

MAREK NAHOTKO¹

CZYM SĄ METADANE?

Metadane – „dane o danych” to ustrukturyzowane dane, które opisują charakterystyki źródeł. Określają one atrybuty i zawartość oryginalnego dokumentu lub dzieła. W projekcie DESIRE² metadane zdefiniowano jako „związane z obiektami dane, które uwalniają potencjalnych użytkowników od potrzeby posiadania pełnej, apriorycznej wiedzy o istnieniu i charakterystykach tych obiektów”. Innymi słowy standardowa informacja bibliograficzna, streszczenia, terminy indeksowe oraz abstrakty są to wszystko namiastki (surogaty) oryginałów, a więc metadane.

Termin ten najczęściej stosuje się do źródeł elektronicznych (choć niekiedy) i odnosi się do „danych” w najszerszym sensie – m. in. baz danych, informacji tekstowej, grafiki, muzyki. Ponieważ koncepcja ta obejmuje także dane nie związane z opisem źródła, odchodzi ona znacznie od konwencjonalnej reprezentacji dokumentu, takiej jak rekordy MARC.

Metadane spełniają szczególną rolę w sieciach rozległych. Ponieważ większość wyszukiwarek przeszukuje dane tekstowe, bardzo ważne jest dodanie opisu tekstowego do nietekstowych plików, co umożliwi ich wyszukiwanie. Dopiero niedawno rozpoczęto próby przeszukiwania baz danych obrazów środkami poza-

¹ Absolwent BIN UJ, doktorat w dziedzinie nauk humanistycznych w Uniwersytecie Wrocławskim, specjalista w zakresie informacji naukowej, bibliotekarz dyplomowany.

² Tryb dostępu: <http://www.ukoln.ac.uk/metadata/desire/> [odczyt: 03.08.2005].

tekstowymi, takimi jak wykresy kolorów lub przez porównywanie twarzy czy innych elementów dzieł sztuk plastycznych. Ilość nitekstowych baz

danych rośnie: przykładem może być informacja przestrzenna, nie tylko po prostu nazwy geograficzne czy administracyjne, ale także współrzędne długości i szerokości geograficznej, wysokości i głębokości; dane opisujące siłę tornado; obrazy zasobów ziemi w podczerwieni; obrazy NASA z Marsa lub księżycy; bazy danych słynnych tematów muzycznych lub kolekcji muzealnych.

Idea metadanych jest nową koncepcją powstałą poza obszarem tradycyjnych tekstów i opisów bibliograficznych. Gdy tworzono zasoby danych cyfrowych – szczególnie danych o przestrzeni geograficznej – stało się jasne, że niezbędne są surogaty pozwalające na dostarczenie dodatkowej informacji o zawartości danych. Wzrastająca ilość i różnorodność obiektów udostępnianych w formie cyfrowej doprowadziła do stwierdzenia, że same dane pozbawione informacji o tym, jak i w jakim celu zostały zgromadzone, w jakich są zapisane formatach, jakiej wymagają platformy dla przeglądania i obróbki, czy istnieją zastrzeżenia co do ich reprodukcji i wykorzystania oraz konwencjonalnej identyfikacji, takiej jak autor lub producent, tytuł, hasła przedmiotowe, posiadają bardzo małą wartość.

Tworzenie metadanych w przypadku dokumentów elektronicznych jest jeszcze istotniejsze niż indeksowanie i katalogowanie źródeł drukowanych. Nie ma szansy na efektywne przeszukiwanie milionów dokumentów umiejscowionych w tysiącach serwerów na całym świecie bez wspomagania tego procesu przez metadane.

Systemy służące ocenie lub filtracji informacji także posługują się metadanymi. Gromadzą one metadane o autentyczności zasobów, możliwości uzyskania dostępu, podpisy cyfrowe, dane o prawie autorskim, możliwości tworzenia reprodukcji itp. Opis danych w zestawie danych umożliwia określenie przydatności zestawu dla określonych celów. Podobnie, dla pewnych rodzajów obiektów, ważne jest określenie pochodzenia i transformacji, które zostały wykonane na oryginalnym obiekcie elektronicznym w jego tzw. „cyklu życia”.

Niektóre metadane używane są tylko przez komputery. Określają one na przykład format, w którym utworzono źródło, dzięki czemu komputer może otworzyć zarówno aplikację jak i dokument automatycznie. Wskaźniki o lokalizacji lustrzanych danych (ang. mirrored) ułatwiają optymalny wybór sposobu dostępu. Metadane administracyjne wspomagają zarządzanie obiektem poprzez udostępnianie informacji o dacie jego utworzenia, kolejnych modyfikacji lub odpowiedzialności za zarządzanie nim.

Metadane pełnią więc kilka funkcji. Po pierwsze działają jako surogat większej całości. Muszą one charakteryzować we właściwy sposób oryginalny dokument tak, aby użytkownik zrozumiał jego zawartość, źródło pochodzenia, a także warunki wykorzystania. Jest to podstawowy problem dla gigantycznych zbiorów danych, takich jak np. bazy NASA.

Poza funkcją opisową poprawnie skonstruowany schemat metadanych musi także stanowić standaryzowaną strukturę i zasób terminologii. Tworzenie zestawu elementów o etykietach typu „twórca”, „autor”, „rzeźbiarz” lub „kompozytor”, nie przyda się na wiele, jeżeli pola te, spełniając tę samą funkcję (określają one tzw. „agenta”, czyli czynnik powodujący powstanie dzieła), nie będą sprowadzone do jednego, prostego pojęcia. Podobnie, jeżeli utworzy się „pole autorskie”, ale dopuści używanie różnych nazw dla tego samego twórcy, to nie uzyska się możliwości prostego i skutecznego wyszukania wszystkich dzieł interesującego nas autora. Niezbędne staje się utrzymywanie list autorytarnych lub standardowych słowników ujednoliconych nazw i odsyłaczy od form odrzuconych.

Zasadniczym celem metadanych jest poprawa jakości przeszukiwania zasobów, jednak spełniają one także inne funkcje:

- ułatwiają administrowanie danymi;
- zapewniają bezpieczeństwo i ochronę zasobów;
- ułatwiają zarządzanie informacją;
- pozwalają na ocenę treści danych;
- identyfikują zasoby;
- ułatwiają obsługę praw autorskich;

– ułatwiają archiwizację obiektów cyfrowych.

Podobnie do zasad stosowanych w tradycyjnych katalogach i indeksach, niezbędne jest tworzenie schematów metadanych pozwalających na tworzenie opisów różnego poziomu, w zależności od rodzaju obiektów i zamierzonego sposobu ich wykorzystania. Źródła, których wartość jest chwilowa, mogą otrzymywać tylko minimalny opis, podczas gdy te, które służą permanentnie lub mają wartość komercyjną, wymagają na ogół znacznie pełniejszego opracowania.

Metadane mogą być przygotowane podczas tworzenia obiektu cyfrowego (np. fotografii), przez jego twórcę lub z jego upoważnienia. Mogą one także być dodane później, jako efekt tradycyjnego procesu katalogowania. Początkowo twórcy schematów metadanych oczekiwali, że autorami metadanych będą autorzy obiektów cyfrowych. Obecnie rozwinęło się wiele inicjatyw służących „skatalogowaniu Internetu”, często na bazie niekomercyjnych wysiłków bibliotek i organizacji bibliotekarskich, bądź specjalistów z pewnych dziedzin wiedzy. Zwykle wysiłki te ograniczają się do tworzenia opisów wyselekcjonowanych źródeł, stanowiących zbiory danych używanych do jasno określonych celów. Zdarza się, że w uzupełnieniu do typowej informacji opisowej kolejne organizacje uzupełniają tę informację, np. o ocenę jakości (czy też wiarygodności) informacji.

Przedstawione punkty widzenia są ważne dla rozwoju sieciowych systemów informacji cyfrowej, wskazując dużą różnorodność koncepcji metadanych i ich praktycznych realizacji. Koncepcje te mogą być lepiej zrozumiane, gdy dokonamy podziału metadanych na odrębne kategorie. Najczęściej wyróżnia się metadane administracyjne, opisowe, archiwizujące, udostępniania i techniczne, choć możliwe są także i inne podziały.

W tabeli 1. opisane zostały wymienione rodzaje metadanych.

Rodzaj	Definicja	Przykłady
Administracyjne	Metadane służące zarządzaniu i administrowaniu zasobami informacyjnymi	– Informacja dot. gromadzenia, – Dokumentacja wymogów dla legalnego dostępu, – Informacja o lokalizacji
Opisowe	Metadane służące opisowi lub identyfikacji zasobów informacyjnych	– Rekordy katalogowe, – Narzędzia wyszukiwawcze – Wyspecjalizowane indeksy
Archiwizujące	Metadane związane z archiwizowaniem danych zasobów informacyjnych	– Dokumentowanie fizycznego stanu zasobów – Dokumentowanie działań podjętych dla archiwizacji fizycznej lub cyfrowej wersji źródła
Udostępniania	Metadane dotyczące poziomu i sposobu użytkowania zasobów informacyjnych	– Dane o udostępnianiu, – Informacja o zmianach treści i wersjach
Techniczne	Metadane dotyczące sposobu funkcjonowania systemu lub użycia metadanych	– Dokumentacja sprzętu i oprogramowania, – Informacja o digitalizacji, np. format, stopień kompresji itp. – Zapewnienie autentyczności i bezpieczeństwo danych

Tab. 1. Rodzaje metadanych

Oprócz wcześniej używanych formatów, takich jak wspomniany już MARC, powstały także nowe, przygotowane z myślą o opisie zasobów elektronicznych. Są ich obecnie setki. Dla ich opisanie i zarządzania nimi powstają specjalne rejestry, takie jak np. SCHEMAS³. Najbardziej znanym formatem jest Dublin Core Metadata Element Set (DCMES) zarządzany przez Dublin Core Metadata Initiative (DCMI)⁴. DCMES jest to zestaw 15 elementów danych (zob. polskie tłumaczenie)⁵, uważanych przez jego twórców za przydatne do opisu dokumentów elektronicznych (zob. tab. 2). Założeniem

³ Tryb dostępu: <http://www.schemas-forum.org> [odczyt: 03.08.2005].

⁴ Tryb dostępu: <http://dublincore.org/> [odczyt: 03.08.2005].

⁵ Tryb dostępu: <http://ebib.oss.wroc.pl/standard/dc.html> [odczyt: 03.08.2005].

twórców DCMES była jego prostota – użytkownikami zestawu mają być autorzy dokumentów elektronicznych. Szybko jednak okazało się, że 15 elementów to za mało i podjęto decyzję o ich kwalifikowaniu, czyli uszczegółowianiu ich znaczenia przez dołączanie kwalifikatorów (zob. tłumaczona lista kwalifikatorów)⁶.

Jest również częstą praktyką tworzenie zestawu kwalifikatorów lokalnie, dzięki czemu DCMES może być dostosowany do indywidualnych potrzeb.

Zawartość (Content)	Własność intelektualna (Intellectual Property)	Dookreślenie – Instantiation
Tytuł (Title)	Twórca (Creator)	Data (Date)
Opis rzeczowy (Subject)	Wydawca (Publisher)	Typ (Type)
Opis (Description)	Współtwórca (Contributor)	Format (Format)
Źródło (Source)	Własność (Rights)	Identyfikator (Identifier)
Język (Language)		
Relacja (Relation)		
Miejsce i Czas (Coverage)		

Tab. 2. Elementy Dublin Core

Od początku DCMES tworzony był w oparciu o kilka przyjętych zasad:

- Elementy muszą być proste, zrozumiałe i łatwe do wykorzystania, dzięki czemu każdy twórca zasobów sieciowych będzie miał możliwość opisanie własnego dokumentu bez potrzeby odbywania specjalnych szkoleń.
- Każdy element jest zarówno opcjonalny jak i powtarzalny.

⁶ Tryb dostępu: <http://ebib.oss.wroc.pl/standard/dcq.html> [odczyt: 03.08.2005].

- Elementy powinny być międzynarodowe i niezależne od danej dyscypliny w odniesieniu do zakresu i zastosowania.
- Zestaw elementów powinien być rozszerzalny, umożliwiając rozbudowę w zależności od potrzeb dziedziny zastosowań.
- Najważniejszym zastosowaniem zestawu elementów jest tworzenie opisów zasobów Web, umieszczanych wewnątrz tych zasobów (ang. embedded), tworzonych przez autorów źródeł.

Twórcy wszystkich formatów metadanych starają się zapewnić sobie możliwość konwersji do DCMES, jako powszechnie znanego i stosowanego standardu metadanych. Taka konwersja możliwa jest także pomiędzy Dublin Core a starszymi formatami danych, głównie MARC.

Na zakończenie warto stwierdzić, że właściwie zaprojektowane metadane powodują efekt w postaci właściwego zarządzania informacją. Spójny zestaw metadanych może być stosowany na wiele różnych sposobów pozwalając zaspokajać niestandardowe potrzeby użytkowników. Jednak do projektowania i zarządzania metadanymi niezbędne jest wydatkowanie pokaźnych funduszy oraz sporo wysiłku intelektualnego i technicznego. Można na przykład przedstawić podstawowe zagadnienia, które muszą być rozwiązywane podczas tworzenia cyfrowych systemów i obiektów:

- Określenie schematu lub schematów, które mogą być zastosowane dla najlepszej realizacji potrzeb twórców, systemu i użytkowników informacji;
- Decyzje o tym, które aspekty metadanych są zasadnicze dla planowanych celów, co oznacza przyjęcie optymalnego zestawu elementów metadanych w stosunku do przyjętych efektów;
- Zapewnienie, że zastosowano schematy metadanych w najnowszych wersjach.

Obecnie metadane mniej służą celowi, dla którego zostały stworzone, czyli katalogowaniu zasobów sieciowych (Internetu), a częściej stosowane są w miejscach, które stanowią w Internecie odpowiednik uporządkowanego świata bibliotek tradycyjnych. Są to głównie portale internetowe oraz wewnętrzne sieci przedsiębiorstw – intranety.

Wykaz innych publikacji na temat metadanych:

1. Nahotko, Marek (2005). Metadane dla obiektów kultury. W: Kocójowa, Maria red. *Informacja o obiektach kultury i Internet*. Kraków: IINiB UJ, s. 29-34. Seria III: ePublikacje Instytutu Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego ; Nr 1.
2. Nahotko, Marek (2004). *Metadane : sposób na uporządkowanie Internetu*. Kraków: Wydaw. Uniw. Jagiellońskiego, 201 s. Zeszyty Naukowe Uniwersytetu Jagiellońskiego MCCLXX. Prace z Bibliotekoznawstwa i Informacji Naukowej pod red. Marii Kocójowej. Z. 6 [8].
3. Nahotko, Marek (2004). Metadane dla oceny jakości zbiorów internetowych. W: Pietruch-Reizes Diana; Babik, Wiesław red. *Usługi – Aplikacje – Treści w gospodarce opartej na wiedzy*. Warszawa: PTIN, s. 226–231.
4. Nahotko, Marek [Dok. elektr.] (2003). Identyfikatory dla dokumentów w sieciach rozległych. W: Szczepańska, Barbara red. *Internet w bibliotekach II: łączność, współpraca, digitalizacja. Wrocław, 23-26 września 2003*. Warszawa: Wydaw. SBP
Tryb dostępu: <http://ebib.oss.wroc.pl/matkonf/iwb2/nahotko.php> [odczyt 16.03.2005].
5. Nahotko, Marek (2002). Metadata – contemporary tool for electronic documents description. In: *10th International seminar Scientific and Technical information in Central and Eastern Europe. Zakopane 9-12 May 2001. Information Society Technologies. Proceedings*. Warsaw: Inform. Process. Cent. for Sci. and Tech. Inform., p. 81-86.
6. Nahotko, Marek (2002). Metadane jako narzędzie opracowania dokumentów elektronicznych. W: Kocójowa, Maria red. *Elektroniczne publikacje w bibliotekach*. Kraków: Wydaw. UJ, s. 137-140.
7. Nahotko, Marek (2001). Functional Requirements for Bibliographic records (FRBR) – model opracowania zbiorów bibliotecznych IFLA. *Bibliotekarz* [nr] 1, s. 13-16.
8. Nahotko, Marek [Dok. elektr.] (2001). Metadane dla czasopism elektronicznych. *EBIB. Elektroniczna Biblioteka* 1 (19). Tryb

- dostępu: <http://ebib.oss.wroc.pl/2001/19/nahotko.html> [odczyt: 21.07.2005].
9. Nahotko, Marek (2000). Metadane – sposób na uporządkowanie Internetu. *Bibliotekarz* [nr] 7/8, s. 11-17.
 10. Nahotko, Marek (2000). RDF [Resource Description Framework] jako narzędzie uporządkowania zasobów internetowych. *Praktyka i Teoria Informacji Naukowej* nr 2, s. 3-6.

Tłumaczenia:

1. *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. [Dok. elektr.] (2000). Tł. z j. ang. Marek Nahotko. *EBIB. Elektroniczna Biblioteka. Standardy*.
Tryb dostępu: <http://ebib.oss.wroc.pl/standard/dc.html> [odczyt 21.07.2005].
2. *Dublin Core Qualifiers*. [KOMPUTERYZACJA... : ks. Gonet K., Dok. elektr.] (2001). Tł. z j. ang. Marek Nahotko. *EBIB. Elektroniczna Biblioteka. Standardy*.
Tryb dostępu: <http://ebib.oss.wroc.pl/standard/dcq.html> [odczyt 21.07.2005].