

PROGETTO DI UNA UNITÀ DI RICERCA - MODELLO B
Anno 2005 - prot. 2005095190_001

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

CATARCI **TIZIANA**

Professore Ordinario *06/11/1961* *CTRTZN61S46H501Y*

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"

06-49918331 *06-49918331* *catarci@dis.uniroma1.it*
(Prefisso e telefono) *(Numero fax)* *(Indirizzo posta elettronica)*

1.5 Curriculum scientifico del Responsabile Scientifico dell'Unità di Ricerca

Testo italiano

Tiziana Catarci ha conseguito il Dottorato di Ricerca in Informatica all'Università di Roma "La Sapienza", dove è attualmente professore ordinario.

I suoi maggiori interessi di ricerca sono rivolti ad aspetti teorici ed applicativi relativi al progetto di basi di dati, alle basi di dati eterogenee, ai sistemi informativi cooperativi, ai linguaggi di interrogazione, ai formalismi visuali per basi di dati, alle interfacce utente, all'usabilità e all'accesso al Web. Su questi argomenti ha pubblicato oltre 100 articoli sulle più importanti riviste e conferenze internazionali e nazionali ed ha scritto od editato 14 libri.

Ha partecipato ad importanti progetti nazionali ed internazionali su sistemi informativi e sistemi visuali per l'interrogazione di basi di dati. Progetti europei particolarmente rilevanti sono KIM, VENUS, FADIVA, LAURIN, SEWASIE, DELOS e INTEROP, vertenti

sull'integrazione di sistemi informativi, l'accesso all'informazione, le biblioteche digitali e l'interazione uomo-macchina. Recenti progetti nazionali sono 'Daquincis' (Cofin 2001), su qualità dei dati in ambienti cooperativi, e "MAIS" (Firb 2001) su sistemi informativi multicanale e dispositivi mobili. Nel 2003 è vincitrice dell'IBM Eclipse Innovation Award per il progetto SEAWER, Semantic Annotation of web Resources.

Tiziana Catarci è stata spesso invitata a tenere relazioni nell'ambito di conferenze internazionali e presso prestigiose università ed istituti di ricerca italiani e stranieri. Per esempio, già nel 1988 tiene un seminario presso il Dipartimento di Computer Sciences and Engineering della University of Florida; Nel 2000 viene invitata ad essere il keynote speaker della conferenza internazionale GIScience 2000, il keynote speaker della Fourth International Conference on Flexible Query Answering Systems.

Da diversi anni svolge attività di revisione per le più importanti riviste scientifiche dei settori di interesse, incluse ACM TOCHI; ACM TOIS; VLDB Journal; IEEE TKDE; Data and Knowledge Engineering; JVLIC; IS; ACM Interact; Computing Surveys.

Nel Settembre 1996 è co-editor di un secondo numero speciale del JVLIC su Visual Query Languages. Nel Dicembre 1996 è co-editor di un numero speciale dell'ACM Sigmod RECORD su Information Visualization. Nel 2004 è editor di un numero speciale di Information Systems su qualità dei dati per sistemi informativi cooperativi e nel 2005 di un numero speciale del WWW Journal su Web Information Systems Engineering.

Dal 1998 fa parte del Comitato Editoriale dell'ACM SIGMOD Digital Symposium Collection (DiSC)

Dal 2000 fa parte del Comitato Editoriale di IEEE Multimedia.

Dal 2001 fa parte del Comitato Editoriale del World Wide Web Journal (Kluwer eds.).

Da diversi anni partecipa alle attività dei comitati di programma delle più importanti conferenze internazionali dei settori di interesse.

Nel 1996 è Program co-Chairperson del terzo Workshop on Advanced Visual Interfaces (AVI'96), Gubbio, Italia.

È Program co-Chairperson della quinta International Conference on Visual Database Systems (VDB5) che si è tenuta in Giappone nel Maggio 2000.

È Program co-Chairperson della quarta conferenza internazionale su Web Information System Engineering (WISE'03) che si è tenuta a Roma nel Dicembre 2003.

È Program co-Chairperson della terza conferenza internazionale su Ontologies, Databases and Applications of SEMantics (ODBASE2004), che si è tenuta a Cipro nell'Ottobre 2004.

È Tutorial Chair della 32nd International Conference on Very Large Databases (VLDB'06) che si terrà a Seoul nel 2006.

Dal 1992 ad oggi è nello Steering Committee delle Working Conferences on Advanced Visual Interfaces che si tengono ogni due anni in Italia.

Nel corso del 1996 è tra i fondatori dell'Italian Chapter dell'ACM SIGCHI (ACM Special Interest Group on Computer-Human Interaction).

Dal Giugno 1997 fa parte dell'IFIP Special Interest Group WG 2.6. (Database), di cui è il coordinatore dal gennaio 2005.

Dal 1997 al 2000 è il coordinatore del gruppo di studio, promosso dall'Autorità per l'Informatica nella Pubblica Amministrazione (AIPA), su Usabilità dei Sistemi Informativi nella Pubblica Amministrazione.

È il coordinatore del comitato organizzatore della 27th International Conference on Very Large Databases (VLDB'01), che si è tenuta a Roma nel Settembre 2001.

È nello Steering Committee della serie di Conferenze Internazionali IFIP su Data Semantics of a Networked World, la prima delle quali (Semantics for Grid Databases) si è tenuta a Parigi nel Giugno 2004.

Testo inglese

Tiziana Catarci received her PhD in Computer Science from the

University of Rome, where she is currently a Full Professor. Tiziana Catarci's main research interests are in theoretical and application oriented aspects of database design, cooperative information systems, data quality, visual formalisms for databases, user interfaces, usability, and Web access. She has published over 100 papers and 14 books in a variety of subjects comprising Heterogeneous and Cooperative Database Systems, Database Integration, Database and Web Access, User Interfaces for Databases, Controlled Studies and Usability Testing, Query Language Expressive Power. Particularly relevant are some European projects, including KIM, VENUS, FADIVA, LAURIN, SEWASIE, DELOS, INTEROP, aiming at investigating different aspects of database integration, digital libraries, interface design and human-computer interaction. Recent national projects are "Daquincis", on data quality, and "MAIS" on multichannel information access. Nel 2003 she has been the recipient of the IBM Eclipse Innovation Award for the SEAWER project on semantic annotation of web resources.

Tiziana Catarci has been often invited to give talks during important international conferences and at prestigious national and international universities and research centers. For instance, at the very beginning of her activity, in 1988 she was invited to give a talk at the Department of Computer Sciences and Engineering of the University of Florida; During 2000 she was the keynote speaker of the international conference GIScience 2000, and of the Fourth International Conference on Flexible Query Answering Systems.

Since many years Tiziana Catarci is one of the reviewers of the most important journals in her areas of research, including TOCHI; ACM TOIS; VLDB Journal; IEEE TKDE; Data and Knowledge Engineering; JVLIC; IS; ACM Interact; Computing Surveys.

She has been guest co-editor of two special issues of the Journal of Visual Languages and Computing (on "Visual Query Languages") and of ACM SIGMOD Record (on "Information Visualization"). In 2004 she will be editor of a special issue of Information Systems on data quality and a special issue of the WWW Journal on Web Systems Engineering.

Since 1998 she is in the editorial board of the SIGMOD Digital Symposium Collection (DiSC).

Since 2000 she is in the editorial board of IEEE Multimedia.

Since 2001 she is in the editorial board of the World Wide Web Journal (Kluwer eds.).

She is regularly involved in the program committees of the most important international conferences in her areas of interest.

She has been the program co-chair of the Fifth IFIP Working Conference on Visual Database Systems (VDB5) and has been the program co-chair of the International Workshop on Advanced Visual Interfaces (AVI) in 1996.

In 2003 she has been the program co-chair of the Int. Workshop on Data Quality in Cooperative Information Systems (DQCIS'03) and the program co-chair of the 4th Int. Conf. on Web Information System Engineering (WISE'03).

In 2004 she has been the program co-chair of the 3rd Int. Conf. on Ontologies, Databases and Applications of SEMantics (ODBASE2004).

She is Tutorial Chair of the 32nd International Conference on Very Large Databases (VLDB'06) which will be held in Seoul in 2006.

Since 1992 she is in steering committee of the Working Conferences on Advanced Visual Interfaces (AVI).

She is the chair of the Organizing Committee of the 27th International Conference on Very Large Databases (VLDB2001) that was

held in Roma in September 2001.

She is in steering committee of the series of IFIP Int. Conferences on Data Semantics of a Networked World, whose first edition has been held in Paris in 2004.

During 1996 she was among the founders of the Italian Chapter of the ACM SIGCHI (ACM Special Interest Group on Computer-Human Interaction).

Since 1997 she is part of the IFIP Special Interest Group WG 2.6. (Database), of which became the coordinator in January 2005.

Since 1997 she is the chair of the working group on Usability of Public Administration Information Systems promoted by the Italian Authority on Computer Systems in Public Administration.

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

1. CATARCI T., MILANO D., SCANNAPIECO, M. (2005). *Peer-to-Peer Data Quality Improvement in the DaQuinCIS System*. *JOURNAL OF DIGITAL INFORMATION MANAGEMENT*. vol. to appear ISSN: 0972-7272
2. BERTINI E, CATARCI T., DI BELLO L., KIMANI S. (2005). *Visualization in Digital Libraries*. In MATTHIAS HEMMJE, CLAUDIA NIEDEREE, THOMAS RISSE *From Integrated Publication and Information Systems to Information and Knowledge Environments* ISBN: 3-540-24551-0
3. CATARCI T. (2004). *Special issue on data quality in cooperative information systems*. *INFORMATION SYSTEMS*. vol. 29 ISSN: 0306-4379
4. CATARCI T., PAOLO DONGILLI, TANIA DI MASCIO, ENRICO FRANCONI, GIUSEPPE SANTUCCI, SERGIO TESSARIS (2004). *An ontology based visual tool for query formulation support*. *16th European Conference on Artificial Intelligence (ECAI2004)*.
5. D.MILANO, M.SCANNAPIECO, CATARCI T. (2004). *A Peer-to-Peer Service Supporting data Quality: Design and Implementation Issues*. *Proceedings dell'Int. Conference on Semantics of a Networked World*.
6. CATARCI T., KATIA P. SYCARA (2004). *Proceedings ODBASE 2004, LNCS 3290*. Springer Verlag.
7. D. MILANO, M. SCANNAPIECO, CATARCI T. (2004). *Quality-driven Query Processing of Xquery Queries*. *International Workshop on Data and Information Quality (DIQ2004)*.
8. BERTINO E, CATARCI T., Elmagarmid AK, Hacid MS (2003). *Quality of Service Specification in Video Databases*. *IEEE MULTIMEDIA*. vol. 10 ISSN: 1070-986X
9. LUCA DE SANTIS, MONICA SCANNAPIECO, CATARCI T. (2003). *Trusting Data Quality in Cooperative Information Systems*. *DOA/CoopIS/ODBASE 2003 Confederated International Conferences*.
10. KIMANI S, Iodi s, CATARCI T., SANTUCCI G, sartori c (2003). *VidaMine: A Visual data Mining Environment*. *JOURNAL OF VISUAL LANGUAGES AND COMPUTING*. ISSN: 1045-926X to appear.
11. CATARCI T., MECELLA M., MYLOPOULOS J., ORLOWSKA M. (2003). *Proceedings of the 4th International Conference on Web Information systems Engineering*. IEEE Press.
12. CATARCI T., MATARAZZO G, Raiss G (2002). *Driving Usability into the Public Administration: the Italian Experience*. *INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES*. vol. 57 pp. 121-138 ISSN: 1071-5819
13. S. FERNANDES SILVA, CATARCI T., U. SCHIEL (2002). *Formalizing Visual Interaction with Historical Databases*. *INFORMATION SYSTEMS*. vol. 27 ISSN: 0306-4379
14. S. KIMANI, CATARCI T., I. CRUZ (2002). *Effective Web Renderings*. In GEROIMENKO V, CHEN C. *Visualising the Semantic Web*: Springer Verlag (UNITED STATES).
15. CATARCI T., SANTUCCI G. (2001). *The prototype of the DARE system*. *ACM SIGMOD2001*.
16. CATARCI T. (2000). *What Happened When Database Researchers Met Usability*. *INFORMATION SYSTEMS*. vol. 25(3) pp. 177-212 ISSN: 0306-4379
17. D.CALVANESE, CATARCI T., G.SANTUCCI (2000). *Building a digital library of newspaper clippings: The LAURIN Project*. *ADL2000*.
18. CATARCI T., M.PASSERI, G.SANTUCCI (2000). *Querying Multiple Databases Dynamically on the World Wide Web*. *WISE2000*.
19. H.ARISAWA, CATARCI T. (2000). *Advances in Visual Information Management*. Kluwer Academic Publishers.
20. CATARCI T., G.SANTUCCI, L.TARANTINO (1999). *Emerging Issues in Visual Interfaces*. *KNOWLEDGE ENGINEERING REVIEW*. vol. 14 pp. 1-5 ISSN: 0269-8889
21. CATARCI T., CHANG S.K., NARDI D., SANTUCCI G. (1998). *WAG: Web-At-a-Glance*. *INTERNATIONAL JOURNAL OF COOPERATIVE INFORMATION SYSTEMS*. vol. 7 pp. 187-214 ISSN: 0218-8430
22. CATARCI T. (1997). *Interacting with Databases in the Global Information Infrastructure*. *IEEE COMMUNICATIONS MAGAZINE*. vol. 35 pp. 72-76 ISSN: 0163-6804
23. CATARCI T., CARDIFF J., SANTUCCI G. (1997). *Graphical Interaction with Heterogeneous Databases*. *VLDB JOURNAL*. vol. 6(2) pp. 97-120 ISSN: 1066-8888

24. CATARCI T., M.F.COSTABILE, S.LEVIALDI, C.BATINI (1997). *Visual Query Systems for Databases: A Survey*. JOURNAL OF VISUAL LANGUAGES AND COMPUTING. vol. 8 pp. 215-260 ISSN: 1045-926X
25. CATARCI T. (1996). *Databases and the Web: New Requirements for an Easy Access*. ACM COMPUTING SURVEYS. vol. 28A ISSN: 0360-0300
26. CATARCI T. (1996). *Interaction with Databases*. IEEE COMPUTER GRAPHICS AND APPLICATIONS. vol. 16 pp. 67-69 ISSN: 0272-1716
27. CATARCI T., S.K.CHANG, M.F.COSTABILE, S.LEVIALDI, G.SANTUCCI (1996). *A Graph-based Framework for Multiparadigmatic Visual Access to Databases*. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING. vol. 8 pp. 455-475 ISSN: 1041-4347
28. CATARCI T., D'ANGIOLINI G., LENZERINI M. (1995). *Statistical Data Modeling based on Concept. Logic*. DATA & KNOWLEDGE ENGINEERING. vol. 7 pp. 93-125 ISSN: 0169-023X
29. CATARCI T., L.TARANTINO (1995). *A Hypergraph-based Framework for Visual Interaction with Databases*. JOURNAL OF VISUAL LANGUAGES AND COMPUTING. vol. 6 pp. 135-166 ISSN: 1045-926X
30. CATARCI T., LENZERINI M. (1993). *Representing and Using Interschema Knowledge in Cooperative Information Systems*. INTERNATIONAL JOURNAL OF COOPERATIVE INFORMATION SYSTEMS. vol. 2(4) pp. 375-398 ISSN: 0218-8430

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.	CATARCI	Tiziana	Dip. INFORMATICA E SISTEMISTICA	Prof. Ordinario	ING-INF/05	6	6
2.	SANTUCCI	Giuseppe	Dip. INFORMATICA E SISTEMISTICA	Prof. Associato	ING-INF/05	4	4
3.	SALZA	Silvio	Dip. INFORMATICA E SISTEMISTICA	Prof. Associato	ING-INF/05	4	4
4.	ROSATI	Riccardo	Dip. INFORMATICA E SISTEMISTICA	Prof. Associato	ING-INF/05	4	4
5.	BALDONI	Roberto	Dip. INFORMATICA E SISTEMISTICA	Prof. Ordinario	ING-INF/05	4	4
TOTALE						22	22

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

n°	Cognome	Nome	Dipartimento	Data di inizio del contratto	Durata(in anni)	Mesi Uomo	
						1° anno	2° anno
1.	SCANNAPIECO	Monica	Dip. INFORMATICA E	01/06/2004	2	5	5

		SISTEMISTICA					
2. LEMBO	Domenico	Dip. INFORMATICA E SISTEMISTICA	01/10/2004	2		5	5
3. VIRGILLITTO	Antonino	Dip. INFORMATICA E SISTEMISTICA	01/06/2004	2		4	4
TOTALE						14	14

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.	milani	alessia	Dip. INFORMATICA E SISTEMISTICA	2003	3	Dottorato	4	4
2.	milano	diego	Dip. INFORMATICA E SISTEMISTICA	2003	3	Dottorato	4	4
3.	poggi	antonella	Dip. INFORMATICA E SISTEMISTICA	2003	3	Dottorato	4	4
4.	querzoni	leonardo	Dip. INFORMATICA E SISTEMISTICA	2003	3	Dottorato	4	4
TOTALE							16	16

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	
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

Semantica Emergente attraverso l'Integrazione Dati basata su Ontologie in Sistemi Peer-to-Peer con Requisiti di Credibilità delle Sorgenti

Testo inglese

Emergence of Semantics through Ontology-based Integration in Trust-aware Peer-to-Peer systems

2.2 Settori scientifico-disciplinari interessati dal Programma di Ricerca

ING-INF/05 - Sistemi di elaborazione delle informazioni

2.3 Parole chiave

Testo italiano

SEMANTICA EMERGENTE ; INTEGRAZIONE DI ONTOLOGIE ; PEER-TO-PEER ; RETI AD-HOC ; CREDIBILITA' ; LOCALIZZAZIONE DI RISORSE DECENTRALIZZATA ; QUALITA' DEI DATI ; ELABORAZIONE DI INTERROGAZIONI

Testo inglese

EMERGENT SEMANTICS ; ONTOLOGY INTEGRATION ; PEER-TO-PEER ; AD-HOC NETWORKING ; TRUST ; DECENTRALIZED RESOURCE LOCATION ; DATA QUALITY ; QUERY PROCESSING

2.4 Base di partenza scientifica nazionale o internazionale

Testo italiano

L'integrazione ed il coordinamento di contenuti eterogenei e distribuiti su sistemi informativi aperti operanti su Web, è una delle sfide cruciali allo stato attuale dell'evoluzione delle infrastrutture per l'IT. La condivisione di informazioni semantiche, piuttosto che l'elaborazione di informazione o il puro scambio sintattico alla XML, è uno dei principali obiettivi della tecnologia dell'informazione. La semantica riveste particolare importanza perché l'informazione deve poter essere condivisa in ambienti aperti dove nodi interagenti, che possono essere chiamati sorgenti, siti, agenti, ecc., e che per semplicità chiameremo peers, non condividono necessariamente una comprensione comune del mondo a disposizione. La mancanza di una base di conoscenza comune genera il bisogno di una guida esplicita alla comprensione dell'esatto significato dei dati, da cui emerge la crescente importanza delle ontologie. Inoltre, significativi scambi tra peers possono verificarsi solo sulla base di accordi affidabili e dinamici su interpretazioni comuni all'interno di un contesto specifico. La particolare forma di interoperabilità semantica descritta sopra viene definita come "semantica emergente" [Aberer et al., 2004].

Di seguito, specificheremo lo stato dell'arte nelle principali aree che riguardano le problematiche di ricerca della semantica emergente: (i) integrazione dei dati tradizionale, (ii) integrazione di informazione scorretta, (iii) integrazione dati basata su ontologie, (iv) integrazione dati peer-to-peer, (v) problematiche di credibilità e (vi) P2P overlay networks e Semantic Grid. (i) Integrazione Dati Tradizionale. L'obiettivo dell'integrazione dei dati è fornire ai clients un accesso virtuale ad un insieme di fonti dati eterogenee ed autonome. La ricerca sull'integrazione dei dati si è concentrata principalmente su approcci dichiarativi [Lenzerini, 2002] (anziché su approcci procedurali), dove il sistema di integrazione dati esporta verso l'utente una visione globale e riconciliata dei dati (definita schema globale o ontologia globale) sulla cui base vengono formulate le richieste dell'utente e su cui il sistema mantiene una specifica dichiarativa (o mapping) delle interrelazioni fra l'ontologia globale ed i dati delle fonti [Lenzerini, 2002]. Tra i vari problemi relativi all'integrazione dati, quello di rispondere ad interrogazioni poste sulla base dello schema globale è stato il più trattato [Duschka et al., 2000; Cali' et al., 2004, Calvanese et al., 1998; 2001].

(ii) Integrazione di Informazione Scorretta. L'integrazione di dati in presenza di informazione non corretta è stata trattata con metodi formali in [Galhardas et al., 2000; Calvanese et al., 1999]. La metodologia proposta all'interno del Progetto Europeo DWQ [Jarke et al., 2000; Calvanese et al., 1999] è basata sulla specifica di mapping per riconciliare dati in diverse sorgenti. Alcuni sistemi di integrazione proposti di recente che trattano il problema dei conflitti a livello di istanze sono: il sistema DaQuinCIS [Scannapieco et al., 2004], e il sistema interattivo proposto in [Sattler et al., 2003]. Anziché risolvere i conflitti tramite tecniche ad hoc, alcuni lavori hanno proposto l'uso di semantiche di riparazione da adottare per poter ottenere risposte consistenti da dati inconsistenti [Bravo and Bertossi, 2003; Cali' et al., 2003a; Cali' et al., 2003b].

(iii) Integrazione dati basata su ontologie. L'uso di ontologie ha assunto negli ultimi anni una crescente importanza per l'integrazione e coordinamento di dati. Tuttavia, lo stato dell'arte delle tecniche per l'integrazione basata su ontologie sembra ancora essere in una fase preliminare. In effetti, la maggior parte dei lavori svolti sulle ontologie ha riguardato quale linguaggio o quale metodo usare per costruire una ontologia integrata globale sulla base delle ontologie gestite da sorgenti locali [Broekstra et al., 2000; Doan et al., 2004], mentre il problema effettivo di come interrogare una ontologia integrata necessita di ulteriori studi.

(iv) Integrazione Dati Peer-to-Peer. Il problema dell'integrazione fra nodi informativi in sistemi distribuiti aperti è stato recentemente affrontato nel contesto del Peer-to-Peer data computing [Halevy et al., 2003]. Un sistema P2P è caratterizzato da una struttura costituita da vari nodi autonomi che mantengono l'informazione e che sono legati ad altri nodi attraverso dei mapping

mentre nessuna ontologia globale a cui il client possa riferirsi è disponibile. La mancanza di uno schema globale condiviso o di una ontologia rendono difficile per le parti che costituiscono il sistema il raggiungimento di un consenso globale sui dati semantici. Approcci iniziali si avvalgono di un corpus predefinito di termini che offrono un contesto iniziale per definire nuovi concetti [McCool and Guha, 2003] o fanno uso di gossiping e mapping a livello di locale per sostenere incrementalmente l'interoperabilità [Aberer et al., 2003], ma vi è ancora una fondamentale mancanza di comprensione sulle problematiche di base di integrazione dati in sistemi P2P. In effetti, dato che nessun singolo attore è responsabile dell'intero sistema, non è realistico assumere restrizioni sulla topologia complessiva dei mapping P2P [Halevy et al., 2003; Fagin et al., 2003]. Per cui, si deve considerare la possibilità che i mapping possano avere una struttura arbitraria, che include cicli fra i vari nodi. Ciò deve essere considerato sia dal punto di vista della modellazione del sistema e caratterizzazione della sua semantica (vedi ad es. [Catarci and Lenzerini, 1993; Halevy et al., 2003;] per una caratterizzazione in logica del primo ordine e [Calvanese et al., 2003;2004] per una proposta di semantica alternativa) che dal punto di vista del calcolo di risposte alle interrogazioni poste ad un peer, che comporta difficoltà nel distribuire la computazione complessiva sui nodi singoli.

(v) *Problematiche di Credibilità.* Un ulteriore importante aspetto è relativo alle misure per valutare il grado di consenso. Tali misure possono riferirsi alla valutazione della credibilità nei confronti di un agente che fornisce informazione e alla valutazione della qualità percepita dell'informazione ricevuta. La credibilità è tipicamente basata sul consenso fra opinioni reciproche che le persone detengono. Il sistema di punteggio di e-bay è un esempio in tal senso. Tuttavia, esso non prende in considerazione l'affidabilità e l'expertise dei valutatori. Vari lavori recenti hanno studiato meccanismi per stabilire il consenso sulla credibilità prendendo in considerazione la reputazione di chi fornisce i referaggi [Kamvar et al., 2003]. La qualità dell'informazione dipende dalle opinioni delle persone o dalle applicazioni che utilizzano tale informazione. Nel primo caso, il calcolo del consenso sarà una componente cruciale nel determinare la qualità dell'informazione [De Santis et al., 2003]. L'informazione di provenienza [Ehrig et al., 2003] consente di mettere in relazione l'informazione alle sue fonti o sorgenti, e da ciò al suo punteggio di credibilità. [Siebes e van Harmelen, 2002] e [Tempich et al., 2003] hanno dimostrato come l'informazione di provenienza possa essere usata per concordare sulla semantica.

(vi) *P2P Overlay Networks e Semantic Grid.* Per quanto riguarda gli aspetti di implementazione fisica del paradigma P2P, la recente crescita di applicazioni P2P in Internet ha motivato l'interesse verso P2P overlay networks general purpose. Un interessante problema di ricerca è lo sviluppo di architetture overlay che possano supportare applicazioni complesse [Baldoni et al., 2004; 2005] senza sovraccaricare le risorse di rete e raggiungere un elevato numero di partecipanti (peers) su larga scala. Lo stato dell'arte corrente identifica due diverse architetture per P2P overlay networks che possono offrire le possibili caratteristiche: (i) Sistemi strutturati, che sfruttano una topologia strutturata di rete overlay e uno specifico posizionamento di risorse per fornire ai clienti delle primitive di scoperta risorse di alta prestazione. Tali sistemi possono essere ulteriormente raggruppati in "debolmente strutturati", in cui la topologia della rete è quasi casuale ma il piazzamento/dislocazione delle risorse è determinato da specifici algoritmi, e "altamente strutturati", in cui sia la topologia della rete che le risorse sono altamente organizzati. Attualmente un forte interesse è centrato sul secondo tipo e cioè su sistemi che implementano distributed hash tables (DHTs) [Ratnasamy et al., 2001; Rowstron and Druschel, 2001; Stoica et al., 2001; Zhuang et al., 2001]; (ii) Sistemi destrutturati, dove la topologia di rete è prevalentemente casuale o costruita usando regole non rigide. Inoltre la dislocazione delle risorse non è affatto controllata. Tali scelte permettono a questi sistemi di gestire i nodi che entrano o escono dal sistema (alto tasso di abbandono) dato che queste operazioni solitamente generano un basso livello di costi aggiuntivi, ma rendono la scoperta di risorse costosa; infatti, i sistemi P2P destrutturati spesso implementano algoritmi basati su flooding o tecniche random walk per la scoperta di risorse [Cohen and Shenker, 2002; Gkantsidis et al., 2004].

Con riferimento alle architetture che realizzano il paradigma P2P, sta assumendo crescente importanza il concetto di Grid semantica [Calvanese et al., 2004b] come un'opportuna infrastruttura per le Organizzazioni Virtuali, basata su servizi standardizzati che nasconde la complessità dovuta a sorgenti dati eterogenee e gestisce la dinamica di un ambiente di rete. Queste motivazioni sono alla base della tendenza attuale a modellare infrastrutture commerciali complesse come Grid, e giustificano perché le tecnologie Grid abbiano un notevole successo nell'industria. Infatti, la Open Grid Services Architecture (OGSA) è il livello base per lo sviluppo di ambienti operativi integrati e virtualizzati (i.e. 'On Demand') nell'ambito di IBM.

Testo inglese

Integrating and coordinating heterogeneous and distributed contents in open information systems over the Web is one of the crucial challenges at the current evolutionary stage of IT infrastructure. Semantic information sharing, rather than information processing or just syntactic exchange à la XML, is one of the information technology primary goals of these years. Semantics is particularly important because information has to be sharable in open environments where interacting nodes, which can be called sources, sites, agents, etc., and that we will simply call peers, do not necessarily share a common understanding of the world at hand. Lack of common background generates the need for explicit guidance in understanding the exact meaning of data, hence the growing importance of ontologies. Furthermore, meaningful exchanges between peers can only occur on the basis of dynamic trustful agreements on common interpretations within the context of a given task. The particular form of semantic interoperability above described is referred to as "emergent semantics"[Aberer et al., 2004].

In the following, we will detail the SoA in the main areas that involve the emergent semantics research issues: (i) traditional data integration, (ii) incorrect information integration, (iii) ontology-based data integration, (iv) peer-to-peer data integration, (v) trust issues and (v) P2P overlay networks and Semantic Grid.

(i) *Traditional Data Integration.* The traditional scientific base for achieving cooperation in open, distributed, multi-knowledge environments is in data integration. The goal of data integration is to provide clients with a virtual access to a set of heterogeneous and autonomous data sources. Research in data integration has mostly focused on declarative approaches [Lenzerini, 2002] (as opposed to procedural approaches), where the data integration system exports to the user a global reconciled view of the data (called global schema or global ontology) in terms of which user requests are formulated, and the system maintains a declarative specification (i.e., a mapping) of the interrelationships between the global ontology and the data at the sources [Lenzerini, 2002]. Among the various problems related to data integration, the one of answering queries posed over the global schema is the one that has been addressed most intensively [Duschka et al., 2000; Cali' et al., 2004, Calvanese et al., 1998; 2001].

(ii) *Incorrect Information Integration.* Data integration in presence of incorrect information has been addressed with formal methods in [Galhardas et al., 2000; Calvanese et al., 1999]. The methodology proposed within the European Project DWQ (Foundations of Datawarehouse Quality) [Jarke et al., 2000; Calvanese et al., 1999] is based on the declarative specification of reconciliation correspondences between data in different sources. Some recently proposed data integration systems that deal with instance level conflicts are: the DaQuinCIS System [Scannapieco et al., 2004], and the interactive system proposed in [Sattler et al., 2003]. Rather than solving conflicts with ad-hoc techniques, some works have proposed the use of repair semantics to be adopted in order get

consistent answers from inconsistent data [Bravo and Bertossi, 2003; Cali' et al., 2003a; Cali' et al., 2003b].

(iii) *Ontology-based Data Integration.* The use of ontologies have assumed in the last years a growing importance in integrating and coordinating data. However, state-of-the-art ontology-based integration techniques seems still to be in a preliminary stage. Indeed, most of the work carried out on ontologies is on which language or which method to use to build a global integrated ontology on the basis of the ontologies managed by local sources [Broekstra et al., 2000; Doan et al., 2004], whereas the problem of actually querying an integrated ontology needs still further investigation.

(iv) *Peer-to-Peer Data Integration.* The issue of cooperation, integration, and coordination between information nodes in open distributed systems has been recently addressed in the context of Peer-to-Peer (P2P) data computing [Halevy et al., 2003]. A P2P system is characterized by a structure constituted by various autonomous nodes that hold information and that are linked to other nodes by means of mappings and no global ontology the client can refer is actually available. The lack of any agreed-upon global schema or ontology makes it very difficult for the participating parties in the system to reach a global consensus on semantic data. Initial approaches rely on some pre-defined corpus of terms serving as an initial context for defining new concepts [McCool and Guha, 2003] or make use of gossiping and local translation mappings to incrementally foster interoperability in the large [Aberer et al., 2003].

There is still a fundamental lack of understanding behind the basic issues of data integration in P2P systems. Indeed, since no single actor is in charge of the whole system, it is unrealistic to assume restrictions on the overall topology of the P2P mappings [Halevy et al., 2003; Fagin et al., 2003]. Hence, one has to take into account that the mappings may have an arbitrary structure, possibly involving cycles among various nodes. This needs to be addressed both from the point of view of modeling the system and characterizing its semantics (see, e.g., [Catarci and Lenzerini, 1993; Halevy et al., 2003;] for a first order logic characterization and [Calvanese et al., 2003;2004] for an alternative semantics proposal) and from the point of view of computing answers to queries posed to a peer, with respect to which difficulties arise from the necessity of distributing the overall computation to the single nodes, exploiting their local processing capabilities and the underlying technological framework.

(v) *Trust Issues.* A further aspect deserving particular attention is the one related to measures for evaluating the degree of consensus. These measures may refer to the assessment of trust into an information providing agent and to the assessment of the perceived quality of information received. Trust is typically based on the consensus of people's opinions about each other. The e-bay rating system is an example of this. However, it doesn't take into account the trustworthiness and the "expertise" of the raters. Various recent works investigated mechanisms to establish consensus on trust taking into account reputation of referrals [Kamvar et al., 2003]. The quality of information is either dependent on the opinions of people or upon the applications consuming that information. In the former case, consensus computation will be a crucial component in determining the quality of information [De Santis et al., 2003]. Provenance information [Ehrig et al., 2003] allows to relate information to its source, and thus to its trust rating. [Siebes and van Harmelen, 2002] and [Tempich et al., 2003] have shown how provenance information can be used for agreeing on semantics.

(vi) *P2P Overlay Networks and Semantic Grid.* As for physical implementation aspects related to the P2P paradigm, the recent growth of P2P applications on the Internet has motivated the interest in general purpose P2P overlay networks. A challenging research problem in this field is the development of overlay architectures that can support complex applications [Baldoni et al., 2004; 2005] without overloading network resources and scale up to a huge number of participants (peers). Current state of the art in this field has identified two different architectures for P2P overlay networks that can offer the envisaged features: (i) Structured systems, which exploit a structured overlay network topology and specific resource placement to provide clients with high performance resource discovery primitives. Such systems can be further grouped in loosely structured, where the network topology is almost random but resource placement is determined by specific algorithms, and highly structured, where both the network topology and resources are highly organized. Currently great interest is focused in the latter type and specifically in systems implementing distributed hash tables (DHTs) [Ratnasamy et al., 2001; Rowstron and Druschel, 2001; Stoica et al., 2001; Zhuang et al., 2001]; (ii) Unstructured systems, where the network topology is mostly random or built using loose rules. Moreover resource placement is not controlled at all. These architectural choices make these systems highly resilient to nodes entering or leaving the system (high churn rate) because these operations usually generate a low overhead, but also render resource discovery expensive; in fact unstructured P2P systems often implement algorithms based on flooding or random walk techniques for resource discovery [Cohen and Shenker, 2002; Gkantsidis et al., 2004].

As for architectures realizing the P2P paradigm, we point out the growing importance of the notion of semantic Grid [Calvanese et al., 2004b] as a suitable infrastructure for Virtual Organizations, based on standardized services which hides the complexity of heterogeneous data sources and handles the dynamics of the networking environment. This motivates the current trend of modelling complex business infrastructures as Grids, and this is why Grid technologies attract so much interest in the industry. In fact, the Open Grid Services Architecture (OGSA) is the foundational layer for developing integrated and virtualized (i.e. "On Demand") operating environments at IBM.

2.4.a Riferimenti bibliografici

[Aberer et al., 2003] K. Aberer, P. Cudré-Mauroux, and M. Hauswirth. *The Chatty Web: Emergent Semantics Through Gossiping.* In *Int. World Wide Web Conf. (WWW 2003)*, 2003.

[Aberer et al., 2004] K. Aberer, P. C. Mauroux, A. M. Ouksel, T. Catarci, M. S. Hacid, A. Illarramendi, V. Kashyap, M. Mecella, E. Mena, E. J. Neuhold, O. De Troyer, T. Risse, M. Scannapieco, F. Saltor, L. de Santis, S. Spaccapietra, S. Staab, and R. Studer. *Emergent Semantics Principles and Issues.* ICSNW 2004, 2004.

[Baldoni et al., 2004] Roberto Baldoni, Roberto Beraldi, Leonardo Querzoni and Antonino Virgillito. *Subscription-Driven Self-Organization in Content-Based Publish/Subscribe.* In *Proc. of the International Conference on Autonomic Computing (ICAC)*, 2004.

[Baldoni et al., 2005] Roberto Baldoni, Carlo Marchetti, Roman Vitenberg and Antonino Virgillito. *Content-based Publish/Subscribe over Structured Overlay Networks.* In *Proc. of the 25th International Conference on Distributed Computing Systems (ICDCS)*, 2005. To appear.

[Bravo and Bertossi, 2003] Loreto Bravo, Leopoldo Bertossi. *Logic programming for consistently querying data integration systems.* In *Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003)*, pages 10-15, 2003.

[Broekstra et al., 2000] J. Broekstra, M. Klein, D. Fensel, I. Horrocks. *Adding Formal Semantics to the Web: Building on top of RDF Schema.* In *Proc. the ECLD 2000, Workshop on Semantic Web 2000.*

[Cali' et al., 2003a] Andrea Cali', Domenico Lembo, Riccardo Rosati. *On the decidability and complexity of query answering over inconsistent and incomplete databases.* In *Proc. of the 22nd ACM SIGACT SIGMOD SIGART Symp. on Principles of Database*

Systems (PODS 2003), pages 260-271, 2003.

[Cali' et al., 2003b] Andrea Cali', Domenico Lembo, Riccardo Rosati. Query rewriting and answering under constraints in data integration systems. In Proc. of the 18th Int. Joint Conf. on Artificial Intelligence (IJCAI 2003), pages 16-21, 2003.

[Cali' et al., 2004] Andrea Cali', Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini. Data Integration under Integrity Constraints. In Information Systems, 29:147-163, 2004.

[Calvanese et al., 1998] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati. Description logic framework for information integration. In Proc. of the KR'98, pages 2-13, 1998.

[Calvanese et al., 1999] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, Riccardo Rosati. A Principled Approach to Data Integration and Reconciliation in Data Warehousing. Workshop on Design and Management of Data Warehouses, 1999.

[Calvanese et al., 2001] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati. Data integration in data warehousing. Int. J. of Cooperative Information Systems, 10(3):237-271, 2001.

[Calvanese et al., 2003] Diego Calvanese, Elio Damaggio, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Semantic data integration in P2P systems. In Proc. of the Int. Workshop on Databases, Information Systems and Peer-to-Peer Computing (DBISP2P 2003), 2003.

[Calvanese et al., 2004] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, and Riccardo Rosati. Logical Foundations of Peer-To-Peer Data Integration. In Proc. of PODS 2004, pages 241-251, 2004.

[Calvanese et al., 2004b] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Riccardo Rosati, Guido Vetere: Hyper: A Framework for Peer-to-Peer Data Integration on Grids. ICSNW 2004: 144-157

[Catarci and Lenzerini, 1993] Tiziana Catarci and Maurizio Lenzerini. Representing and using interschema knowledge in cooperative information systems. J. of Intelligent and Cooperative Information Systems, 2(4):375-398, 1993.

[Clarke et al., 1999] I. Clarke, S.G. Miller, T.W. Hong, O. Sandberg, and B. Wiley, "Freenet: A Distributed Anonymous Information Storage and Retrieval System", in Proc. of Int. Workshop on Design Issues in Anonymity and Unobservability (DIAU 2000), Berkeley, CA, USA, July 2000.

[Cohen and Shenker, 2002] Edith Cohen, Scott Shenker. Replication strategies in unstructured peer-to-peer networks. In Proc. of ACM SIGCOMM Conference, 2002.

[De Santis et al., 2003] L. De Santis, M. Scannapieco, T. Catarci: Trusting Data Quality in Cooperative Information Systems. In Proc. of CoopIS 2003, Catania, Italy, 2003.

[Doan et al., 2004] A. Doan, N. Noy, A. Halevy (editors): Special Issue on Semantic Integration, SIGMOD Record, Vol.33, no.4, 2004.

[Duschka et al., 2000] Oliver M. Duschka, Michael R. Genesereth, and Alon Y. Levy. Recursive query plans for data integration. J. of Logic Programming, 43(1):49-73, 2000.

[Ehrig et al., 2003] M. Ehrig, P. Haase, F. van Harmelen, R. Siebes, S. Staab, H. Stuckenschmidt, R. Studer, and C. Tempich. The swap data and metadata model for semanticsbased peer-to-peer systems. In Proceedings of MATES-2003, pages 144-155, 2003.

[Fagin et al., 2003] Ronald Fagin, Phokion G. Kolaitis, Renee J. Miller, and Lucian Popa. Data exchange: Semantics and query answering. In Proc. of the 9th Int. Conf. on Database Theory (ICDT 2003), pages 207-224, 2003.

[Galhardas et al., 2000] H. Galhardas, D. Florescu, D. Shasha, E. Simon. An Extensible Framework for Data Cleaning. IEEE-ICDE, 2000.

[Gkantsidis et al., 2004] Christos Gkantsidis, Milena Mihail, and Amin Saberi. Random walks in peer-to-peer networks. In Proc. of IEEE INFOCOM, 2004.

[Halevy et al., 2003] Alon Halevy, Zack Ives, Dan Suciu, and Igor Tatarinov. Schema mediation in peer data management systems. In Proc. of the 19th IEEE Int. Conf. on Data Engineering (ICDE 2003), pages 505-516, 2003.

[Jarke et al., 2000] M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis. Fundamentals of Data Warehouses. Springer-Verlag, 2000.

[Kamvar et al., 2003] S. Kamvar, M. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In International World Wide Web Conference (WWW), pages 640-651, 2003.

[Lenzerini, 2002] Maurizio Lenzerini. Data integration: A theoretical perspective. In Proc. of PODS 2002, pages 233-246, 2002.

[McCool and Guha, 2003] R. McCool and R.V. Guha. Tap, building the semantic web.

[Ratnasamy et al., 2001] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp and Scott Shenker. A scalable content-addressable network. In Proc. of ACM SIGCOMM Conference, 2001.

[Rowstron and Druschel, 2001] Antony Rowstron, Peter Druschel. Pastry: Scalable, decentralized object location and routing for large-scale peer-to-peer systems. Lecture Notes in Computer Science 2218-329, 2001.

[Sattler et al., 2003] K.Sattler, S.Conrad, and G.Saake, Interactive Example-Driven Integration and Reconciliation for Accessing Database Integration, Information systems,28, 2003.

[Scannapieco et al., 2004] M. Scannapieco, A.Virgillito, M. Marchetti, M. Mecella, R. Baldoni: The DaQuinCIS Architecture: a Platform for Exchanging and Improving Data Quality in Cooperative Information Systems. Information Systems, vol. 29, no. 7, 2004.

[Siebes and van Harmelen, 2002] R. Siebes and F. van Harmelen. Ranking agent statements for building evolving ontologies. In Proceedings of the AAAI-02 workshop on meaning negotiation, Alberta, Canada, July 28 2002, 2002.

[Stoica et al., 2001] Ion Stoica, Robert Morris, David R. Karger, M. Frans Kaashoek, Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In Proc. of ACM SIGCOMM Conference, 2001.

[Tempich et al., 2003] C. Tempich, S. Staab, and A. Wranik. REMINDIN': Semantic query routing in peer-to-peer networks based on social metaphors, 2003. submitted for publication.

[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.

[Zhuang et al., 2001] Zhuang, S. Q., Zhao, B. Y., Joseph, A. D., Katz, R. & Kubiawicz, J. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical Report UCB/CSD-01-1141, University of California at Berkeley, Computer Science Division, April 2001.

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

Testo italiano

Grandi imprese, organizzazioni commerciali, sistemi di e-government ed, in breve, ogni tipo di comunità lavorativa che utilizza internet, manifesta oggi la necessità di accedere in maniera coordinata e virtuale a risorse di informazione distribuite.

L'integrazione di dati è perciò un compito fondamentale per molte attività lavorative e di vita quotidiana. Le tecniche di integrazione dati sono basate sull'idea di mappare tutti gli schemi delle sorgenti (schemi locali) in un singolo schema globale integrato. La definizione di mapping tra schema locale e globale è un compito statico, parte di una fase preliminare di progettazione del sistema. Oggi, emergono nuovi requisiti che non sono presi in considerazione dalle tradizionali tecniche di integrazione dei dati: le imprese virtuali integrano dati di molte entità commerciali autonome e disponibili su Web, che agiscono come nodi (peers), assumendo sia il ruolo di produttori che di consumatori di dati.

Come scenario applicativo, consideriamo la collaborazione scientifica in campo medico, in cui personale medico manifesta la necessità di condividere dati in formato elettronico, per poter migliorare i processi decisionali e ridurre gli errori. Un medico potrebbe voler ottenere tutte le informazioni pertinenti alla malattia di un certo paziente, dai geni che possono aver contribuito a generarla, sino ai sintomi e trattamenti possibili. Ciò necessita l'integrazione di sorgenti disperse, spesso disponibili sul Web, e non note a priori.

In tale scenario, nuove problematiche di integrazione dati devono essere studiate. Il raggiungimento di un consenso globale su di un singolo schema condiviso diventa impossibile con un numero enorme di sorgenti autonome. Inoltre, questo tipo di ambiente, estremamente dinamico, richiede la verifica di un nuovo accordo globale ad ogni cambiamento del sistema.

Il consenso locale deve essere raggiunto dinamicamente e l'integrazione degli schemi deve avvenire durante l'interazione dei peer, conformemente ai nuovi principi del paradigma della "emergent semantic".

Inoltre, le caratteristiche di questo nuovo setting richiedono nuovi modi di processare interrogazioni dai peer. Si noti che, sebbene vi siano alcuni sforzi iniziali relativi al problema del processing di interrogazioni in sistemi peer-to-peer (P2P), essi sono basati su restrizioni della topologia complessiva dei mapping (ad esempio, assenza di cicli fra i peer). Questa assunzione non sembra realistica, visto che non esiste un singolo attore responsabile dell'intero sistema. Ulteriori problematiche riguardano come scoprire dinamicamente quali sorgenti sono disponibili e possono essere sfruttate per rispondere alle interrogazioni. La disponibilità dei peer nel tempo non è garantita, perciò può essere necessario risolvere situazioni in cui le sorgenti coinvolte in una particolare interrogazione diventano improvvisamente non disponibili. Inoltre, è necessario decidere quale sia il modo più efficiente ed affidabile per indirizzare le interrogazioni sulla rete.

Diversi problemi connessi a queste proprietà delle reti P2P sono già stati studiati nell'area dei sistemi distribuiti. In generale, tuttavia, le soluzioni esistenti non sono adatte all'integrazione dati P2P. Un importante obiettivo di questa ricerca sarà comprendere come sfruttare tali soluzioni nel contesto dell'integrazione dati.

A causa della natura aperta dei sistemi P2P, chiunque può entrare a far parte di una rete di peer. Ciò comporta due problemi diversi ma tra loro collegati: la gestione della credibilità e la qualità dei dati. La gestione della credibilità consente ai peer di identificare altri peer di cui si fidano, e interagire solo con tali peer. Ogni sorgente dati può in principio influenzare il risultato finale integrato e dunque deve essere controllata. Per esempio, nello scenario applicativo sopra descritto, è vitale proteggere gli utenti dalla diffusione di informazione derivante da sorgenti non affidabili. I meccanismi di credibilità sono strettamente legati alla qualità dei dati disponibili ai peer. I dati sono spesso affetti da problemi di qualità, che vanno da una scarsa accuratezza a livello sintattico (i dati sono errati, per esempio in conseguenza di errori durante la loro raccolta o manipolazione), alla presenza di vari formati di dati o di inconsistenze, sia in una singola sorgente sia in diverse sorgenti, sino a problemi relativi alla loro diffusione.

Data la natura aperta di queste reti, nessun tipo di restrizione può essere applicata sulla qualità dei dati condivisi dai peers. Perciò, è essenziale che il controllo della qualità dei dati sia applicato sia attraverso i meccanismi di credibilità descritti sopra, che attraverso adeguate misure che prendano esplicitamente in considerazione la qualità nel processing delle interrogazioni. Nello scenario di riferimento, è importante per i medici e ricercatori essere in grado di accedere a "dati di elevata qualità" o se possibile ai dati migliori (ad esempio i più aggiornati ed accurati) disponibili nella rete.

L'unità UNIROMA affronterà i problemi di ricerca sopra descritti attraverso lo sforzo congiunto del gruppo di Basi di Dati e di Conoscenza e del gruppo di Sistemi Distribuiti. In particolare, UNIROMA identificherà soluzioni al problema di integrazione dati in un ambiente P2P, prendendo in considerazione le problematiche di credibilità e di qualità dei dati. UNIROMA inoltre progetterà e svilupperà uno strato P2P relativo su cui tutti i servizi ESTEEM verranno dispiegati.

Il piano di lavoro consiste di due fasi. Ogni fase è costituita da attività e relativi deliverable. Tali fasi sono descritte nel seguito.

Fase 1

Attività 1.1

La prima fase riguarderà uno studio comparato sull'integrazione dati P2P, per evidenziare punti di forza e di debolezza. Saranno esaminate, nell'ambito dell'"emergent semantics", le tecniche per la comparazione di ontologie, la ricerca di mapping tra ontologie, la rappresentazione e la specifica dei mapping. Tutte queste caratteristiche verranno analizzate tenendo in considerazione parametri di qualità relativi a ciascun peer, quali accuratezza, completezza, consistenza, etc., la caratterizzazione della credibilità dei peer e tenendo in considerazione requisiti specifici del paradigma P2P, quali flessibilità dinamicità, e modularità.

Riguardo al query answering, saranno studiati i metodi esistenti per la rappresentazione dell'informazione all'interno di un singolo peer nell'ottica di caratterizzarne il potere espressivo e le associate capacità computazionali del peer. Inoltre, verranno presi in considerazione gli approcci esistenti per la rappresentazione di mapping tra peer e per caratterizzare la semantica del processo complessivo di query answering. Particolare enfasi verrà data agli aspetti concernenti l'effettiva realizzazione dei metodi di query answering come servizi offerti dai peer, attraverso l'uso di infrastrutture tecnologiche esistenti in vari contesti, come architetture Grid.

Le infrastrutture di comunicazione P2P esistenti verranno analizzate, con una particolare enfasi sui sistemi P2P strutturati. La letteratura presenta diversi sistemi in cui dati complessi vengono mappati sul semplice spazio logico offerto della overlay networks strutturate (ad esempio sistemi publish/subscribe content-based). Tali sistemi verranno analizzati per ricavarne metodi e tecniche di validità generale per la costruzione di mappature applicabili agli scopi del progetto.

Risultati dell'Attività 1.1

D1.1: Rapporto tecnico sullo studio comparato dei: (i) metodi per la specifica di peer e di mapping in sistemi di integrazione dati P2P; (ii) modelli per la qualità dei dati e per la credibilità in tali sistemi, (iii) sistemi P2P strutturati ed applicazioni basate su di essi.

Attività 1.2

La seconda attività sarà dedicata a collezionare requisiti dettagliati per lo scenario applicativo di ESTEEM (collaborazione scientifica nel campo medico). L'insieme dei requisiti sarà raccolto attraverso interviste e questionari ai partner utenti del progetto. Si prevede che i partner utenti del progetto saranno medici del Policlinico Universitario "A. Gemelli" e del Policlinico Universitario "Umberto I" di Roma. Ulteriori partner potranno essere individuati durante lo svolgimento del progetto. I requisiti raccolti saranno sfruttati per la realizzazione del dimostratore integrato che sarà implementato nella seconda fase del progetto. Si sottolinea che, anche se il dimostratore sarà implementato su uno scenario applicativo specifico, la piattaforma ESTEEM è di impiego generico, e può essere di supporto per diverse comunità semantiche.

Risultati dell'Attività 2.2

DALL.1 (tutti i partner): Rapporto tecnico nel quale verranno riportati i requisiti raccolti per lo scenario applicativo.

Attività 1.3

La terza attività sarà dedicata allo sviluppo di tecniche per la specifica e per il processamento di interrogazioni in sistemi di integrazione dati P2P trust-aware.

Verrà identificato un modello semantico adeguato sia per il singolo peer, che per il query answering nei sistemi P2P. In particolare, considerando aspetti quali la qualità dei dati e la credibilità, verranno identificate soluzioni ai seguenti problemi: (i) caratterizzazione della struttura e della semantica di ogni singolo peer, (ii) definizione della semantica per il mapping tra peer, e della semantica dell'intero sistema P2P, (iii) risoluzione delle query in un sistema P2P trust-aware, (iv) sviluppo di algoritmi efficienti per il query answering. Tecniche e metodi sviluppati in questa attività verranno implementati in un prototipo software nella fase 2.

Risultati dell'Attività 1.3

D1.2: Rapporto tecnico contenente la definizione di un framework formale per sistemi di integrazione dati P2P trust-aware, la specifica della loro semantica, la definizione di tecniche per il query answering coerenti con il modello semantico scelto.

Attività 1.4

La quarta attività sarà dedicata alla progettazione di una infrastruttura P2P basata su una overlay network strutturata. Tale infrastruttura fornirà ai peer primitive di base, quali (i) localizzazione di risorse, (ii) group membership, (iii) comunicazione uno-a-uno e uno-a-molti. Inoltre la overlay network sarà progettata per accomodare tutti i requisiti necessari per supportare le tecniche di processamento delle query definite nell'Attività 1.2. Da questo punto di vista saranno investigati metodi per mappare i dati oggetto delle query sullo spazio logico fornito dalla overlay network strutturata, per realizzare una corrispondenza diretta tra le operazioni di query e le operazioni a livello di overlay network.

Risultati dell'Attività 1.4

D1.3: Rapporto tecnico contenente l'architettura di un sistema P2P basato su una overlay network strutturata, assieme alle tecniche necessarie per poter fornire ai peer primitive per la localizzazione di risorse, la group membership e la comunicazione.

Attività 1.5

Questa attività sarà incentrata sulla specifica di alto livello dell'architettura di ESTEEM, da definirsi in cooperazione con gli altri partner del progetto. L'architettura fornirà un insieme di servizi per: l'accesso integrato e trust-aware ai dati distribuiti tra i peer di una comunità semantica, la localizzazione ed l'identificazione di web services all'interno della stessa comunità, e la definizione del consenso e la gestione di comunità semantiche. Tali servizi saranno accessibili attraverso uno strato software adattivo e multicanale. I servizi di ESTEEM saranno sviluppati sotto forma di web services accessibili attraverso uno strato di cooperazione P2P. L'architettura di ESTEEM sarà il risultato della convergenza tra web services, P2P e grid. Una possibile soluzione per tale convergenza è l'uso dell'Open Grid Services Architecture (OGSA). Essa infatti permette agli sviluppatori di integrare servizi e risorse attraverso ambienti e comunità distribuite, eterogenee e dinamiche. Il modello OGSA adotta il Web Services Description Language (WSDL) per definire il concetto di grid service. Come compito specifico, UNIROMA definirà le interfacce WSDL dell'insieme dei servizi che garantiranno l'accesso integrato e trust-aware ai dati del sistema P2P.

Risultati dell'Attività 1.5

DALL.2 (tutti i partner): Rapporto tecnico contenente la definizione ad alto livello dell'architettura di ESTEEM, e le specifiche WSDL di tutti i servizi da essa offerti. UNIROMA è responsabile per la specifica sia dello strato di cooperazione P2P, sia dei servizi di integrazione dei dati.

Fase 2

Attività 2.1

Il raggiungimento degli obiettivi del progetto richiede la definizione di un'architettura comprensiva per ESTEEM, che tenga conto di tutti gli aspetti relativi ai problemi precedentemente discussi. UNIROMA si occuperà di definire un'architettura basata su servizi per l'accesso integrato ai dati in un sistema P2P, con gestione orientata alla qualità ed alla credibilità, e guiderà l'integrazione tra i vari partner durante il processo di definizione delle rispettive parti dell'architettura di ESTEEM.

Per supportare la progettazione dell'architettura, e per facilitare l'individuazione di possibili problemi derivanti dall'interazione tra i servizi che saranno sviluppati dai diversi partner, l'attività di progettazione prevede lo sviluppo di un prototipo freddo (mock-up) per i servizi sviluppati da UNIROMA per l'accesso integrato e trust-aware ai dati in sistemi P2P. Concorrentemente, gli altri partner implementeranno mock-up per i rispettivi servizi. Tutti i partner saranno quindi coinvolti in un processo di verifica comune durante il quale tali prototipi verranno integrati. Ciò fornirà a ciascun partner un'visione completa dei possibili problemi. Il risultato di tali test verrà considerato durante lo svolgimento dell'Attività 2.2.

Risultati dell'attività 2.1

DALL.3 (tutti i partner): Rapporto tecnico contenente la definizione dettagliata dell'architettura di ESTEEM.

DALL.4 (tutti i partner): Prototipi e dati raccolti durante la fase di test.

Attività 2.2

In questa attività verranno sviluppati i servizi di UNIROMA tenendo conto delle specifiche definite nelle precedenti attività. La fase di implementazione seguirà un approccio a spirale/iterativo e sarà basato su tecnologie Grid e Web Services.

Risultati dell'Attività 2.2

D1.4: Implementazione di un insieme di servizi trust- e quality-aware per l'accesso integrato ai dati in sistemi P2P.

D1.5: Implementazione dello strato di cooperazione P2P.

Attività 2.3

Tale attività verrà sviluppata congiuntamente da tutti partner. Essa consiste nella implementazione di un dimostratore integrato, tramite il quale verrà mostrata l'utilizzo della piattaforma ESTEEM per la collaborazione scientifica in campo medico. L'implementazione del dimostratore integrato servirà come piattaforma di test per verificare l'efficacia dei risultati della ricerca su uno scenario di uso reale. L'implementazione del dimostratore di ESTEEM sarà basata sui servizi sviluppati nell'Attività 2.2. UNIROMA sarà responsabile dello sviluppo dei moduli del dimostratore inerenti le attività di integrazione dati trust-aware.

Risultati dell'Attività 2.3

DALL.5 (tutti i partner): Rapporto tecnico nel quale verranno illustrati il dimostratore unificato ed i risultati degli esperimenti condotti nel contesto dello scenario applicativo.

Testo inglese

Large enterprises, business organizations, e-government systems, and, in short, any kind of internetworking community, need today a coordinated and virtualized access to distributed information resources. Data integration is thus a fundamental task for many business or life-critical activities.

Integration techniques were based on the idea of mapping all the schemas of the sources (local schemas), to a single, integrated global schema. Mapping definition between local and global schemas was a static task, part of a preliminary system design phase. Today, new requirements emerge that are not taken into account by traditional data integration: virtual enterprises integrate data coming from many, Web-available, disparate autonomous business entities that act as peers, assuming both the roles of data producers and consumers at the same time.

As application scenario, we consider scientific collaboration in medicine, i.e., a scenario 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 integration of disparate sources often from the web and that are not known in advance. In such a scenario, the achievement of a global consensus on a single shared schema becomes impossible with a huge number of sources, each driven by its own purposes and internal policies. Furthermore, this kind of environment is extremely dynamic, and would require the auditing of a new global agreement at every change of the system. Local consensus must be reached dynamically: mapping and integration of local schemas must occur during interaction among peers, according to the new principles of the emergent semantics paradigm. Furthermore, the highly dynamic features of this new setting require new ways to process queries issued by peers. While acknowledging that there are some initial efforts to peer-to-peer (P2P) query processing [Halevy et al., 2003; Fagin et al., 2003], they are based on restrictions on the overall topology of the P2P mappings (e.g., absence of cycles), which is an unrealistic assumption since no single actor is in charge of the whole system. Performing query processing in P2P systems raises some further issues that need to be addressed, including how to dynamically discover which sources are available and can be exploited to answer queries. Availability of peers over time is not guaranteed, therefore it may be necessary to recover from situations in which sources involved in a particular query suddenly become unavailable. Furthermore, it is necessary to decide which is the most efficient and reliable way to route queries over the network.

Several problems connected to these properties of P2P networks have already been investigated by the distributed systems research community. In general, though, existing solutions are not tailored for P2P data integration. One important goal of this research is to understand how existing solutions can be exploited and adapted to a P2P data integration setting, and which problems require instead new solutions.

Other critical problems arise from the intrinsically open nature of P2P systems. In principle, everyone can join a network of peers. This fact poses serious issues regarding two different but related problems: trust and data quality management.

Trust management enables peers to identify other peers that they trust, and limit their data exchanges to those peers only.

Controlling the sources in a P2P data integration setting is crucial because each source can in principle influence the final, integrated result. As an example, in the medical application scenario described above, it is vital to protect users from the diffusion of information coming from unreliable sources.

Trust mechanisms are closely related to the quality of data available at peers. Data stored electronically is usually affected by a number of quality problems, ranging from poor accuracy at a syntactic level (data is wrong, for example as a consequence of errors during data collection or manipulation), to various forms of inconsistency, both in a single source or across multiple sources, to problems related to their currency, and so on. Due to the open nature of P2P networks, no restriction at all can be enforced on the quality of the data shared by the peers. Thus, it is essential that data quality control is enforced both through the trust mechanisms described above, and with proper measures at query processing time that explicitly take quality into account. Going back to the reference scenario, it is important for the doctors and researchers to be able to access 'high quality data' or if possible the best quality data (e.g. the most up-to-date and accurate) available in the network.

The UNIROMA unit will address the research problems described above with the joint effort of both the Data and Knowledge Bases group and the Distributed Systems group. In particular, UNIROMA will identify solutions to the problem of integrating data in a P2P environment, by taking into account trust and data quality issues. UNIROMA will further design and develop a P2P layer of the overall ESTEEM architecture on which all the ESTEEM services will be deployed.

The workplan includes two phases. Each phase consists of activities and related deliverables. Phases are described in the following.

Phase 1

Activity 1.1

The first activity will regard a comparative study on Peer-to-Peer data integration solutions, focused on highlighting weaknesses and strengths. The current techniques for ontology matching in a multi-ontology environment will be surveyed, in particular for analyzing the crucial aspects of emergent semantics, i.e. comparing ontologies, discovering mappings between ontologies, and representing and specifying such mappings. All the above characteristics will be analyzed considering quality parameters within each peer, such as accuracy, completeness, consistency, etc., and a peer-level trust characterization, and taking into account specific requirements imposed by the P2P paradigm, namely flexibility, dynamicity and modularity. As for query answering, starting from the observation that in the P2P context queries posed by clients need to be answered by retrieving the data residing in a set of peers, we will survey the methods for representing information within a peer, with the goal of characterizing their expressive power, and the associated computational capabilities of the peer in relation to the overall query answering process. Furthermore, existing approaches to representing mappings among peers and to characterizing the semantics of the overall query answering process in P2P system will be considered, with the goal of accessing their relative weaknesses and strengths. Particular emphasis will be given to those aspects related to the effective deployment of the query answering methods, exploiting available technological infrastructure in various contexts, such as Grid architectures.

Existing P2P communication infrastructures will be studied, with a particular emphasis on structured P2P systems. Various systems have been developed where complex data is mapped on the simple logical space offered by structured overlay networks (for example content-based pub/sub systems); these systems will be thoroughly analyzed to devise methods and techniques to build similar mappings for our usages.

Result of Activity 1.1

D1.1: Technical Report illustrating a comparative study on (i) the existing methods for the specification of both peers and peer mappings in P2P data integration systems, (ii) data quality and trust models of such systems (iii) structured P2P systems and applications built on them.

Activity 1.2

The second activity will be devoted to the collection of detailed requirements of the ESTEEM application scenario (scientific collaboration in medicine).

Collection of requirements will be done by interviews and questionnaires to the user partners of the project. Foreseen user partners of the project will be doctors from both the Policlinico Universitario "Agostino Gemelli" (Gemelli Hospital) and Policlinico Universitario "Umberto I" of Rome. Further user partners will be identified during the project. Collected requirements will be exploited for the integrated demonstrator that will be implemented in phase 2. We point out that, even if prototype demonstration is on a specific application scenario, the ESTEEM platform is general purpose and can support different semantic communities.

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

The third activity will be concerned with the design of techniques for the specification of trust-aware P2P data integration systems and for query processing in such systems.

A suitable semantic model for our systems will be identified, so as to specify well-defined semantics both for the single peers, and for query answering in P2P systems. In particular, solutions to the following problems will be devised, taking into account the quality of data and the level of trust: (i) characterization of the structure of a single peer in a P2P system, and the semantics of its specification, (ii) definition of the semantics of mappings between peers, and therefore the semantics of the overall P2P data integration system, (iii) answering queries posed to a peer in a trust-aware P2P system, (iv) deployment of efficient query answering algorithms. Techniques and methods defined in this activity will be implemented in a software prototype in Phase 2.

Results of Activity 1.3

D1.2: Technical Report defining a formal framework for trust-aware P2P data integration systems, and specifying their semantics, and illustrating the techniques for query answering in such systems coherent with the chosen semantic model.

Activity 1.4

The fourth activity will be devoted to the design of a P2P infrastructure in the form of a structured overlay network. This infrastructure will provide peers with basic primitives like (i) resource location, (ii) group membership, (iii) point-to-point and point-to-multipoint communication. Moreover the overlay network will be designed to accommodate all the requirements needed to support the query processing techniques defined in Activity 1.2. From this point of view we will investigate methods to map the data target of queries on the logical space provided by the structured overlay network in order to realize a direct mapping between query operations and network operations.

Results of Activity 1.4

D1.3: Technical report defining the architecture of a P2P system based on a structured overlay network along with algorithms and techniques needed to provide peers with resource location, group membership and communication primitives.

Activity 1.5

This activity will be concerned 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 formation and management of semantic communities. Such services will be accessible through a multi-channel and adaptive software layer. ESTEEM services will be developed as Web Services to be deployed onto a P2P cooperation layer. The ESTEEM architecture will be the result of the convergence of web services, P2P and grid. One possibility is to align P2P, Grid and Web Services by using the Open Grid Services Architecture (OGSA - <http://www.globus.org/ogsa/>). Indeed, OGSA lets developers to integrate services and resources across distributed, heterogeneous, dynamic environments and communities. The OGSA model adopts the Web Services Description Language (WSDL - <http://www.w3.org/TR/wsdl>) to define the

concept of a grid service using principles and technologies from both the grid and Web services communities. As a specific task, UNIROMA will define the WSDL interfaces of the set of services for integrated and trust-aware access to data in P2P systems.

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.

UNIROMA coordinates the deliverable and is responsible for the specification of the P2P cooperation layer and of the data integration services.

Phase 2

Activity 2.1

The accomplishment of the project goals requires designing a comprehensive architecture for ESTEEM, which will take into account all the aspects of the problem previously highlighted. UNIROMA will be involved in this process in two ways. First, it will design an architecture consisting of services for the integrated access to data (query processing) in a P2P system, with trust and quality oriented management. Then, UNIROMA will drive the integration among the various partners during the process of designing the respective parts of the architecture of ESTEEM.

In order to support the design of the ESTEEM architecture, and to bring to the surface all possible problems arising from the interaction between the services to be developed by the various partners, the design activity will involve the development of a mock-up for the services to be developed by UNIROMA for integrated and trust-aware access to data in P2P systems. Concurrently, other partners will implement mock-ups for the respective services. All the partners will then be involved in a common testing process during which such prototypes will be integrated. This will provide each partner with an insight of possible problems, and each of them will provide their own view and assessment. 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 task the implementation of the UNIROMA services will take place based on the specification and design developed in the previous activities. The implementation approach will follow a spiral/evolutionary approach and will be based on Web services and grid technologies.

Results of Activity 2.2

D1.4: Implementation of a set of services for integrated trust and quality aware access to data in a P2P environment.

D1.5: Implementation of the P2P cooperation layer.

Activity 2.3

This activity will be jointly developed by all partners. It consists of the implementation of an integrated demonstrator, showing the usage and feasibility of the ESTEEM platform in a real case. As application scenario, scientific collaboration in medicine will be considered, on the basis of the collection of requirements carried out during Activity 2.2. The implementation of the integrated demonstrator will serve as a test bed to verify the effectiveness of the research result accomplished during the entire lifetime of the ESTEEM project in a real usage scenario.

The implementation of the ESTEEM demonstrator will be based on the services developed in Activity 2.2. UNIROMA will be in charge of developing those modules of the demonstrator involving the trust-aware data integration activities.

Results of Activity 2.3

DALL.5 (jointly developed by all partners): Technical Report illustrating the unified 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
<i>Personale universitario dell'Università sede dell'Unità di Ricerca</i>		5	22	22	44
<i>Personale universitario di altre Università</i>		0	0	0	0
<i>Titolari di assegni di ricerca</i>		3	14	14	28
<i>Titolari di borse</i>	<i>Dottorato</i>	4	16	16	32
	<i>Post-dottorato</i>	0			
	<i>Scuola di Specializzazione</i>	0			
<i>Personale a contratto</i>	<i>Assegnisti</i>	1	11	11	22
	<i>Borsisti</i>	0			
	<i>Dottorandi</i>	0			
	<i>Altre tipologie</i>	0			
<i>Personale extrauniversitario</i>		0			
TOTALE		13	63	63	126

Testo inglese

		Numero	Mesi uomo 1° anno	Mesi uomo 2° anno	Totale mesi uomo
<i>University Personnel</i>		5	22	22	44
<i>Other University Personnel</i>		0	0	0	0
<i>Work contract (research grants, free lance contracts)</i>		3	14	14	28
<i>PHD Fellows & PHD Students</i>	<i>PHD Students</i>	4	16	16	32
	<i>Post-Doctoral Fellows</i>	0			
	<i>Specialization School</i>	0			
<i>Personnel to be hired</i>	<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 tipologie</i>	0			
<i>No cost Non University Personnel</i>		0			
TOTALE		13	63	63	126

PARTE III

3.1 Costo complessivo del Programma dell'Unità di Ricerca

Testo italiano

Voce di spesa	Spesa in Euro	Descrizione
Materiale inventariabile	16.000	Acquisto 1 elaboratore server e 2 PC client per lo sviluppo di alcune funzionalità del prototipo dell'architettura software. Libri.
Grandi Attrezzature		
Materiale di consumo e funzionamento	13.000	Toner, memoria flash, CD riscrivibili, sw di base. Costi dipartimentali di gestione del fondo.
Spese per calcolo ed elaborazione dati		
Personale a contratto	40.000	Contratto a progetto biennale. Il personale a contratto collaborerà alla ricerca per la definizione, progettazione e successiva implementazione di alcune funzionalità della architettura software P2P a servizi per la cooperazione dinamica.
Servizi esterni	3.000	Manutenzione hw e sw.
Missioni	22.000	Riunioni di coordinamento e partecipazioni a congressi
Pubblicazioni		
Partecipazione / Organizzazione convegni	20.000	Partecipazione a congressi per la presentazione dei risultati del progetto ed organizzazione di un workshop conclusivo del progetto.
Altro	6.000	Incarichi ad-hoc per lo sviluppo di particolari componenti sw.
TOTALE	120.000	

Testo inglese

Voce di spesa	Spesa in Euro	Descrizione
Materiale inventariabile	16.000	1 server and 2 PC clients for prototype development of the cooperative software platform. Books.
Grandi Attrezzature		
Materiale di consumo e funzionamento	13.000	Toner, flash memory, accessories, basic software. Department management costs.
Spese per calcolo ed elaborazione dati		
Personale a contratto	40.000	Research contractors will cooperate to define, design and implement key components of the P2P service-based software platform for dynamic cooperation.
Servizi esterni	3.000	Hw and sw maintenance.
Missioni	22.000	Project meetings and area conference participation.
Pubblicazioni		
Partecipazione / Organizzazione convegni	20.000	Participation in conferences to present project results and organization of a closing project workshop.
Altro	6.000	Specific software development.
TOTALE	120.000	

3.2 Costo complessivo del Programma di Ricerca

		Descrizione
Costo complessivo del Programma dell'Unità di Ricerca	120.000	
Fondi disponibili (RD + RA) comprensivi dell'8% max per spese di gestione	36.000	European NoE INTEROP (EC IST-508011) e DELOS (EC G038-507618).

		<i>Responsabile T.Catarci.</i>
Cofinanziamento di altre amministrazioni	0	
Cofinanziamento richiesto al MIUR	84.000	

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 18:50