

MOMIS: un sistema di Description Logics per l'integrazione del sistema informativo d'impresa

D. Beneventano^{1,2}, S. Bergamaschi^{1,2}, A. Corni^{1,2} and M. Vincini¹

(1) Università di Modena e Reggio Emilia (2) CSITE-CNR Bologna
DSI - Via Campi 213/B V.le Risorgimento, 2
41100 Modena 40136 Bologna

e-mail: [domenico.beneventano, sonia.bergamaschi, alberto.corni, maurizio.vincini]@unimo.it

Nel corso degli ultimi anni le imprese si sono via via dotate di sistemi per l'archiviazione delle informazioni costruendo sistemi informativi contenenti dati correlati tra loro ma spesso ridondanti, eterogenei e non sempre consistenti. D'altra parte l'esplosione delle reti di calcolatori ha accelerato l'esigenza di condivisione e del reperimento delle informazioni. Si sente la necessità di accedere a tutte le informazioni aziendali con strumenti omogenei e costruire applicazioni che utilizzino tali informazioni come se si lavorasse su di un'unica base di dati. Ad esempio consideriamo un'azienda manifatturiera nel campo dei personal computer che ha acquisito due nuove divisioni e che si vuole dotare di un controllo centralizzato della produzione, in particolare consideriamo la gestione delle scorte nei vari magazzini (Fig. 1.a). Una delle due nuove divisioni acquisite gestisce i prodotti grazie ad una base di dati semistutturata (MS) caratterizzata dalla capacità di gestire ordini a fornitori, l'altra divisione è caratterizzata dal possedere un magazzino automatizzato MR che è gestito grazie ad un DBMS relazionale. La terza sorgente, che chiameremo MO, è una base di dati ad oggetti che mantiene la lista di tutti i prodotti trattati dalla multinazionale che ha acquisito i magazzini MS e MR.

Da questo semplice esempio si può intuire che per creare una *vista* unica che permetta un accesso trasparente alle sorgenti occorre affrontare problemi dovuti ad eterogeneità strutturali e applicative, nonché alla mancanza di una ontologia comune, che porta a differenze semantiche tra le fonti di informazione. In questo lavoro proponiamo un sistema intelligente per l'integrazione delle informazioni chiamato MOMIS (Mediator environment for Multiple Information Sources, [1], [2]) che permette di costruire ed interagire con una vista integrata delle sorgenti.

Come altri sistemi proposti in letteratura, il sistema MOMIS segue un *approccio semantico* proponendo l'integrazione a partire dagli schemi concettuali ed utilizzando il *mediatore* per l'esecuzione delle query. A differenza di altri progetti simili, MOMIS fornisce un effettivo supporto al *progettista dell'integrazione* grazie allo strumento *SI-Designer*, un'interfaccia grafica che guida il progettista nel percorso di integrazione. MOMIS è stato progettato per fornire un accesso integrato ad informazioni eterogenee memorizzate sia in database di tipo tradizionale (e.g. relazionali, object-oriented) o file system, sia in sorgenti di tipo semistutturato.

Seguendo l'architettura di riferimento di MOMIS (Fig. 1.b) si possono distinguere i componenti principali Mediatore (formato dai moduli *Global Schema*, *Query Manager*), *Wrappers* e strumenti esterni a MOMIS (WordNet e ODB-Tools). Un importante componente di MOMIS è *SI-Designer*, l'interfaccia di interazione con l'utente nella fase di creazione della vista globale (lo strumento per il clustering *ARTEMIS* è stato integrato in *SI-Designer*). Tutti questi componenti comunicano tra loro mediante CORBA.

Allo scopo di supportare l'integrazione semantica di informazioni introduciamo un linguaggio object-oriented, chiamato ODL_{I^3} , per la rappresentazione di schemi concettuali che descrivono le sorgenti dati semistutturate e non. ODL_{I^3} riprende le specifiche del linguaggio ODL (dell'ODMG) introducendo estensioni utili per modellare sorgenti semistutturate e informazioni utilizzate dal motore inferenziale

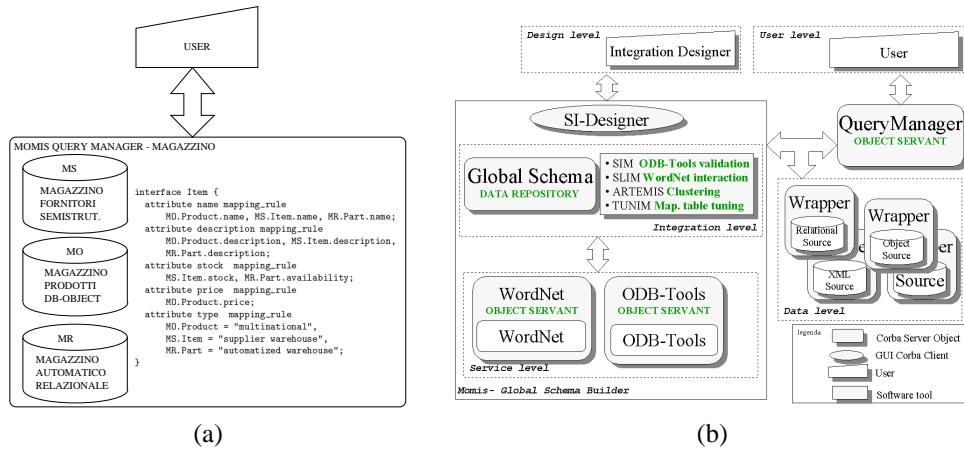


Figura 1: (a) Esempio di applicazione di MOMIS, (b) Architettura del sistema MOMIS

basato su *Description Logics* (ODB-Tools [3]). ODL_{T3} è il linguaggio unico utilizzato da MOMIS, ed è utilizzato, ad esempio, dai Wrappers per presentare la descrizione delle sorgenti a cui sono collegati.

Il primo passo nell'estrazione di informazioni è la costruzione di una rappresentazione semantica ricca delle sorgenti da integrare attraverso un modello di dati comune. Nell'approccio semantico ciò viene realizzato descrivendo in ODL_{T3} lo schema concettuale delle singole sorgenti. Per quanto riguarda le sorgenti strutturate convenzionali (database relazionali, object-oriented, file) la descrizione è disponibile ed è automaticamente convertita in ODL_{T3} dai wrappers (Fig. 1.b). Per le sorgenti semistrustrate (oggi lo standard de facto è XML) non è disponibile un vero e proprio schema, tuttavia, MOMIS, interagendo con il progettista, tenta di estrarre dalla struttura della sorgente il maggior numero di informazioni possibile, mantenendo quelle informazioni tipiche delle sorgenti semistrustrate (ad esempio l'opzionalità di attributi).

Per ottenere l'ontologia alla base dell'integrazione vengono utilizzate tecniche di ragionamento delle *Description Logics* (DL_S) ed il sistema lessicale *WordNet* per costruire un *Common Thesaurus* [4]. Il progettista può comunque definire relazioni (intensionali ed estensionali) mediante un'apposita interfaccia grafica. Le relazioni e la conoscenza degli schemi vengono poi espresse in linguaggio OLC (Object Language with Complements allowing Descriptive cycles) ed elaborate tramite un motore inferenziale basato su DL_S per la validazione e l'inferenza di nuove relazioni.

Nella fase del *clustering* si ottiene una proposta di schema globale integrato partendo dalla conoscenza presente nel *Common Thesaurus*. Vengono identificate in modo semi-automatico classi semanticamente simili e raggruppate in cluster. Il clustering avviene in base alle affinità tra classi, l'affinità tra due classi, valutata dal modulo ARTEMIS, dipende dal numero ed il tipo di relazioni di thesaurus che riguardano le due classi.

Lo schema globale costituisce il punto di accesso ai dati delle sorgenti integrate da parte degli utenti di MOMIS. Le classi dello schema sono definite a partire dai cluster presenti nell'*albero di affinità*; la generazione di tali classi (denominate g_{ci}) è interattiva e supportata dal modulo Global Schema Builder. A questo punto MOMIS ha la conoscenza sufficiente per rispondere a query eseguite sullo schema integrato. Quando un utente invia una query a MOMIS il modulo mediatore *Query Manager* (QM) produce un insieme di sotto-query da inviare alle sorgenti coinvolte dalla query. Ciò avviene attraverso le fasi di *ottimizzazione semantica* e *formulazione del piano di interrogazione*.

In conclusione, il sistema MOMIS supporta il progettista durante tutto il processo di integrazione del sistema informativo aziendale, partendo dall'acquisizione degli schemi preesistenti, permettendo la definizione di relazioni tra le varie strutture dati, definendo un cluster di classi semanticamente simili dal quale generare lo schema virtuale globale che rappresenta la visione unica del patrimonio informativo aziendale. L'altro compito fondamentale di MOMIS è quello di rendere trasparente l'accesso ai dati, permettendo all'utente di interagire direttamente con lo schema globale virtuale.

Riferimenti bibliografici

- [1] S. Bergamaschi, S. Castano, A. Corni, R. Guidetti, G. Malvezzi, M. Melchiori, and M. Vincini. Information integration: the MOMIS project demonstration. In *26th VLDB: Int. Conf. On Very Large Databases*, Cairo, Egypt, September 2000.
- [2] S. Bergamaschi, S. Castano, and M. Vincini. Semantic integration of semistructured and structured data sources. *SIGMOD Record Special Issue on Semantic Interoperability in Global Information*, 28(1), March 1999.
- [3] Domenico Beneventano, Sonia Bergamaschi, Claudio Sartori, and Maurizio Vincini. ODB-QOPTIMIZER: A tool for semantic query optimization in oodb. In *Int. Conference on Data Engineering - ICDE97*, 1997. <http://sparc20.dsi.unimo.it>.
- [4] D. Beneventano, A. Corni S. Bergamaschi, R. Guidetti, and G. Malvezzi. SI-Designer: un tool di ausilio all'integrazione intelligente di sorgenti di informazione. In *SEBD: Sistemi Evoluti per Basi di Dati*, Palazzo dell'Emiciclo L'Aquila, Italy, June 2000.