nad automatycznym klasyfikowaniem zasobów Internetu za pomocą UKD, duński serwis BitHit (<http://www.bithit.dk/>) oraz serwis powstały w ramach 5. Programu Ramowego Komisji Europejskiej – EASEL (Educator Access to Services in Electronic Landscape; <http://www.fdggroup.com/easel/>), gromadzący wszelkie materiały elektroniczne wykorzystywane w kształceniu i nauczaniu.

KBK stosowana jest głównie do organizacji katalogów amerykańskich serwisów informacji o kontrolowanej jakości, np. w CYBERSTACKSSM – wirtualnej bibliotece Iowa State University (<http://www.public.iastate.edu/~CYBERSTACKS/>), czy w serwisie ICRC – Internet Collegiate Reference Collection, prowadzonym w Harvey A. Andruss Library w Bloomsburg University (<http://www.library.bloomu.edu/reference>). W obu tych serwisach KBK stosowana jest w wersji uproszczonej i służy strukturyzacji i przeglądaniu katalogu tematycznego.

5. ZAKOŃCZENIE

Jeszcze niedawno kwestionowana użyteczność struktur klasyfikacyjnych do organizowania zasobów Internetu, dzisiaj wydaje się być poza dyskusją. Z jednej strony, dynamicznie rozwijają się hierarchiczne ontologie specjalistyczne; w ramach sztucznej inteligencji powstała nowa subdyscyplina zwana „inżynierią ontologiczną” (ang. *ontological engineering*). Z drugiej strony, coraz szerzej jako instrument organizacji zasobów Internetu wykorzystuje się znane klasyfikacje piśmiennictwa, których popularyzację w sieci wspierają instytucje zarządzające ich rozwojem – Biblioteka Kongresu, OCLC, Konsorcjum UKD i British Standards Institution oraz organizacje nadzorujące rozwój klasyfikacji specjalistycznych. Pytanie, które dzisiaj należy postawić nie dotyczy więc sensu tworzenia struktur klasyfikacyjnych zasobów sieci, ale tego, jak efektywnie te struktury generować, jak je sprawnie aktualizować i jak efektywnie operować nimi w opracowaniu ogromnych, silnie zróżnicowanych i bezustannie zmieniających się zasobów sieci. Jest oczywiste, że metody manualnej klasyfikacji nie mogą rozwiązać problemu. Kwestią kluczową pozostaje zatem wypracowanie efektywnych metod automatycznej klasyfikacji informacyjnej zawartości sieci. Rośnie liczba podejmowanych projektów badawczych i eksperymentalnych systemów, których celem jest testowanie programów automatycznej strukturyzacji zasobów Internetu według kryteriów tematycznych.^{††††} Większość z tych projektów wykorzystuje metody indeksowania derywacyjnego, tj. ekstrahuje wyrażenia z tekstów dokumentów i na ich podstawie kwalifikuje dokumenty do predefiniowanych klas tematycznych. Stosunkowo niewiele badań dotyczy automatyzacji klasyfikowania za pomocą tradycyjnych klasyfikacji piśmiennictwa. Ten nurt poszukiwań reprezentuje wspomniany projekt DESIRE, w którego drugiej fazie przygotowany jest program automatycznej klasyfikacji według UKD, testowany w serwisie SOSIG (<http://www.lub.lu.se/desire>), a także niemiecki projekt GER-

^{††††} Charakterystykę wielu takich projektów i systemów znaleźć można w serwisie CYBERSTACKSSM – Project Aristotle(sm): Automated Categorization of Web Resources (<http://www.public.iastate.edu/%7ECYBERSTACKS/Aristotle.htm>).

HARD wykorzystujący programy analizy lingwistycznej do automatycznego klasyfikowania również na podstawie UKD i amerykański projekt Scorpion realizowany przez OCLC, którego celem jest opracowanie systemu automatycznego katalogowania i klasyfikacji zasobów elektronicznych za pomocą KDD (<http://orc.rsch.oclc.org:6109>). Warto też wspomnieć o próbach wykorzystania technologii sieci neuronowych do automatycznego klasyfikowania zasobów Internetu; jednym z przykładów jest system KBS-CROSS opracowany w Bibliotece uniwersytetu w Lund, automatycznie generujący tablice przejścia między UKD i KBK oraz przyporządkowujący dokumentom symbole UKD na podstawie zawartych w ich metadanych symboli KBK (<http://delphi.kstr.lth.se/kbs/projects/kbscross.html>).

LITERATURA

- [1] BSO. (2001). *BSO-BROAD SYSTEM OF ORDERING: a systematic overview of knowledge*. BSO Home Page. <http://www.classbso.demon.co.uk/wwwpgaa.htm>
- [2] DDC. (2001). *Dewey Decimal Classification. Summaries*. http://www.oclc.org/dewey/about/ddc_21_summaries.htm
- [3] Gradinarov, P. (2000). *BSO: Broad System of Knowledge Ordering*. "RSCI Journal. Academic Resources Channel" vol. 2 nr 1 (February 29); <http://www.realsci.com/infobox.cfm/Key/2000022921390>
- [4] Indyka-Piasecka, A. (2000). Możliwości zastosowania tradycyjnych metod wyszukiwania informacji w sieci WWW. W: MISSI 2000: II Krajowa Konferencja – Multimedialne i sieciowe systemy informacyjne 2000. <http://www.zsi.pwr.wroc.pl/missi2000/referat34.htm>
- [5] Janssen, H. (1998). *Gütersloh: Neues aus der Elternbibliothek*. "BuB-Journal" vol. 50 nr 9: 546-547
- [6] Koch, T., Day, M. (1997). *The Role of Classification Schemes in Internet Resource Description and Discovery. Project DESIRE – Development of a European Service for Information on Research and Education*. Work Package 3 of Telematics for Research Project DESIRE (RE 1004) <http://www.ukoln.ac.uk/metadata/desire/classification>
- [7] LCC. (2001). *LCC Outline*. <http://www.loc.gov/catdir/cpsol/lcco/lcco.html>
- [8] McKiernan, G. (2001). *Beyond Bookmarks: Schemes for Organizing the Web*. CyberStacks(sm), Iowa State University, Ames, IA 50011; <http://www.public.iastate.edu/~CYBERSTACKS/CTW.htm>
- [9] Pollitt, A.S. (1998). *The key role of classification and indexing in view-based searching* <http://www.hud.ac.uk/schools/cedar/pollifla.html>
- [10] Steinberg, S.G. (1996). *Seek and ye shall find (maybe)*. "Wired [US]" 4.05 May <http://www.hotwired.com/wired/4.05/features/indexweb.html>
- [11] UDC. (2001). *Universal Decimal Classification. Outline*. <http://www.udcc.org/outline/outline.htm>
- [12] Weinberg, B.H. (1996). *Complexity in indexing systems. Abandonment and failure: implications for organizing the Internet*. "ASIS 1996 Annual Conference Proceedings, October 19 – 24, 1996" <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>