

---

# WISDOM

*Programma di ricerca (cofinanziato dal MIUR, esercizio 2004)*  
**Ricerca Intelligente su Web basata su Ontologie di Dominio**  
**Web Intelligent Search based on DOMain ontologies**

---

## **Rapporto sulla architettura metodologica e funzionale di riferimento**

S. BERGAMASCHI, D. BENEVENTANO, L. CABIBBO, P. CIACCIA, V. CRESCENZI, P. MERIALDO

**D0.R1**

**20 Giugno 2005**

### **Sommario**

L'enorme quantità di dati e la crescente disponibilità di servizi presenti su Web rendono sempre più importante lo sviluppo di infrastrutture e sistemi software che permettano ai clienti collegati alla rete di "ricaricarsi" di dati di interesse per i propri bisogni informativi. Il progetto WISDOM ha come obiettivo principale la definizione di tecniche, e strumenti per la ricerca, la localizzazione e la fruizione personalizzata di risorse informative disponibili su Web. Le soluzioni proposte sono basate su ontologie di dominio, e su una architettura distribuita e decentralizzata. Obiettivo di questo documento è la definizione di una architettura metodologica e funzionale di riferimento al fine di garantire coerenza tra le soluzioni che verranno studiate dalle varie unità di ricerca che partecipano al progetto.

<b>Tema</b>	Comune ai tre temi
<b>Codice</b>	D0.R1
<b>Data</b>	20 Giugno 2005
<b>Tipo di prodotto</b>	Rapporto tecnico
<b>Numero di pagine</b>	8
<b>Unità responsabile</b>	RM
<b>Unità coinvolte</b>	MO, TN, BO, RM
<b>Autori</b>	S. Bergamaschi, D. Beneventano, L. Cabibbo, P. Ciaccia, V. Crescenzi, P. Merialdo
<b>Autore da contattare</b>	Paolo Merialdo, Dipartimento di informatica e automazione – Via della Vasca Navale 79, 00149 Roma

# Rapporto sulla architettura metodologica e funzionale di riferimento

S. Bergamaschi, D. Beneventano, L. Cabibbo, P. Ciaccia, V. Crescenzi, P. Merialdo

20 Giugno 2005

## Abstract

L'enorme quantità di dati e la crescente disponibilità di servizi presenti su Web rendono sempre più importante lo sviluppo di infrastrutture e sistemi software che permettano ai clienti collegati alla rete di "ricaricarsi" di dati di interesse per i propri bisogni informativi. Il progetto WISDOMha come obiettivo principale la definizione di tecniche, e strumenti per la ricerca, la localizzazione e la fruizione personalizzata di risorse informative disponibili su Web. Le soluzioni proposte sono basate su ontologie di dominio, e su una architettura distribuita e decentralizzata. Obiettivo di questo documento è la definizione di una architettura metodologica e funzionale di riferimento al fine di garantire coerenza tra le soluzioni che verranno studiate dalle varie unità di ricerca che partecipano al progetto.

## 1 Introduzione

Obiettivo principale del progetto WISDOMè lo studio di tecniche e strumenti per la localizzazione e la fruizione personalizzata di risorse informative disponibili su Web. Lo scopo è quello di definire una infrastruttura software che permetta alle applicazioni di trarre vantaggio dalla enorme disponibilità di informazioni presenti sul Web. Elementi caratterizzanti la proposta sono l'adozione di una architettura distribuita e decentralizzata e il riferimento a ontologie di dominio per supportare il processamento distribuito di query.

I principali problemi che il progetto affronta sono legati all'elevato numero di sorgenti informative che potrebbero essere coinvolte in una interrogazione e alla dinamicità dell'ambiente. Questi aspetti infatti richiedono di affrontare considerevoli problemi di integrazione poichè, dovendo assumere che le sorgenti siano autonome (tipicamente siti Web) si manifesteranno eterogeneità a più livelli di astrazione (di formato, logico, semantico). La dinamicità dell'ambiente rende necessario lo studio di soluzioni che consentano sia di far fronte a modifiche che possono occorrere, anche a livello intensionale, sulle sorgenti, sia alla possibilità di allargare l'insieme delle sorgenti di interesse. Inoltre, l'efficienza del sistema potrebbe essere compromessa se non si adottano opportune tecniche di ottimizzazione delle interrogazioni. Anche il questo caso l'elevato numero di sorgenti autonome potenzialmente coinvolte in una query risulta un fattore critico.

L'idea principale del progetto è quella di sfruttare un'integrazione flessibile del contenuto semantico. A tal fine sono previsti due livelli di integrazione: una *integrazione forte* di un insieme di sorgenti in un nodo informativo (che nel seguito chiameremo super-peer); una *integrazione debole* basata su mapping semantici tra i concetti offerti da differenti nodi informativi. In questo modo le interrogazioni poste su un nodo informativo possono essere processate facendo riferimento sia alle informazioni disponibili localmente che a quelle offerte da altri nodi informativi.

Obiettivo di questo documento è la definizione dei principali elementi archiettrali e la descrizione delle loro interazioni in situazioni di riferimento al fine di garantire coerenza tra le soluzioni che verranno studiate nel progetto.

Il documento è organizzato come segue. La Sezione 2 presenta la letteratura di riferimento per i sistemi peer-to-peer, con particolare attenzione ai cosiddetti P2P Data Management Sys-

tems (PDMS) ai quali la proposta architetturale per WISDOM si ispira. La Sezione 3 descrive i principali componenti architeturali, e la terminologia di riferimento. Successivamente, la Sezione 4 illustra l'interazione tra i vari elementi della architettura con riferimento a tre casi d'uso.

## 2 Stato dell'arte

Le attuali reti peer-to-peer (P2P) supportano solo un insieme limitato di metadati, come ad esempio i nomi dei file. Recentemente è stata proposta una nuova classe di reti P2P, dette reti P2P "schema-based" (vedi [1, 3, 2, 6]), che combinano approcci provenienti dalle aree di ricerca dei sistemi P2P, dei database e del semantic web. Queste reti sono costruite su peer che usano metadati per descrivere i loro contenuti.

A seconda del grado di centralizzazione è possibile individuare diversi tipi di sistemi di P2P. Nei sistemi cosiddetti *puri*, quali Gnutella<sup>1</sup> e Freenet<sup>2</sup>, tutti i peer hanno contemporaneamente funzioni di client e di server senza nessun mediatore o server centralizzato. Quindi ogni peer ha ruoli e responsabilità in tutte le funzionalità: query, download, ecc.. Nei sistemi *ibridi*, quale Napster<sup>3</sup>, la ricerca è effettuata su una directory centralizzata, ma il download avviene come in un sistema P2P puro – quindi, i peer sono uguali solo nelle loro funzionalità di download.

Le reti *super-peer*, quale ad esempio KaZaA<sup>4</sup> che è attualmente uno dei sistemi più popolari, rappresentano un incrocio tra i sistemi ibridi e i sistemi puri. Un super-peer è un nodo che funge come un server centralizzato per un sottoinsieme di client. I client effettuano interrogazioni ai loro super-peer e ricevono i risultati da esso, come in un sistema ibrido. Tuttavia, i super-peer sono collegati fra loro come in un sistema puro, propagano messaggi su questa rete sovrapposta ed effettuano e rispondono alle interrogazioni a nome dei loro client e di essi stessi. Quindi, i super-peer sono uguali nelle funzionalità di ricerca e tutti i peer (client inclusi) sono uguali in termini di download.

Come riportato in un recente survey sui sistemi peer-to-peer [7], un'area di ricerca che ha attratto considerevole attenzione riguarda il raggruppamento e l'organizzazione semantica del contenuto in reti super-peer [8], poiché i metadati di un piccolo gruppo di peer possono essere centralizzati su un singolo super-peer; un super-peer è un nodo che funge da server centralizzato per un sottoinsieme di client. I client inviano le interrogazioni e ricevono il risultato dal loro super-peer; inoltre, i super-peer sono collegati tra di loro e si scambiano messaggi su questa *overlay network*, elaborando le interrogazioni sia proprie che dei relativi client. Quindi una *super-peer network* è una P2P network, con un'architettura parzialmente centralizzata, formata da questi super-peer e dai loro client.

Un esempio concreto di super-peer network è costituito dal progetto Edutella project<sup>5</sup>, dove ogni super-peer è un sistema a mediatore che integra i metadati ed effettua la fusione dei dati provenienti da differenti sorgenti, riconciliando le informazioni in conflitto. Nella rete, un peer invia un'interrogazione al mediatore del suo super-peer, che la elabora sia utilizzando il suo schema integrato sia interrogando altri super-peer.

Un altro approccio per l'integrazione semantica di sorgenti informative eterogenee nel contesto di una struttura distribuita è stato proposto nei sistemi Peer Data Management System (PDMS) [3, 4]. La proposta dei sistemi PDMS è basata su un'architettura di gestione dei dati decentralizzata e facilmente estendibile nella quale ogni utente può contribuire con nuovi dati, nuovi schemi di dati o anche mapping tra gli schemi di altri peer. I sistemi PDMS rappresentano un'evoluzione naturale dei sistemi di integrazione dei dati, nei quali il singolo schema

---

<sup>1</sup><http://www.gnutella.com>

<sup>2</sup><http://freenet.sourceforge.net>

<sup>3</sup>[www.napster.com](http://www.napster.com)

<sup>4</sup><http://www.kazaa.com>

<sup>5</sup><http://edutella.jxta.org>

logico dei dati è sostituito da un insieme di schemi (gli schemi individuali dei peer) correlati con mapping semantici. In un sistema PDMS ogni nodo può essere una sorgente informativa, un sistema a mediatore, o entrambi; in particolare, un nodo mediatore (che è un nodo della rete con elevate risorse di calcolo) realizza l'integrazione semantica di un insieme di sorgenti informative derivando uno schema globale delle informazioni acquisite. Ogni nodo ha un proprio schema ed i mapping semantici (specificati tra coppie di schemi) consentono, data un'interrogazione formulata rispetto ad un certo nodo, di riformulare tale interrogazione sui nodi vicini; in questo modo, i nodi possono condividere i dati senza uno schema logico centrale. Più in dettaglio, data un'interrogazione su un particolare nodo, l'elaborazione dell'interrogazione procede riformulando iterativamente tale interrogazione, sulla base dei mapping semantici, fino a quando tutti le sorgenti informative rilevanti sono state raggiunte. L'approccio descritto nei citati articoli è implementato in un sistema denominato PIAZZA. In [5] vengono discussi aspetti semantici ed algoritmici relativi al problema di mapping tra i dati nei sistemi peer-to-peer; i sistemi proposti si basano sull'uso di tabelle di mapping che elencano coppie di valori corrispondenti per ricercare dati memorizzati nei differenti peer.

### 3 Architettura P2P per WISDOM

L'architettura per il progetto WISDOM fa riferimento ai sistemi ibridi, includendo sia peer (le sorgenti informative) che super-peer. L'idea è quella di integrare i dati di un piccolo gruppo di sorgenti su un singolo super-peer, al quale i client inviano le interrogazioni e dal quale ricevono il risultato. Il super-peer è basato su una architettura a strati wrapper-mediatore: lo strato di wrapper si preoccupa di mascherare l'accesso fisico ai dati, mentre lo strato mediatore offre una vista integrata ed effettua la fusione dei dati provenienti da differenti sorgenti, riconciliando le informazioni in conflitto.

Affinché un insieme di super-peer possano collaborare nel processamento delle interrogazioni, contribuendo con le informazioni disponibili presso le proprie sorgenti informative anche a interrogazioni poste su altri super-peer, è prevista la costituzione di una overlay network. I nodi della rete sono i super-peer, ai quali è richiesto di esporre il proprio contenuto informativo attraverso una ontologia di dominio; gli archi sono formati da mapping finalizzati a stabilire relazioni tra le ontologie di una coppia di super-peer.

Nel seguito illustriamo i principali elementi della architettura proposta ed introduciamo la terminologia di riferimento per il progetto. Dapprima ci concentriamo nella descrizione degli elementi interni al super-peer; successivamente mostriamo quali sono gli elementi che consentono di costituire la overlay network di super-peer e di processare (efficientemente) interrogazioni distribuite.

#### 3.1 Sorgenti e Super-peer

La Figura 1 mostra i principali elementi architetturali di un generico super-peer  $P_i$ . Il super-peer è composto da un insieme di sorgenti informative  $S_{i1}, S_{i2}, \dots, S_{in}$ . Ogni sorgente  $S_{ij}$  rappresenta una risorsa di informazioni disponibile su Web; tipicamente l'insieme (eventualmente un sottoinsieme) delle pagine di un sito Web.

**Wrapper** Ad ogni sorgente informativa  $S_{ij}$  è associato un wrapper  $W_{ij}$ , il cui scopo è quello di rendere trasparenti agli strati superiori i metodi di accesso ai dati offerti dalla sorgente. Un wrapper infatti consente di accedere ai dati della sorgente offrendo uno schema  $S_{ij}$ , attraverso il quale gli strati sovrastanti possono porre interrogazioni.

Nel caso di una singola pagina Web, un wrapper consiste in un programma che estrae i dati di interesse dal codice HTML e li organizza in un formato strutturato, ad esempio in un documento XML. Nel caso di un sito Web, o di una porzione di esso, un wrapper offre una

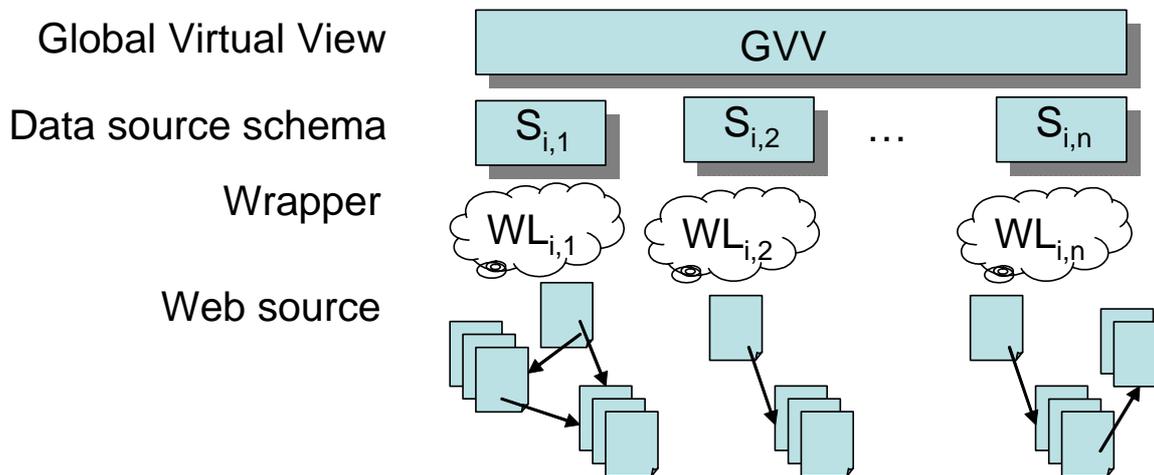


Figure 1: Elementi architetturali del generico super-peer  $P_i$

vista strutturata ed uniforme dei dati pubblicati nel sito e codifica non solo le primitive per l'estrazione dei dati, ma anche i percorsi ipertestuali da seguire per raggiungere le informazioni di interesse. In definitiva il wrapper mostra una vista uniforme dei dati (ad esempio sotto forma di un documento XML, o di uno schema relazionale) e si preoccupa di tradurre le interrogazioni poste su questa vista in istruzioni finalizzate a individuare le pagine di interesse ed estrarre e organizzare nel formato di destinazione le informazioni richieste.

**Global Virtual View** L'insieme degli schemi offerti dai wrapper associati alle sorgenti che partecipano ad un super-peer sono conciliati e integrati in una Global Virtual View. Lo scopo della Global Virtual View è quello di offrire una vista integrata, coerente e consistente delle informazioni presenti nelle sorgenti del super-peer. La vista integrata è descritta da uno schema globale, attraverso il quale è possibile effettuare interrogazioni per recuperare le informazioni offerte dalle sorgenti.

Nel nostro contesto la Global Virtual View di un super-peer ha molteplici funzionalità. In primo luogo essa gestisce i metadati che consentono di descrivere gli schemi delle sorgenti integrate e le loro relazioni con lo schema globale. In particolare l'integrazione è effettuata secondo una strategia GAV (Global as a View): ogni elemento dello schema globale è descritto come una vista sugli schemi delle sorgenti. Inoltre, la Global Virtual View accetta le interrogazioni poste sullo schema globale e ne consente il processamento, che consiste in una riscrittura dell'interrogazione posta sullo schema globale in un insieme di interrogazioni espresse sugli schemi locali; in base alla strategia GAV, le interrogazioni sono processate tramite un'operazione di "unfolding": ciascun atomo dello schema globale viene espanso, sostituendolo con la sua definizione nel mapping con lo schema locale. Infine, la Global Virtual View ha la responsabilità di elaborare i risultati ottenuti dalle interrogazioni sugli schemi locali, effettuando la fusione dei dati provenienti da differenti sorgenti e riconciliando le informazioni in conflitto.

### 3.2 Rete di super-peer

Per consentire il reperimento delle informazioni di interesse anche al di fuori del super-peer che accetta l'interrogazione, un insieme di super-peer sono messi in relazione attraverso una opportuna rete. La Figura 2 mostra i principali elementi architetturali che consentono di dare luogo ad una rete di super-peer.

Come affermato in Sezione 1 L'idea principale è quella di associare ad ogni super-peer  $P_i$  una *ontologia*, denotata  $Ont_i$ , che descrive l'offerta informativa del super-peer. Una rete di super-

peer è costruita definendo relazioni (mapping) tra le ontologie di un insieme di super-peer.

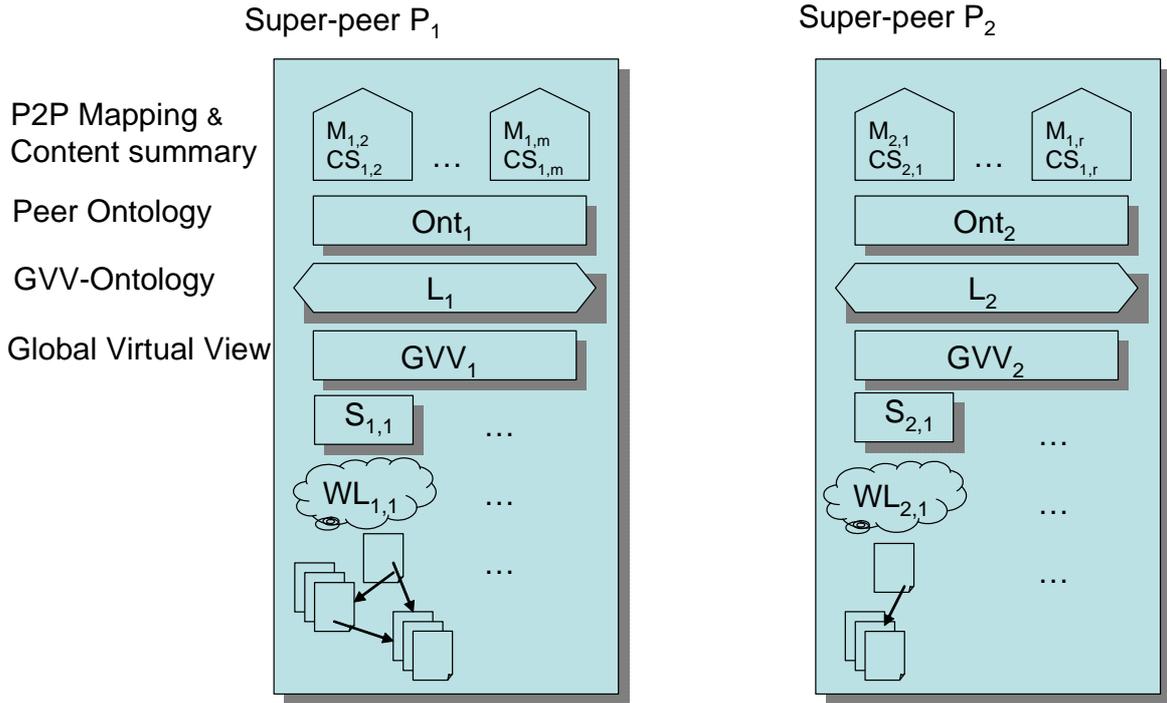


Figure 2: Elementi architetturali per la costituzione di una rete di super-peer in WISDOM

**Ontologia del super-peer** L'ontologia di un super-peer descrive a livello intensionale il contenuto informativo del super-peer in termini di entità concettuali e delle relazioni esistenti fra di esse. L'ontologia in pratica costituisce l'interfaccia da e verso l'esterno di ogni super-peer. È infatti attraverso l'ontologia che vengono stabilite le relazioni con altri super-peer. Inoltre, è importante sottolineare che la principale forma di interazione tra l'utente e il super-peer avviene attraverso l'ontologia.

**Mapping GVV-Ontology** Come illustrato in Figura 2, ad ogni ontologia  $Ont_i$  è associato un mapping  $L_i$  che stabilisce una corrispondenza tra i concetti della ontologia e le strutture della Global Virtual View. Attraverso questo mapping, interrogazioni espresse sulla ontologia che facciano riferimento ad aspetti estensionali possono essere tradotte in interrogazioni sulla Global Virtual View.

**P2P Mapping** La rete di super-peer è costruita, oltre che dai super-peer, da un insieme di *P2P mapping*, denotati  $M_{i,j}$ , che definiscono una relazione tra l'ontologia  $Ont_i$  del super-peer  $P_i$ , e l'ontologia  $Ont_j$  del super-peer  $P_j$ . Dato il mapping  $M_{i,j}$  definiamo *target ontology* l'ontologia  $Ont_i$  e *source ontology* l'ontologia  $Ont_j$ . Intuitivamente, il mapping  $M_{i,j}$  permette di riscrivere query espresse sulla ontologia  $Ont_i$  in query espresse sulla ontologia  $Ont_j$ . Come vedremo nel seguito, per il processamento delle interrogazioni è conveniente considerare che per ogni super-peer  $P_i$  esista un *trivial mapping*,  $M_{i,i}$ , che mappa i concetti della propria ontologia  $Ont_i$  su se stessa.

**Content Summary** I P2P mapping consentono ad un super-peer di estendere le interrogazioni che riceve ad altri super-peer, idealmente a tutti i super-peer per i quali ha definito un mapping.

Tuttavia, non sempre può risultare conveniente propagare una interrogazione verso altri super-peer. Ad esempio sarebbe inefficiente coinvolgere super-peer che hanno una limitata estensione dei concetti oggetto della interrogazione.

Per evitare che una interrogazione venga propagata verso super-peer “non rilevanti” è quindi utile associare a ciascun mapping  $M_{i,j}$  informazioni quantitative di natura statistica sulla estensione dei concetti del super-peer di destinazione. Queste informazioni vengono chiamate *content summaries*, e denotate  $CS_{i,j}$ . Il content summary  $CS_{i,j}$  ha lo scopo di dare informazioni quantitative sull'estensione dei concetti di  $Ont_i$  che possono essere trovati attraverso il mapping  $M_{i,j}$  con il super-peer  $P_j$ . Un semplice esempio del tipo di informazioni che può offrire un content summary è la cardinalità.

Osserviamo che i content summary vengono calcolati a partire dal content summary associato al trivial mapping, offrendo questi informazioni quantitative sulla estensione della ontologia di ciascun super-peer. A loro volta i content summary dei trivial mapping sono ottenuti informazioni aggregate, che definiamo *atomic content summary*, relative alle estensioni delle singole sorgenti che partecipano al super-peer. Denotiamo  $cs_{ik}$  l'atomic content summary associato alla sorgente  $k$  del super-peer  $P_i$ .

**User Interface** L'utente finale interagisce con la rete di super-peer mediante una opportuna interfaccia utente. Poiché assumiamo che le interrogazioni vengano poste con riferimento alle ontologie associate ai super-peer, l'interfaccia utente adotta paradigmi di interazione specificatamente pensati per rappresentare graficamente una ontologia di dominio.

In particolare, una funzionalità importante che l'interfaccia utente fornisce è quella di utilizzare i P2P mapping per consentire all'utente di “navigare” non solo sui concetti presenti nel super-peer locale, ma anche su quelli disponibili su altri super-peer della rete.

## 4 Casi d'uso

In questa sezione descriviamo l'architettura funzionale del sistema con riferimento a tre principali casi d'uso, illustrando le interazioni dei vari componenti architetturali descritti nella precedente sezione:

1. *UC1-Query Processing*: una interrogazione viene posta sulla ontologia  $Ont_i$  del super-peer  $P_i$
2. *UC2-Aggiunta di una sorgente*: una nuova sorgente viene aggiunta ad un super-peer  $P_i$
3. *UC3-Aggiunta di un super-peer*: un nuovo super-peer  $P_n + 1$  viene aggiunto alla rete

### 4.1 UC1: Query Processing

Le interrogazioni vengono poste sulla ontologia di un super-peer. È importante ricordare che in generale quando il super-peer  $P_i$  è chiamato a rispondere ad una interrogazione, questa potrebbe venire direttamente dalla interfaccia utente oppure da un altro super-peer  $P_j$ , collegato attraverso il mapping P2P  $M_{i,j}$ .

Infatti, quando il generico super-peer  $P_i$  riceve una query, questa viene processata in parte *localmente*, per recuperare le informazioni offerte dalle sorgenti sotto il controllo del super-peer stesso, in parte *esternamente*, ovvero propagata verso i super-peer  $P_j$ , per i quali esiste un mapping P2P  $M_{i,j}$ , con l'obiettivo di ottenere anche informazioni rilevanti alla interrogazione anche da sorgenti che partecipano ad altri super-peer.

Il processamento locale prevede la riscrittura della interrogazione sullo schema della GVV del super-peer. Questa riscrittura viene effettuata sulla base del mapping GVV-ontology associato ad ogni super-peer. Al fine di ottimizzare il processamento della interrogazione verranno prese

in considerazione le informazioni offerte dagli atomic content summary associati a ciascuna sorgente.

Il processamento esterno consiste invece nel riscrivere la interrogazione sulle ontologie di super-peer collegati al super-peer che ha accettato la query con un P2P mapping. Anche in questo caso, uno degli obiettivi che si vogliono perseguire è quello di non coinvolgere nel processamento super-peer che potrebbero rilevarsi poco promettenti alla definizione del risultato. L'interrogazione verrà quindi propagata verso i super-peer confinanti con una politica di routing che miri a privilegiare i super-peer più promettenti. La valutazione sarà basata sulle informazioni fornite dai content summary associati ai P2P mapping.

## 5 UC2: Aggiunta di una sorgente ad un peer

L'aggiunta di una nuova sorgente informativa  $S_{i,m+1}$  all'interno del super-peer  $P_i$  comporta numerosi interventi, come segue:

1. Inferenza del wrapper: se la sorgente informativa è costituita da un sito Web (o eventualmente da una porzione di un esso), è necessario definirne il wrapper. Il wrapper descriverà la sorgente attraverso uno schema che viene esposto agli strati superiori.
2. Calcolo atomic content summary: il processo di inferenza del wrapper richiede che una porzione limitata ma significativa della sorgente sia esplorata. Durante questa fase vengono elaborate informazioni quantitative relative informazioni offerte dalla sorgente, utili a produrre una stima dell'atomic content summary della sorgente.
3. Aggiornamento della Global Virtual View: lo schema offerto dal wrapper viene essere integrato nella Global Virtual View del super-peer. Questa operazione comportare la risoluzione di eterogeneità semantiche e strutturali.
4. Aggiornamento del mapping GVV-Ontology: questa operazione deve essere effettuata solo nel caso in cui la GVV abbia subito modifiche anche a livello intensionale (ad esempio include informazioni relative a nuove entità concettuali, oppure a nuove relazioni).
5. Aggiornamento della ontologia del super-peer: come nel caso precedente anche questa operazione deve essere effettuata solo nel caso in cui la GVV abbia subito modifiche anche a livello intensionale.
6. Aggiornamento dei P2P mapping: dopo aver aggiornato il content summary associato al trivial mapping, a tutti i super-peer che abbiano  $Ont_i$  come source ontology di un P2P mapping viene mandato un messaggio per attivare l'aggiornamento del corrispondente content summary. Questa operazione richiede il solo aggiornamento dei content summary associati ai P2P mapping nel caso in cui non ci siano state modifiche a livello intensionale; altrimenti è necessario allineare alla nuova ontologia tutti i mapping P2P.

## 6 UC3: Aggiunta di un peer alla rete

Questo caso d'uso corrisponde ad includere un nuovo super-peer nella rete. A tal fine è necessario stabilire i P2P mapping tra alcuni super-peer noti e il nuovo super-peer. Individuati i super-peer con i quali si vuole stabilire un collegamento, si definiscono i P2P mapping, che coinvolgono l'ontologia del nuovo super-peer come source ontology, ed eventualmente vengono inviati messaggi ad altri super-peer affinché calcolino P2P mapping che coinvolgano l'ontologia del nuovo super-peer come target ontology.

## References

- [1] K. Aberer, P. Cudre-Mauroux, and M. Hauswirth. *The chatty web: Emergent semantics through gossiping*. Int. World Wide Web Conf. (WWW2003), Budapest, Hungary, 2003.
- [2] P. A. Bernstein, F. Giunchiglia, A. Kementsietsidis, J. Mylopoulos, L. Serafini, and I. Zaihrayeu. *Data management for peer-to-peer computing: A vision. Work*. The Web and Databases, Madison, 2002.
- [3] A. Y. Halevy, Z. G. Ives, P. Mork, and I. Tatarinov. *Piazza: Data management infrastructure for semantic web applications*. Int. World Wide Web Conf. (WWW2003), Budapest, 2003.
- [4] A. Halevy, Z. Ives, D. Suci, and I. Tatarinov. *Schema mediation in peer data management systems*. International Conference on Data Engineering (ICDE'03), Bangalore, India, March 2003.
- [5] A. Kementsietsidis, M. Arenas, and R. J. Miller. *Mapping data in peer-to-peer systems: semantics and algorithmic issues*. SIGMOD 2003, 325-336, 2003
- [6] A. Loser, W. Nejdl, M. Wolpers, and W. Siberski. *Information Integration in Schema-Based Peer-To-Peer Networks*. Conference of Advanced Information Systems Engineering (CAiSE 03), Klagenfurt, June 2003.
- [7] S. A. Theotokis, D. Spinellis. *A survey of peer-to-peer content distribution technologies*. ACM Comput. Surv. 36(4): 335-371 (2004)
- [8] B. Yang and H. Garcia-Molina. *Designing a Super-Peer network*. Int. Conf. on Data Engineering (ICDE03), Bangalore, 2003.