

PROGETTO DI UNA UNITÀ DI RICERCA - MODELLO B
Anno 2005 - prot. 2005095190_003

PARTE I

1.1 Programma di Ricerca afferente a

Area Scientifico Disciplinare 09: Ingegneria industriale e dell'informazione 100%

1.2 Durata del Programma di Ricerca

24 Mesi

1.3 Coordinatore Scientifico del Programma di Ricerca

CATARCI TIZIANA catarci@dis.uniroma1.it

ING-INF/05 - Sistemi di elaborazione delle informazioni

Università degli Studi di ROMA "La Sapienza"

Facoltà di INGEGNERIA

Dipartimento di INFORMATICA E SISTEMISTICA "Antonio Ruberti"

1.4 Responsabile Scientifico dell'Unità di Ricerca

CASTANO SILVANA
Professore Straordinario 25/10/1965 CSTSVN65R65D872D

INF/01 - Informatica

Università degli Studi di MILANO

Facoltà di SCIENZE MATEMATICHE FISICHE e NATURALI

Dipartimento di INFORMATICA E COMUNICAZIONE

02/50316319 (Prefisso e telefono) 02/50316229 (Numero fax) castano@dico.unimi.it (Indirizzo posta elettronica)

1.5 Curriculum scientifico del Responsabile Scientifico dell'Unità di Ricerca

Testo italiano

Silvana Castano è professore ordinario di basi di dati e sistemi informativi presso il Dipartimento di Informatica e Comunicazione dell'Università degli Studi di Milano, dove è attualmente responsabile del gruppo Information Systems & knowledge management (ISLab) e dove è stata professore associato (1998-2001) e prima ancora ricercatore (1993-1998). Nel 1993 ha conseguito il titolo di dottore di ricerca in Ingegneria Informatica e Automatica presso il Politecnico di Milano. Principali interessi di ricerca sono nell'area delle basi di dati e sistemi informativi e delle ontologie e Semantic Web, con particolare riguardo a interoperabilità semantica, condivisione di conoscenza in sistemi P2P, integrazione di informazioni, sicurezza di basi di dati e XML. Su questi argomenti ha pubblicato numerosi articoli a riviste e conferenze internazionali e nazionali e ha partecipato a numerosi progetti di ricerca sia nazionali che internazionali, fra cui i recenti progetti EU FP6 NoE INTEROP (Interoperability Research for Networked

Enterprises Applications and Software), dove è responsabile dell'unità dell'Università di Milano, e FIRB WEB MINDS, dove è responsabile del Gruppo di Lavoro su Ontologie e metadati. E' co-autore del libro Database Security (Addison Wesley-ACM Press, 1995). Dal giugno 1997, e' Coordinatore del Gruppo di lavoro AICA "Basi di dati".

Testo inglese

Silvana Castano is full professor of database and information systems at University of Milano, where she currently chairs the Information systems & knowledge management (ISLab) group. Previously, she has been with the University of Milano as associate professor (1998-2001) and assistant professor (1993-1998), respectively. She received the Ph.D. degree in Computer Science from Politecnico di Milano, in 1993. Her main research interests are in the area of databases and information systems and ontologies and Semantic Web, with current focus on knowledge sharing in peer-based systems, information integration, semantic interoperability, database and XML security. She has published her research results in the major journals and in the refereed proceedings of the major conferences of the field. On these topics, she has been working in several national and international research projects, including recent the EU FP6 NoE INTEROP (Interoperability Research for Networked Enterprises Applications and Software) project, where she is responsible of the University of Milano group, and the FIRB WEB MINDS project, where she is responsible of the Working Group on Metadata and ontologies. She is an author of the book Database Security (Addison Wesley-ACM Press, 1995). Since June 1997, she is chair of the AICA Working Group on Databases.

1.6 Pubblicazioni scientifiche più significative del Responsabile Scientifico dell'Unità di Ricerca

1. CASTANO S., FERRARA A., MONTANELLI S., RACCA G. (2004). *Semantic Information Interoperability in Open Networked Systems*. In M. BOUZEGHOUB, C.A. GOBLE, V. KASHYAP, S. SPACCAPIETRA EDS. *Semantics for Grid Databases, First International IFIP Conference on Semantics of a Networked World: ICSNW 2004. Revised papers*. (vol. LNCS 3226 pp. 215-230). ISBN: 3-540-23609-0: Springer
2. CASTANO S., FERRARA A., MONTANELLI S., RACCA G. (2004). *Matching Techniques for Resource Discovery in Distributed Systems Using Heterogeneous Ontology Descriptions*. *IEEE Proc. of the Int. Conference on Coding and Computing (ITCC04)*. Las Vegas, Nevada, USA, April
3. CASTANO S., FERRARA A., MONTANELLI S. (2003). *H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems*. *1st Int. VLDB Workshop on Semantic Web and Databases (SWDB)*. Berlin, September
4. CASTANO S., FERRARA A., MONTANELLI S., ZUCHELLI D. (2003). *HELIOS: a General Framework for Ontology-based Knowledge Sharing and Evolution in P2P Systems*. *DEXA WEBS Workshop*. Prague, Czech Republic, September (pp. 597-603). IEEE Computer Society.
5. CASTANO S., FERRARA A. (2003). *Knowledge Representation and Transformation in Ontology-based Data Integration*. In BORYS OMELAYENKO, MICHEL C. A. KLEIN EDS. *Knowledge Transformation for the Semantic Web. Frontiers in Artificial Intelligence and Applications* (vol. Vol. 95). ISBN: 1-58603-325-5: IOS Press
6. CASTANO S., FERRARI E. (2003). *Protecting Datasources over the Web: Policies, Models, and Mechanisms*. In D. TANIAR, J.W. RAHAYU EDS. *Web-Powered Databases* (pp. 299-330). ISBN: 1-59140035-X: Idea Group
7. CASTANO S., BERTINO E., FERRARI E., MESITI M. (2002). *Protection and administration of XML data sources*. *DATA & KNOWLEDGE ENGINEERING*. vol. 43(3) pp. 237-260 ISSN: 0169-023X
8. O. BOUCELMA, CASTANO S., C.A. GOBLE, V. JOSIFOVSKI, Z. LACROIX, B. LUDSCHER (2002). *Report on the EDBT'02 Panel on Scientific Data Integration*. *SIGMOD RECORD*. vol. 31(4) pp. 107-112 ISSN: 0163-5808
9. BERGAMASCHI S., CASTANO S., VINCINI M., BENEVENTANO D. (2001). *Semantic Integration of Heterogeneous Information Sources*. *DATA & KNOWLEDGE ENGINEERING*. vol. 36(3) ISSN: 0169-023X
10. CASTANO S., V. DE ANTONELLIS, S. DE CAPITANI DI VIMERCATI (2001). *Global Viewing of Heterogeneous Data Sources*. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*. vol. 13(2) ISSN: 1041-4347
11. BERGAMASCHI S., CASTANO S., VINCINI M. (1999). *Semantic Integration of Semistructured and Structured Data Sources*. *SIGMOD RECORD*. vol. 28(1) ISSN: 0163-5808
12. CASTANO S., DE ANTONELLIS V. (1999). *A Discovery-Based Approach to Database Ontology Design*. *DISTRIBUTED AND PARALLEL DATABASES*. vol. 7(1) ISSN: 0926-8782

1.7 Risorse umane impegnabili nel Programma dell'Unità di Ricerca

1.7.1 Personale universitario dell'Università sede dell'Unità di Ricerca

Personale docente

n°	Cognome	Nome	Dipartimento	Qualifica	Settore Disc.	Mesi Uomo
----	---------	------	--------------	-----------	---------------	-----------

						1° anno	2° anno
1.	CASTANO	Silvana	Dip. INFORMATICA E COMUNICAZIONE	Prof. Ordinario	INF/01	5	5
2.	FERRARA	Alfio	Dip. INFORMATICA E COMUNICAZIONE	Ricercatore Universitario	ING-INF/05	4	4
3.	AGOSTINI	Alessandra	Dip. INFORMATICA E COMUNICAZIONE	Ricercatore Universitario	INF/01	4	4
4.	MESITI	Marco	Dip. INFORMATICA E COMUNICAZIONE	Ricercatore Universitario	INF/01	4	4
TOTALE						17	17

Altro personale

Nessuno

1.7.2 Personale universitario di altre Università

Personale docente

Nessuno

Altro personale

Nessuno

1.7.3 Titolari di assegni di ricerca

Nessuno

1.7.4 Titolari di borse

n°	Cognome	Nome	Dipartimento	Anno di inizio borsa	Durata(in anni)	Tipologia	Mesi Uomo	
							1° anno	2° anno
1.	MONTANELLI	Stefano	Dip. INFORMATICA E COMUNICAZIONE	2004	3	Dottorato	4	4
TOTALE							4	4

1.7.5 Personale a contratto da destinare a questo specifico programma

n°	Qualifica	Costo previsto	Mesi Uomo		Note
			1° anno	2° anno	
1.	Assegnista	40.000	11	11	assegnista di ricerca
TOTALE		40.000	11	11	

1.7.6 Personale extrauniversitario indipendente o dipendente da altri Enti

Nessuno

PARTE II

2.1 Titolo specifico del programma svolto dall'Unità di Ricerca

Testo italiano

Emergenza della semantica attraverso la costituzione guidata da consenso di comunità semantiche in ambienti P2P

Testo inglese

The emergence of semantics through consensus-driven formation of semantic communities in P2P multi-knowledge environments

2.2 Settori scientifico-disciplinari interessati dal Programma di Ricerca

ING-INF/05 - Sistemi di elaborazione delle informazioni

INF/01 - Informatica

2.3 Parole chiave

Testo italiano

COMUNITA' SEMANTICHE ; NEGOZIAZIONE E CONSENSO ; DISTRIBUZIONE SEMANTICA DI QUERY ; POLITICHE DI CONDIVISIONE ; MATCHING DINAMICO DI ONTOLOGIE

Testo inglese

SEMANTIC COMMUNITIES ; NEGOTIATION AND CONSENSUS ; SEMANTIC QUERY DISTRIBUTION ; SHARING POLICIES ; DYNAMIC ONTOLOGY MATCHING

2.4 Base di partenza scientifica nazionale o internazionale

Testo italiano

La costituzione guidata da consenso di comunità semantiche di peer e la disponibilità di tecniche per la propagazione di query su base semantica rappresenta una sfida nello stadio attuale di sviluppo di sistemi aperti a rete. Questo diventa ancora più rilevante in contesti di semantica emergente, dove requisiti di collaborazione semantica fra peer che si manifestano in maniera spontanea e imprevista richiedono capacità dinamiche di negoziazione per la definizione di accordi su interpretazioni comuni nell'ambito di un certo task. La base scientifica di interesse per la formazione di comunità semantiche in ambienti P2P è costituita in letteratura dai sistemi basati su peer, ed in particolare dal tema delle comunità P2P basate su interessi, dal routing semantico in ambiente P2P, da temi di reputazione e trust, e da lavori che trattano il tema del matching di ontologie.

Per quanto riguarda i sistemi P2P, recenti contributi sono dedicati alla descrizione di tecniche per evolvere dalle tradizionali reti P2P che supportano lo scambio di file basato su nome [Waterhouse et al., 2002], a sistemi più complessi, capaci di supportare lo scambio di contenuti strutturati (come per esempio documenti e dati relazionali) sfruttando RDF e ontologie tematiche come metadati. Esempi significativi di questi ultimi sistemi P2P sono descritti in [Nejdl et al., 2002, Broekstra et al., 2003] metadati basati su RDF(S) vengono utilizzati per descrivere e interrogare le risorse in modo semanticamente ricco.

Tuttavia, molti aspetti riguardanti i sistemi P2P costituiscono ancora temi aperti e devono essere approfonditi per scopi di collaborazione semantica. In particolare, algoritmi di propagazione semantica delle interrogazioni sono richiesti per migliorare le prestazioni e l'efficacia dei processi di ricerca necessari alla condivisione delle risorse. Alcuni risultati iniziali in questo senso sono descritti in [Staab et al., 2004], dove viene descritto il protocollo REMINDIN' per la propagazione di interrogazioni. Tale protocollo agisce osservando quali interrogazioni sono gestite con successo dagli altri nodi, immagazzinando queste osservazioni e successivamente utilizzando queste informazioni per la selezione del destinatari delle interrogazioni. In [Crespo e Garcia-Molina, 2003] si presenta un approccio in cui i nodi semanticamente collegati formano una Semantic Overlay Network (SON) e le interrogazioni sono indirizzate alle SON opportune, aumentando le probabilità che i file di interesse vengano trovati rapidamente e riducendo il carico di ricerca per i peer che non possiedono contenuti rilevanti. In [Haase et al., 2004], i nodi pubblicizzano la loro esperienza sulla rete P2P. Sulla base di una comune ontologia condivisa e della somiglianza semantica fra l'argomento dell'interrogazione e l'esperienza degli altri peer, un nodo può selezionare i nodi adatti a cui inoltrare l'interrogazione. Gli attuali algoritmi di propagazione delle interrogazioni sono essenzialmente basati su osservazioni statistiche ed utilizzano, in alcuni casi, un'ontologia comune, spesso soltanto una tassonomia. Uno degli obiettivi della nostra ricerca all'interno del progetto sarà quello di sviluppare tecniche semantiche di propagazione di interrogazioni per scopi di semantica emergente, estendendo le tecniche attuali per adattare a contesti con ontologie multiple, rilassando il vincolo di considerare una comune ontologia condivisa. Tali tecniche saranno in grado di supportare la formazione dinamica di accordi e verranno utilizzate per la propagazione di interrogazioni.

L'idea di realizzare comunità di nodi per mezzo di un approccio basato sulla semantica è in una fase di ricerca iniziale e pochi risultati sono apparsi in letteratura. Per esempio, in [Bonifacio et al., 2003] gli autori presentano un'architettura P2P per supportare federazioni di peer in cui la condivisione di conoscenza è basata su un modello sociale di collaborazione. Peer singoli e federazioni utilizzano una piattaforma di scambio di conoscenza per organizzare la conoscenza secondo una prospettiva individuale o di gruppo gestendo significati differenti per mezzo di un algoritmo di matching semantico. La formazione di coalizioni dinamiche

per la riduzione dello spazio di ricerca e per realizzare una diffusione migliore delle informazioni ai peer partecipanti è discussa in [Khambatti et al., 2003]. In questo approccio, le comunità emergono autonomamente secondo gli interessi dichiarati dai peer e ogni nodo è caratterizzato da un attributo di "coinvolgimento" che definisce il suo livello di partecipazione nella comunità.

Un contributo originale che sarà curato nel progetto di ricerca è relativo allo sviluppo di tecniche di aggregazione di comunità che potranno combinare le descrizioni ontologiche degli interessi dei nodi con tecniche di matching dinamico di ontologie, che risolvano le limitazioni del matching esatto impiegato negli attuali approcci e offrano capacità di matchmaking semantico nella formazione di accordi. Obiettivo delle tecniche di matching di ontologie è quello di confrontare differenti descrizioni ontologiche per identificare i concetti che hanno affinità semantica fra di loro. Uno studio delle tecniche di matching per ontologie è descritto in [Shvaiko, 2004], dove vengono classificati approcci formali ed euristici. I primi sono fondati sulla semantica di modelli teorici e su tecniche di matching automatizzate [Madhavan et al., 2002, Bouquet et al., 2003, Giunchiglia e Shvaiko, 2004] mentre i secondi sono basati sull'idea di identificare i mapping che sussistono fra le strutture simili di un grafo mediante la combinazione di tecniche di analisi, matching e apprendimento [Doan et al., 2003]. In [Castano et al., 2003b], proponiamo tecniche euristiche per effettuare matching dinamico di ontologie in contesti aperti basati su peer. Le caratteristiche peculiari delle nostre tecniche sono la possibilità di realizzare il processo di matching in modo flessibile ed adattabile, configurando dinamicamente l'algoritmo con il modello di matching più adatto al caso in esame. Questo viene effettuato considerando il livello di ricchezza delle descrizioni ontologiche e pesando l'importanza delle caratteristiche linguistiche e contestuali dei concetti nella valutazione del matching. Le tecniche di matching dinamico di ontologie verranno usate come base per guidare la fase di negoziazione del consenso e dell'aggregazione di peer durante il processo di formazione di comunità, per identificare i mapping fra i concetti delle diverse ontologie dei peer e mantenerli all'interno di una data comunità.

La natura intrinsecamente aperta delle comunità P2P pone alcune rilevanti questioni riguardanti il mantenimento delle informazioni relative alla reputazione dei membri. Il problema dello scambio di informazioni di trust e reputazione in sistemi peer-based è stato affrontato in [Aberer and Despotovic, 2001, Abdul-Rahman and Hales, 1999, Marti and Garcia-Molina, 2004]. Un altro requisito importante nella condivisione di risorse e nella gestione di comunità in sistemi P2P è legato alla sicurezza e al trust, in modo da permettere a un peer di mantenere il controllo dei propri contenuti e delle proprie capacità di elaborazione, e di concederne l'accesso ad altri peer sotto specifiche condizioni. Gli attuali sistemi P2P affrontano tali problemi ad uno stadio primordiale; per esempio in Seti@home un nodo può impostare la sua disponibilità sulla base del suo stato interno o della rete. Tuttavia, aspetti più sofisticati come la combinazione di credenziali o condizioni temporali dovrebbero essere considerati in contesti di semantica emergente richiedendo la capacità di specificare politiche articolate [Kagal et al., 2003, Martinez Perez et al., 2005]. In ambiente GRID, alcune proposte sono in via di sviluppo anche se ancora ad uno stadio iniziale [Sundaram and Chapman, 2002]. Si noti che tali politiche di condivisione rappresentano un livello di sicurezza addizionale rispetto all'autenticazione e crittografia, assicurati dai tradizionali meccanismi di sicurezza di rete [Stallings, 2000]. Il contributo della ricerca su questi aspetti è legato allo sviluppo di un linguaggio basato su ontologie per la specifica di politiche dichiarative, allo scopo di rendere possibile la definizione di politiche di condivisione che possano essere applicate a risorse molteplici, con una interpretazione consistente sia all'interno di una comunità sia tra comunità differenti. Particolare attenzione sarà posta alla progettazione di politiche compatibili con gli standard per i certificati digitali in modo da poter essere efficacemente utilizzate nell'ambito di ambienti interconnessi [OASIS, 2004]. L'unità di ricerca di UNIMI ha prodotto contributi su temi di condivisione di conoscenza in sistemi interconnessi aperti e comunità distribuite [Agostini et al., 2003, Castano et al., 2003a, Castano et al., 2003b]. Contributi specifici che costituiscono la base scientifica di partenza per questo progetto riguardano: (i) tecniche dinamiche di matching dinamico di ontologie sviluppate nel sistema P2P HELIOS [Castano et al., 2004a, Castano et al., 2004d] sulla base della precedente esperienza nello sviluppo di tecniche di matching di schemi implementate nello strumento ARTEMIS per l'integrazione semantica di informazioni [Castano et al., 2001]; (ii) architetture di collaborazione orientata ai servizi per la scoperta dinamica di conoscenza e tecniche di interoperabilità semantica in sistemi distribuiti aperti [Castano et al., 2004b, Castano et al., 2005]; (iii) tecniche di indirizzamento di "query indagine" nel sistema HELIOS per la ricerca di conoscenza [Castano et al., 2004c]; (iv) linguaggi per la definizione di politiche basate su XML per l'accesso selettivo e controllato e la disseminazione di documenti Web [Castano and Ferrari, 2003, Bertino et al., 2001] e per la personalizzazione di servizi in contesti distribuiti [Bertino et al., 2004].

Testo inglese

The consensus-driven formation and management of semantic communities and the availability of advanced techniques for query propagation on a semantic basis is a challenging issue in the current stage of development of open networked system architectures, to provide sharing of distributed resources and to enforce semantic collaboration in an effective way. This is even more true in emergent semantics contexts, where unpredictable collaboration requirements that emerge among peers require dynamic capabilities of negotiating agreements on common interpretations within the context of a given task [Aberer et al., 2004]. In particular, the possibility for the parties to interact peer-to-peer to negotiate local agreements and the capability to build global agreements (and thus communities) by aggregation of local agreements is strongly required, to make the process scalable and suited for open networked contexts in general and to emergent semantics contexts in particular.

The main scientific base relevant for the formation of semantic communities in multi-knowledge environments is in literature P2P interest-based communities, P2P semantic routing, peer reputation and trust and in work on ontology matching.

With respect to P2P systems, recent research work is devoted to provide techniques for evolving from basic P2P networks supporting only file exchanges using simple filenames as metadata [Waterhouse et al., 2002], to more complex systems, capable of supporting the exchange of structured contents (e.g., documents, relational data) by exploiting explicit schemas to describe knowledge, usually using RDF and thematic ontologies as metadata. Significant schema-based P2P networks are described in [Nejdl et al., 2002, Broekstra et al., 2003] where a RDF(S) metadata model is used for semantic resource description and querying.

However, several aspects concerning P2P systems remain still open issues and need to be further investigated for semantic collaboration. In particular, semantic query propagation algorithms are required to improve the performance and the effectiveness of discovery and search processes for resource sharing. Some initial results in this direction can be found in [Staab et al., 2004]. Here, the REMINDIN' multi-step query propagation protocol is described, to enforce selected propagation of queries by observing which queries are successfully answered by other peers, by storing these observations, and by subsequently using this information for peer selection. A similar approach is presented in [Crespo and Garcia-Molina, 2003], where the node connections of a P2P system are content-influenced. Thus, semantically related nodes form a Semantic Overlay Network (SON) and queries are routed to the appropriate SONs, increasing the chances that matching files will be found quickly, and reducing the search load on nodes that have unrelated content. In [Haase et al., 2004], peers advertise their experience in the P2P network. Based on a shared common

ontology, the semantic similarity between the subject of a query and the expertise of other nodes, a peer can select appropriate peers for query forward.

Current query propagation algorithms are essentially based on statistical observations and exploit, in some cases, a shared ontology, often just a taxonomy. Further research work is required to enhance current query propagation techniques in order to drive the selection of the best recipients for a given query by exploiting ontology knowledge of the sending peer. One important goal of our research in the project will be to develop query propagation techniques that address emergent semantics requirements, by extending current techniques to work in multi-ontology contexts and thus releasing the constraint of having an initial common shared ontology. Such techniques will be thus able to support dynamic agreements formation to be used for query propagation.

The idea of supporting peer communities by means of a semantics-based approach is at an initial stage of research, and very few proposals have appeared in the literature. For instance, in [Bonifacio et al., 2003] the authors present a P2P architecture for supporting peer federations in which knowledge sharing is based on a social collaboration model. Single peers and federations use the Kex platform (Knowledge Exchange System) for organizing knowledge from an individual- or community-based perspective managing different meanings by means of a semantic matching algorithm. The formation of dynamic coalitions for pruning the search space and allow a better dissemination of information to participating peers is discussed in [Khambatti et al., 2003]. In this approach, communities emerge autonomously according to the interests advertised by peers and each peer is characterized by an involvement attribute which state the level of participation of the peer in the community.

A novel contribution that will be given in the research project is related to the development of community aggregation techniques that will be able to combine ontological descriptions of peer interests with dynamic ontology matching techniques, to overcome the limitations of exact matching of current approaches and to provide semantic matchmaking capabilities in agreement formation.

The general goal of ontology matching techniques is to compare different ontological descriptions for finding concepts that have a semantic affinity. A survey of ontology matching techniques is provided in [Shvaiko, 2004], where formal and heuristic approaches are classified. The former are founded on model theoretic semantics and automated matching techniques [Madhavan et al., 2002, Bouquet et al., 2003, Giunchiglia and Shvaiko, 2004] while the latter are based on the idea of guessing mappings that may hold between similar graph structures through a combination of analysis, matching, and learning techniques [Doan et al., 2003]. In [Castano et al., 2003b], we have proposed heuristic techniques for performing dynamic ontology matching in open, peer-based networked contexts. Peculiar features of our techniques are the capability of performing the matching process in a flexible and adaptable way, by dynamically configuring the matching algorithm with the most appropriate matching model for the specific matching case. This is done by taking into account the level of richness of ontological knowledge description as well as the requested degree of mix between the importance of linguistic and contextual features of concepts in the evaluation of their matching degree. Dynamic ontology matching techniques will be used in the project as the basis for driving the negotiation of agreements and the aggregation of peers during the community formation process, in order to discover the mappings between concepts of different peer ontologies and maintain them for a given community, once their member have committed agreements.

The intrinsically open nature of P2P communities poses serious issues regarding the maintenance of reputation information about the members of the communities. Some work on models to support the exchange of trust/reputation information in peer-based can be found in [Aberer and Despotovic, 2001, AbdulRahman and Hailes, 1999, Marti and GarciaMolina, 2004]. Another important requirement in P2P community management and resources sharing is related to security and trust, to allow a peer to retain ownership of its own contents and processing capabilities, and to allow other peers to access them under the appropriate terms and conditions. Current P2P systems generally seldom address such problems; for example in Seti@home a peer can set its availability just based on its internal state or on the network state. However, more sophisticated aspects like combination of credentials or temporal conditions need to be considered in emergent semantics contexts, and this requires specification capabilities [Kagal et al., 2003, Martinez Perez et al., 2005]. In the GRID environments, some proposals are under development but they are still at an early stage [Sundaram and Chapman, 2002]. It has to be noted that such policies represent an additional security level on top of authentication and encryption, which are ensured by conventional network security mechanisms [Stallings, 2000]. Contribution of our research is related to the development of an ontology-based language for declarative policy specification, to support specification of sharing policies which can be applied to multiple resources and with consistent interpretation within a community but also across communities. We believe that it is important that new proposals of policy sharing be designed to be compliant with digital certificate standards for effective use of developed policies in networked environments [OASIS, 2004].

The UNIMI unit has provided contributions about the general problem of knowledge sharing in open networked systems and distributed communities [Agostini et al., 2003, Castano et al., 2003a, Castano et al., 2003b]. Specific contributions constituting the starting scientific base for this project regards: (i) dynamic ontology matching techniques developed in the HELIOS P2P system [Castano et al., 2004a, Castano et al., 2004d] on the ground of previous experience of developing schema matching techniques in the framework of the semantic information integration tool environment ARTEMIS [Castano et al., 2001]; (ii) service-based collaboration architecture for dynamic knowledge discovery techniques and semantic interoperability in open networked systems [Castano et al., 2004b, Castano et al., 2005]; (iii) probe query routing in the HELIOS P2P system for knowledge discovery [Castano et al., 2004c]; (iv) XML-based policy specification formalisms and techniques for controlled and selective access and dissemination of Web documents [Castano and Ferrari, 2003, Bertino et al., 2001] and for service personalization in distributed contexts [Bertino et al., 2004].

2.4.a Riferimenti bibliografici

[Abdul-Rahman and Hailes, 1999] A. Abdul-Rahman and S. Hailes, "Relying on Trust to Find Reliable Information", in Proc. of 1999 Int. Symposium on Database, Web and Cooperative Systems (DWACOS'99), Baden-Baden, Germany, August 1999.

[Aberer and Despotovic, 2001] K. Aberer and Z. Despotovic, "Managing Trust in a Peer-2-Peer Information System", in Proc. of the 2001 ACM CIKM Int. Conference on Information and Knowledge Management, Atlanta, Georgia, USA, November 2001.

[Aberer et al., 2004] K. Aberer et al., "Emergent Semantics Principles and Issues", in Proc. of the 9th Int. Conference on Database Systems for Advances Applications - DASFAA 2004, Jeju Island, Korea, March 2004.

- [Agostini et al., 2003] A. Agostini, S. Albolino, R. Boselli, G. De Michelis, F. De Paoli, and R. Dondi, "Stimulating Knowledge Discovery and Sharing", in *Proc. of the ACM Conference GROUP 2003*, Sanibel Island, Florida, USA, November 2003.
- [Bertino et al., 2001] E. Bertino, S. Castano, E. Ferrari, and M. Mesiti, "Specifying and Enforcing Access Control Policies for XML Document Sources", *World Wide Web Journal*, 3(3) ISSN: 1085-2301, 2001.
- [Bertino et al., 2004] E. Bertino, M. Mesiti, and M. Cochinala, "A Trigger-Based Approach for Communication Personalization", in *Proc. of 8th Int. Database Engineering & Applications Symposium (IDEAS 2004)*, Coimbra, Portugal, July 2004.
- [Bonifacio et al., 2003] M. Bonifacio, P. Bouquet, G. Marni and M. Nori, "Peer - Mediated Distributed Knowledge Management", in *Proc. of Int. Symposium on Agent Mediated Knowledge Management (AMKM 2003)*, Stanford, CA, USA, March 2003.
- [Bouquet et al., 2003] P. Bouquet, B. Magnini, L. Serafini, and S. Zanobini, "A SAT-based Algorithm for Context Matching", in *Proc. of the 4th Int. and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2003)*, Stanford, CA, USA, June 2003.
- [Broekstra et al., 2003] J. Broekstra et al., "A Metadata Model for Semantics-based Peer-to-Peer Systems", in *Proc. of the 1st WWW Int. Workshop on Semantics in Peer-to-Peer and Grid Computing (SemPGRID 2003)*, Budapest, Hungary, May 2003.
- [Castano and Ferrari, 2003] S. Castano, E. Ferrari, "Protecting Datasources over the Web: Policies, Models, and Mechanisms", in *Web-Powered Databases*, ISBN: 1-59140035-X, pages 299-330, Idea Group, 2003.
- [Castano et al., 2001] S. Castano, V. De Antonellis, and S. De Capitani Di Vimercati, "Global Viewing of Heterogeneous Data Sources", *IEEE Transaction on Knowledge and Data Engineering*, 13(2), March/April 2001.
- [Castano et al., 2003a] S. Castano, A. Ferrara, S. Montanelli, and D. Zucchelli, "HELIOS: a General Framework for Ontology-based Knowledge Sharing and Evolution in P2P Systems", in *Proc. of the 2nd IEEE DEXA Int. Workshop on Web Semantics (WEBS 2003)*, Prague, Czech Republic, September 2003.
- [Castano et al., 2003b] S. Castano, A. Ferrara, and S. Montanelli, "H-MATCH: an Algorithm for Dynamically Matching Ontologies in Peer-based Systems", in *Proc. of the 1st VLDB Int. Workshop on Semantic Web and Databases (SWDB 2003)*, Berlin, Germany, September 2003.
- [Castano et al., 2004a] S. Castano, A. Ferrara, S. Montanelli, and G. Racca, "Matching Techniques for Resource Discovery in Distributed Systems Using Heterogeneous Ontology Descriptions", in *Proc. of the IEEE Int. Conference on Coding and Computing (ITCC 2004)*, Las Vegas, Nevada, USA, April 2004.
- [Castano et al., 2004b] S. Castano, A. Ferrara, S. Montanelli, and G. Racca, "Semantic Information Interoperability in Open Networked Systems", in *Proc. of the Int. Conference on Semantics of a Networked World (ICSNW)*, in cooperation with ACM SIGMOD 2004, Paris, France, June 2004.
- [Castano et al., 2004c] S. Castano, A. Ferrara, S. Montanelli, E. Pagani, G.P. Rossi, and S. Tebaldi, "On Combining a Semantic Engine and Flexible Network Policies for P2P Knowledge Sharing Networks", in *Proc. of the 1st IEEE DEXA Workshop on Grid and Peer-to-Peer Computing Impacts on Large Scale Heterogeneous Distributed Database Systems (GLOBE 2004)*, Zaragoza, Spain, September 2004.
- [Castano et al., 2004d] S. Castano, A. Ferrara, S. Montanelli, and G. Racca, "From Surface to Intensive Matching of Semantic Web Ontologies", in *Proc. of the 3rd IEEE DEXA Int. Workshop on Web Semantics (WEBS 2004)*, Zaragoza, Spain, September 2004.
- [Castano et al., 2005] S. Castano, A. Ferrara, and S. Montanelli, "Ontology-based Interoperability Services for Semantic Collaboration in Open Networked Systems", in *Proc. of the 1st Int. Conference on Interoperability of Enterprise Software and Applications (INTEROP-ESA 2005)*, Geneva, Switzerland, February 2005.
- [Crespo and Garcia-Molina, 2003] A. Crespo and H. Garcia-Molina, "Semantic Overlay Networks for P2P Systems", Stanford University, Technical Report, 2003.
- [Doan et al., 2002] A. Doan, J. Madhavan, P. Domingos, and A. Halevy, "Learning to Map between Ontologies on the Semantic Web", in *Proc. of the 11th Int. World Wide Web Conference (WWW 2002)*, Honolulu, Hawaii, USA, May 2002.
- [Giunchiglia and Shvaiko, 2004] F. Giunchiglia, P. Shvaiko, "Semantic Matching", *The Knowledge Engineering Review Journal*, 3, pages 256-280, 2004.
- [Haase et al., 2004] P. Haase, R. Siebes, and F. van Harmelen, "Peer Selection in Peer-to-Peer Networks with Semantic Topologies", in *Proc. of the Int. Conference on Semantics of a Networked World (ICSNW)*, in cooperation with ACM SIGMOD 2004, Paris, France, June 2004.
- [Kagal et al., 2003] L. Kagal, T. Finin, and A. Johshi, "A Policy Language for Pervasive Computing Environment", in *Proc. of the 4th IEEE Int. Workshop on Policies for Distributed Systems and Networks*, Como, Italy, June 2003.
- [Khambatti et al., 2003] M. Khambatti, K. Dong Ryu, and P. Dasgupta, "Structuring Peer-to-Peer Networks Using Interest-Based Communities", in *Proc. of the 1st Int. Workshop on Databases, Information Systems, and Peer-to-Peer Computing (DBISP2P 2003)*, Berlin Germany, September 2003.

[Madhavan et al., 2002] J. Madhavan, P.A. Bernstein, P. Domingos, A. Halevy, "Representing and Reasoning about Mappings between Domain Models", in Proc. of the 18th National Conference on Artificial Intelligence and 14th Conference on Innovative Applications of Artificial Intelligence, Edmonton, Alberta, Canada, AAAI Press, 2002.

[Marti and Garcia-Molina, 2004] S. Marti and H. Garcia-Molina, "Limited Reputation Sharing in P2P Systems", in Proc. of the 5th ACM Conference on Electronic Commerce ACM-EC 2004, New York, NY, USA, May 2004.

[Martinez Perez et al., 2005] G. Martinez Perez, F.J. Garcia Clemente, and A.F. Gomez Skarmeta, "Policy-based Management of Web and Information Systems Security: an Emerging Technology", Idea Group Inc., 2005, In press.

[Nejdl et al., 2002] W. Nejdl et al., "EDUTELLA: a P2P Networking Infrastructure Based on RDF", in Proc. of the 11th Int. World Wide Web Conference (WWW 2002), Honolulu, Hawaii, USA, May 2002.

[OASIS, 2004] OASIS consortium, "Extensible Access Control Markup Language (XACML)", November 2004.

[Shvaiko, 2004] P. Shvaiko, "A Classification of Schema-based Matching Approaches", in Proc. of the 1st ISWC-04 Workshop on Meaning Coordination and Negotiation (MCN-04), 2004.

[Staab et al., 2004] S. Staab, C. Tempich, and A. Wranik, "REMINDIN: Semantic Query Routing in Peer-to-Peer Networks based on Social Metaphors", in Proc. of the 13th Int. conference on World Wide Web, WWW 2004, New York, NY, USA, May 2004.

[Stallings, 2000] W. Stallings, "Network Security Essentials: Applications and Standards", Prentice Hall, 2000.

[Sundaram and Chapman, 2002] B. Sundaram and B.M. Chapman, "XML-based Policy Engine Framework for Usage Policy Management in Grids", in Proc. of the Int. Workshop on GRID, Lecture Notes in Computer Science, 2536, 2002.

[Waterhouse et al., 2002] S.R. Waterhouse, D.M. Doolin, G. Kan, and Y. Faybishenko, "Distributed Search in P2P Networks", IEEE Internet Computing, 6(1), pages 68-72, 2002.

2.5 Descrizione del programma e dei compiti dell'Unità di Ricerca

Testo italiano

Obiettivo dell'attività del gruppo di ricerca dell'Università di Milano (UNIMI) è lo sviluppo di metodi, tecniche e strumenti per la costituzione guidata da consenso di comunità semantiche in sistemi P2P, al fine di supportare efficacemente la condivisione di risorse distribuite e di collaborazione semantica in generale.

La costituzione guidata da consenso di comunità semantiche di peer e la disponibilità di tecniche per la propagazione di query su base semantica rappresenta una sfida nello stadio attuale di sviluppo di sistemi aperti a rete. Questo diventa ancora più rilevante in contesti di semantica emergente, dove requisiti di collaborazione semantica fra peer che si manifestano in maniera spontanea e imprevista richiedono capacità dinamiche di negoziazione per la definizione di accordi su interpretazioni comuni nell'ambito di un certo task. Questo è tipico ad esempio di contesti basati su peer e grid semantico, caratterizzati da un insieme di nodi paritetici senza reciproca conoscenza a priori e senza alcun grado di relazione mutua, che necessitano di cooperare attraverso la condivisione delle loro risorse (dati, documenti, servizi). A causa dell'elevata dinamicità e variabilità dei requisiti di collaborazione e condivisione, in questi contesti di collaborazione non sono previste autorità centralizzate che gestiscano descrizioni comuni delle risorse condivise da tutti i nodi del sistema. Piuttosto, ciascun peer fornisce la propria descrizione della conoscenza relativa alle risorse che condivide attraverso una sua ontologia (o peer ontology). I contesti di interesse saranno quindi caratterizzati da molteplici ontologie indipendenti. Nel progetto, si considera uno scenario applicativo che coinvolge personale medico di diversi paesi a livello internazionale che hanno esigenze di collaborare e condividere informazioni mediche al fine di migliorare i processi decisionali e ridurre le probabilità di errore. Un medico vorrebbe poter specificare la malattia del paziente e ottenere le informazioni rilevanti inerenti, dai geni che possono contribuire a causarla fino ai sintomi e alle possibili cure. Questo richiede la capacità di identificare a livello mondiale una serie di peer diversi e sparsi e non conosciuti a priori che sono in possesso di informazioni rilevanti di interesse.

Allo scopo di facilitare la scoperta di risorse e l'elaborazione di interrogazioni in questi contesti fortemente dinamici e aperti, la formazione di comunità autonome e auto-organizzanti di peer pone nuove problematiche da investigare, al fine di gestire il problema dell'elevato traffico di rete dovuto a singole interazioni fra peer e di fornire un meccanismo di coordinamento per il processamento e la propagazione delle richieste di risorse su base semantica, sfruttando le descrizioni ontologiche disponibili per le risorse da condividere. In particolare, al fine di rendere il processo scalabile e adeguato alle caratteristiche dei contesti aperti a rete, è fortemente richiesta la capacità di interagire peer-to-peer fra le diverse parti per la negoziazione di accordi locali e la capacità di costruire accordi globali (e quindi comunità) per aggregazione di accordi locali. Inoltre, la natura intrinsecamente aperta dei sistemi P2P, e quindi anche delle comunità, pone seri problemi circa la necessità di mantenere informazioni sulla reputazione dei membri di una comunità. A tal fine, sono necessarie politiche e meccanismi per la specifica delle regole sotto cui deve avvenire la condivisione e sotto cui un peer dichiara la propria disponibilità a processare le richieste che riceve.

Il gruppo di ricerca di UNIMI affronterà le problematiche di ricerca descritte in precedenza sviluppando un framework unificato, basato su servizi per la costituzione guidata da consenso di comunità semantiche di peer per supportare efficacemente la collaborazione semantica e la scoperta di risorse di interesse. Il framework sarà basato su: i) tecniche di aggregazione/negoziazione, che sfruttano descrizioni ontologiche di risorse e tecniche di ontology matching per la costituzione, mantenimento e scioglimento di comunità semantiche di peer, ii) politiche di condivisione, per la specifica di disponibilità condizionata di un peer a rispondere a richieste entranti e di condivisione condizionata di risorse messe a disposizione nell'ambito di una comunità semantica e iii) tecniche e politiche di reputazione/propagazione per la distribuzione efficace delle richieste su

base semantica al fine di realizzare la condivisione coordinata di risorse distribuite all'interno di una comunità semantica confermata (committed) con le sue politiche. Tali tecniche e politiche saranno fornite sottoforma di servizi Web sulla architettura P2P di ESTEEM e dimostrate sull'applicazione di collaborazione scientifica in ambito medico.

Le attività di ricerca di UNIMI sono organizzate in due fasi come segue.

Fase 1

Attività 1.1

La prima attività sarà dedicata allo studio dello stato dell'arte e all'identificazione dei problemi aperti che verranno affrontati con lo sviluppo delle tecniche per la formazione di comunità semantiche in contesti P2P. Particolare attenzione sarà posta ai recenti sviluppi nell'area dei sistemi P2P basati su schemi, che sfruttano metadati basati su RDF(S) o ontologie tematiche per la descrizione delle risorse. I sistemi P2P puri e schema-based saranno passati in rassegna criticamente con il duplice obiettivi di: (i) analizzare criticamente e classificare le funzionalità di base attualmente messe a disposizione per la gestione di comunità di peer, specialmente comunità basate su interessi dei peer, discutendone i limiti dal punto di vista degli aspetti semantici; (ii) analizzare criticamente e classificare le attuali tecniche di propagazione di query. Entrambe le tecniche ai punti (i) and (ii) saranno analizzate con lo scopo specifico di evidenziare il ruolo delle ontologie (in molti casi molto limitato, se non assente) nella gestione degli aspetti semantici per i diversi scopi.

Ulteriore lavoro di rassegna di UNIMI in questa prima attività riguarda l'analisi critica dei modelli e politiche esistenti per l'accesso controllato e selettivo a risorse informative in sistemi distribuiti su Web. Tale analisi sarà rivolta ad evidenziarne i punti di forza e debolezza, alla luce dei nuovi requisiti posti dalla definizione di politiche di condivisione per comunità semantiche in contesto P2P.

Risultati dell'Attività 1.1

D2.1: Rapporto tecnico che illustra lo stato dell'arte su (i) attuali tecniche e meccanismi per la gestione di comunità di in reti P2P, (ii) propagazione di query in sistemi P2P, e (iii) politiche e modelli per accesso selettivo e disseminazione controllata di risorse informative in sistemi distribuiti aperti su Web.

Attività 1.2

UNIMI eseguirà questa attività in collaborazione con gli partner del progetto. L'attività riguarda la raccolta di requisiti dettagliati per lo scenario applicativo di ESTEEM.

La raccolta dei requisiti verrà eseguita mediante interviste e questionari dai partner del progetto.

Risultati dell'Attività 1.2

DALL.1 (prodotto congiuntamente da tutte le unità): Rapporto tecnico che descrive i requisiti raccolti per lo scenario applicativo.

Attività 1.3

UNIMI svilupperà metodi e tecniche evolute per la formazione dinamica, basata su consenso di comunità semantiche di peer intorno a comuni interessi per scopi di collaborazione semantica e scoperta di risorse. A tale scopo, UNIMI svilupperà tecniche di negoziazione/agggregazione, che sfruttano descrizioni ontologiche delle risorse dei vari peer e tecniche di matching dinamico di ontologie per la costituzione, mantenimento e scioglimento di comunità semantiche di peer, sulla base di criteri ristretti di appartenenza e regole di funzionamento. Sarà obiettivo delle tecniche di negoziazione quello di gestire i significati delle risorse e le relative politiche di condivisione (si veda l'Attività 1.4) specificate da un peer al fine di raggiungere un consenso e un accordo condiviso sul significato e sulle politiche nella comunità in questione. Obiettivo delle tecniche di aggregazione l'organizzazione i membri di una comunità confermata secondo una organizzazione strutturata (ad es., struttura gerarchica) per propagazione di query e condivisione di risorse su base semantica. In questo ambito, si utilizzeranno ed adatteranno le tecniche sviluppate da UNIMI nell'ambito del sistema HELIOS per il matching dinamico di ontologie in contesti aperti, al fine di organizzare i membri di una comunità semantica in base al livello di matching delle descrizioni ontologiche delle risorse che hanno concordato di condividere nella comunità. Come risultato, si stabiliranno mapping fra le ontologie dei membri di una comunità semantica, che riflettono gli accordi globali della comunità. Obiettivo delle politiche e tecniche di reputazione/negoziazione quello di definire strategie e regole per l'instradamento efficace e mirato delle richieste a livello intra e inter comunità, considerando l'organizzazione strutturata della comunità e le politiche di condivisione (si veda l'Attività 1.4). In particolare, le politiche di propagazione sfrutteranno le primitive delle reti di overlay semantico dell'infrastruttura P2P infrastruttura sviluppata nel progetto dall'unità UNIROMA per definire regole di routing semantico per indirizzare richieste a un insieme di destinatari che condividono risorse rilevanti (cioè, a vicini semantici) e che possiedono la reputazione di esperti sugli argomenti indicati nella richiesta di risorse. In questo senso, lo scopo delle politiche e delle tecniche di reputazione è quello di fornire modelli e tecniche per asserire il livello di affidabilità e competenza di un peer rispetto ad un certo argomento e di mantenere e scambiare tali informazioni all'interno di una comunità e fra comunità diverse, anche durante il processo di formazione di una comunità.

Risultati dell'Attività 1.3

D2.2: Rapporto tecnico sul framework unificato per la costituzione guidata da consenso e la gestione di comunità semantiche, con descrizione delle tecniche di negoziazione/agggregazione per la gestione del consenso e la formazione di accordi e delle tecniche e politiche di reputazione/propagazione per la propagazione semantica di richieste.

Attività 1.4

Negli attuali sistemi P2P si assume che i peer siano sempre disponibili a rispondere a richieste in arrivo da altri peer in qualunque

momento. Tuttavia, questa ipotesi non è sempre vera, poiché i peer possono voler restringere la propria disponibilità in base a condizioni che devono essere soddisfatte che dipendono da diversi fattori quali condizioni temporali (ad es., l'istante di arrivo di una richiesta) oppure condizioni sullo stato interno del peer (ad es., carico di lavoro). UNIMI definirà un linguaggio basato su ontologie per la specifica dichiarativa di politiche di condivisione per esprimere la disponibilità condizionata all'elaborazione di richieste in arrivo e la condivisione condizionata delle risorse all'interno di una comunità semantica. L'uso di un linguaggio ontologico promuoverà l'interoperabilità semantica nella gestione di politiche all'interno e fra comunità. Le politiche consentiranno di specificare regole e restrizioni sulle risorse che un peer proprietario condivide nella comunità e sulle sue disponibilità di processamento di richieste. Le politiche si applicheranno sia a livello di comunità, per specificare il comportamento comune di tutta la comunità, sia a livello di peer, per restringere ulteriormente o sovrascrivere le politiche di comunità. Le politiche consentiranno di considerare, separatamente o in combinazione, un ampio insieme di restrizioni alla condivisione e disponibilità per adattarsi ai requisiti di flessibilità e dinamicità tipici dei sistemi aperti P2P, quali ad esempio, condizioni temporali, condizioni relative al tipo di connessione di rete, condizioni sullo stato interno di un peer, credenziali e preferenze degli utenti. Ad esempio, considerando l'applicazione di collaborazione scientifica in ambito medico, sarà possibile specificare una politica per la condivisione di dati clinici solamente durante l'orario d'ufficio e solamente da utenti che dimostrano di essere medici ospedalieri.

Risultati dell'Attività 1.4

D2.3: Rapporto tecnico sul linguaggio di specifica e sulle politiche di condivisione e la loro applicazione a comunità semantiche.

Attività 1.5

Questa attività riguarda la specifica di alto livello dell'architettura ESTEEM, definita in collaborazione con gli partner del progetto. L'architettura fornirà un insieme di servizi per l'accesso integrato e trust-aware ai dati distribuiti presso i vari peer di una comunità semantica, un insieme di servizi per la scoperta e il matching di servizi all'interno della comunità, e un insieme di servizi per la formazione guidata da consenso di comunità semantiche e la loro gestione. In particolare, l'attività di UNIMI sarà dedicata alla definizione delle interfacce WSDL interfaces dell'insieme di servizi per la formazione di comunità semantiche e la distribuzione semantica di richieste.

Risultati dell'Attività 1.5

DALL.2 (prodotto congiuntamente da tutte le unità): Rapporto tecnico di definizione dell'architettura di alto livello di ESTEEM, e della specifica WSDL di tutti i servizi forniti.

Fase 2

Attività 2.1

Obiettivo di questa attività di UNIMI è la specifica dell'architettura di dettaglio dei servizi di supporto per la formazione guidata da consenso di comunità semantiche e la loro gestione e per la distribuzione semantica di richieste su un'infrastruttura P2P, usando peer ontology come spazi di conoscenza per la descrizione semantica delle risorse che i peer dichiarano di condividere e tecniche per il matching dinamico di ontologie per interoperabilità semantica nella condivisione di risorse distribuite. Questa attività progettuale coinvolge lo sviluppo di un mock-up per i servizi sviluppati da UNIMI a supporto di comunità semantiche e distribuzione semantica di richieste basati su politiche di condivisione. UNIMI sarà coinvolta con gli altri partner in processo comune di testing, specificamente per gli aspetti legati ai propri prototipi. I risultati di questo test saranno considerati nella successiva Attività 2.2.

Risultati dell'Attività 2.1

DALL.3 (prodotto congiuntamente da tutte le unità): Rapporto tecnico di definizione dell'architettura di dettaglio di ESTEEM.
DALL.4 (prodotto congiuntamente da tutte le unità): Prototipi mock-up e dati raccolti durante la fase di testing.

Attività 2.2

In questa attività, UNIMI implementerà i propri servizi; l'implementazione si baserà sulle specifiche prodotte nelle attività precedenti. L'implementazione si baserà su Web service e tecnologie grid.

Risultati dell'Attività 2.2

D2.4: Implementazione dei servizi per formazione guidata da consenso di comunità semantiche e loro gestione e distribuzione semantica di richieste in sistemi P2P, considerando politiche di condivisione.

Attività 2.3

UNIMI eseguirà questa attività in collaborazione con gli altri partner del progetto. L'attività riguarda l'implementazione di un dimostratore integrato che mostri l'uso e la fattibilità della piattaforma ESTEEM in un caso reale. L'implementazione del dimostratore sarà basata sui servizi dell'architettura ESTEEM. UNIMI avrà il compito di sviluppare i moduli del dimostratore che riguardano la formazione guidata da consenso di comunità semantiche e la distribuzione semantica di richieste P2P.

Risultati dell'Attività 2.3

DALL.5 (prodotto congiuntamente da tutte le unità): Rapporto tecnico che descrive il dimostratore integrato e i risultati della sperimentazione nell'ambito dello scenario applicativo.

Testo inglese

Main goal of UNIMI research activity is the development methods, techniques, and tools for consensus-driven formation of semantic communities and their management in multi-knowledge environments to enforce distributed resource sharing and semantic

collaboration in an effective way. The consensus-driven formation and management of semantic communities and the availability of advanced techniques for query propagation on a semantic basis is a challenging issue in the current stage of development of open networked systems. This is even more true in emergent semantics contexts, where the emergence of collaboration requirements among peers requires dynamic capabilities of negotiating agreements on common interpretations within the context of a given task. This is typical for instance of semantic Grids and peer-based systems, characterized by a set of independent peer parties without prior reciprocal knowledge and no degree of relationship, that dynamically need to cooperate by sharing their resources (such as data, documents, services). These collaboration contexts considered in the project are multi-knowledge, in that no centralized authorities are defined managing a comprehensive view of the resources shared by all the nodes in the system, due to the high dynamicity and variability of collaboration and sharing requirements; rather that each peer is responsible of providing the knowledge description of the resources to be shared through its own ontology. An application scenario is given involving medical personnel across several countries that need to share electronic medical records, in order to enhance decision-making processes and decrease medical errors. A doctor would like to input a patient's illness and obtain all relevant information about it, from the genes that may contribute to causing it, to the symptoms and possible treatments. This requires the identification of worldwide disparate peers that can provide information resources of interest, and that are not known in advance.

In order to facilitate resource discovery and query processing processes in such highly dynamic and open contexts, the formation of autonomous and self-organizing communities of parties poses new issues to be investigated, to handle the problem of high network traffic due to single-peer interactions and to provide a coordination mechanism for processing and forwarding resource query requests on a semantic basis, by exploiting available ontologies describing resources to be shared. In particular, the possibility for the parties to interact peer-to-peer to negotiate local agreements and the capability to build global agreements (and thus communities) by aggregation of local agreements is strongly required, to make the process scalable and suited to open networked contexts. Furthermore, the intrinsically open nature of P2P systems, and thus of the communities, poses serious issues regarding the maintenance of reputation information about the members of the communities and requires policies and mechanisms to specify the rules regulating resource sharing and the conditions under which a peer is available to process incoming queries.

The UNIMI unit will address the research problems described above by developing a unified service-based framework for the consensus-driven formation of dynamic semantic communities among disparate peer parties for effective semantic collaboration and resource discovery. The framework will be based on: i) aggregation/negotiation techniques, which exploit ontological resource descriptions and ontology matching in order to form, maintain, and, disband semantic communities of peers, ii) sharing policies for the specification of conditional availability of the peers in answering incoming requests and of conditional resource sharing within a semantic community and iii) reputation/propagation techniques and policies for effective distribution of resource requests on a semantic basis to enforce coordinated sharing of distributed resources within a committed community, with its policies. Such techniques and related policies will be deployed as Web services on the ESTEEM P2P middleware architecture and demonstrated in a given scientific collaboration application in medicine.

The research activities of UNIMI are articulated in 2 phases as follows.

Phase 1

Activity 1.1

The first activity will be concerned with the study of the state of the art and identification of open problems to be addressed for semantic community formation and management and controlled resource sharing in multi-ontology P2P contexts. Particular attention will be given to recent developments in the area of schema-based P2P systems which exploit RDF(S)-based metadata schemas or thematic ontologies for resource description. Basic and schema-based P2P systems will be critically surveyed with the twofold goal of (i) critically analyzing and classifying the current basic functionalities commonly available for community management, specially interest-based P2P communities, by discussing their limitations specifically from the semantics point of view and (ii) critically analyzing and classifying current query propagation functionalities. Both techniques at (i) and (ii) will be analyzed with the purpose of putting in evidence the actual role (in most cases very limited, if none) of ontologies for management of semantic aspects for each respective functionality.

Further work of UNIMI in this first activity will be devoted to the critical analysis of existing policies and models for selective and controlled access to information resources in open distributed systems over the Web. Such analysis will be devoted to highlight their weaknesses and strengths in view of the new requirements posed by the definition of sharing policies for semantic communities in P2P context.

Result of activity 1.1

D2.1: Technical report illustrating the state of the art on (i) the existing techniques and mechanisms for community management in P2P networks, (ii) query propagation in P2P systems, and (iii) access control policies and models for controlled and selective access and dissemination of information resource in open distributed systems over the web.

Activity 1.2

UNIMI will perform this activity jointly with other partners. The activity will consist of the collection of detailed requirements for the ESTEEM application scenario (scientific collaboration in medicine).

The collection of requirements will be done by interviews and questionnaires to the user partners of the project.

Results of Activity 1.2

DALL.1 (jointly developed by all partners): Technical Report illustrating all the requirements collected for the application scenario.

Activity 1.3

UNIMI will develop advanced methods and techniques for the dynamic, consensus-driven formation of semantic communities of peers around common interests for semantic collaboration and resources discovery. To this end, UNIMI will develop negotiation/aggregation techniques which will exploit ontological resource descriptions provided by the parties and dynamic ontology techniques in order to form, maintain, and disband aggregations of parties based on restricted membership criteria and rules of operation. It will be the goal of negotiation techniques to be developed to deal with the different resource meanings and sharing policies (see Activity 1.4) provided by different peers in order to gain consensus on commonly agreed sharing policy and resource meaning within the community under formation. It will be goal of the aggregation techniques to organize the members of a committed community according to a structured organization (e.g., hierarchical structure) for effective query propagation and resource sharing on a semantic basis. To this end, techniques developed by UNIMI in the framework of the HELIOS project for dynamic matching of heterogeneous peer ontologies will be used and adapted to this context, in order to organize the semantic community members according to the level of matching of ontology descriptions of the resources they agreed to share in the community. As a result, mappings between ontologies of community members will be discovered and maintained in the structured community organization, that reflect community agreements. It will be the goal of the reputation/propagation policies to define strategies and rules for resource query requests dissemination at intra/inter community level, by taking into consideration the committed communities structured organizations and sharing policies (see Activity 1.4). In particular, the propagation policies will exploit the semantic overlay network primitives of the P2P infrastructure developed by UNIROMA in the project to define semantic routing rules for addressing each resource request to a set of recipients which share relevant resources (i.e., semantic neighbours) and which are trusted to be expert on the topics expressed by the resource request. In this sense, the goal of the reputation policies is to provide models and techniques for assessing the level of reliability and expertise of a peer party with respect to a specific topic, and to allow the maintenance and exchange of such reliability information within and across semantic communities, and also during the community formation process.

Results of Activity 1.3

D2.2: Technical report on the framework for consensus-driven formation and management of semantic communities, provided with techniques for negotiation/aggregation for consensus management and agreement negotiation and reputation/propagation policies and techniques for semantic query propagation.

Activity 1.4

Current P2P systems assume that peers are always willing to answer incoming queries from any peer, at any time. This assumption however is not always true, as peers might want to restrict their availability to process incoming requests according to the satisfaction of given conditions. UNIMI will define an ontology-based language to specify sharing policies for conditional availability of the peers in answering incoming requests and for conditional resource sharing within a semantic community. The ontology-based language will enable semantic interoperability in policy management within and across communities. Sharing policies will be able to specify the sharing restrictions and rules that must hold on the resources that a owner party shares in a semantic community as well as the restrictions and rules on its capability to process incoming requests. Such policies can apply either at the community level, to state a common behavior of the community, or at the peer level, to further restrict community policies. Sharing policies will enable the specification, separately or in combination, of a wide range of conditional sharing restrictions to cope with the flexibility and dynamicity requirements typical of open, peer-based contexts, ranging from temporal conditions, to conditions on the internal state of the receiving peer, to conditions on the kind of network connection, to conditions on the credentials of the requesting user. For example, if we consider the semantic community formed to enable scientific collaboration in the medicine application, it will be for instance possible to specify a policy for the sharing of clinical data only during office time and only by users proven to be hospital doctors.

Results of Activity 1.4

D2.3: Technical report on the specification formalisms and on sharing policies and their application to semantic communities.

Activity 1.5

This activity will concern with the high-level specification of the ESTEEM architecture, to be defined in cooperation with the other partners of the project. The architecture will provide a set of services for integrated and trust-aware access to data distributed over peers of a semantic community, a set of services for the discovery and matching of web services within the community, and a set of services for consensus-driven formation and management of semantic communities. In particular, the UNIMI activity will be devoted to the definition of the WSDL interfaces of the set of services for consensus-driven semantic community formation and semantic query distribution.

Results of Activity 1.5

DALL.2 (jointly developed by all partners): Technical report defining the high-level architecture of ESTEEM, and the WSDL specification of all the provided services.

Phase 2

Activity 2.1

Goal of UNIMI activity is the design of a detailed architecture of support services for consensus-driven formation and management of semantic communities and for semantic query distribution in a P2P infrastructure, using peer ontologies as knowledge spaces for semantically describing resources that peers are willing to share, and dynamic ontology matching techniques for semantic interoperability in distributed resource sharing. The design activity will involve the development of a mock-up for the services to be developed by UNIMI for community formation and management and for query propagation, based on sharing policies. UNIMI will be involved with other partners in a common testing process, specifically for the aspects related to his prototypes. The results of this test shall be fully taken into account in the next activity 2.2.

Results of Activity 2.1

DALL.3 (jointly developed by all partners): Technical report defining the detailed architecture of ESTEEM.

DALL.4 (jointly developed by all partners): Mock-up prototypes and data collected during the testing phase.

Activity 2.2

In this activity, UNIMI will implement its services; the implementation will take place based on the specification and design developed in the previous activities. The implementation will be based on Web services and grid technologies.

Results of Activity 2.2

D2.4: Implementation of a set of services for consensus-driven semantic communities formation and their management and for semantic query distribution in a P2P environment based on sharing policies.

Activity 2.3

UNIMI will perform this activity jointly with other partners. The activity will consist of the implementation of an integrated demonstrator, showing the usage and feasibility of the ESTEEM platform in a real case. The implementation of the ESTEEM demonstrator will be based on the services of the ESTEEM architecture. UNIMI will be in charge of developing the modules of the demonstrator involving the consensus-driven formation of semantic communities issues.

Results of Activity 2.3

DALL.5 (jointly developed by all partners): Technical report illustrating the integrated demonstrator, as well as the results of the experimentation in the context of the application scenario.

2.6 Descrizione delle attrezzature già disponibili ed utilizzabili per la ricerca proposta con valore patrimoniale superiore a 25.000 Euro

Testo italiano

Nessuna

Testo inglese

Nessuna

2.7 Descrizione delle Grandi attrezzature da acquisire (GA)

Testo italiano

Nessuna

Testo inglese

Nessuna

2.8 Mesi uomo complessivi dedicati al programma

Testo italiano

		Numero	Mesi uomo 1° anno	Mesi uomo 2° anno	Totale mesi uomo
Personale universitario dell'Università sede dell'Unità di Ricerca		4	17	17	34
Personale universitario di altre Università		0	0	0	0
Titolari di assegni di ricerca		0			
Titolari di borse	Dottorato	1	4	4	8
	Post-dottorato	0			
	Scuola di Specializzazione	0			

Personale a contratto	<i>Assegnisti</i>	1	11	11	22
	<i>Borsisti</i>	0			
	<i>Dottorandi</i>	0			
	<i>Altre tipologie</i>	0			
Personale extrauniversitario		0			
TOTALE		6	32	32	64

Testo inglese

		Numero	Mesi uomo 1° anno	Mesi uomo 2° anno	Totale mesi uomo
University Personnel		4	17	17	34
Other University Personnel		0	0	0	0
Work contract (research grants, free lance contracts)		0			
PHD Fellows & PHD Students	<i>PHD Students</i>	1	4	4	8
	<i>Post-Doctoral Fellows</i>	0			
	<i>Specialization School</i>	0			
Personnel to be hired	<i>Work contract (research grants, free lance contracts)</i>	1	11	11	22
	<i>PHD Fellows & PHD Students</i>	0			
	<i>PHD Students</i>	0			
	<i>Other tipologies</i>	0			
No cost Non University Personnel		0			
TOTALE		6	32	32	64

PARTE III

3.1 Costo complessivo del Programma dell'Unità di Ricerca

Testo italiano

Voce di spesa	Spesa in Euro	Descrizione
Materiale inventariabile	9.000	Acquisizione di stazioni di lavoro e periferiche
Grandi Attrezzature		
Materiale di consumo e funzionamento	7.000	Oneri per la gestione amministrativa del fondo e materiale vario accessorio per ufficio e laboratorio
Spese per calcolo ed elaborazione dati		
Personale a contratto	40.000	Istituzione di un assegno di ricerca di durata biennale sui temi del progetto (progettazione e implementazione di prototipi e servizi)
Servizi esterni		
Missioni	15.000	Missioni per riunioni di progetto, seminari e convegni/workshop nazionali e internazionali inerenti il progetto
Pubblicazioni		
Partecipazione / Organizzazione convegni	5.000	Spese per partecipazione/organizzazione di convegni e workshop inerenti il progetto
Altro		
TOTALE	76.000	

Testo inglese

Voce di spesa	Spesa in Euro	Descrizione
Materiale inventariabile	9.000	PC workstations and peripherals
Grandi Attrezzature		
Materiale di consumo e funzionamento	7.000	Fund management expenses plus consumables (e.g., toner, CD)
Spese per calcolo ed elaborazione dati		
Personale a contratto	40.000	Research position for prototype tool and service design and implementation
Servizi esterni		
Missioni	15.000	Travel expenses for participation to project meetings, seminars and international/national conferences/workshops related to the project
Pubblicazioni		
Partecipazione / Organizzazione convegni	5.000	Participation to/organization of conferences and workshops related to the project
Altro		
TOTALE	76.000	

3.2 Costo complessivo del Programma di Ricerca

		Descrizione
Costo complessivo del Programma dell'Unità di Ricerca	76.000	
Fondi disponibili (RD + RA) comprensivi dell'8% max per spese di gestione	22.800	Finanziamento dell'Università di Milano disponibile in caso di accettazione della domanda
Cofinanziamento di altre amministrazioni		
Cofinanziamento richiesto al MIUR	53.200	

3.3.1 Certifico la dichiarata disponibilità e l'utilizzabilità dei fondi di Ateneo (RD e RA)

SI

(per la copia da depositare presso l'Ateneo e per l'assenso alla diffusione via Internet delle informazioni riguardanti i programmi finanziati e la loro elaborazione necessaria alle valutazioni; D. Lgs, 196 del 30.6.2003 sulla "Tutela dei dati personali")

Firma _____

Data 05/04/2005 ore 20:38