

dr hab. Dariusz Put¹

Katedra Systemów Obliczeniowych, Wydział Zarządzania
Uniwersytet Ekonomiczny w Krakowie

Projektowanie systemów integracyjnych w oparciu o współdzieloną ontologię

WSTĘP

Współcześnie projektowane systemy zintegrowane są zwykle oparte na centralnej bazie danych. Tego typu rozwiązania są mało elastyczne i skalowalne, a modyfikacja ich struktury, konieczna po pewnym okresie ich eksploatacji, jest czasochłonna i angażuje zasoby organizacji. Celowe wydaje się więc prowadzenie badań nad opracowaniem rozwiązań, w których możliwa będzie dynamiczna modyfikacja struktury repozytorium zasobów informacyjnych, nieskomplikowane dołączanie nowych źródeł danych oraz formułowanie zapytań *ad hoc*. Na potrzebę poszukiwania systemów integracyjnych, stanowiących alternatywę dla rozwiązań opartych na centralnej bazie danych, wskazują m.in. M. Fernandez i in. [1998, 2000], proponując system STRUDEL do zarządzania informacjami zamieszczanymi na stronach internetowych. T. Novotný [2007] zauważa potrzebę poszukiwania rozwiązań dla integracji danych semistrukturalnych. J. Hammer, J. McHugh, H. Garcia-Molina [1997] przedstawiają propozycję modelu do zarządzania tego typu danymi w postaci systemu TSIMMIS. Wykorzystanie ontologii do zarządzania zasobami informacyjnymi proponują V. Katifori i in. [2005] opisując rozwiązanie o nazwie OntoPIM, a także A. Maedche i in. [2002] oraz H. Xiao, I.F. Cruz [2005]. Z kolei H. Kozankiewicz, K. Stencel, K. Subieta [2005] prezentują system modułowy oparty na warstwie pośredniej.

Celem artykułu jest omówienie pożądaných własności systemu integracyjnego oraz propozycja modelu integracji heterogenicznych zasobów informacyjnych opartego na współdzielonej ontologii, posiadającego zidentyfikowane własności. W artykule omówiono architekturę modelu, zadania realizowane przez poszczególne komponenty w procesie integracji, sposób formułowania i wykonywania zapytań oraz proces projektowania systemu opartego na proponowanym modelu.

¹ Adres korespondencyjny: Uniwersytet Ekonomiczny w Krakowie, ul. Rakowicka 27, 31-510 Kraków; e-mail: putd@uek.krakow.pl; tel. +4812 2935265.

SYSTEMY INTEGRACYJNE JAKO ALTERNATYWA DLA SYSTEMÓW ZINTEGROWANYCH

We współczesnych systemach transakcyjnych, włączając rozwiązania klasyczne oparte na centralnej bazie danych, nie występuje problem heterogeniczności: dane mają jednolitą budowę, struktury i atrybuty posiadają jednoznaczne nazwy, zasady normalizacji i reguły integralności minimalizują redundancję, nie ma potrzeby reformułowania zapytań, a dane pobrane z systemów składowych nie muszą być łączone. Głównymi wadami takich systemów są brak elastyczności, skalowalności, możliwości formułowania zapytań *ad hoc* oraz konieczność modyfikacji systemu w odpowiedzi na zachodzące zmiany. Celowe wydaje się poszukiwanie bardziej adekwatnych rozwiązań, które będą miały własność elastycznego i dynamicznego dostosowania struktury repozytoriów zasobów informacyjnych w odpowiedzi na zmiany zachodzące w otoczeniu organizacji.

W systemach integrujących heterogeniczne zasoby informacyjne sytuacja jest odmienna. W tego typu rozwiązaniach występuje wieloaspektowa różnorodność: modeli danych, systemów bazodanowych, języków zapytań, kategorii i form zasobów informacyjnych, nazw instancji i atrybutów, metod modyfikacji zasobów. Istnieje wiele propozycji rozwiązania niektórych z wymienionych problemów, w tym głównie sposobu projektowania systemów integrujących (zob. [Chawathe i in., 1994; Dittrich i in., 2007; Fernandez i in., 1998, 2000; Hammer, McHugh, Garcia-Molina, 1997; Katifori V. i in., 2005; Kozankiewicz, Stencel, Subieta, 2005; Lahiri, Abiteboul, Widom, 1999; Maedche i in., 2002; McHugh i in., 1997; Novotný, 2007; Papakonstantinou, Garcia-Molina, Widom, 1995; Stenbit, 2003; Xiao, Cruz, 2005]). Wynika z nich m.in., że model będący podstawą budowy systemów integrujących musi posiadać komponenty odpowiedzialne za przeprowadzenie tego złożonego procesu, pozwalające na nieskomplikowane wyszukiwanie i wybieranie potrzebnych zasobów informacyjnych. Systemy integracyjne muszą realizować następujące zadania składające się na proces integracji w heterogenicznym środowisku rozproszonym:

- komunikacja z zewnętrznymi repozytoriami danych;
- reformułowanie zapytań w języku zaimplementowanym w systemie integrującym na zapytania wykonywane w systemach składowych;
- łączenie i unifikacja zasobów pobranych z niejednorodnych systemów;
- udostępnienie metainformacji o współdzielonych zasobach informacyjnych umożliwiającej tworzenie zapytań *ad hoc*;
- podział zadania integracji na możliwie niezależne podzadania;
- udostępnienie różnorodnych form wizualizacji zasobów informacyjnych.

Opracowanie rozwiązania do zarządzania rozproszonymi i niejednorodnymi zasobami informacyjnymi jest bardziej złożone, niż w przypadku tradycyjnych systemów opartych na centralnej bazie danych. Biorąc pod uwagę zadania, jakie

muszą być realizowane w tego typu systemach opartych na współdzielonej ontologii, można wskazać następujące ich własności:

- łączą niezależne systemy składowe (podsystemy);
- dostarczają użytkownikom aktualnych lub historycznych informacji;
- jednym z ich składników jest współdzielona, automatycznie modyfikowalna ontologia zawierająca metainformacje o integrowanych zasobach;
- umożliwiają zarówno tworzenie zapytań predefiniowanych, zidentyfikowanych w procesie projektowania systemu, jak i zapytań *ad hoc*;
- proces tworzenia zapytań jest możliwie nieskomplikowany;
- język zapytań jest maksymalnie uproszczony;
- język zapytań jest możliwie elastyczny i umożliwia tworzenie precyzyjnych zapytań;
- pozwalają na integrację wszystkich istniejących kategorii informacji;
- mogą być wykorzystywane do integracji wewnątrz i międzyorganizacyjnej;
- są skalowalne, dzięki czemu możliwe będzie dołączenie nowych systemów składowych podczas eksploatacji;
- nie wpływają na efektywność działania integrowanych systemów;
- charakteryzują się łatwością projektowania, implementacji i konfiguracji;
- posiadają warstwowo-modułową strukturę.

Zaprojektowanie systemu posiadającego wszystkie powyższe własności jest zadaniem złożonym, choćby ze względu na fakt, że niektóre z wymienionych postulatów są ze sobą sprzeczne (np. elastyczność języka zapytań i jednoczesna jego prostota, nieskomplikowana struktura w sytuacji złożoności zadania integracji). Przystępując do realizacji tego typu przedsięwzięcia należy się przygotować na konieczność rezygnacji z niektórych własności.

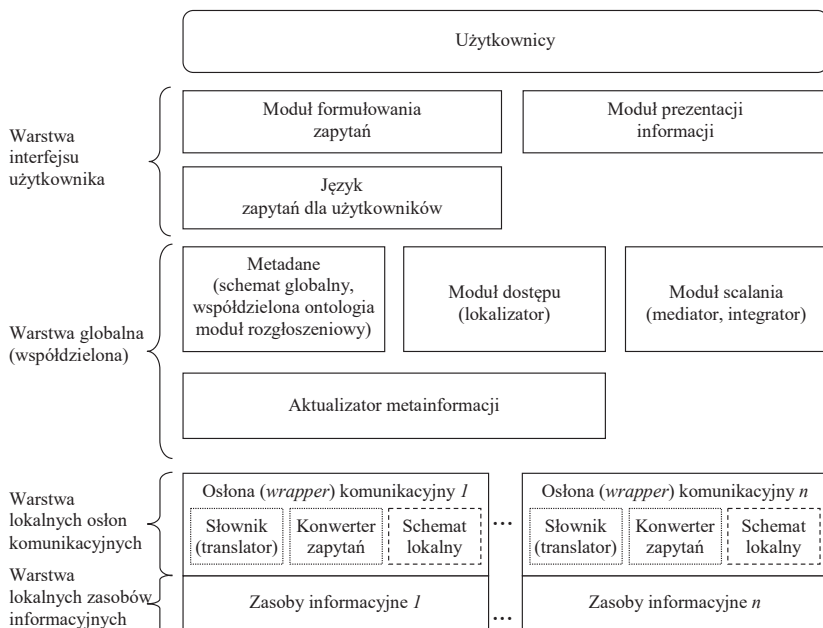
ARCHITEKTURA SYSTEMU I ZADANIA REALIZOWANE PRZEZ POSZCZEGÓLNE KOMPONENTY

W systemie integrującym zasoby pochodzące z wielu heterogenicznych źródeł, w którym istnieje możliwość definiowania zapytań przez użytkowników, należy uwzględnić konieczność realizacji zadań, które nie występują w systemach tradycyjnych. Ponieważ problem integracji jest zadaniem wieloaspektowo złożonym, dobrym rozwiązaniem jest jego podział na podzadania. Proponowany model integracji heterogenicznych rozproszonych zasobów informacyjnych HIRIM (*Heterogeneous Information Resources Integrating Model*) składa się z czterech warstw, w których zdefiniowano moduły realizujące zadania cząstkowe w procesie integracji. Są to warstwy (rys. 1):

- lokalnych zasobów informacyjnych;
- lokalnych osłon (*wrapperów*) komunikacyjnych;

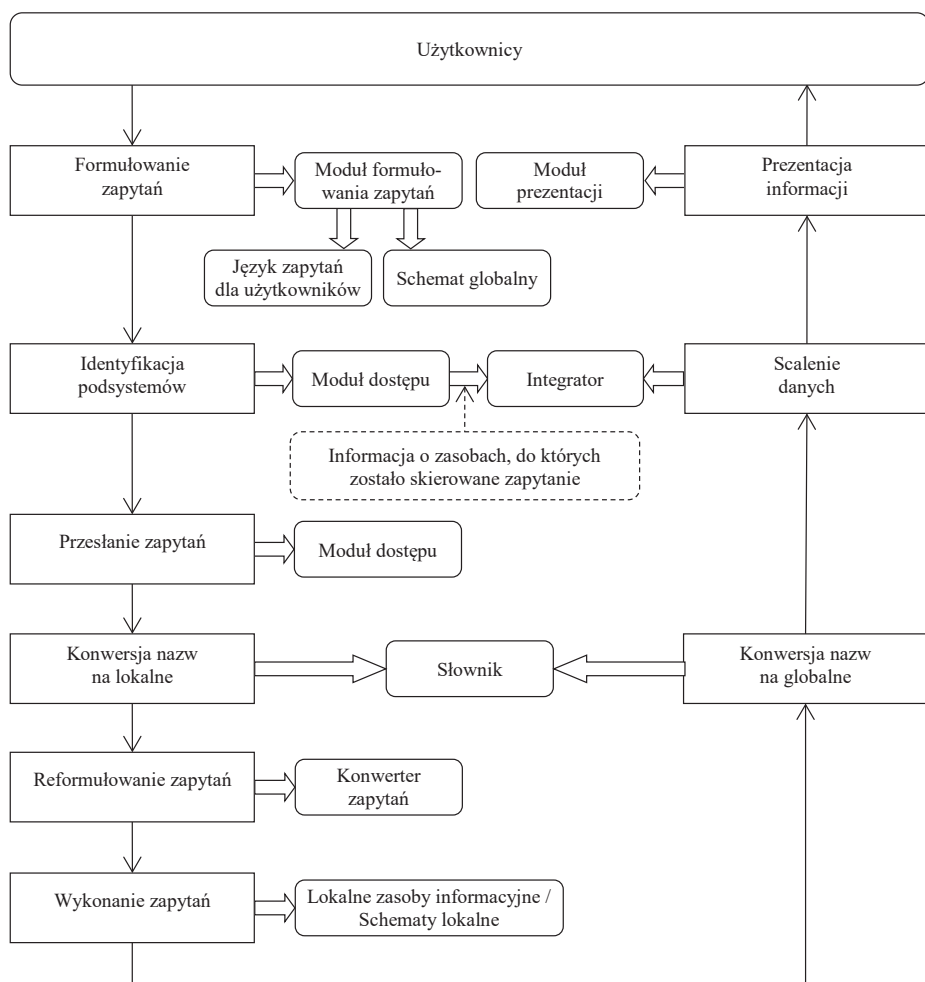
- globalna (współdzielona);
- interfejsu użytkownika.

Proces przetwarzania zapytań oraz zadania realizowane przez poszczególne komponenty modelu w tym procesie zobrazowano na rys. 2. Użytkownicy, tworząc zapytania, wykorzystują moduł formułowania zapytań, korzystając z aktualnej zawartości współdzielonej ontologii. Następnie moduł dostępu identyfikuje systemy składowe posiadające poszukiwaną informację reprezentowaną przez koncepty wybrane przez użytkownika w procesie formułowania zapytań. Moduł ten kieruje zapytania tylko do wybranych w ten sposób systemów składowych. W kolejnym etapie słowniki znajdujące się w poszczególnych systemach składowych zmieniają nazwy instancji i atrybutów na lokalne ekwiwalenty, a konwertery zapytań tłumaczą zapytania na język zapytań wykorzystywany w danym systemie składowym. Następnie zapytania są wykonywane w systemach lokalnych. Tak wybrane zasoby informacyjne są przesyłane do lokalnych słowników, które dokonują konwersji nazw instancji i atrybutów do postaci globalnej. W takiej formie zasoby informacyjne są przesyłane do integratora w warstwie globalnej. Po otrzymaniu informacji od wszystkich podsystemów integrator łączy zasoby informacyjne i przesyła do systemu użytkownika, który sformułował zapytanie. Tam moduł prezentacji informacji przygotowuje informacje dla użytkownika od strony wizualizacyjnej i prezentuje je w wybranej formie.



Rys. 1. Architektura modelu HIRIM

Źródło: opracowanie własne.



Rys. 2. Proces wykonywania zapytań w systemie opartym na modelu HIRIM

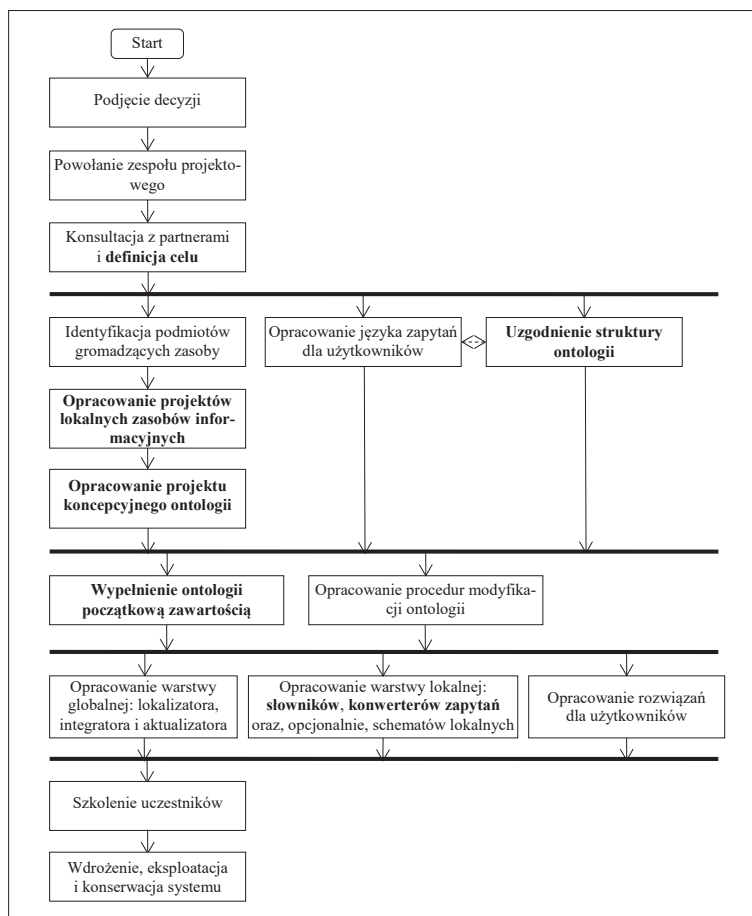
Źródło: opracowanie własne.

PROJEKTOWANIE SYSTEMU INTEGRACYJNEGO OPARTEGO NA PROPONOWANYM MODELU

Współdzielona ontologia może być modyfikowana, a więc może być pusta bezpośrednio po wdrożeniu, jednakże wypełnianie jej, podczas projektowania, metainformacją o dostępnych zasobach sprawi, że bezpośrednio po wdrożeniu system będzie spełniał swoją funkcję. Współpraca przedstawicieli właścicieli systemów składowych umożliwi także uzgodnienie zasad tworzenia i modyfikacji

oraz struktury ontologii. Proces projektowania systemu opartego na modelu HI-RIM powinien się składać z następujących etapów (rys. 3):

1. Podjęcie decyzji o stworzeniu systemu integracyjnego. Proces projektowania, a następnie wdrożenia systemu, nie będzie miał wpływu na dotychczasową działalność organizacji, jednak jego przeprowadzenie, wdrożenie oraz utrzymanie będą angażować zasoby organizacji.
2. Powołanie zespołu projektowego. Zespół projektowy przydziela zadania, czuwa nad całością prac i jest odpowiedzialny za przeprowadzenie projektu.
3. Konsultacja z partnerami na temat możliwości ich uczestnictwa w przedsięwzięciu. Istotne jest uzgodnienie, które informacje będą współdzielone.
4. Identyfikacja podmiotów gromadzących zasoby na własne potrzeby. Należy zbierać informacje o działach, a także poszczególnych pracownikach, którzy gromadzą dane i informacje mogące być przedmiotem współdzielenia.
5. Opracowanie języka zapytań o maksymalnie nieskomplikowanej składni, aby możliwe było formułowanie zapytań *ad hoc*.
6. Uzgodnienie struktury ontologii, zdefiniowanie ram dotyczących jej budowy.
7. Opracowanie projektów lokalnych zasobów informacyjnych. Identyfikowane są koncepty, atrybuty oraz powiązania między nimi.
8. Opracowanie projektu koncepcyjnego współdzielonej ontologii. Zbiór wszystkich projektów koncepcyjnych stanowi podstawę do opracowania współdzielonej ontologii.
9. Wypełnienie ontologii początkową zawartością. Należy stworzyć projekt logiczny, a następnie fizyczny metadanych. Projekt ten musi mieścić się w ramach uzgodnionej struktury.
10. Opracowanie procedur modyfikacji ontologii.
11. Opracowanie aplikacji działających w warstwie globalnej.
12. Opracowanie warstwy lokalnych osłon komunikacyjnych. Administratorzy podsystemów przygotowują rozwiązania umożliwiające dostęp do lokalnie przechowywanych zasobów współdzielonych. Na wejściu do warstwy lokalnej znajdują się zapytania formułowane przez użytkowników, a na wyjściu współdzielone zasoby informacyjne organizacji odpowiadające zapytaniu.
13. Opracowanie rozwiązań dla użytkowników: metody formułowania zapytań oraz aplikacji odpowiedzialnych za wizualizację zasobów otrzymanych z systemu.
14. Szkolenie. Powinno ono obejmować omówienie procedur korzystania z języka zapytań dla użytkowników, modułu wizualizacji informacji, a także możliwość predefiniowania zapytań i zapisywania ich w systemie do ponownego wykorzystania.
15. Wdrożenie, eksploatacja i konserwacja systemu. Proces wdrożenia nie ma wpływu na dotychczasowe działanie systemów składowych. Pojawia się jedynie nowa funkcjonalność – możliwość formułowania zapytań do systemu współdzielonego.



Rys. 3. Proces projektowania systemu opartego na modelu HIRIM

Źródło: opracowanie własne.

PODSUMOWANIE

Systemy integracyjne stanowią alternatywę dla rozwiązań zintegrowanych opartych na centralnej bazie danych. Posiadają szereg własności, które decydują o ich przydatności do integracji często modyfikowanych, heterogenicznych zasobów informacyjnych. Proces ich projektowania jest złożony, ale prawidłowo przeprowadzony może dać organizacji wymierne korzyści w zakresie zarządzania własnymi zasobami informacyjnymi oraz tworzenia rozwiązań umożliwiających automatyczną wymianę dokumentów i informacji z partnerami biznesowymi. W artykule starano się zidentyfikować pożądane własności modeli integracyjnych

oraz zaproponowano model składający się z czterech warstw zbudowanych z modułów realizujących zadania cząstkowe składające się na proces integracji. Taka modułowa budowa systemu opartego na proponowanym modelu ułatwi zarówno jego tworzenie i implementację, jak i późniejszą eksploatację i modyfikację. Ze względu na to, że proces wykonywania zapytań w tego typu rozwiązaniu jest bardziej złożony niż w przypadku systemów tradycyjnych, zaproponowano metodę prowadzącą do pobierania danych z rozproszonego i heterogenicznego środowiska, jaki stanowią systemy źródłowe będące przedmiotem integracji. Omówiono także proces projektowania systemu opartego na proponowanym modelu.

BIBLIOGRAFIA

- Chawathe S. i in., 1994, *The TSIMMIS Project: Integration of Heterogeneous Information Sources* [w:] *Proceedings of IPSJ Conference*, Tokio.
- Dittrich J.-P. i in., 2007, *From Personal Desktops to Personal Dataspace: A Report on Building the iMeMex Personal Dataspace Management System*, GI-Fachtagung für Datenbanksysteme in Business, Technologie und Web (BTW), Aachen, Germany.
- Fernandez M. i in., 1998, *Catching the Boat with STRUDEL: Experiences with a Web-Site Management System*, SIGMOD Conference, <http://dx.doi.org/10.1145/276304.276341>.
- Fernandez M. i in., 2000, *Declarative Specification of Web Sites with STRUDEL*, "VLDB Journal", 9 (1).
- Hammer J., McHugh J., Garcia-Molina H., 1997, *Semistructured Data: The TSIMMIS Experience*, Proceedings of the ADBIS'97 Conference, St. Petersburg, Russia.
- Katifori V. i in., 2005, *OntoPIM: How to Rely on a Personal Ontology for Personal Information Management*, Proceedings of the First Workshop on the Semantic Desktop, International Semantic Web Conference, Galway, Ireland.
- Kozankiewicz H., Stencel K., Subieta K., 2005, *Intelligent Data Integration Middleware Based on Updateable Views*, Springer Berlin/Heidelberg, 3490/2005, http://dx.doi.org/10.1007/11558637_4.
- Lahiri, T., Abiteboul, S., Widom, J., 1999, *Ozone: Integrating Structured and Semistructured Data* [w:] *Proceedings of the Seventh International Conference on Database Programming Languages*, 297–323, Kinloch Rannoch, Scotland, http://dx.doi.org/10.1007/3-540-44543-9_18.
- Maedche A. i in., 2002, *SEAL – Tying up Information Integration and Web Site Management by Ontologies*, Technical Report, Institute AIFB, University of Karlsruhe, Germany.
- McHugh J. i in., 1997, *Lore: A Database Management System for Semistructured Data*, SIGMOD Record, 26 (3), <http://dx.doi.org/10.1145/262762.262770>.
- Novotný T., 2007, *A Content-Oriented Data Model for Semistructured Data* [w:] *Proceedings of the Dateso 2007 Workshop*, red. Pokorný J., Snášel V., Richta K, Amphora Research Group, Czech Republic.
- Papakonstantinou Y., Garcia-Molina H., Widom J., 1995, *Object Exchange across Heterogeneous Information Sources* [w:] *Proceedings of the IEEE International Conference on Data Engineering*, Taipei, Taiwan, <http://dx.doi.org/10.1109/ICDE.1995.380386>.

- Stenbit J.P., 2003, *DoD Net-centric Data Strategy*, Department of Defense, Washington.
- Xiao H., Cruz I.F., 2005, *A Multi-Ontology Approach for Personal Information Management* [w:] *Proceedings of 4-th International Semantic Web Conference*, 6–10.11.2005, Galway, Irland.

Streszczenie

Artykuł dotyczy zagadnienia projektowania elastycznych, skalowalnych i przyjaznych dla użytkownika systemów integrujących rozproszone, heterogeniczne zasoby informacyjne. W artykule starano się wykazać, że opracowanie takiego systemu jest bardziej złożone niż zaprojektowanie rozwiązania tradycyjnego opartego na centralnej bazie danych. Omówiono architekturę modelu integracyjnego oraz proces projektowania systemu opartego na tym modelu. Implementacja takiego rozwiązania może przynieść organizacji szereg korzyści: łatwiejszy dostęp do wielu źródeł informacji, dostęp dla uprzednio zdefiniowanych, jak i niezdefiniowanych użytkowników, możliwość formułowania zapytań *ad hoc*, standaryzację dostępu do zasobów informacyjnych, dostęp do informacji przechowywanych w różnych systemach w czasie rzeczywistym, ułatwienie kooperacji pomiędzy partnerami biznesowymi. Jednak proces przygotowania systemu integrującego wymaga opracowania modułów realizujących poszczególne zadania cząstkowe składające się na proces integracji w niejednorodnym i rozproszonym środowisku, m.in.: opracowanie przyjaznego dla użytkownika języka zapytań i sposobu formułowania zapytań, opracowanie struktury ontologii i metod jej modyfikacji, przygotowania słowników, konwerterów zapytań czy integratorów.

Słowa kluczowe: integracja zasobów informacyjnych, modele integracyjne, systemy integracyjne, proces integracji zasobów informacyjnych, ontologie

Designing integrating systems based on co-shared ontology

Summary

In the paper the problem of designing flexible, scalable and user-friendly systems integrating distributed and heterogeneous information resources is described. It is justified that the elaboration of such systems is more complex than in case of traditional solutions based on central database. The example of the architecture of integrating model and the process of the creation of a system based on the model are discussed. Implementation of such integrating system may give a considerable number of benefits to organisation: easier access to various information resources, access for previously defined and undefined users, the possibility to formulate *ad hoc* queries, standardisation of an access to information resources, an access to data and information stored in various systems in real time, the facilitation of cooperation between business partners. However, the process requires elaboration of modules that perform individual tasks constituting the whole process of searching information resources in heterogeneous and distributed sources, e.g.: elaboration of a user query language and the method of queries formulation, establishing the structure of the ontology and methods of its on-line modification, preparing dictionaries, queries converters and integrator module.

Keywords: information resources integration, integrating models, integrating systems, information resources integration process, ontologies

JEL: D83, L15, L86