POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

TECHNOLOGIE INFORMATYCZNE W ZARZĄDZANIU SYSTEMY WSPOMAGANIA DECYZJI

pod redakcją:

Jana Studzińskiego,

Ludosława Drelichowskiego,

Olgierda Hryniewicza,

Janusza Kacprzyka

TECHNOLOGIE INFORMATYCZNE W ZARZĄDZANIU
SYSTEMY WSPOMAGANIA DECYZJI

Polska Akademia Nauk • Instytut Badań Systemowych

Seria: BADANIA SYSTEMOWE
tom 26

Redaktor naukowy:

Prof. dr hab. Jakub Gutenbaum

Warszawa 2000

TECHNOLOGIE INFORMATYCZNE W ZARZĄDZANIU SYSTEMY WSPOMAGANIA DECYZJI

pod redakcją
Jana Studzińskiego, Ludosława Drelichowskiego
Olgierda Hryniewicza i Janusza Kacprzyka

Książka zawiera wybór referatów przedstawionych na konferencji "Komputerowe systemy wielodostepne KSW'2000" w Ciechocinku w 2000 r. Konferencja pod patronatem Komitetu Badań Naukowych została zorganizowana przez Akademię Techniczno-Rolniczą w Bydgoszczy, Instytut Badań Systemowych PAN, Komisję Informatyki PAN - Oddział w Gdańsku oraz Bydgoskie Zakłady Elektromechaniczne "BELAM" S.A. w Bydgoszczy.

ISBN 83-85847-53-7
ISSN 0208-8028

Rozdział 1

Metodologia systemów informatycznych zarządzania

TWORZENIE PROFILI HYPERSDI W OPARCIU O METADANE HURTOWNI DANYCH

Witold Abramowicz, Krzysztof Węcel
Katedra Informatyki Ekonomicznej
Akademia Ekonomiczna w Poznaniu
e-mail: {W.Abramowicz,K.Wecel}@kie.ae.poznan.pl

Koncepcja hurtowni danych przewiduje przechowywanie przede wszystkim danych liczbowych, które pozwalają na uzyskanie dowolnych raportów obejmujących całą organizację. Interpretacja danych liczbowych opiera się w głównej mierze na wiedzy na temat biznesu oraz doświadczeniu analityków. Taka informacja ma charakter wewnętrzny. Rozwój zasobów Internetu stwarza nowe możliwości w zakresie pozyskiwania informacji zewnętrznych. Artykuł ten opisuje sposób budowania profili HyperSDI, służących do filtrowania informacji z Internetu. Głównym źródłem informacji o potrzebach informacyjnych użytkowników są metadane. Do tworzenia profili na podstawie metadanych wykorzystywane są dobrze znane techniki służące do indeksowania dokumentów. W celu ułatwienia komunikacji różnych aplikacji, profile hurtowni zapisywane są w XML.

1. Wprowadzenie

Czy możliwe jest skonstruowanie takiego filtru informacyjnego, aby hurtownia danych była zasilana odpowiednimi informacjami?

W ciągu ostatnich paru lat hurtownie danych znalazły się w centrum zainteresowania twórców systemów informacyjnych oraz menedżerów. Przedsiębiorstwa przekonują się, że właściwa informacja dostarczona osobom podejmującym decyzje jest potężnym narzędziem. Dane wyciągnięte z systemów operacyjnych i umieszczone w hurtowni danych nie są wystarczające, aby sprostać niemal niezaspokojonemu apetytowi menedżerów na informację.

Jeszcze do niedawna pozyskanie surowej informacji było dosyć trudne i dlatego była ona wysoko ceniona. Obecnie możemy pozyskiwać wszelkie informacje z Internetu. Dane stały się obfitym zasobem i właśnie z tego powodu, podobnie jak każdy inny towar, tracą swoją wartość. Stajemy przed

problemem oddzielenia istotnych faktów od pozostałych, a więc filtrowania informacji.

Największą i najbardziej złożoną kolekcją informacji jest sieć WWW. Składają się na nią setki tysięcy źródeł informacji. Sieć może posłużyć jako model tego, co powinno zostać osiągnięte wewnątrz przedsiębiorstwa, jeśli chodzi o organizację informacji w celu ułatwienia dostępu do niej. Dane numeryczne zawarte w hurtowni danych reprezentują treść ustrukturalizowaną. Z kolei sieć WWW składa się głównie ze statycznych stron z tekstem i obrazkami, określanych często jako treść nieustrukturalizowana. Jako że systemy informacyjne rozwijają się w kierunku lepszego wspierania podejmowania decyzji, dodanie nieustrukturalizowanych zasobów informacyjnych do ustrukturalizowanej zawartości hurtowni danych staje się niezbędne.

Struktura niniejszego artykułu jest następująca. W następnej części znajduje się krótkie wprowadzenie do hurtowni danych, w którym podkreślone są dwa ważne z punktu widzenia tego artykułu aspekty: tematyczność hurtowni oraz opis przy pomocy metadanych. Kolejna część wprowadza do tematyki filtrów informacyjnych oraz profili użytkowników. W części 4. została opisana nasza propozycja tworzenia profili HyperSDI w hurtowniach danych. Część 5. zawiera uwagi implementacyjne. Na końcu artykułu zostały zawarte wnioski oraz kierunki dalszych badań.

2. Rola hurtowni danych

W pierwotnej koncepcji hurtownia danych była repozytorium obejmującym całe przedsiębiorstwo, które ujednolicało i scalało wszystkie dane organizacji w jedną zunifikowaną strukturę. Z tego repozytorium wszystkie oddziały mogły uzyskać jasny i zrozumiały obraz całej organizacji. Hurtownia danych jest zazwyczaj podstawowym elementem systemów wspomagania decyzji (SDW) wykorzystywanych przez menedżerów.

Hurtownie danych można spotkać w najróżniejszych kształtach i rozmiarach. Niektóre są małe i trywialne. Niektóre są rozproszone i skomplikowane. Inne z kolei obejmują całą heterogeniczną sieć przedsiębiorstwa. Ale niezależnie od ich złożoności, hurtownie obiecują zamienić dane w wiedzę i pomóc przedsiębiorstwu zwiększyć swoją konkurencyjność.

2.1 Definicja hurtowni danych

Hurtownia danych jest to uporządkowany tematycznie, zintegrowany, zawierający wymiar czasowy, nieulotny zbiór danych wykorzystywany do wspomagania procesu podejmowania strategicznych decyzji (Inmon, Hackathorn, 1994).

2.2 Znaczenie metadanych

Metadane, które są zazwyczaj swobodnie definiowane jako „dane o danych", zyskują na znaczeniu, gdy organizacje zmagają się ze złożonym procesem dostarczania zintegrowanych rozwiązań swoim użytkownikom końcowym. Dokładniej należałoby je zdefiniować jako informacje opisujące właściwości zgromadzonych danych. Bez metadanych nie można mówić o hurtowni danych. Umożliwiają one użytkownikom poruszanie się po hurtowni, pozwalają na zapanowanie nad ilością danych przechowywanych w hurtowni, informują, jakie dane są aktualnie dostępne. Dopiero dokładnie opisane dane przedstawiają w przedsiębiorstwie określoną wartość.

W literaturze rozróżnia się dwa typy metadanych: metadane biznesowe i techniczne.

Tabela 1. Metadane biznesowe i techniczne

	Biznesowe	**Techniczne**
	najczęściej nieustrukturalizowane	najczęściej ustrukturalizowane
Definicja (definiuje lub opisuje dane)	• Co oznacza ta liczba? • Gdzie mogę znaleźć tę informację?	• format liczby, długość pola • baza danych, katalog
Transformacja (jak uzyskano dane)	• Jak to zostało obliczone? • Jakie reguły biznesowe zastosowano?	• filtry, agregaty • obliczenia, wyrażenia
Zarządzanie (jak dane są używane)	• Kto jest kierownikiem zespołu? • Jak aktualne są dane?	• planowanie zasobów, przydzielanie miejsca • indeksowanie, stopień zapełnienia dysków

Źródło: B. Moncla, *Business Meta Data Integration*, DM Review, September 1999, http://www.dmreview.com/

Metadane biznesowe są zazwyczaj słabo ustrukturalizowane bądź w ogóle nie posiadają struktury. Nie mogą być opisane przy pomocy rekordów i atrybutów. Częściej przyjmują formę dokumentów tekstowych, schematów organizacyjnych, katalogów członków, rozkładów szkoleń, obrazów i prawdopodobnie mogą również zawierać dane audio i video.

Można wyróżnić trzy poziomy metadanych w hurtowni danych (Singh, 1999, s. 227):

- Poziom aplikacji. Te metadane definiują strukturę danych przechowywanych w bazach operacyjnych oraz wykorzystywanych przez aplikacje. Są one dość złożone i zorientowane na aplikacje.
- Rdzeń hurtowni. Przechowywane w systemie bazodanowym hurtowni. Są zorientowane na tematy, bazują na jednostkach abstrak-

cyjnych ze świata rzeczywistego, np. „produkt", „klient", „dostawca".

- Poziom użytkownika. Mapują rdzenne metadane hurtowni na koncepcje biznesowe, które są bardziej użyteczne i lepiej rozumiane przez użytkowników końcowych.

W naszym projekcie skupiamy się na metadanych biznesowych. Najbardziej istotne są rdzenne metadane hurtowni, opisujące poszczególne tematy. Dodatkowo będą wykorzystywane metadane z poziomu użytkownika, w szczególności do tworzenia hiperpołączeń między hurtownią danych a pozyskanymi z Internetu dokumentami.

2.3 Logiczna organizacja hurtowni danych

W opisie logicznej organizacji hurtowni danych oraz w elementach implementacyjnych opieramy się na architekturze hurtowni danych firmy SAS Institute. Do zarządzania hurtownią wykorzystywany jest SAS/Warehouse Administrator.

Nadrzędną jednostką jest środowisko hurtowni. W ramach środowiska można zdefiniować wiele hurtowni danych. Dzięki temu każda z hurtowni dziedziczy takie metadane, jak właściciele i administratorzy, serwery, na których mają być składowane dane, biblioteki ze zbiorami. W ramach poszczególnych hurtowni również definiowane są metadane, które są współdzielone przez wszystkie obiekty zawarte w danej hurtowni.

Zgodnie z postulatami hurtownia danych zorganizowana jest w wiele grup tematycznych. Obszar tematyczny (*subject*) jest definiowany jako zbiór danych dotyczących jednego zagadnienia. W każdym obszarze tematycznym musi występować dokładnie jedna tabela logiczna (*detail logical table*). Składa się ona z jednej tabeli szczegółowej w postaci wielowymiarowej bazy danych lub wielu tabel szczegółowych połączonych w schemat gwiazdy. W obszarze tematycznym mogą być również umieszczone tabele przejściowe oraz grupy podsumowujące (*summary groups*) z różnymi poziomami agregacji.

Z tematem może być też skojarzony *info mart*, w którym umieszczane są informacje wygenerowane z tabel szczegółowych i sumarycznych. Z reguły są to raporty i wykresy, aczkolwiek mogą tu się znaleźć zapytania, dokumenty lub nawet całe aplikacje.

Szczegółowy opis logicznej organizacji hurtowni wybranej przez nas implementacji można znaleźć w (SAS Institute Inc., 1997a).

2.4 Hurtownia danych a Internet

Rozwój Internetu oraz integracja technologii hurtowni danych z technologią WWW spowodowały, że większą uwagę poświęca się potrzebom informacyjnym wszystkich użytkowników w ramach przedsiębiorstwa. Innymi słowy, zawartość hurtowni danych powinna być dostępna z przeglądarki, wyszukiwarki, poprzez e-mail czy Lotus Notes. Użytkownik zawsze powinien dotrzeć do interesującej go treści niezależnie od tego, w jaki sposób wyszukiwał informację. Oprócz tworzenia powiązań metadanych między aplikacjami a bazami danych, powinno się tworzyć połączenia między wspólnymi narzędziami wykorzystywanymi w Internecie a interfejsem przeglądarek (Singh, 1999, s. 228).

Istniejące rozwiązania pozwalają publikować dane z hurtowni danych w Internecie. Niniejszy artykuł jest częścią większego projektu, opisanego szerzej w (Abramowicz i inni, 2000), którego celem jest odwrócenie kierunku przepływu informacji. Nie hurtownia ma być źródłem informacji dla Internetu, ale Internet ma zasilać hurtownię danych.

3. Filtry informacyjne

3.1 HyperSDI

Koncepcja systemu HyperSDI powstała w 1995 roku w Katedrze Informatyki Ekonomicznej na Akademii Ekonomicznej w Poznaniu, a jej podstawę stanowią prace (Abramowicz, 1990, 1991). Opiera się ona na połączeniu idei systemów wyszukiwawczych, filtrów informacyjnych oraz hipertekstowej organizacji dokumentów.

System wyszukiwawczy jest to system służący do przechowywania jednostek informacji, które są przetwarzane, przeszukiwane, pozyskiwane i rozsyłane do różnych grup użytkowników. Tradycyjne systemy wyszukiwawcze zostały opisane w (van Rijsbergen, 1979, Salton, McGill, 1983).

Selektywna dystrybucja informacji polega na dostarczaniu określonym odbiorcom (zwanym konsumentami informacji) informacji zgodnych z ich zainteresowaniami, a odrzuceniu dokumentów, które ich nie interesują. Pierwotnie pojęcie to zostało wprowadzone przez H.P. Luhna (Luhn, 1958) w celu udoskonalenia komunikacji naukowej. Nową koncepcję zaproponował Housmann w (Greiff, 1998). Zakładała ona, że celem SDI nie jest znalezienie dokumentów, kiedy zajdzie taka potrzeba, ale ciągłe informowanie użytkowników o nowo pojawiających się dokumentach.

HyperSDI, podobnie jak inne systemy SDI, jest filtrem informacyjnym. Informacje przychodzące lub pozyskiwane są filtrowane w oparciu o potrzeby informacyjne zapisane w profilach użytkowników. Profile HyperSDI oraz sposoby ich doskonalenia zostały opisane w (Ceglarek, 1997).

3.2 Profile użytkowników

Użytkownik systemu HyperSDI, czyli konsument informacji, określa swoje zainteresowania poprzez zdefiniowanie *profilu konsumenta*. W klasycznych systemach wyszukiwawczych profil taki zawiera ważoną listę słów, zwanych termami.

Wszystkie postulaty związane z tworzeniem profili, przedstawione w (Ceglarek, 1997) zostały uwzględnione przy definiowaniu formatu zapisu profilu w formacie XML (Extended Markup Language).

3.3 Definicja SDIProfile w XML

Profile HyperSDI są zapisywane w XML. Na potrzeby niniejszego projektu została stworzona specjalna definicja typu dokumentu (DTD). Pełne źródło tego formatu można znaleźć w (Węcel).

Kodowanie profili HyperSDI w XML wydaje się być słusznym posunięciem wobec wzrastającej popularności tego języka. Podczas kilku miesięcy pracy nad projektem pojawiło się na rynku dużo narzędzi wspomagających tworzenie i wyświetlanie dokumentów XML. Firma Microsoft w najnowszej przeglądarce Internet Explorer 5.0 wprowadziła pełną obsługę języka XML. W przyszłości ma on szansę stać się podstawowym językiem opisu dokumentów w sieci World Wide Web.

Użycie języka XML powoduje, że implementacje SDIProfile mogą działać w dowolnym systemie operacyjnym. Dzięki temu łatwiejsze będzie tworzenie aplikacji filtrujących informacje, korzystających z uniwersalnego formatu.

4. Tworzenie profili

4.1 Terminologia

Profile tematów hurtowni danych służą do reprezentowania potrzeb informacyjnych grupy użytkowników związanych z danym tematem hurtowni danych. Nie wszystkie profile będą wykorzystywane do pozyskiwania dokumentów.

W naszym rozwiązaniu przewidujemy dwa typy profili ze względu na konsumenta informacji:

- profil użytkownika związany z osobą,
- profil tematu hurtowni danych.

Profil tematu hurtowni reprezentuje potrzeby informacyjne wirtualnego użytkownika i jest związany z wybranym obszarem tematycznym hur-

towni danych. Każdy profil tematu jest zarządzany przez osobę, określaną dalej jako administrator tematu.

W ramach koncepcji tworzenia profili tematów przewiduje się tworzenie kilku typów profili, z których każdy będzie miał swoje zastosowanie. Dla ich rozróżnienia została wprowadzona następująca terminologia:

- metaprofil – profil tematu hurtowni danych uzyskany w wyniku analizy metadanych, nie zmienia się, jest tylko profilem wyjściowym,
- bieżący profil tematu (zwany po prostu profilem tematu) – najistotniejszy profil, służący do pozyskiwania dokumentów, doskonalony w trakcie eksploatacji systemu,
- multiprofil – profil tematu hurtowni danych uzyskany poprzez złożenie profili wszystkich użytkowników tego tematu.

4.2 Indeksowanie

Jak już zostało to wcześniej wspomniane, w hurtowni danych znajduje się wiele tematów. Temat grupuje różne obiekty dotyczące jednego zagadnienia, takie jak tabele, wykresy, raporty. Każdy obiekt wśród innych atrybutów zawiera swoją nazwę, opis oraz komentarze, zapisane w metadanych. Naszym celem jest wykorzystanie tych metadanych do ekstrakcji użytecznej informacji do zbudowania profili.

Informacja tekstowa o każdym obiekcie może być potraktowana jako dokument. Następnie każdy taki dokument jest indeksowany z wykorzystanie tradycyjnych metod. Indeksy powinny dobrze reprezentować dany temat hurtowni danych i wyróżniać go spośród innych. W ten sposób otrzymujemy termy, które znajdą się w metaprofilu hurtowni.

W tym algorytmie nie zakłada się rozłączności tematycznej struktur hurtowni danych. Podobne dokumenty indeksujemy tymi samymi deskryptorami. Analogicznie będzie ze strukturami hurtowni danych. To, że indeks dobrze reprezentuje jakiś temat hurtowni danych nie oznacza, że reprezentuje on tylko ten jeden temat.

Podstawowym założeniem jest to, że metadane właściwie opisują informacje zawarte w hurtowni danych. Zakłada się również, że wszelkim zmianom dokonanym w hurtowni będą towarzyszyć stosowne zmiany w metadanych. Prawidłowe metadane są warunkiem koniecznym tworzenia poprawnych profili tematów hurtowni danych.

4.3 Wyszukiwanie słów charakterystycznych

Postulaty, które powinien spełniać dobrze skonstruowany metaprofil, są dokładnie takie same jak postulaty, które dotyczą dobrego profilu dokumentu. Poszczególne tematy hurtowni danych za pomocą metaprofili powinny być:

- dobrze reprezentowane – oddanie informacji gromadzonych w danym wycinku hurtowni danych,
- dobrze odróżnialne od innych tematyk,
- reprezentowane w sposób zwięzły (przez ograniczoną liczbę termów).

Poniżej prezentujemy sposoby na wyszukanie słów charakterystycznych dla poszczególnych tematów. Wykorzystywane są powszechnie znane statystyki, takie jak współczynnik koncentracji termu oraz odwrotna częstość dokumentu, które mogą również służyć do określenia wag wyszukanych słów.

4.3.1 Współczynnik koncentracji termu

Współczynnik koncentracji *i*-tego termu (tu deskryptora) można wyliczyć z poniższego wzoru:

$$\rho_i^{(y)} = \frac{\max^{(y)}\left(F^i\right)}{sum\left(F^i\right)}, \qquad (1)$$

gdzie:

F^i – wektor reprezentujący częstość *i*-tego deskryptora w poszczególnych tematach,

$\max^{(y)}(v)$ – suma y największych elementów wektora v,

$sum(v)$ – suma wszystkich elementów wektora v.

Jeśli współczynnik koncentracji *i*-tego deskryptora $\rho_i^{(y)}$ ma małą wartość, oznacza to, że deskryptor ten jest rozproszony po wielu tematach hurtowni danych. Z tego wynika też, że jest on mało różnicujący. Deskryptory takie nie są brane pod uwagę przy tworzeniu metaprofilu tematu. Duża wartość współczynnika koncentracji deskryptora oznacza, że jest on mocno skupiony w y tematach hurtowni danych. Zatem jest dla nich charakterystyczny.

Pozostaje jeszcze wyznaczenie wartości parametru y. Powinien on rosnąć wraz ze wzrostem liczby tematów w hurtowni. Nie można bliżej określić wartości tego parametru bez przeprowadzenia odpowiednich prób.

Zatem tworzenie profilu tematu hurtowni danych sprowadza się do znalezienia deskryptorów o dużych wartościach $\rho_i^{(y)}$ i ich włączeniu do odpowiednich profili.

4.3.2 Odwrotna częstość dokumentu

Idea tej metody opiera się na hipotezie, że do indeksowania szczególnie przydatne są deskryptory, które często występują w analizowanym dokumencie i rzadko we wszystkich pozostałych dokumentach. Teza ta jest prawdziwa dla zamkniętego zbioru dokumentów. Z uwagi na względną stabilność metadanych, hurtownię można uznać za zamknięty zbiór „dokumentów".

Zaletą tej metody są zadowalające wyniki i prosta implementacja. Wadą jest konieczność zmian wag w przypadku aktualizacji metadanych. Nie jest to problemem, gdyż właśnie doskonalenie profilu polega na podążaniu za zmianami w hurtowni.

Inspiracją były wzory wykorzystywane do automatycznego indeksowania dokumentów, przedstawione w (Salton, McGill, 1983, s. 63).

Przyjmuje się założenie, że znaczenie termu jest odwrotnie proporcjonalne do liczby tematów hurtowni danych, w których został znaleziony (S_i). Zmodyfikowany wzór:

$$IDF_i = \log_2 \frac{N}{S_i} + 1, \tag{2}$$

gdzie:

IDF_i – odwrotna częstość dokumentu *i*-tego deskryptora,

N – ogólna liczba tematów hurtowni danych,

S_i – liczba tematów hurtowni danych, w których wystąpił deskryptor *i*.

Drugim etapem indeksowania jest nadanie odpowiednich wag. Zmodyfikowany wzór do nadawania wag deskryptorom wybranym z metadanych hurtowni mógłby wyglądać następująco (również zaadaptowany z (Salton, McGill, 1983,)):

$$IDFW_i^k = F_i^k \cdot \left(\log_2 N - \log_2 S_i + 1\right), \tag{3}$$

gdzie:

$IDFW_i^k$ – waga *i*-tego deskryptora w *k*-tym temacie hurtowni,

F_i^k – liczba wystąpień deskryptora *i*-tego w *k*-tym temacie hurtowni.

Funkcja ta przypisuje duże znaczenie deskryptorom występującym tylko w niewielkiej liczbie tematów hurtowni danych.

4.4 Algorytm

Do utworzenia metaprofilu proponowany jest następujący algorytm:

- Wyszukanie słów (deskryptorów) w metadanych.
- Ograniczenie zbioru deskryptorów.
- Utworzenie metaprofili.

Z powodu ograniczonej ilości miejsca w artykule nie będziemy szerzej omawiać powyższego algorytmu. Szczegółowy opis można znaleźć w (Węcel).

5. Implementacja

Jako środowisko do budowy hurtowni danych wykorzystano produkt firmy SAS Institute. Jest to kompleksowy system służący do wspomagania podejmowania decyzji, na który składa się wiele modułów. SAS/Warehouse Administrator integruje dane i metadane, zatem nie ma potrzeby tworzenia osobnych repozytoriów. Dodatkowo SAS zawiera narzędzia umożliwiające publikację zawartości hurtowni danych w Internecie – SAS/Intrnet.

Struktura metadanych oraz funkcje umożliwiające modyfikowanie metadanych są opisane w podręczniku SAS/WA (SAS Institute Inc., 1997b).

System SAS zawiera wiele wbudowanych języków (np. 4GL, SCL). Z wielu powodów nie zdecydowaliśmy się jednak na implementację sugerowanych algorytmów wewnątrz systemu SAS. Za bardziej elastyczne rozwiązanie uznana została Java. Dzięki temu można uzyskać niezależność od systemu operacyjnego oraz producenta hurtowni danych. Dodatkowym atutem jest lepsze przystosowanie Javy do operacji na tekście oraz pracy w środowisku rozproszonym. Nie jest problemem dostęp do baz danych SAS z poziomu języka Java, gdyż SAS dostarcza odpowiedni sterownik JDBC (Java Database Connectivity) – SAS/Sharenet.

Z poziomu SAS/WA metadane mogą być eksportowane do pojedynczej tabeli. Tabela ta jest potem dostępna poprzez JDBC. Odpowiednia aplikacja pobiera eksportowane metadane i analizuje je wykorzystując algorytm przedstawiony w sekcji 4. W wyniku tego otrzymujemy profile hurtowni danych, zapisywane w zdefiniowanym formacie XML. Do obsługi plików

XML wykorzystywana jest biblioteka XML dostarczana przez Sun Microsystems w ramach Java Project X[1].

6. Zakończenie

Wzrost ilości informacji przekracza możliwości jej przetwarzania przez menedżerów. Z pomocą przychodzą im systemy informacyjne, których zadaniem jest wspieranie kierownictwa w podejmowaniu decyzji. Ostatnio popularnym elementem takich systemów są hurtownie danych. Samo wykorzystanie danych liczbowych z hurtowni danych nie wystarczy jednak do zdobycia przewagi konkurencyjnej.

Niniejszy artykuł jest częścią większego projektu, którego celem jest zasilanie hurtowni danych informacjami nieustrukturalizowanymi. Zostało tu jedynie przedstawione tworzenie pierwszego profilu na podstawie metadanych. Z powodu ograniczeń objętościowych artykułu nie było możliwe bardziej szczegółowe rozwinięcie wielu interesujących aspektów ciągle rozwijanego modelu. Jeśli chodzi o tworzenie pierwszych profili, dalsze badania powinny brać pod uwagę wykorzystanie technik text mining.

Do tej pory nic nie zostało napisane o modyfikacji profili. Przez formułowanie profili rozumiemy nie tylko utworzenie pierwszego profilu, ale także jego ciągłe doskonalenie. Takie metody obecnie istnieją i powstały jako rozszerzenie funkcjonalności systemu HyperSDI. Wykorzystywana jest informacja zwrotna od użytkowników oraz techniki uczenia podczas filtrowania. Potrzebne są również nowe narzędzia do zarządzania użytkownikami oraz zadaniami w środowisku hurtowni danych.

Innym ważnym aspektem jest standaryzacja. Większa interoperacyjność może być osiągnięta poprzez standaryzację metadanych. Obecnie jest już dostępnych kilka metod umożliwiających współdzielenie metadanych pomiędzy różnymi aplikacjami. Pojawiło się kilka standardów, wiele firm pracuje nad własnym podejściem do wymienialności metadanych (np. Common Warehouse Metadata Interchange, Metadata Coalitions). Również wykorzystanie XML do zapisu profili ułatwia dostęp do nich z różnych aplikacji. Dzięki temu tworzenie różnych filtrów informacyjnych, korzystających ze wspólnych profili, powinno być łatwiejsze.

---

[1] Sun Microsystems, *Java Technology and XML*, http://java.sun.com/xml/

Literatura

Abramowicz W. (1990) Information Dissemination to Users with Heterogenous Interests, J. Grabowski (ed.), *Computers in Science and Higher Education, Mathematical Research,* Vol. 57, Akademie-Verlag, Berlin 1990, s. 62-71.

Abramowicz W. (1991) Hypertexte und ihre IR-basierte Verbreitung, *Humboldt Universität, Fachbereich Informatik*, Berlin, 292+VII.

Abramowicz W., Kalczyński P., Węcel K. (2000) Information Filters Supplying Data Warehouses with Benchmarking Information, *W. Abramowicz, J. Zurada (eds.), Selected Aspects of Knowledge Discovery in Business Information Systems.*

Ceglarek D. (1997) Applying Taxonomous Methods in Selective Distribution of Information (SDI) Systems Supplying Users with Business Information, *Rozprawa doktorska, Akademia Ekonomiczna w Poznaniu*, Wydział Ekonomii.

Greiff, W.R. (1998) A Theory of Term Weighting Based on Exploratory Data Analysis, *21st International Conference on Research and Development in Information Retrieval.*

Housman, E.M., Kaskela E.D. (1970) State of the art in Selective dissemination of information, *IEEE Transactions on Engeeneering Writing and Speech*, Vol. 13, s. 78-83.

Inmon, W.H. Hackathorn R.D., (1994) Using the Data Warehouse, *John Wiley & Sons Inc.*, New York.

Kimball, R. (1996) The Data Warehouse Toolkit - Practical Techniques for Building Dimensional Data Warehouses, *John Wiley & Sons Inc.*, New York.

Luhn, H.P. (1958) A Business Intelligence System, *IBM Journal of Research and Development*, Vol. 2, No. 2, s. 159-165.

Moncla, B. (1999) Business Meta Data Integration, *DM Review*, September 1999, http://www.dmreview.com/.

van Rijsbergen, C.J. (1979) Information Retrieval, *Butterworths*, London, http://www.dcs.gla.ac.uk/Keith/Preface.html.

Salton, G., McGill M.J. (1983) Introduction to Modern Information Retrieval, *McGraw-Hill Book Company.*

SAS Institute Inc., (1997) SAS/Warehouse Administrator™ User's Guide, Release 1.1, First Edition, *Cary, NC, SAS Institute Inc.*, 142 pp.

SAS Institute Inc., (1997) SAS/Warehouse Administrator™ Metadata API Reference, Release 1.2, *Cary, NC: SAS Institute Inc.* 86 pp.

Singh, H. (1999) Interactive Data Warehousing, *Prentice Hall.*

Sprague, R.J., Freudenreich L.B. (1978) Building Better SDI Profiles for Users of Large, Multidisciplinary Data Bases, *Journal of the American Society for Information Science, John Wiley & Sons*, November 1978, Vol. 29, No. 6, s. 278-282.

Węcel, K., Odkrywanie wiedzy dla doskonalenia profili HyperSDI w hurtowniach danych, *Praca magisterska, Akademia Ekonomiczna w Poznaniu*, Wydział Ekonomii.

ISSN 0208-8029
ISBN 83-85847-53-7

W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy prosimy o kontakt z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 837-35-78 w. 241 e-mail: bibliote@ibspan.waw.pl