PROGRAMMA DI RICERCA - MODELLO A
Anno 2004 - prot. 2004095494

1.1 Programma di Ricerca di tipo

Interuniversitario

Area scientifico disciplinare *Ingegneria industriale e dell'informazione (100%)*

1.2 Titolo del Programma di Ricerca

Testo italiano
WISDOM: Ricerca Intelligente su Web basata su Ontologie di Dominio

Testo inglese
WISDOM: Web Intelligent Search based on DOMain ontologies

1.3 Abstract del Programma di Ricerca

Testo italiano
L'enorme quantità di dati e la crescente disponibilità di servizi sul Web rendono sempre più importante lo sviluppo di infrastrutture e sistemi software che, fornendo strumenti per l'integrazione delle risorse informative, per la loro localizzazione e per la fruizione personalizzata delle stesse, permettano ai clienti collegati alla rete di "ricaricarsi" di dati di interesse per i propri bisogni informativi, evitando il problema di "information overloading" che si riscontra usando i comuni motori di ricerca.
WISDOM si pone nell'ambito di ricerca del Semantic Web e ha come obiettivo principale lo sviluppo di tecniche e strumenti intelligenti, basati su ontologie di dominio, per la ricerca di informazione su Web. In particolare, si vuole reperire informazione in modo integrato ed efficiente sia da siti di tipo data-intensive che da siti e pagine Web con contenuto scarsamente strutturato. Il progetto si articolerà in tre temi tra loro sinergici e complementari e definira' un'architettura metodologica e funzionale di riferimento al fine di garantire coerenza tra le soluzioni che verranno messe a punto nei tre temi.
L'obiettivo del primo tema (Creazione ed Estensione di una Ontologia di Dominio) è lo studio di soluzioni per la rappresentazione semantica dei contenuti delle sorgenti informative in ambito Web, con particolare riferimento ai siti data-intensive e ai siti/pagine Web con contenuto scarsamente strutturato. La rappresentazione ed integrazione di tali sorgenti informative portera' alla creazione dinamica di ontologie di dominio per effetto della scoperta/integrazione di nuove sorgenti informative. L'obiettivo del secondo tema (Semantica Emergente: Scoperta di Mapping Semantici tra Ontologie di Dominio) è lo sviluppo di tecniche e strumenti di supporto alla identificazione, scoperta, validazione e memorizzazione di relazioni semantiche fra ontologie di dominio in ambito Web . Il tipo di relazioni semantiche che sarà investigato dovra` contenere gli elementi necessari per risolvere una interrogazione rispetto a piu` ontologie e permettere di sviluppare tecniche di mapping basate sulla semantica del linguaggio, le catene lessicali e la deduzione logica. L'obiettivo del terzo tema (Elaborazione di Interrogazioni) è lo sviluppo di tecniche di ricerca di informazione su web in grado di utilizzare l'infrastruttura semantica sviluppata dai temi 1 e 2. Considerando l'eterogeneita' dei dati/siti trattati e i vincoli imposti dall'ambiente distribuito, verranno studiati e sviluppati meccanismi efficaci ed efficienti di elaborazione delle interrogazioni che usano la caratterizzazione delle sorgenti per selezionare le sorgenti utili, risolvono problemi di riscrittura e integrazione dei risultati sulle diverse sorgenti. Le problematiche affrontate nel progetto sono di estrema attualità e rivestono grande importanza applicativa e industriale, in particolare per lo sviluppo di nuove applicazioni fortemente personalizzate che possano sfruttare a pieno le potenzialità offerte dal Web.
Al progetto partecipano 4 unita' universitarie, con 18 fra professori e ricercatori (per un totale di 137 mesi uomo), 7 dottorandi (72 mesi uomo) e personale a contratto per 86 mesi uomo. Il costo del progetto e' di 393.500 Euro, di cui 130.500 per personale a contratto. Le unita' vantano una lunga esperienza di collaborazione a progetti, sia nazionali che internazionali. Il coordinamento del progetto verra' assicurato attraverso l'individuazione di un coordinatore per ciascun tema, che interagira' con il responsabile nazionale, al fine di monitorare lo stato di avanzamento relativo. E' prevista una riunione collegiale dopo ognuna delle 3 fasi in cui si articola il progetto. I risultati previsti sono di natura scientifico-metodologica, descritti in rapporti tecnici e in pubblicazioni, e realizzativa (sviluppo di strumenti a livello prototipale). I metodi e gli strumenti proposti saranno validati attraverso attivita' sperimentale.

Testo inglese
The huge amount of data and services available on the Web requires the development of systems that, overcome the "information overloading" problem of traditional search engines. In particular, there is the need of developing novel tools for the integration, the localization and the customizable fruition of informative resources which allow the clients to "recharge" with interesting data.
WISDOM goal is to develop intelligent techniques and tools, based on domain ontologies, to perform effective and efficient information search on the WEB. In particular, we aim at developing systems for retrieving information both from data-intensive and unstructured site/web pages, in an integrated and efficient way. The project will be articulated in three synergic and complementary themes and will define a reference methodological and functional architecture in order to ensure compatibility of

the solutions offered by the three themes. The goal of Theme1(Creation and extension of a domain ontology) is the study and development of solutions to represent the semantics of the contents of Web sources. The goal of Theme2 (Emergent Semantics: Discovering semantic mappings among domain ontologies) is the study and development of techniques and tools to support identification, discovery and storage of semantic mappings among domain ontologies. The investigated semantic mappings will support the rewriting of a query with respect to more ontologies and will be based on language semantics, lexical chains and logic inferences. The goal of Theme3 (Query processing) is the development of techniques for searching information, exploiting the semantic infrastructure developed within Themes 2 and 3. Efficient and effective query processing mechanisms, considering data/sites hetereogeneity and constraints imposed by the distributed environment, will be developed. In particular, these techniques will rely on sources characterization to individuate useful sources, solve rewriting problems and integrate results from different sources.
The issues addressed in WISDOM are relevant with a high applicative and industrial impact able to effectively exploit the potentialities of the Web.
4 units coming from different universities participate to the project, with 18 professors and researchers (for a total of 137 man-months), 7 PhD (72 man-months) and external people under contract (86 man-months). The
total cost of the project is 393500 Euro, with 135500 Euro set aside for external people. The units involved in the project have a long experience in collaborations in both national and international projects. The project management will be guaranteed by a coordinator for each theme, who will cooperate with the project leader, with the aim of monitoring relative progress. A collegiate meeting is expected at the end of each of the three phases in which the project is articulated. Project results will be both of scientific-methodological nature, documented through a series of technical reports, and implementative, in the form of prototype tools. Methods and tools proposed in the project will be validated through experimental activity.

1.4 Durata del Programma di Ricerca

24 Mesi

1.5 Settori scientifico-disciplinari interessati dal Programma di Ricerca

ING-INF/05 - Sistemi di elaborazione delle informazioni

1.6 Parole chiave

Testo italiano
ONTOLOGIA DI DOMINIO ; SISTEMI INFORMATIVI SU WEB ; SELEZIONE DI SORGENTI WEB ; INTERROGAZIONE SU ARCHITETTURE DISTRIBUITE ; MAPPING TRA ONTOLOGIE

Testo inglese
DOMAIN ONTOLOGY ; WEB INFORMATION SYSTEMS ; WEB SOURCE SELECTION ; QUERYING ON DISTRIBUTED ARCHITECTURES ; MAPPING AMONG ONTOLOGIES

1.7 Coordinatore Scientifico del Programma di Ricerca

BERGAMASCHI ***SONIA***

Professore Ordinario ***01/07/1953*** ***BRGSNO53L41F257K***

ING-INF/05 - Sistemi di elaborazione delle informazioni

Università degli Studi di MODENA e REGGIO EMILIA

Facoltà di INGEGNERIA

Dipartimento di INGEGNERIA DELL'INFORMAZIONE

059 2056132	***059 2056126***	***sonia.bergamaschi@unimo.it***
(Prefisso e telefono)	*(Numero fax)*	*(Email)*

1.8 Curriculum scientifico

Testo italiano
Sonia Bergamaschi è nata a Modena ed ha ricevuto la Laurea in Matematica presso l'Università degli Studi di Modena nell'anno 1977. È professore ordinario di "Sistemi di Elaborazione delle Informazioni" presso la Facolta` di Ingegneria dell'Università di Modena e Reggio Emilia (sede di Modena) e guida il gruppo di ricerca su Database, "DBGROUP", presso il Dipartimento di Ingegneria dell'Informazione (www.dbgroup.unimo.it.) La sua attività di ricerca è stata principalmente rivolta alla rappresentazione ed alla gestione della conoscenza nelle Basi di Dati di elevate dimensioni, con particolare attenzione sia agli aspetti teorici e formali sia a quelli implementativi. E` stata molto attiva nell'area dell'accoppiamento di tecniche di Intelligenza Artificiale, e Basi di Dati al fine di sviluppare Sistemi di Basi di Dati Intelligenti. Su tali argomenti sono stati ottenuti rilevanti risultati teorici ed è stato sviluppato il sistema ODB-Tools per il controllo di consistenza di schemi e l'ottimizzazione semantica delle interrogazioni. Recentemente si è occupata di Integrazione Intelligente di Informazioni, proponendo un sistema a mediatore, chiamato MOMIS, per fornire un accesso integrato a sorgenti di informazioni strutturate e semistrutturate che consenta all'utente di formulare una singola interrogazione e di ricevere una risposta unificata.
Dal 2001 e` coordinatore del SIG "Agenti Intelligenti" di AgentLinkII e dal 2002 e` coordinatore del progetto di ricerca europeo SEWASIE che ha come obiettivo lo sviluppo di un motore di ricerca semantico. Ha pubblicato piu` di novanta articoli su riviste e conferenze internazionali e le sue ricerche sono state finanziate da MURST, CNR, ASI e Comunità Europea. È stata membro nel comitato di programma di numerose conferenze nazionali ed internazionali di Basi di Dati e Intelligenza Artificiale. E` membro di IEEE Computer Society e di ACM.
Per una descrizione dettagliata dell'attivita` di ricerca si veda: www.dbgroup.unimo.it.

Testo inglese
Sonia Bergamaschi was born in Modena (Italy) and received her Laurea degree in Mathematics from Università di Modena on 1977. She is currently full professor of Computer Engineering in the Engineering Faculty at the Università di Modena e Reggio Emilia (associate professor from 1992 to 1999) and leads the "DBGROUP" , i.e. the database research group, at the Dipartimento di Ingegneria dell'informazione (www.dbgroup.unimo.it).
Her research activity has been mainly devoted to knowledge representation and management in the context of very large databases facing both theorical and implementation aspects.
She was very active in the area of coupling artificial intelligence (Description Logics) and database techniques to develop Intelligent Database Systems. On this topic very relevant theoretical results have been obtained and a system ODB-Tools has been developed. More recently, her research efforts have been devoted to the Intelligent Information Integration (I3) topic. An I3 system, called MOMIS, to provide an integrated access to structured and semistructured data sources and to allow a user to pose a single query and to receive a single unified answer has been proposed.
Sonia Bergamaschi is coordinator since 2001 of the Intelligent Information Agents group of the european network of excellence AgentLinkII and since 2002 of the European Research project SEWASIE whose aim is to develop a semantic search engine. She has published about ninety international journal and conference papers and her researches have been founded by the Italian MURST, CNR, ASI institutions and by European Community projects. She has served on the committees of international and national database and AI conferences.
She is a member of the IEEE Computer Society and of the ACM.
For a detailed description of the research activity and of the developed systems see: www.dbgroup.unimo.it.

1.9 Pubblicazioni scientifiche più significative del Coordinatore del Programma di Ricerca

1. *D. BENEVENTANO; BERGAMASCHI S.; C. SARTORI (2003).* ***Description Logics for Semantic Query Optimization in Object-Oriented Database Systems*** *ACM TRANSACTIONS ON DATABASE SYSTEMS. (March 2003).*
2. *D. BENEVENTANO; BERGAMASCHI S.; F. GUERRA; M. VINCINI (2003).* ***Synthesizing an Integrated Ontology*** *IEEE INTERNET COMPUTING. (vol. 7 pp. 42-51)*
3. *BERGAMASCHI S.; I. BENETTI; D. BENEVENTANO; F. GUERRA; M. VINCINI (2002).* ***An Information Integration Framework for E-Commerce*** *IEEE INTELLIGENT SYSTEMS.*
4. *BERGAMASCHI S.; S. CASTANO; D. BENEVENTANO; M. VINCINI (2001).* ***Semantic Integration of Heterogeneous Information Sources*** *DATA & KNOWLEDGE ENGINEERING. (vol. 36 pp. 215-249) Special Issue on Intelligent Information Integration, Elsevier Science B.V.*
5. *BENEVENTANO D; BERGAMASCHI S.; LODI S.; SARTORI C. (1998).* ***Consistency Checking in Complex Object Database Schemata with Integrity Constraints*** *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING. (vol. 10 (4) pp. 576-598)*

1.10 Elenco delle Unità di Ricerca

n°	Responsabile Scientifico	Qualifica	Settore Disc.	Università	Dipartimento	Mesi Uomo
1.	*BERGAMASCHI SONIA*	*Professore Ordinario*	*ING-INF/05*	*MODENA e REGGIO EMILIA*	*INGEGNERIA DELL'INFORMAZIONE*	*12*
2.	*BOUQUET PAOLO*	*Ricercatore Universitario*	*ING-INF/05*	*TRENTO*	*INFORMATICA E TELECOMUNICAZIONI*	*12*
3.	*CIACCIA PAOLO*	*Professore Ordinario*	*ING-INF/05*	*BOLOGNA*	*ELETTRONICA, INFORMATICA E SISTEMISTICA-DEIS*	*12*

4.	*MERIALDO PAOLO*	*Ricercatore Universitario*	*ING-INF/05*	*ROMA TRE*	*INFORMATICA E AUTOMAZIONE*	*12*

1.11 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		*18*	*69*	*68*	*137*
Personale universitario di altre Università		*0*	*0*	*0*	*0*
Titolari di assegni di ricerca		*0*			
Titolari di borse	*Dottorato*	*7*	*36*	*36*	*72*
	Post-dottorato	*0*			
	Scuola di Specializzazione	*0*			
Personale a contratto	*Assegnisti*	*1*	*11*	*11*	*22*
	Borsisti	*4*	*21*	*31*	*52*
	Dottorandi	*0*			
	Altre tipologie	*2*	*0*	*12*	*12*
Personale extrauniversitario		*2*	*8*	*8*	*16*
TOTALE		**34**	**145**	**166**	**311**

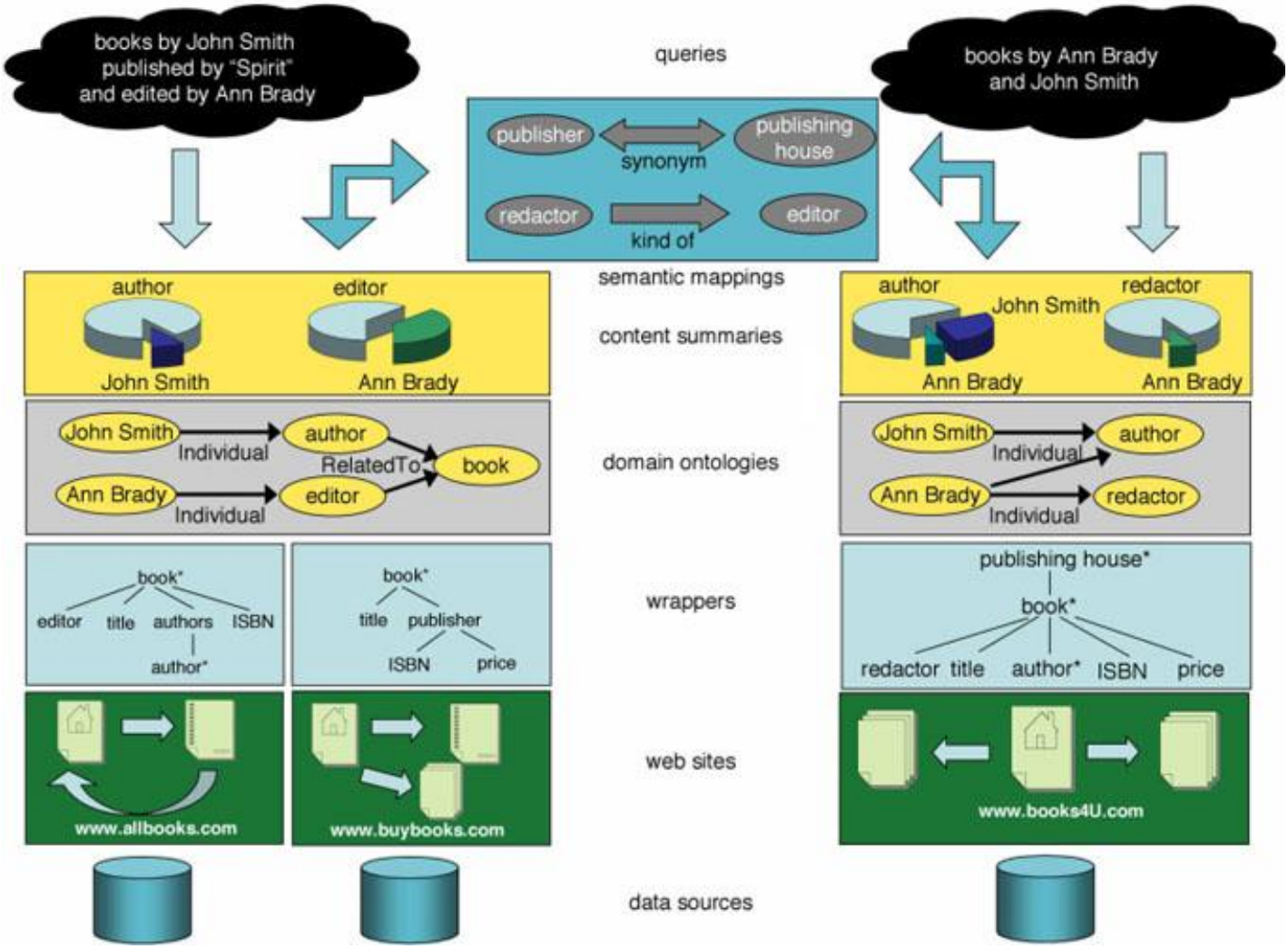
Testo inglese

		Numero	Mesi uomo 1° anno	Mesi uomo 2° anno	Totale mesi uomo
University Personnel		*18*	*69*	*68*	*137*
Other University Personnel		*0*	*0*	*0*	*0*
Work contract (research grants, free lance contracts)		*0*			
PHD Fellows & PHD Students	*PHD Students*	*7*	*36*	*36*	*72*
	Post-Doctoral Fellows	*0*			
	Specialization School	*0*			
Personnel to be hired	*Work contract*	*1*	*11*	*11*	*22*
	PHD Fellows & PHD Students	*4*	*21*	*31*	*52*
	PHD Students	*0*			
	Other tipologies	*2*	*0*	*12*	*12*
No cost Non University Personnel		*2*	*8*	*8*	*16*
TOTALE		**34**	**145**	**166**	**311**

2.1 Obiettivo del Programma di Ricerca

Testo italiano

L'enorme quantità di dati e la crescente disponibilità di servizi sul Web rendono sempre più importante lo sviluppo di infrastrutture e sistemi software che, fornendo strumenti per l'integrazione delle risorse informative, per la loro localizzazione e per la fruizione personalizzata delle stesse, permetta ai clienti (sia umani che artificiali) collegati alla rete di "ricaricarsi" delle informazioni di interesse, evitando i problemi di "information overloading" che si riscontrano usando i comuni motori di ricerca.
Il progetto WISDOM ha come obiettivo principale lo sviluppo di tecniche, e strumenti, basati su ontologie di dominio, per la ricerca efficace ed efficiente di informazione su Web e si colloca quindi nell'ambito di ricerca del Semantic Web. Si articolerà in tre temi tra loro sinergici e complementari e definira' un'architettura metodologica e funzionale di riferimento al fine di garantire coerenza tra le soluzioni che verranno messe a punto nei tre temi.
TEMA 1: Creazione ed Estensione di una Ontologia di Dominio
TEMA 2: Semantica Emergente: Scoperta di Mapping Semantici tra Ontologie di Dominio
TEMA 3: Elaborazione di Interrogazioni
L'obiettivo del primo tema è lo studio e lo sviluppo di soluzioni per la rappresentazione semantica dei contenuti delle sorgenti informative in ambito Web, con particolare riferimento ai siti data-intensive e ai siti/pagine Web con contenuto scarsamente strutturato. La rappresentazione ed integrazione di tali sorgenti informative portera' alla creazione di ontologie di dominio e alla loro eventuale modifica per effetto della scoperta/integrazione di nuove sorgenti informative. L'obiettivo del secondo tema è lo sviluppo di soluzioni per realizzare il mapping semantico fra ontologie di dominio in ambito Web, con particolare riferimento allo sviluppo di tecniche e strumenti di supporto alla identificazione, scoperta, validazione e memorizzazione di relazioni semantiche. L'obiettivo del terzo tema è lo sviluppo di tecniche di ricerca di informazione su web in grado di utilizzare l'infrastruttura semantica sviluppata dai temi 1 e 2. Considerando l'eterogeneita' dei dati/siti trattati e i vincoli imposti dall'ambiente distribuito, verranno studiati e sviluppati di meccanismi efficaci ed efficienti di elaborazione delle interrogazioni che usano la caratterizzazione delle sorgenti per selezionare le sorgenti utili, che risolvono i problemi di riscrittura e di integrazione dei risultati sulle sorgenti.
In figura e` rappresentato uno scenario di riferimento per il progetto: due diverse ontologie riferite ad uno stesso dominio, che rappresentano anche conoscenza estensionale, messe in relazione con semplici mapping semantici. Le due interrogazioni, Query1 e Query2, sono poste ciacuna con riferimento ad una ontologia: le tecniche sviluppate nel progetto permetteranno di rispondere a ciascuna query interrogando, se rilevanti, tutti i siti che sono riferiti alle ontologie presenti nella rete.

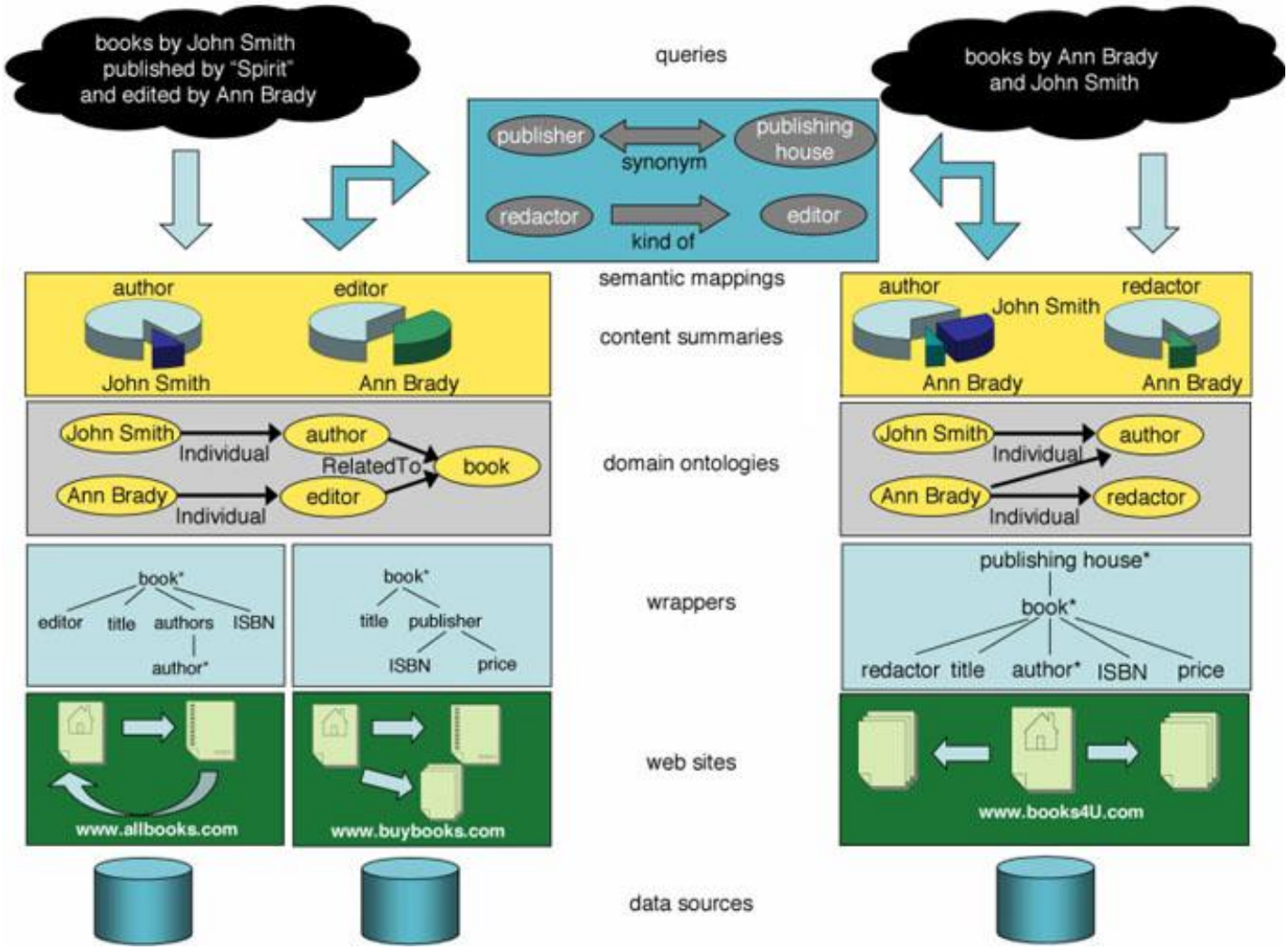


Relativamene al TEMA 1, un primo obiettivo è la definizione di un linguaggio di ontologia per la descrizione strutturale e semantica dei contenuti delle sorgenti, in termini di metadati, compatibile con standard W3C (XML, RDF, RDFS, XML Schema, OWL). In particolare, per far fronte a query specifiche, tale linguaggio deve consentire una caratterizzazione sintetica del contenuto (istanze) delle sorgenti informative.
Una ontologia di dominio è rappresentata come una vista globale virtuale (GVV - Global Virtual View) di un insieme di sorgenti informative relative allo stesso dominio. Per i siti data-intensive, il primo problema da affrontare è l'estrazione dello schema tramite opportuni wrapper generati automaticamente. Un secondo problema è quello di dare una semantica ai dati estratti da wrapper generati automaticamente. Per tale problema si valuteranno estensioni alle tecniche per la annotazione dei dati estratti da wrapper con approcci basati sulla semantica dell'ontologia di dominio. Problemi di natura diversa riguardano i siti e le pagine Web con contenuto scarsamente strutturato. In questo caso l'approccio è sfruttare la tecnologia dei Web search engine (es. Google), opportunamente estesa/complementata con strumenti di natura semantica. Si proporranno tecniche mirate alla costruzione di schemi di classificazione - tipicamente gerarchici - dei documenti disponibili. Un ultimo obiettivo è lo sviluppo di tecniche per estendere una ontologia di dominio tramite l'aggiunta di una nuova sorgente informativa.

Relativamente al TEMA 2, un primo obiettivo è la definizione di un linguaggio per la rappresentazione di mapping complessi tra ontologie di dominio. Tale linguaggio dovrà permettere di rappresentare tutti quegli elementi del mapping che sono necessari per potere risolvere una query rispetto a differenti ontologie, cioè la riscrittura della query, e l'individuazione di sorgenti utili. Un altro obiettivo è l'analisi e lo sviluppo di tecniche di mapping semantico tra ontologie, compresa una valutazione del contributo che ognuna di queste tecniche può portare alla computazione delle differenti tipologie di mapping definite nel linguaggio. Tecniche innovative per scoprire mappings tra ontologie di dominio saranno definite. In particolare, verranno considerate tecniche basate sulla semantica del linguaggio e catene lessicali e le tecniche basate sulla deduzione logica. Inoltre, verranno considerate anche tecniche per inferire mapping in base alla similarità tra i dati contenuti nelle sorgenti in quanto la caratterizzazione sintetica del contenuto (istanze) è un aspetto rilevante del progetto.
L'ultimo obiettivo è la definizione di un'architettura generale per la scoperta e la gestione di mapping semantici tra ontologie, che includa i principali moduli concettuali necessari per la computazione di mapping semantici, come per esempio risorse lessicali (ad esempio, WordNet) e moduli di ragionamento automatico.

Relativamente al TEMA 3, un primo obiettivo è quello di sfruttare la caratterizzazione delle sorgenti per indirizzare l'esecuzione verso le sole sorgenti ritenute più rilevanti. A tale scopo un ruolo fondamentale viene giocato dai mapping semantici tra le ontologie di dominio e dalla definizione di una "distanza semantica" tra i concetti coinvolti nei mapping. Relativamente agli aspetti di esecuzione si intendono definire tecniche per la riscrittura automatica di interrogazioni che, sfruttando le informazioni sulla semantica dei singoli concetti descritti nelle ontologie di riferimento e il contesto in cui sono inseriti, riscrivano l'interrogazione verso le altre ontologie in una forma che sia il più possibile simile a quella originaria. La determinazione del risultato di un'interrogazione richiede di ricostruire ogni oggetto coinvolto a partire dalle informazioni relative che lo caratterizzano e che si trovano distribuite su più sorgenti ("object fusion"). In questo caso l'obiettivo è estendere i metodi noti di "full disjunction" al caso di match approssimati e di eterogeneità semantica (presenza di valori diversi per stessi attributi gestiti da più sorgenti). Ulteriore obiettivo è lo sviluppo di tecniche, corrette ed efficienti, anche al variare del criterio di combinazione (ad es., somma pesata) dei vari fattori che influenzano la rilevanza degli oggetti, per la determinazione dei "migliori" N oggetti per una data interrogazione. Ultimo obiettivo è lo sviluppo di meccanismi che permettano una navigazione interattiva del risultato, rispettando i livelli di astrazione offerti dalle ontologie. A tal fine si studieranno opportuni operatori che consentano di fruire del risultato a diversi livelli, favorendo l'individuazione di pattern significativi nei dati da parte dell'utente.

Testo inglese
The huge amount of data and services available on the Web requires the development of systems that, overcome the "information overloading" problem of traditional search engines. In particular, there is the need of developing novel tools for the integration, the localization and the customizable fruition of informative resources which allow the clients to "recharge" with interesting data. WISDOM goal is to develop intelligent techniques and tools, based on domain ontologies, to perform effective and efficient information search on the WEB. In particular, we want to retrieve information in an integrated and efficient way both by data-intensive and unstructured site/web pages. The project will be articulated in three synergic and complementary themes and will define a reference methodological and functional architecture in order to ensure compatibility of the solutions offered by the three themes. The goal of Theme1(Creation and extension of a domain ontology) is the study and development of solutions to represent the semantics of the contents of Web sources, with particular reference to data-intensive and unstructured Websites/pages. The goal of Theme2 (Emergent Semantics: Discovering semantic mappings among domain ontologies) is the study and development of techniques and tools to support identification, discovery and storage of semantic mappings among domain ontologies. Semantic mappings that will be investigated will include the elements necessary to solve a query with respect to more ontologies and allow the development of techniques based on language semantics, lexical chains and logic inferences. The goal of Theme3 (Query processing) is the development of techniques for searching information, able to exploit the semantic infrasctructure developed within Themes 2 and 3. Considering data/sites hetereogeneity and the constraints imposed by the distributed environment, efficient and effective query processing mechanisms, using sources characterization to individuateuseful sources, solving rewriting problems and integrating results from different sources, will be studied and developed.
A reference scenario for the project is represented in figure: two different ontologies referred to the same domain, which represent, besides intensional, extensional knowledge and which are related by means of simple semantic mappings. The two queries, Query1 and Query2, are formulated with reference to a specific ontology: the techniques developed in the project will allow to answer both by accessing, if relevant, all the sites referred to the ontologies which are in the net.



Theme1: the first goal is the definition of an ontology language able to express the structural and semantic descriptions of a source contents, in terms of metadata, compliant with the W3C (XML, RDF, RDFS, XML Schema, OWL) standard. In particular, in order to be able to support the individuation of useful sources to solve a query, the language has to support the synthetic characterization of the source contents (instances). The study of emerging standards will be focused on the issue of managing the evolution of an ontology, and on the integrated management of heterogeneous and autonomously developed ontologies. Another issue is the generation of a wrapper to extract data from the source. To address this issue, innovative and scalable techniques for automatically generating a wrapper will be developed; in particular, algorithms for inferring the schema of a data intensive web site will be used to generate a set of wrappers for extracting data from the whole site. For web sites offering unstructured contents, the research will concentrate on the classification of documents in hierarchical representations of concepts (taxonomies) and on the discovery of mappings among taxonomies.

Theme2: the first goal is the definition of a language for representing complex mappings among domain ontologies. Such a language will be used to represent all the elements of mappings which need in order to solve a query with respect to different ontologies, that is the query rewriting, and the identification of useful sources.
Another goal is the analysis and the development of semantic mapping techniques between ontologies, including an assessment of the contribution that each technique may provide to the discovery of the different kind of mappings defined in language. Innovative techniques to discover mappings among domain ontologies will be defined. In particular, we will consider discovery techniques based both on the language semantics plus lexical chains and logic inference. Moreover, since the synthetic description of the contents (instance) is a relevant aspect of the project, we will analyze techniques to infer mappings exploiting the similarity among the information sources data.
The last goal is the definition of a general architecture for the discovery and the management of mappings between ontologies, which includes the main conceptual modules used in the semantic mappings framework, such as Lexical resources (e.g. WordNet) and automatic reasoning modules.

Theme3: the first objective of is to exploit the characterization of sources to bias execution only towards the most relevant sources. To this end, a basic role is played by semantic mappings between domain ontologies and by the definition of a "semantic distance" between the concepts involved in mappings. As to execution, we will define techniques for automatic rewriting of queries which, by exploiting the information on the semantics of the single concepts described in the reference ontologies and the context where they are placed, rewrite the query against the other ontologies in a form that closely matched the original one. Determining the result of a query requires that each object involved is rebuilt starting from the relative information that characterize it, that is distributed on several sources ("object fusion"). In this case the target is to extend the known full-disjunction methods (based on exact matching

between consistent components of the objects) to the case of approximate matches and semantic heterogeneity (presence of different values for the same attributes managed by different sources). Besides, we will study techniques for automatically defining "join maps" (identification, within local sources, of the objects corresponding to the same real-world object). Further objective is to develop techniques for determining the "best" N objects for a given query; such techniques should be correct and efficient independently of the criteria chosen for combining the different factors that impact on the object relevance (e.g., weighted sum). Last target is to develop methods that allow the result to be interactively navigated, according to the abstraction levels offered by ontologies. To this end, we will investigate proper operators for seeing the results at different levels, in order to support the user in recognizing significant patterns in data.

2.2 Base di partenza scientifica nazionale o internazionale

Testo italiano

La crescente disponibilita' di informazioni pubblicate sul web e i limiti dei tradizionali motori di ricerca hanno portato allo sviluppo di una nuova area di ricerca, chiamata Semantic Web (Berners-Lee, 2001), il cui obiettivo e' quello di rendere i contenuti delle pagine Web riconoscibili attraverso l'introduzione di opportuni markup semantici (metadati). Attualmente, gli approcci al Semantic Web consentono l'annotazione semantica di risorse ipotizzando l'esistenza a-priori di ontologie in grado di descrivere il dominio di interesse. Maggiore è l'accuratezza dell'ontologia, maggiore è la precisione dell'annotazione.
Una delle sfide chiave nello sviluppo di sistemi distribuiti aperti, come il Web, una intranet aziendale o il Semantic Web, è di rendere possibile lo scambio di informazione attraverso applicazioni che utilizzano schemi ed ontologie di dominio autonomamente sviluppate per organizzare localmente le informazioni.
L'interoperabilità tra queste applicazioni dipende essenzialmente dall'abilità di scoprire o utilizzare mapping tra tali schemi ed ontologie di dominio eterogenee. In particolare, nel contesto del Semantic Web dove il numero di ontologie cresce a dismisura, poter disporre di tecniche ed algoritmi che permettano mapping tra di esse diviene un fattore cruciale per una soluzione.
Una ulteriore sfida e` quella di fornire un risultato compiuto e sintetico ad una interrogazione su un sistema aperto e distribuito quale e` quello del Web.
Nel seguito verranno illustrate, per ciascuna delle tematiche citate, lo stato dell'arte internazionale e nazionale, assieme ad una visione sintetica delle competenze delle unita` di ricerca coinvolte.

TEMA1: CREAZIONE ED ESTENSIONE DI UNA ONTOLOGIA DI DOMINIO
In questa sezione descriviamo sinteticamente la base di partenza relativa alle problematiche di rappresentazione, creazione ed estensione di un'ontologia. Per una trattazione piu' completa dello stato dell'arte rimandiamo alla Base di partenza dei Modelli B delle unita' coinvolte.

Il sistema MOMIS (Mediator Environment for Multiple Information Sources) sviluppato dall'unità di Modena si pone l'obiettivo di generare una descrizione sintetica ed integrata delle informazioni provenienti da sorgenti eterogenee, in modo che l'utente abbia a disposizione una vista globale virtuale (GVV) sulle sorgenti coinvolte senza conoscerne l'effettivo grado di eterogeneità (Bergamaschi, 2001). La GVV rappresenta una concettualizzazione del dominio di interesse, cioè una Ontologia di Dominio, ottenuta a partire dalle sorgenti stesse.

Per quanto riguarda la creazione e la popolazione di una ontologia e' utile distinguere le tecniche proposte per i siti data-intensive da quelle per i siti con contenuto scarsamente strutturato.
Per i siti data-intensive l'estrazione di informazioni dalle pagine web viene realizzata attraverso opportuni programmi, detti wrapper. In letteratura esistono numerosi formalismi per la scrittura di wrapper (Atzeni, 1997; Crescenzi 1998; Sahuguet, 1999); successivamente a questi sono stati sviluppati sistemi, basati su tecniche di machine learning supervisionato, per la generazione semi-automatica di wrapper (Kushmerick, 1997; Muslea, 1999; Soderland, 1999; Adelberg, 1998; Embley,1999). Queste proposte richiedono un significativo intervento umano. Recentemente, nel progetto RoadRunner (Crescenzi, 2001), sviluppato dalla unita' di Roma Tre, e nelle proposte di (Arasu, 2003) e (Chang, 2001) sono state sviluppate tecniche che consentono di automatizzare la generazione di un wrapper. Queste tecniche inferiscono un wrapper per un insieme di pagine strutturalmente simili, analizzandone similitudini e differenze.
Per adottare queste tecniche al fine di costruire nuove ontologie e' necessario affrontare alcuni interessanti problemi. In primo luogo e' necessario inferire la struttura (o schema) di un sito web: in sostanza e' necessario individuare gli insiemi di pagine simili offerti dal sito. In letteratura il problema e' stato studiato solo con riferimento ad un particolare dominio (siti di notizie), e assumendo l'esistenza di due tipologie di pagine definite a priori (pagine indice e pagine di contenuto) (Liu, 2004; Kao, 2004). Un ulteriore problema e' quello di dare una semantica ai dati estratti da wrapper generati automaticamente. Tecniche per la annotazione dei dati estratti da wrapper sono state proposte in (Arlotta, 2003) e in (Wang, 2003). Nel progetto si intende estendere queste tecniche con approcci basati su analisi linguistica. Ad esempio, TUCUXI (Benassi, 2004) è un sistema che sfrutta la teorizzazione linguistica delle proprietà di coesione e coerenza (Halliday, 1976) per costruire gruppi di parole (catene lessicali) fra loro semanticamente correlate. Le catene lessicali possono essere costruite con l'ausilio di una ontologia di lessico, ad esempio WordNet (Miller, 1995) (Galley, 2003).

Per quanto riguarda i siti con contenuti scarsamente strutturati, per esempio collezioni di documenti su un dominio comune, esistono vari metodi per clusterizzarli. Questi metodi possono essere raggruppati in due principali categorie: metodi bottom-up e metodi top-down. I primi, data una collezione di documenti, utilizzano tecniche ˆ tipicamente di text mining ˆ per analizzare i contenuti dei documenti e raggrupparli in categorie; tali categorie sono poi organizzate gerarchicamente in modo semi-automatico o manuale, a seconda della precisione richiesta e della complessità del dominio. I metodi top down, dato uno schema di classificazione gerarchico (magari già esistente, per esempio quello di web directory definito dal progetto dmoz.org), utilizzano tecniche per popolare lo schema con i documenti appartenenti a una certa collezione.
Per il progetto, entrambi i tipi di tecniche verranno prese in esame, con una preferenza per le tecniche top-down, le quali privilegiamo la ricchezza e la precisione dello schema. Questo perché avere uno schema ricco e ben definito facilita la sua successiva integrazione in una ontologia di dominio (GVV).
Per la gestione dell,estensione e della modifica di una ontologia, in letteratura sono stati proposti due approcci. Il primo e' basato

sull'evoluzione (Motik, 2002) e mira ad adattare i concetti di una ontologia alle variazioni del dominio modellato. Il secondo e' basato sul versioning (Klein, 2001): i cambiamenti vengono gestiti creando differenti versioni della stessa ontologia. In WISDOM il problema dell'estensione di una ontologia verra` affrontato seguendo l'approccio basato sull'evoluzione applicato ad una GVV sviluppata con MOMIS.

TEMA 2: SEMANTICA EMERGENTE: SCOPERTA DI MAPPING SEMANTICI TRA ONTOLOGIE DI DOMINIO
Allo stato attuale, i mapping tra ontologie vengono definiti per lo più a mano, con un processo molto dispendioso (in termini di risorse e tempo) e suscettibile di frequenti errori; queste considerazioni hanno motivato numerose attività di ricerca sui metodi per descrivere mapping, manipolarli e generarli (semi)automaticamente.
Gli approcci proposti in letteratura per la definizione e la generazione dei mapping possono essere analizzati secondo due dimensioni principali: l'architettura generale e le tecniche di generazione dei mapping.
Per quanto concerne l'architettura generale, sono due essenzialmente gli approcci proposti: "Global Schema" e "Peer-to-Peer" (P2P). Dati due schemi (locali) da mappare, il primo approccio ha lo scopo principale di crearne un terzo (possibilmente virtuale), detto schema globale, atto ad integrare i primi due. I metodi per creare lo schema globale sono principalmente due (Fagin, 2003): GAV (Global as View) e LAV (Local as View). Essi differiscono nel modo in cui i mapping sono definiti: nel GAV, ogni elemento dello schema globale è definito tramite una query sugli schemi locali. Nel LAV, ogni elemento di uno schema locale è definito con una query sullo schema globale. Recentemente, un nuovo metodo, denominato GLAV (Global Local As View) è stato proposto in (Fagin, 2003), dove i mapping mettono in relazione una query sugli schemi locali con una query sullo schema globale.
L'approccio "Peer-to-Peer" non presuppone l'esistenza di alcuno schema globale e si basa sulla generazione di mapping "diretti" tra elementi di schemi differenti (Madhavan, 2001; Bouquet, 2003a; Giunchiglia, 2003). Tale approccio appare particolarmente vantaggioso quando i mapping tra ontologie/strutture devono essere computati a run-time, ovvero non esiste modo di integrare a priori gli schemi da mappare in uno schema globale unico.

Le tecniche di generazione dei mapping possono essere suddivise essenzialmente in quattro classi (Rahm, 2001; Giunchiglia, 2003): graph matching, schema matching, semantic matching e instance-based matching.
- Graph matching: in queste tecniche, uno schema è visto come un insieme di nodi uniti da un insieme di archi (un grafo) ed i mapping vengono generati considerano solo conoscenza strutturale, ignorano completamente altre fonti di informazione (Zhang 1995; Wang, 1994, Pelillo, 1998; Milo, 1998; Carroll, 2002; Valtchev, 2003].
- Schema matching: queste tecniche hanno come scopo principale la determinazione della similarità tra nodi appartenenti a schemi eterogenei per mezzo di tecniche di graph matching, con in aggiunta alcune informazioni di tipo "linguistico". In particolare, in tali tecniche viene largamente utilizzato un Lessico (o Thesaurus) per interpretare le etichette dei nodi del grafo al fine di riuscire a gestire casi di sinonimia ed ipernimia (Madhavan, 2001; Bergamaschi, 2001).
- Semantic matching: un mapping è detto "semantico" se possiede una chiara interpretazione "model-theoretic", come p.e. relazioni del tipo "equivalenza logica" oppure "implicazione logica". Questi mapping sono dedotti per mezzo di tecniche di ragionamento automatico su formule che rappresentano il significato dei singoli nodi di uno schema. Una tale formula è costruita utilizzando informazione proveniente da un Lessico (p.e. WordNet) e da una ontologia di dominio (Bouquet, 2003a; Giunchiglia, 2003).
- Instance based matching: diversamente dalle precedenti tecniche, i mapping vengono inferiti in base alla similarità tra i dati contenuti negli schemi stessi (Doan, 2002; Honiden, 2003).

Tra i sistemi più interessanti per che usano tecniche di generazione dei mapping citiamo COMA (Do, 2002), che supporta la combinazione di diverse tecniche di matching, Cupid (Madhavan, 2001), che combina algoritmi di matching per nomi e strutture, Similarity Flooding (Melnik, 2002), che fornisce un algoritmo di grande versatilità per il graph-matching, GLUE (Doan, 2002), che sfrutta tecniche di machine learning per creare i mapping tra gli schemi utilizzando in modo particolare il concetto di distribuzione della probabilità di unione, DIKE (Palopoli, 2003), che implementa un algoritmo che inferisce automaticamente i mapping attraverso l'analisi strutturale delle sorgenti, LSD (Doan, 2000), che usa tecniche di machine learning per inferire da esempi forniti dall'utente delle regole di matching generali e MOMIS (Beneventano, 2003), che utilizza per la generazione dei mapping delle relazioni derivate dall'analisi degli schemi, dal lessico e inferite tramite l'uso di logica descrittiva.

TEMA 3: ELABORAZIONE DI INTERROGAZIONI
Nell'ambito del progetto, le problematiche di interesse per il Tema 3 sono la riscrittura delle interrogazioni in base ai mapping semantici, la selezione delle sorgenti rilevanti per una data interrogazione, il recupero efficiente dei risultati, il problema della object fusion e la navigazione dei risultati.
I mapping semantici tra le ontologie, trattati nel tema precedente, hanno un ruolo fondamentale nella fase di pre-processing delle interrogazioni in quanto permettono la riscrittura delle stesse utilizzando le ontologie specifiche per ciascun dominio.
Al fine di selezionare le sorgenti più rilevanti per una data interrogazione è necessario caratterizzare strutturalmente, semanticamente e statisticamente ciascuna sorgente. Mentre per i primi due aspetti si può fare affidamento sulle descrizioni fornite da wrapper e ontologie di dominio, la descrizione statistica necessita di informazioni in grado di riassumere il contenuto della sorgente in termini di dati (istanze) gestiti. Le soluzioni attualmente esistenti in letteratura (Gravano, 1999; Ipeirotis, 2002; Gravano 2003) si basano essenzialmente sull'estrazione di un insieme di parole chiave con associate frequenze di occorrenza, e non sono quindi in grado di tenere conto delle relazioni semantiche esistenti tra i termini (concetti e valori) presenti nell'interrogazione e quelli propri della sorgente (Ganesan, 2003).
Riguardo al recupero dei risultati, il collezionare tutti i dati minimamente rilevanti dalle sorgenti selezionate per effettuarne il vaglio in base a criteri di importanza in un secondo momento non rappresenta evidentemente una soluzione efficiente. Si rendono quindi necessari meccanismi di esecuzione che limitino al minimo le risorse necessarie garantendo, nel contempo, la correttezza del risultato. Tali meccanismi devono inoltre tenere in considerazione sia la rilevanza delle sorgenti rispetto all'interrogazione, basandosi sui mapping, che le modalità di accesso alle sorgenti stesse (Fagin, 2001). In generale, occorre "pesare" opportunamente diversi aspetti che incidono sulla rilevanza dei risultati (rilevanza di sorgenti e/o istanze, completezza rispetto alla richiesta, ecc.) e sull'efficienza della risoluzione dell'interrogazione (es., tempi di risposta delle sorgenti). In letteratura il problema è stato affrontato unicamente per i casi di una singola sorgente strutturata (Ciaccia, 2000; Bruno, 2002) o di più sorgenti strutturate, accessibili via Web (Bruno, 2004).

Il problema della object fusion riguarda il raggruppamento di informazioni relative allo stesso oggetto del mondo reale memorizzate nelle differenti sorgenti. Requisito base perché tale fusione sia fattibile è che le differenti rappresentazioni dello stesso oggetto siano identificabili (Naumann 2002). Una volta individuato l'oggetto, occorre quindi affrontare problemi di eventuali inconsistenze tra le sorgenti (Bertossi, 2003; Greco, 2003; Naumann 2002; Lin, 1998). Infine, allo scopo di sintetizzare in un unico risultato tutte le informazioni relative allo stesso oggetto provenienti dalle differenti sorgenti, un operatore particolarmente promettente proposto in letteratura è quello di full-disjunction (Galindo-Legaria, 1994; Ullman 1996).
Infine, per restituire all'utente un risultato significativo e facilmente fruibile, assume notevole importanza la possibilità di navigare e sintetizzare efficacemente i dati ottenuti dall'interrogazione. A tal proposito, nell'ambito della business intelligence e dei database multi-dimensionali sono state studiate tecniche per la fruizione del contenuto informativo a differenti livelli di aggregazione attraverso l'applicazione di operatori OLAP (Gyssens, 1997). Nel settore del data mining e della knowledge extraction sono state studiate tecniche per la rappresentazione di pattern tipici del data mining, quali regole associative e cluster (Imielinski, 1996) e per la creazione di modelli general-purpose, estensibili e riusabili, per la rappresentazione di pattern (Rizzi, 2003). In entrambi i casi, comunque, la sintesi dell'informazione non è guidata da ontologie, quindi tali modelli non sono direttamente applicabili al nostro caso.

Testo inglese
The huge amount of information published on the web and the limitation of traditional search engines have motivated the development of a new research area, called Semantic Web (Berners-Lee, 2001). People in the area are working to make the contents of web pages electronically processable by means of appropriate semantic mark-up (metadata). Current approaches for the Semantic Web allow the semantic annotation of resources but they suppose the availability of an ontology describing the domain of interest. The higher the accuracy of the ontology, the higher the precision of the annotation.

One of the key challenges in the development of open distributed systems, including today's World-Wide Web, organizational intranets, and the emerging Semantic Web, is enabling the exchange of meaningful information across applications which may use autonomously developed schemas for organizing locally available information. Interoperability among these applications depends critically on the ability to discover/use mappings between these heterogeneous schemas/domain ontologies. In particular, within the Semantic Web area, where the ontologies number is excessively increasing, a crucial point for a solution is to have techniques and algorithms allowing mappings among them.

A further challenge in the Web context, and more generally with reference to distributed and heterogeneous open systems, is the capability to answer a query in a complete and synthetic way.

A description of the international and national state of the art , together with a synthetic view of the research units competences in the above cited research topics is given in the following.

THEME 1: DESIGN AND EXTENSION OF A DOMAIN ONTOLOGY
We briefly describe the building base about the representation, the creation and the extension of an ontology. For a more complete description of the state of the art, we refer to scientific base described in the Model B.

The MOMIS system (Mediator Environment for Multiple Information Sources) developed by the Modena research unit, aims at generating a synthetic and integrated description of information from heterogeneous sources. The user is provided with a global virtual view (GVV) on the involved sources and he does not see their heterogeneities (Bergamaschi, 2001). The GVV represents a conceptualization of the domain of interest, i. e. a domain ontology, obtained from the sources themselves.

Considering the creation and the population of ontologies, it is worth distinguishing the techniques developed for data-intensive web sites from those developed for loosely structured web sites. In the former case, information extraction is performed by means of appropriate programs, called wrappers. In literature first appeared several formalisms for writing wrappers (Atzeni, 1997; Crescenzi 1998; Sahuguet, 1999); then semi-automatic wrapper generator systems based on supervised machine-learning algorithms have been proposed (Kusmerick, 1997; Muslea, 1999; Soderland, 1999; Adelberg, 1998; Embley, 1999). All these proposals requires significant human intervention for generating wrappers. Recently, in the framework of an ongoing project, RoadRunner (Crescenzi, 2001), at the "Università Roma Tre", and in other proposals (Arasu, 2003; Chang, 2001), other techniques have been developed to automatize the generation of wrappers. All these techniques infer a wrapper by analyzing similarities and differences between structurally similar sample pages. These techniques raises a few interesting issues. First, since it is necessary to infer the structure (i.e. the schema) of a web site, suitable clusters of similar pages offered by the site have to be identified. In literature, this issue has been studied only for a specific domain (news web sites), and assuming only two types of pages (index pages and content pages) (Liu, 2004; Kao, 2004). Another issue is that of giving a semantic to data extracted by automatically generated wrappers. The automatic annotation of data extracted by wrappers has already been studied in (Arlotta, 2003) and in (Wang, 2003). In the project we aim at combining the technique for automatic annotation of data extracted by wrappers with approached based on linguistic analysis. For instance, TUCUXI (Benassi, 2004) is a system exploiting the linguistic theoretic properties of coherence and cohesion (Halliday, 1976) to build semantically related word groups (lexical chains). Lexical chains can be built by means of a lexical ontology, such as WordNet (Miller, 1995) (Galley, 2003).

There are several methods for clustering documents of the same domain taken from a loosely structured web site. These methods can be grouped in two main categories: bottom-up methods and top-down methods. The former use text mining techniques for analyzing a collection of documents in order to group them in categories; these categories are then either semi-automatically or manually organized into a hierarchy, depending both on the required precision level and on the complexity of the domain. Top-down methods work on a given hierarchical classification (even already existing, such as the web directories defined by the project dmoz.ord) and populate the schema with documents from a given collection. Both methods will be considered, but the richness and the precision of schemas produced by top-down methods ease the integration into a domain ontology (GVV).

In literature there are two proposals for managing the extensibility of an ontology: the first is based on the evolution of the ontology (Motik, 2002) and aims at adapting the concepts of an ontology to the changes in the domain of interest; the second is based on visioning (Klein, 2001): the changes are handled by creating different versions of the same ontology. During this project, the issue of extensibility of an ontology will be faced by following the approach based on the evolution applied to a GVV developed with MOMIS.

THEME 2: EMERGING SEMANTICS: SEMANTIC MAPPINGS DISCOVERY AMONG ONTOLOGIES
Today, mappings are still largely done by hand, in a labor-intensive and error-prone process; these considerations have motivated numerous research activities on methods for describing mappings, manipulating them, and generating them automatically (or semi-automatically).
The proposed approaches to define and to generate mappings can be analyzed according two main dimensions: the general architecture and the mapping generation techniques.
Concerning the general architecture, there are two proposed approaches: Global schema and P2P approaches.
Given two schemas to be matched, the first approach has the main aim to create a third (possibly virtual) schema, called global schema, integrating the first two ones. In literature, there are mainly two of these global approaches (Fagin, 2003): GAV (Global as View) and LAV (Local as View). They differ in the way mappings are defined: in a GAV approach, each global schema element is defined by means of a query over the local schemas. In a LAV approach, each local schema element is defined by means of a query over the global schema. Recently, a new approach, called GLAV (Global Local As View) has been proposed in (Fagin, 2003), where mappings relates a query over the local schemas to a query over the global schema.
The "peer-to-peer" approach starts from the idea that there is no global schema, and is based on the generation of "direct" mappings between elements of different schemas (Madhavan, 2001; Bouquet, 2003a; Giunchiglia, 2003). This approach seems especially suited for situation in which mappings between ontologies/schemas need to be computed at run-time, namely there is no way of integrating beforehand local schemas into a global one.

The techniques used for generating mappings can be essentially divided in four families (Rahm, 2001; Giunchiglia, 2003): graph matching, schema matching, semantic matching, and instance based matching.

Graph matching: In these techniques, a schema is viewed as a set of labeled nodes linked by a set of edges (a graph) and mappings are generated only on the basis of structural knowledge and by completely ignoring other sources of information (Zhang 1995; Wang, 1994, Pelillo, 1998; Milo, 1998; Carroll, 2002; Valtchev, 2003).

- Schema matching: these techniques have the main goal of determining similarity between nodes of heterogeneous schemas by means of graph matching techniques and some linguistic information. In particular they use Lexicons (or Thesauri) for interpreting labels of the graph in order to handle synonymy (Madhavan, 2001; Bergamaschi, 2001).

- Semantic matching: a mapping is called semantic if it has a clear model-theoretic interpretation, e.g. logically equivalent or logically implied relationships. These mappings are deduced by applying automated reasoning techniques to formulae that represent the meaning of each node in a schema. Such a formula is built by using information provided in Lexical resources (e.g. WordNet) and in domain ontologies (Bouquet, 2003a; Giunchiglia, 2003)

- Instance based matching: unlike the first three techniques mapping are inferred on the basis ofsimilarity relations between data contained in those schemas. (Doan, 2002; Honiden, 2003).

Among the most interesting systems that can be applied to the problem of schema/graphs matching we cite COMA (Do, 2002), which supports the combination of different schema matching techniques, Cupid (Madhavan, 2001), which combines matching algorithms based on name and structural matching and Similarity Flooding (SF) (Melnik, 2002), which exploits a particularly versatile graph matching algorithm, GLUE (Doan, 2002) which exploits machine learning techniques to build mapping among schemas by using in particular the joint probability distribution, DIKE (Palopoli, 2003), which implements an algorithm to automatically infer mappings by means of a sources structural analysis, LSD (Doan, 2000),, which exploits machine learning techniques to infer general matching rules starting form user-provided examples, and MOMIS (Beneventano, 2003), which uses schema-derived relationships, lexicon-derived relationships and relationships inferred by means of description logics techniques to generate mappings.

THEME 3: QUERY PROCESSING
Within the project, issues that are relevant to Theme 3 include the rewriting of queries based on semantic mappings, the selection of data sources that are relevant for a given query, the efficient retrieval of results, the problem of object fusion and the navigation of results.
Semantic mappings established between ontologies play a fundamental role in the query pre-processing phase, as they constitute the starting point for rewriting queries using the specific ontologies for the particular domain at hand.
In order to select the most "relevant" data sources for a given query, a structural, semantic, and statistical characterization is necessary for each data source. While for the first two aspects we can rely, respectively, on the descriptions obtained from the wrappers and on the domain ontologies, statistical characterization requires information concerning the data instances. The approaches available in the literature (Gravano, 1999; Ipeirotis, 2002; Gravano 2003) are based on the extraction of a set of keywords and their frequencies, thus they are not able to keep into account the semantic relationships between terms (concepts and values) present in a query and those that characterize the data source (Ganesan, 2003).
As for the retrieval of query results, a simple solution consists in first retrieving all the objects in the selected data sources that are, even marginally, relevant, and then determine the subset of objects that best satisfies the query. Obviously, such solution is not efficient at all. Rather, execution mechanisms have to be developed that ensure the result correctness and, at the same time, require a minimal amount of resources. Such techniques have to properly take into account both the relevance of data sources, based on the mappings, and their access modalities (Fagin, 2001). The global scenario, thus requires to appropriately "weigh" and combine all the different aspects that impact on both the relevance of results (e.g., relevance of data sources/instances, completeness with respect

to the desired answer) and the efficient query processing (e.g., data sources response times). This issue has been addressed in the literature only for the simple cases when a single structured data source (Ciaccia, 2000; Bruno, 2002) or multiple structured data sources accessible on the Web (Bruno, 2004) are present.
The problem of object fusion concerns the grouping of information related to the same object which are stored in different sources. This requires that the different instantiations of the same object can be identified (Naumann 2002). Once the object has been identified, we have to face the problem of possible inconsistencies of information among sources (Bertossi, 2003, Greco, 2003, Naumann 2002, Lin, 1998). Finally, in order to synthesize in a unique result all the information, coming from different sources, related to the same object, an operator proposed in the literature which seems promising is that of full-disjunction (Galindo-Legaria, 1994, Ullman 1996).
Returning the user a significant and easy to use result requires to effectively navigate and summarize data obtained from the sources. For this problem, in the business intelligence and in the multidimensional database field many techniques have been studied that allow the data to be analyzed at different aggregation levels using OLAP operators (Gyssens, 1997). In the domain of data mining and knowledge extraction many techniques have been studied for modeling typical data mining patterns, as association rules and clusters, (Imielinski, 1996) and to derive general-purpose models, characterized by extensibility and reusability, for the representation of patterns (Rizzi, 2003). In both cases, however, the summary of information is not driven by domain ontologies, thus such models cannot be directly applied to our case.

2.2.a Riferimenti bibliografici

(Adelberg, 1998) B. Adelberg. "NoDoSE a tool for semi-automatically extracting structured and semistructured data from text documents". SIGMOD'98.
(Atzeni, 1997) Mecca, G. and P. Atzeni "Cut and Paste", Journal of Computing and System Sciences, Special issue on PODS'97.
(Atzeni, 2002) P. Atzeni, G. Mecca, P. Merialdo: Managing Web-Based Data: Database Models and Transformations. IEEE Internet Computing 6(4): 33-37 (2002).
(Benassi, 2004) R. Benassi, S. Bergamaschi, M. Vincini: Web Semantic Search with TUCUXI, SEBD'04.
(Beneventano, 2003) D Beneventano, S. Bergamaschi, F. Guerra, M. Vincini: Synthesizing an Integrated Ontology, IEEE Internet Computing Magazine, 2003.
(Bergamaschi, 2001) S. Bergamaschi, S. Castano, D. Beneventano, M.Vincini: Semantic Integration of Heterogeneous Information Sources, DKE, Vol. 36(1), 2001.
(Berners-Lee, 2001) T. Berners-Lee, J. Hendler, O. Lassila: The Semantic Web.
Scientific American 2001.
(Bertino, 2004) E. Bertino, G. Guerrini, M. Mesiti: A matching algorithm for measuring the structural similarity between an XML document and a DTD and its applications. Inf. Syst. 29(1), 2004.
(Bouquet, 2003a) Bouquet, L. Serafini & S. Zanobini: Semantic Coordination: a new approach and an application, 2nd International Semantic Web Conference (ISWC'2003), October 2003, Sanibel Island, Florida, (USA).
(Bruno, 2002) N. Bruno, S. Chaudhuri, L. Gravano: Top-k selection queries over relational databases: Mapping strategies and performance evaluation. ACM TODS 27(2): 153-187 (2002).
(Bruno, 2004) N. Bruno, L. Gravano, A. Marian: Evaluating Top-k Queries over Web-Accessible Databases. To appear in ACM TODS (2004).
(Carroll 2002) J. Carroll, Hewlett-Packard: Matching rdf graphs, ISWC'02.
(Chang, 2001) Chang, Lui "IEPAD: information extraction based on pattern discovery". WWW 2001: 681-688
(Ciaccia, 2000) P. Ciaccia, D. Montesi, W. Penzo, A. Trombetta: Imprecision and User Preferences in Multimedia Queries: A Generic Algebraic Approach. FoIKS 2000: 50-71.
(Crescenzi, 1998) Crescenzi, V. and Mecca, G. "Grammars have exceptions". Information Systems 23(8): 539-565 (1998).
(Crescenzi, 2001) Crescenzi, V., Mecca, G. and Merialdo, P. "RoadRunner: Towards Automatic Data Extraction from Large Web Sites" VLDB 2001: 109-118
(Do et al., 2002) H. Do, E. Rahm: COMA - A system for flexible combination of schema matching approaches. In VLDB 2002: 610-621.
(Doan, 2000) A. Doan, P. Domingos, A. Halevy. Learning Source Description for Data Integration, WebDB'00.
(Doan, 2002) A. Doan, J. Madhavan, P. Domingos, A. Halevy: "Learning to map between ontologies on the semantic web", WWW'02.
(Fagin, 2001) R. Fagin, A. Lotem, M. Naor: Optimal Aggregation Algorithms for Middleware. PODS 2001: 102-113.
(Fagin, 2003) R. Fagin, P. Kolaitis, R. Miller, L. Popa: Data exchange: Semantics and query answering, ICDT'03
(Flesca et al. 2002) Flesca, S., Manco, G., Masciari, E., Pontieri, L., Pugliese, A. "Detecting structural similarities between xml documents". In WebDB'02, pages 55-60.
(Galindo-Legaria, 1994) C. Galindo-Legaria. Outerjoins as disjunctions. In SIGMOD 1994: 348-358.
(Galley, 2003) M. Galley, K. McKeown: Improving Word Sense Disambiguation in Lexical Chaining. IJCAI'03.
(Ganesan, 2003) P. Ganesan, H. Garcia-Molina, J. Widom: Exploiting hierarchical domain structure to compute similarity. ACM TOIS 21(1): 64-93 (2003).
(Giunchiglia 2003) F. Giunchiglia, P. Shvaiko: Semantic Matching. ISWC'03.
(Gravano, 1999) L. Gravano, H. Garcia-Molina, A. Tomasic: GlOSS: Text-Source Discovery over the Internet. ACM TODS 24(2): 229-264 (1999).
(Gravano, 2003) L. Gravano, P.G. Ipeirotis, M. Sahami: QProber: A system for automatic classification of hidden-Web databases. ACM TOIS 21(1): 1-41 (2003).
(Gyssens, 1997) M. Gyssens, L.V .S. Lakshmanan: A Foundation for Multi-Dimensional Databases. VLDB 1997: 106-115.
(Halliday, 1976) M.A.K. Halliday, R. Hasan: Cohesion in English, Longman 1976.
(Honiden, 2003) S. Honiden, R. Ichisem e H. Takeda: Integrating multiple internet directories by instance—base learning, AI and Data Integration, 2003.
(Imielinski, 1996) T. Imielinski, H. Mannila: A Database Perspective on Knowledge Discovery, CACM 39(11):58-64 (1996).
(Ipeirotis, 2002) P.G. Ipeirotis, L. Gravano: Distributed Search over the Hidden Web: Hierarchical Database Sampling and Selection. VLDB 2002: 394-405.
(Klein, 2001) M. Klein, D. Fensel: Ontology Versioning on the Semantic Web, 1th Int'l Semantic Web Working Symp, 2001.

(Kushmerick, 1997) Kushmerick, N., Weld, D. S., and Doorenbos, R. (1997). "Wrapper induction for information extraction". IJCAI'97
(Lin, 1998) J. Lin, A. O. Mendelzon: Merging Databases Under Constraints. Int. J. Cooperative Inf. Syst. 7(1): 55-76 (1998)
(Liu, 2002) Z. Liu, F. Li, W.K. Ng: Wiccap Data Model: Mapping Physical Websites to Logical Views, ER'02.
(Madhavan, 2001) J. Madhavan, P.A. Bernstein, E. Rahm: Generic Schema Matching with Cupid. In VLDB 2001: 49-58.
(Melnik, 2002) H. Garcia-Molina, S. Melnik, E. Rahm: Similarity Flooding: A Versatile Graph Matching Algorithm and its Application to Schema Matching. In ICDE 2002: 117-128.
(Milo, 1998) T. Milo, S. Zohar: Using schema matching to simplify heterogeneous data translation, VLDB'98.
(Motik, 2002) B. Motik at al: User-driven Ontology Evolution Management, EKAW'02.
(Naumann 2002) F. Naumann, M. Haussler: Declarative Data Merging with Conflict Resolution. International Conference on Information Quality (IQ 2002): 212-224.
(Muslea, 1999) Muslea, I., Minton, S., and Knoblock, C. A. (1999). "A hierarchical approach to wrapper induction". Conference on Autonomous Agents, pages 190--197.
(Palopoli, 2003) L. Palopoli, G. Terracina, D. Ursino Experiences using DIKE, a system for supporting cooperative information system and data warehouse design, IEEE Transaction on Knowledge and Data Engineering 15(2), 2003.
(Pelillo 1998) M. Pelillo, K. Siddiqi, e S. W. Zucker. 'Matching hierarchical structures using association graphs', in LCNS 98.
(Rahm 2001) E. Rahm, P.A. Bernstein. 'A survey of approaches to automatic schema matching', in VLDB Journal, 10(4), 2001.
(Rizzi, 2003) S. Rizzi, E. Bertino, B. Catania, M. Golfarelli, M. Halkidi, M. Terrovitis, P. Vassiliadis, M. Vazirgiannis, E. Vrachnos: Towards a logical model for patterns. ER 2003: 77-90.
(Sahuguet, 1999) Sahuguet, A. and Azavant, F. "Web ecology: Recycling HTML pages as XML documents using W4F". WebDB'99
(Soderland, 1997) Soderland, S. "Learning to extract text-based information from the World Wide Web". In KDD'97, pages 251--254.
(Ullman, 1996) J. D. Ullman, A. Rajaraman: Integrating Information by Outerjoins and Full Disjunctions. PODS 1996: 238-248.
(Valtchev, 2003) P. Valtchev e J. Euzenat. 'An integrative proximity measure for ontology alignment', in Proceedings of the workshop on Semantic Integration'03.
(Wang, 2003) Wang, Lochowsky "Data extraction and label assignment for web databases." In WWW 2003: 187-196.
(Zhang, 1995) K. Zhang, J. T. L. Wang, e D. Shasha: On the editing distance between undirected acyclic graphs and related problems, 6th Annual Symp. On Combinatorial Pattern Matching, 1995.
(Wang, 1994) J. Wang, K. Zhang, K. Jeong, e D. Shasha. 'A system for approximate tree matching', in Knowledge and Data Engineering, 6(4), 1994.

2.3 Numero di fasi del Programma di Ricerca:

3

2.4 Descrizione del Programma di Ricerca
Fase 1

Durata e costo previsto

Durata *Mesi 6* **Costo previsto** *Euro 100.000*

Descrizione

Testo italiano
Nel suo complesso, la prima fase del progetto sarà dedicata all'analisi critica delle soluzioni attualmente disponibili per i problemi di interesse e alla definizione dettagliata dei requisiti che il contesto generale del progetto propone sui vari temi di ricerca.
Al fine di garantire coerenza fra le soluzioni che verranno messe a punto nei tre temi, tutte le unità collaboreranno congiuntamente alla definizione di un'architettura metodologia e funzionale di riferimento (prodotto D0.R1), alle specifiche delle interfacce (D0.R2) del prototipo integrato (prodotto D0.P1). La fase si concluderà con un incontro collegiale in cui le elaborazioni sui singoli temi verranno condivise da tutte le UO. Le attività specifiche previste per i singoli temi vengono descritte di seguito.
Relativamente ai tre temi del progetto, l'attività di ricerca si articolerà nel seguente modo:

TEMA 1: CREAZIONE ED ESTENSIONE DI UNA ONTOLOGIA DI DOMINIO
L'attività durante la fase preliminare sarà svolta congiuntamente dalle 4 unità e sarà dedicata all'analisi critica delle soluzioni proposte in letteratura per la definizione di linguaggi di ontologia (prodotto D1.R1). Lo studio degli standard emergenti sarà particolarmente focalizzato sul problema dell'evoluzione delle ontologie e del trattamento integrato di ontologie eterogenee indipendentemente sviluppate.

TEMA 2: SEMANTICA EMERGENTE: SCOPERTA DI MAPPING SEMANTICI TRA ONTOLOGIE
Questa prima fase si concentrerà sull'analisi e sul confronto tra le tecniche allo stato dell'arte per il mapping semantico (prodotto D2.R1), compresa una valutazione del contributo che ognuna di queste tecniche può portare alla computazione di mapping nel contesto specifico dell'architettura di WISDOM, nella quale sono rilevanti mapping tra un certo numero di ontologie di dominio, ognuna delle quali integra una certa collezione di sorgenti informative in uno schema globale virtuale.
Nell'analizzare in modo critico i principali algoritmi di matching presenti in letteratura, particolare attenzione sarà rivolta a quelli

che prevedono tecniche per la risoluzione dei conflitti tra le varie rappresentazioni di uno stesso concetto.
Verranno inoltre analizzate eventuali proposte di standard per il mapping tra ontologie, allo scopo di valutarne il loro impiego nel contesto del progetto WISDOM.
Sulla base dei risultati di queste analisi si delineerà un framework comune per il mapping di ontologie di dominio, con particolare riferimento alla specifica di quale tipo di informazione debba essere rappresentato in un mapping.

TEMA 3: ELABORAZIONE DI INTERROGAZIONI
La prima fase del progetto sarà innanzitutto dedicata all'analisi critica dello stato dell'arte, allo scopo di definire compiutamente i limiti delle soluzioni attualmente disponibili per i problemi di interesse. Si procederà quindi alla formulazione dei requisiti specifici per i diversi argomenti di ricerca pertinenti al Tema 3. Nello specifico:
- Verrà condotta un'analisi critica dei linguaggi di interrogazione e delle tecniche di riscrittura di interrogazione basati su ontologie (prodotto D3.R1), allo scopo di evidenziarne i limiti e definire compiutamente i requisiti per gli strumenti e le tecniche che si andranno a sviluppare nelle fasi successive del progetto.
- Partendo da un'analisi delle principali tipologie di elaborazione di interrogazioni in ambiente distribuito ed eterogeneo, si definiranno compiutamente i limiti delle stesse in relazione all'architettura di WISDOM (nella quale, si ricorda, una sorgente è visibile esternamente solo attraverso l'ontologia di dominio (GVV) che la integra). In particolare, considerando i vari aspetti che possono contribuire a determinare la rilevanza di un risultato, si analizzerà se e come tali aspetti sono influenzati dall'architettura di WISDOM. Si analizzerà inoltre la possibilità di elaborare i dati restituiti dalle interrogazioni al fine di presentarli all'utente in forma compatta e facilmente fruibile, valutando in che misura sia opportuno abbinare tecniche di navigazione e sintesi proprie della business intelligence a forme di rappresentazione di pattern proprie del data mining. Si analizzeranno inoltre i paradigmi di interrogazione visuale di basi di dati alllo scopo di caratterizzarne i limiti nel caso di sistemi, quali WISDOM, basati sull'utilizzo di ontologie.

Testo inglese
The first phase of the project will be devoted to the study of the state of the art and to the definition of the requirements of the project along the various research themes.
In order to guarantee coherence among the solutions developed in the three themes, all the research units will collaborate together in the definition of a methodological and functional reference architecture (deliverable D0.R1) and will define the interfaces (deliverable D0.R2) of the integrated prototype (deliverable D0.P1). A joint meeting will conclude the first phase; thus all the research units will share the studies of each specific theme.

With respect to the three themes of the project, the research activities will be articulated as follows:

THEME 1: CREATION AND EXTENSION OF A DOMAIN ONTOLOGY
During the first phase the research activities will be jointly conducted by all the research units. The objective of this phase is to study the solutions proposed in the literature about languages for the definition of ontologies (deliverable D1.R1).
The study of emerging standards will be focused on the issue of managing the evolution of an ontology, and on the integrated management of heterogeneous and autonomously developed ontologies.

THEME 2: EMERGING SEMANTICS: SEMANTIC MAPPINGS DISCOVERY AMONG ONTOLOGIES
The research activity will focus on the analysis of the semantic mapping techniques at the state of the art and on their comparison with each other (product D2.R1). This activity will include an evaluation of each technique contribution to the mappings computation, within the specific contest of WISDOM where the mappings among the domain ontologies, integrating an information sources collection into a global virtual schema, are relevant.
The matching algorithms proposed in literature will be analyzed to the ones including techniques to solve conflicts among the different representations of the same concept.
Furthermore, standard proposals of ontologies mappings will be analyzed in order to evaluate their use within the WISDOM project. On the basis of the results of these analyses, we will develop a common framework to support domain ontology mapping, with specific reference to the definition of which kind of information has to be included in a mapping.

THEME 3: QUERY PROCESSING
The first phase of the project will be devoted to the critical analysis of state-of-the-art approaches, in order to completely define the limitations of existing solutions for the problems at hand. Then, we will articulate the specific requisites for each research issue relevant for Theme 3. In particular:
- We will perform a critical analysis of query languages and of query rewriting techniques based on ontologies (product D3.R1), with the aim of stressing their limitations and completely defining the requisites for tools and techniques that will be developed in the following phases.
- Starting from an analysis of query processing techniques for distributed and heterogeneous environments, we will identify the limits of such techniques with respect to the WISDOM architecture (where, we remind, a data source can be viewed from the outside only through the domain ontology (GVV) that includes it). In particular, considering the different aspects that can contribute in determining the relevance of a result, we will analyze if, and how, such aspects are affected by the WISDOM architecture. Moreover, we will investigate the opportunity of processing results so as to present them to the user in a compact and easy to use form, assessing whether navigation and summarization techniques borrowed from business intelligence can be coupled with pattern models typical of the data mining domain. Finally, we will consider some visual querying paradigms for databases, in order to characterize their main limitations when they are applied to systems based on ontologies, like WISDOM.

Risultati parziali attesi

Testo italiano
I prodotti attesi in questa fase del progetto sono sia di tipo rapporto tecnico (sigla R) che di tipo prototipo software (sigla P). Il numero dopo il "D" rappresenta il numero del tema (0 significa che il rapporto è comune a tutti i temi). La lista tra parentesi denota le unità coinvolte nella realizzazione del prodotto (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.R1: Rapporto sull'architettura metodologica e funzionale di riferimento (BO, MO, RM, TN)
D1.R1: Analisi Critica dei linguaggi e standard emergenti per le ontologie (BO,MO,RM,TN)
D2.R1: Analisi Critica di linguaggi e tecniche di mapping (MO, TN)
D3.R1: Analisi critica di linguaggi di interrogazione e tecniche di riscrittura basati su ontologie (BO, MO, TN)
D3.R2: Analisi critica delle tecniche di esecuzione di interrogazioni in ambiente eterogeneo (BO)

Testo inglese
All the products in this phase are technical reports (R). The number after the "D" is the theme number (0 means that the report is in common to all themes). The list in parentheses denotes the units involved in delivering the product (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.R1: Technical Report describing the methodological and functional reference architecture (BO, MO, RM, TN)
D1.R1: Technical Report describing a critical analysis of languages and emerging standard for ontologies (BO,MO,RM,TN)
D2.R1: Technical Report describing a critical analysis of mapping languages and techniques (MO, TN)
D3.R1: Critical analysis on query languages and rewriting techniques based on ontologies (BO, MO, TN)
D3.R2: Critical analysis on query processing techniques for heterogeneous environments (BO)

Unità di Ricerca impegnate

Unità n. 1
Unità n. 2
Unità n. 3
Unità n. 4

Fase 2

Durata e costo previsto

Durata *Mesi 6* **Costo previsto** *Euro 120.000*

Descrizione

Testo italiano
A livello generale, la seconda fase del progetto si caratterizza per i seguenti aspetti rilevanti:
1) verrà concordata, di comune accordo tra tutte le UO, la definizione puntuale dei metadati che descrivono le sorgenti informative ed i mapping, al fine di garantire compatibilità tra le soluzioni sviluppate nei tre temi(D0.R2: Specifiche delle interfacce dei componenti del prototipo integrato (BO, MO, RM, TN));
2) verranno prodotti risultati per ognuno degli argomenti specifici trattati dalle UO nell'ambito dei rispettivi temi.
La fase si concludera' con un incontro collegiale in cui i risultati scientifici prodotti verranno condivisi da tutte le UO. Il programma di ricerca specifico sui singoli temi e' descritto di seguito.

TEMA 1: CREAZIONE ED ESTENSIONE DI UNA ONTOLOGIA DI DOMINIO
In questa fase, le unità coinvolte definiranno il linguaggio per la specifica dell'ontologia di dominio (prodotto D1.R2). Il linguaggio di ontologia verrà sviluppato a partire dal linguaggio ODLI3, già utilizzato in MOMIS, e dovrà rispettare una serie di requisiti: in primo luogo dovrà essere compatibile con gli standard W3C, in secondo luogo dovrà essere sufficientemente espressivo per consentire il trattamento integrato di sorgenti informative. Inoltre, tale linguaggio dovrà essere in grado di rappresentare concetti estensionali, al fine di facilitare il compito di reperimento delle sorgenti rilevanti per l'esecuzione di una query.
Un'altra attività di questa fase sarà lo studio del problema dell'evoluzione dell'ontologia di dominio in seguito all'introduzione di una nuova sorgente informativa, con particolare attenzione al fatto che una modifica in uno o più concetti dell'ontologia può causare inconsistenze sia in concetti collegati, sia in altre ontologie in relazione attraverso i mapping. Tale attività prevede la progettazione e lo sviluppo di un prototipo che permetterà l'integrazione di una descrizione di una nuova sorgente tramite un processo semiautomatico basato sul lessico (prodotto D1.P1).
Relativamente al problema della generazione automatica dei wrapper, in questa fase si studieranno tecniche innovative efficaci ed efficienti per inferire la descrizione di un sito data-intensive (prodotto D1.R5). La descrizione (o schema) del sito mira ad individuare classi di pagine, le cui istanze sono gruppi di pagine che condividono la stessa struttura e che offrono le stesse informazioni intensionali. I collegamenti fra le classi rappresentano associazioni concettuali. Per i siti a contenuto scarsamente strutturato l'attività si focalizzerà principalmente sulla classificazione di documenti in rappresentazioni gerarchiche di concetti (tassonomie) e sulla scoperta di mapping fra tassonomie. Un'ulteriore attività sarà relativa alla costruzione di "content summaries", al fine di fornire una caratterizzazione ("profilo") delle sorgenti dal punto di vista statistico che permetta una più

precisa valutazione della rilevanza delle sorgenti stesse relativamente a una data interrogazione e, conseguentemente, la selezione delle sorgenti più significative (prodotto D1.R3).
Come ultima attività si effettuerà un'analisi critica delle tecniche esistenti per l'estrazione di catene lessicali, al fine di sviluppare strumenti di natura semantica in grado di migliorare l'efficacia delle tecniche attualmente utilizzate dai motori di ricerca keyword-based (prodotto D1.R4).

TEMA 2: SEMANTICA EMERGENTE: SCOPERTA DI MAPPING SEMANTICI TRA ONTOLOGIE
L'obiettivo di questa fase è duplice. Prima di tutto verrà definito un linguaggio per rappresentare mapping complessi tra ontologie di dominio eterogenee. Quindi verranno sviluppate tecniche innovative per scoprire mapping tra ontologie di dominio. In particolare, verranno considerate tecniche di scoperta basate sia sulla semantica del linguaggio e le catene lessicali (ricerca di sorgenti) che sulla deduzione logica (Context Matching). Inoltre verranno prese in esame anche tecniche per inferire mapping in base alla similarità tra i dati contenuti nelle sorgenti informative in quanto, come stabilito nel TEMA1, la caratterizzazione sintetica del contenuto (istanze) di una nuova sorgente informativa è un aspetto caratterizante del progetto.

TEMA 3: ELABORAZIONE DI INTERROGAZIONI IN AMBIENTE DISTRIBUITO
Nella seconda fase del progetto verranno messe a punto le soluzioni per gli argomenti oggetto d'indagine del Tema 3.
Verrà definito il linguaggio di interrogazione basato su ontologie di dominio e scelto l'approccio per la riscrittura di interrogazioni.L'idea di base è partire dalle similarità individuate tra i concetti nelle diverse ontologie per riscrivere in una forma che sia il più possibile equivalente a quella originaria l'interrogazione e i valori (costanti) presenti nella stessa. A tale scopo verrà definita una "distanza semantica" tra concetti di diverse ontologie legati da mapping di tipo semantico (Tema 2). Questa distanza sarà uno dei criteri che verranno usati per definire la nozione di sorgente rilevante per una data interrogazione (intuitivamente, mapping con elevata distanza semantica rendono una sorgente poco rilevante) e di "risposta buona" (in particolare, nel caso in cui un concetto mappi su più di un concetto di un'altra ontologia, non è detto che tutti tali mapping siano parimenti rilevanti).
Per la determinazione delle sorgenti rilevanti, le informazioni di natura semantica verranno combinate con quelle di natura strutturale e statistica. Per le prime l'idea è sfruttarle in modo da definire compiutamente il contesto in cui un dato concetto si colloca, per le seconde l'obiettivo è introdurre informazioni di natura quantitativa legate alle sorgenti in quanto tali (in particolare: qualità/affidabilità dei dati forniti dalla sorgente, frequenza di aggiornamento dei dati stessi) e alle istanze in esse presenti. In quest'ultimo caso l'idea è sfruttare l'arricchimento delle ontologie di dominio mediante "content summaries" (Tema 1) in modo da poter attribuire a ciascuna sorgente uno "score" di rilevanza relativo ai valori usati nell'interrogazione (intuitivamente, una sorgente può essere rilevante a livello semantico, ma non a livello di istanze, e quindi non in grado di restituire risultati che soddisfano le condizioni dell'interrogazione).

Per quanto riguarda i problemi legati all'esecuzione dell'interrogazione e alla determinazione del risultato, in questa fase si perverrà alla definizione di tecniche per l'esecuzione di interrogazioni distribuite che, considerando i limiti imposti dall'architettura di WISDOM, siano in grado di restituire i risultati ritenuti più rilevanti minimizzando le risorse necessarie. Poiché la rilevanza di un oggetto può dipendere da vari fattori e dal modo come tali fattori sono tra loro combinati, le tecniche che si svilupperanno saranno di tipo generalizzato, ovvero in grado di funzionare correttamente ed efficientemente anche al variare del criterio di combinazione. Per tale criterio, che nel caso base può ridursi ad una somma pesata dei vari fattori, verrà anche considerato il caso più generale di tipo "qualitativo", ovvero definito non necessariamente mediante tecniche numeriche. Verranno inoltre fornite soluzioni per il problema della "object fusion" estendendo il metodo della "full disjunction", con l'obiettivo di ottenere risposte complete e minimali. Inoltre, verranno definite le tecniche per la determinazione semiautomatica delle Join Map (identificazione nelle sorgenti locali degli oggetti che corrispondono allo stesso oggetto del mondo reale).

Relativamente alla fruizione dei risultati, verranno messi a punto metodi per permettere all'utente di specificare in maniera puntuale il livello di risoluzione desiderato. In particolare, si definiranno tecniche per rappresentare l'informazione in forma compatta e ricca di semantica a differente livelli di astrazione, e si individueranno operatori per la navigazione interattiva dell'informazione sui vari livelli in accordo con l'ontologia di dominio.

Testo inglese
The second phase is characterized by the following aspects: 1) the research units will define the common set of metadata to to describe the information sources and the mapping; this definition will guarantee the compatibility of the solutions developed in the three themes (deliverable D0.R2); 2) for every theme, specific results will be produced by the involved research units.

With respect to the three themes of the project, the research activities will e articulated as follows:

THEME 1: CREATION AND EXTENSION OF A DOMAIN ONTOLOGY

During this phase, the involved research units will define the language for the specification of the domain ontology (deliverable D1.R2). The ontology language will be developed from the ODLI3 language, which has been defined in MOMIS. The proposed language will satisfy the following requirements: first, it will be compliant to the W3C standards; second, it will be sufficiently expressive to allow the integrated management of heterogeneous data sources; also, it will be able to represent extensional concepts in order to ease the task of querying relevant sources.
Another issue studied during this phase will be the evolution of the ontology domain due to the insertion of a new information source. The insertion of a new information source implies two main issues. The first issue is the generation of a wrapper to extract data from the source. To address this issue, innovative and scalable techniques for automatically generating a wrapper will be developed; in particular, the research unit of Roma Tre will define algorithms for inferring the schema of a data intensive web site (deliverable D1.R5); the schema will describe the main classes of pages offered by the site, and it will then be used to generate a set of wrappers for extracting data from the whole site. For web sites offering unstructured contents, the research will concentrate on the classification of documents in hierarchical representations of concepts (taxonomies) and on the discovery of mappings among taxonomies.

The second issue to address when adding a new information source is the management of the inconsistencies that the new source can introduce. In particular, a change in the concepts of the ontology can introduce inconsistencies both among related concepts, both with respect to the mappings to other sources. the reasearch activities will focus on the design and development of prototype to integrate a new source by means of a lexicon-based semi-automatic process (deliverable D1.P1).

A further activity will be the development of techniques to produce "content summaries", in order to provide a "profile" of the information sources. Such a profile will focus on statistical properties that characterize the information source for querying purposes (deliverable D1.R3).

Finally, a critical analysis of techniques for the extraction of lexical chains will be conducted. The aim is to develop semantic methods that improve the effectiveness of traditional keyword-based search engines (deliverable D1.R4).

THEME 2: EMERGING SEMANTICS: SEMANTIC MAPPINGS DISCOVERY AMONG ONTOLOGIES
The goal of this phase is twofold. First, we will define a language able to represent complex mappings among heterogeneous domain ontologies. Then, innovative techniques to discover mappings among domain ontologies will be defined. In particular, we will consider discovery techniques based both on the language semantics plus lexical chains (sources discovery) and logic inference (Context Matching). Moreover, since the synthetic description of the contents (instance) is a relevant aspect of the project (according with THEME1), we will analyze techniques to infer mappings exploiting the similarity among the information sources data.

THEME 3: QUERY PROCESSING
In the second project phase, we will provide solutions for research issues related to Theme 3.
We will define the query language based on domain ontologies and the best approach for query rewriting will be selected. The main idea is to start from the matching obtained for concepts within the different ontologies, writing the query in a new form which is equivalent (as much as possible) to the original one. For this, we will define a "semantic distance" between concepts of different ontologies which are correlated by semantic mappings (Theme 2). Such distance will be one of the criteria used to define the relevance of a data source for a given query (intuitively, mappings with a low semantic distance involve a very relevant data source) and the "goodness" of results (in particular, when a concept is mapped, within an ontology, to several concepts, one has to consider the relevance for each of such mappings).
In order to establish which data sources are relevant for a query, semantic information will be combined with structural and statistic information. The former will be used to completely define the context where a given concept belongs; for the latter, the goal is to introduce quantitative information related to the data sources (in particular: quality/reliability of data within the source, frequency of data updates) and to the data instances therein contained. In this case, the basic rationale is to exploit the enhancement of domain ontologies using "content summaries" (Theme 1) so as to provide a relevance "score" for each data source, based on values supplied by the query (intuitively, a data source can be relevant from a semantic point of view, but not when its instances are considered, thus it cannot return results satisfying the query predicates).
With respect to issues related to query processing and to the retrieval of the result, in this phase we will define techniques for distributed query processing that, considering the limitations imposed by the WISDOM architecture, are able to return the most relevant results using a minimal amount of resources. Since the actual relevance for a given object depends on different factors, and also on how such factors are combined, we will develop general techniques, that will be able to operate correctly and efficiently also when the combination criteria are changed. For such criteria, that in the basic case reduce to a weighted sum of the different factors, we will examine also the more general case of qualitative definitions, which are not based on numerical a characterization. Solutions for the object fusion problem will also be provided, by extending the method based on full-disjunction, with the main goal of obtaining complete and minimal results. Moreover, we will define techniques for semi-automatic assessment of Join-Maps (identification of local sources for objects corresponding to the same real-world object).
To ease the usability of results, we will develop methods that allow the user to specify, in a precise way, the required level of details. In particular, we will define techniques for representing the information in a compact and rich in semantics way at different abstraction levels, and we will supply operators for interactive navigation of information on different layers, in agreement with the domain ontology.

Risultati parziali attesi

Testo italiano
I prodotti attesi in questa fase del progetto sono sia di tipo rapporto tecnico (sigla R) che di tipo prototipo software (sigla P). Il numero dopo il "D" rappresenta il numero del tema (0 significa che il rapporto è comune a tutti i temi). La lista tra parentesi denota le unità coinvolte nella realizzazione del prodotto (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.R2: Specifiche delle interfacce dei componenti del prototipo integrato (BO, MO, RM, TN)
D1.R2: Definizione del linguaggio per la specifica di una ontologia di dominio (BO, MO, TN)
D1.R3: Definizione di tecniche per la creazione di "content summaries" (BO)
D1.R4: Analisi critica delle tecniche esistenti per l'estrazione delle catene lessicali (MO)
D1.R5: Definizione di tecniche per inferire automaticamente lo schema di un sito data-intensive (RM)
D1.P1: Prototipo per l'aggiunta di una nuova sorgente informativa alla Ontologia di Dominio (MO)
D2.R2: Definizione del linguaggio per la specifica di mapping semantici (MO, TN)
D2.R3: Valutazione empirica di misure di similarità semantica (MO)
D3.R3: Definizione del linguaggio di interrogazione e delle tecniche di riscrittura basate su ontologie (BO, MO, TN)
D3.R4 Definizione di tecniche per l'esecuzione di interrogazioni in WISDOM (BO)

Testo inglese
All the products in this phase are technical reports (R) and software prototypes (P). The number after the "D" is the theme number (0 means that the report is in common to all themes). The list in parentheses denotes the units involved in delivering the product (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.R2: Definition of interfaces for the components of integrated prototype (BO, MO, RM, TN)
D1.R2: Definition of the language for the specification of an domain ontology (BO, MO, TN)
D1.R3: Definition of techniques for the creation of "content summaries" (BO)
D1.R4: Critical analysis of the existing techniques for the extraction of lexical chains (MO)
D1.R5: Definition of techniques for the automatic inference of the schema of a data-intensive web site (RM)
D1.P1: Prototype for the insertion of a new information source in into the domain Ontology (MO)
D2.R2: Definition of the language for specification of semantic mappings (MO, TN)
D2.R3: Empirical evaluation of semantic similarities measures (MO)
D3.R3: Definition on query languages and rewriting techniques based on ontologies (BO, MO, TN)
D3.R4 Definition of query processing techniques in WISDOM (BO)

Unità di Ricerca impegnate

Unità n. 1
Unità n. 2
Unità n. 3
Unità n. 4

Fase 3

Durata e costo previsto

Durata *Mesi 12* **Costo previsto** *Euro 173.500*

Descrizione

Testo italiano
Nella terza fase del progetto quanto sviluppato nella fase precedente verrà opportunamente esteso considerando i problemi specifici. In questa fase verranno anche sviluppati sei prototipi software, che permetteranno di condurre un'attività di sperimentazione sui metodi e gli strumenti sviluppati. La fase si concluderà con un incontro finale del progetto in cui i risultati delle sperimentazioni verranno discussi e valutati collegialmente. Verrà inoltre preparato, a cura di tutte le UO, un prototipo software corredato da un rapporto tecnico, sintesi conclusiva sui risultati ottenuti dal progetto (D0.P1 Prototipo integrato di sistema).
Il programma delle attività previste per i singoli temi viene descritto di seguito.

TEMA 1: CREAZIONE ED ESTENSIONE DI UNA ONTOLOGIA DI DOMINIO
Come prima attività verrà studiato in che modo arricchire semanticamente lo schema di un sito data-intensive tramite la tecnica delle catene lessicali, per la quale verranno sviluppati nuovi algoritmi a complessità computazionale lineare adatti a rappresentare efficacemente documenti Web (prodotto D1.R6).

In quest'ultima fase del Tema 1 saranno sviluppati quattto prototipi software.
Il primo sarà un prototipo che, a partire da un'ontologia di dominio esistente, implementerà tecniche di probing (interrogazione) delle sorgenti (considerando le informazioni di natura ontologica e i vincoli che da tali informazioni sono desumibili) e produrrà, a partire dai risultati ottenuti, i relativi "content summaries" (prodotto D1.P2).
Il secondo prototipo avrà come obiettivo la costruzione di catene lessicali estratte dall'analisi di siti web (prodotto D1.P3).
Il terzo prototipo che verrà prodotto sarà quello per associare documenti di risorse Web poco strutturate a schemi di classificazione predefiniti (prodotto D1.P5).
Il quarto prototipo servirà per inferire automaticamente lo schema di un sito data intensive (prodotto D1.P4).

TEMA 2: SEMANTICA EMERGENTE: SCOPERTA DI MAPPING SEMANTICI TRA ONTOLOGIE
Una prima attività sarà rivolta all'estrazione della rappresentazione sintetica delle sorgenti tramite la tecnica delle catene lessicali. In particolare, verranno valutati gli algoritmi proposti nel TEMA 1 secondo vari parametri, tra i quali alcuni di natura tecnologica (robustezza del processo estrattivo, complessità computazionale,…) altri di natura qualitativa (espressività delle catene lessicali come metodologia descrittiva delle sorgenti, possibilità di estensione in ambito multilinguistico, efficacia delle tecniche proposte in relazione all'assegnazione di semantica ai dati estratti). Verranno studiate e sviluppate misure di similarità semantica fra le catene lessicali e l'ontologia di riferimento stessa.

Verrà realizzato un prototipo software che implementerà in collaborazione tra le unità di Modena e Trento una piattaforma per la gestione e la scoperta di mapping semantici tra ontologie di dominio (prodotto D2.P1). La piattaforma delineata in tale prototipo può essere vista come un servizio che può essere invocato per generare tali mapping. Essa costituisce un sistema altamente

modulare e indipendente dal dominio, in cui diversi componenti funzionali possono essere inseriti "plug and play" o customizzati. A seconda delle scelte architetturali che verranno fatte per il progetto WISDOM, la piattaforma potrà essere usata come un servizio condiviso a livello globale, o usato localmente in modalità "peer-to-peer". Il prototipo sarà oggetto di attività sperimentale utilizzando sorgenti informative su Web.

TEMA 3: ELABORAZIONE DI INTERROGAZIONI IN AMBIENTE DISTRIBUITO
La terza fase del progetto sarà prevalentemente dedicata allo sviluppo di prototipi e alla loro integrazione, oltre che alla sperimentazione dei prototipi stessi.
I prototipi del Tema 3 saranno 2.
Il primo prototipo (prodotto D3.P1) si farà carico dell'acquisizione e analisi delle interrogazioni, oltre che della determinazione delle sorgenti rilevanti per l'interrogazione stessa e della riscrittura dell'interrogazione.
Il secondo prototipo (prodotto D3.P2) implementerà le tecniche di esecuzione di interrogazione messe a punto durante la fase 2, e includerà un'interfaccia per la navigazione interattiva dell'informazione a diversi livelli di astrazione sulla base dell'ontologia di dominio.

Testo inglese
In the third phase of the project, the results developed in the preceding phase will be appropriately extended to consider specific issues. In this phase, six software prototypes will be developed to test the methods and tools developed. The phase will be concluded with a final meeting to collegially discuss and evaluate the experimental results. All the research units will also prepare a software prototype, together with a technical report, to summarize the results obtained (D0.P1 Integrated System Prototype). The schedule of the activities planned for each theme is described in the following.

THEME 1: CREATION AND EXTENSION OF A DOMAIN ONTOLOGY
The first activity consists of semantically enriching the scheme of data-intensive web site. It will be based on the technique of lexical chains for which novel linear algorithms will be developed to efficiently represent web documents (deliverable D1.R6).
In the last phase of Theme 1, four software prototypes are released. The first prototype will implement probing (querying) techniques of sources and it will produce content summaries of results obtained (deliverable D1.P2); it will consider the ontological information of sources and the constraints they entail. The second prototype will aim at building lexical chains extracted from the analysis of website (deliverable D1.P3). The third prototype will associate loosely structured web documents to given classifying schemes (deliverable D1.P5). The fourth prototype will automatically infer the scheme of a data-intensive web site (deliverable D1.P4).

THEME 2: EMERGING SEMANTICS: SEMANTIC MAPPINGS DISCOVERY AMONG ONTOLOGIES
The first activity will address the extraction of the sources synthetic representation by means of the lexical chains technique. In particular, the algorithms, proposed in THEME 1, will be evaluated on the basis of different parameters: some technological (soundness of the extraction process, computational complexity), some others qualitative (lexical chains expressiveness to describe the sources, extension capability in the multi-linguistic area, techniques effectiveness with respect to the semantics assigned to the extracted data). Semantic similarity measures among lexical chains and the reference ontology will be studied and developed.
A software prototype implementing a platform to manage and discover semantic mappings among domain ontologies will be developed in collaboration with the units of Modena e Trento (product D2.P1). The prototype can be viewed as a service to be invoked to generate such mappings. It will constitute a highly modular system, domain independent, where the different components can be inserted "plug and play" or customized. On the basis of the architectural choices made in the WISDOM project, the prototype will be used as a global shared service, or locally used in "peer-to-peer" modality. The prototype will be tested by means of web information sources.

THEME 3: QUERY PROCESSING
The third phase of the project will be mainly devoted to the development and the integration of prototypes, and to their experimental evaluation.
Theme 3 will provide 2 different prototypes:
- The first prototype (product D3.P1) will be responsible for acquisition and analysis of queries, for determining which data sources are relevant for each query, and for query rewriting.
- The second prototype (product D3.P2) will implement the query processing techniques devised during phase 2, and will include an interface for interactive navigation of information at different abstraction levels, in agreement with the domain ontology.

Risultati parziali attesi

Testo italiano
I prodotti attesi in questa fase del progetto sono sia di tipo rapporto tecnico (sigla R) che di tipo prototipo software (sigla P). Il numero dopo il "D" rappresenta il numero del tema (0 significa che il rapporto è comune a tutti i temi). La lista tra parentesi denota le unità coinvolte nella realizzazione del prodotto (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.P1: Prototipo integrato di sistema (BO, MO, RM, TN)
D1.R6: Definizione di tecniche per associare semantica allo schema di un sito data-intensive basate su catene lessicali (RM, MO)
D1.P2: Prototipo per la creazione di "content summaries" (BO)
D1.P3: Prototipo per l'estrazione di catene lessicali da siti web (MO)
D1.P4: Prototipo per inferire automaticamente lo schema di un sito data-intensive (RM)
D1.P5: Prototipo per il popolamento automatico di classificazioni (TN).
D2.P1: Prototipo della piattaforma per la generazione/gestione automatica di mapping tra ontologie di domino eterogenee (MO,

TN)
D3.P1: Prototipo per la formulazione di interrogazioni (BO, MO)
D3.P2 Prototipo per l'esecuzione di interrogazioni distribuite in WISDOM (BO)

Testo inglese
All the products in this phase are technical reports (R) and software prototypes (P). The number after the "D" is the theme number (0 means that the report is in common to all themes). The list in parentheses denotes the units involved in delivering the product (BO - Bologna, MO - Modena, RM - Roma, TN - Trento).

D0.P1: Integrated system prototype (BO, MO, RM, TN)
D1.R6: Definition of techniques based on lexical chains to associate semantics to the schema of a data-intensive web site (RM, MO)
D1.P2: Prototype for the creation of "content summaries" (BO)
D1.P3: Prototype for the extraction of lexical chains from web sites (MO)
D1.P4: Prototype for the automatic inference of the schema of a data intensive web site (RM)
D1.P5: Prototype for the automatic population of classification schemes (TN).
D2.P1: Prototype of the framework for the automatic generation/management of mappings between heterogeneous domain ontologies (MO, TN)
D3.P1: Prototype for query formulation (BO, MO)
D3.P2: Prototype for distributed query processing in WISDOM (BO)

Unità di Ricerca impegnate

Unità n. 1
Unità n. 2
Unità n. 3
Unità n. 4

2.5 Criteri suggeriti per la valutazione globale e delle singole fasi

Testo italiano
Per ciascuna fase, e per ognuno dei tre temi in cui il progetto si articola, sono state indicate le attivita' previste e i relativi prodotti. Le attivita' basilari di valutazione dei risultati del progetto potranno essere effettuate verificando la produzione, esaminando la qualita' dei rapporti tecnici redatti e dei prototipi realizzati, e considerando la qualita' delle sedi di pubblicazione (congressi e riviste scientifiche) dei risultati stessi. Per favorire queste attivita' di controllo sulla qualita' dei rapporti e dei prototipi, il coordinatore realizzera', mantendendolo aggiornato, un sito web dove verranno resi disponibili i risultati del progetto e dove verra' pubblicizzato il calendario degli incontri e delle riunioni. Inoltre, le riunioni previste alla fine delle varie fasi del progetto verranno organizzate in modo che vi sia una parte dedicata alle presentazione dei risultati scientifici, aperta alla partecipazione di esterni interessati al progetto stesso.

Testo inglese
For each phase of the project, and for the three themes, all the research activities and the corresponding deliverables have been listed in the proposal. Therefore, the basic evaluation activities could be carried out by analyzing the results, the quality of the deliverables (both technical reports and software prototypes), and the quality of the conferences and the journals where the results will be published. In order to make these evaluation activities easier, the project leader will manage a web site of the project, where all the deliverables and the schedule of the meetings of the project will be available. Moreover, each of the meetings scheduled at the end of the phases will be open to external participants and will include sections devoted to illustrate the major scientific results.

3.1 Spese delle Unità di Ricerca

Unità di Ricerca	Voce di spesa										TOTALE
	Materiale inventariabile	Grandi Attrezzature	Materiale di consumo e funzionamento	Spese per calcolo ed elaborazione dati	Personale a contratto	Servizi esterni	Missioni	Partecipazione / Organizzazione convegni	Pubblicazioni	Altro	
Unità nº 1	*20.000*	*0*	*4.000*	*0*	*36.000*	*0*	*40.000*	*10.000*	*5.000*	*0*	115.000
Unità nº 2	*14.000*	*0*	*1.000*	*0*	*36.000*	*0*	*40.000*	*14.000*	*0*	*0*	105.000
Unità nº 3	*20.000*	*0*	*5.000*	*0*	*18.000*	*0*	*40.000*	*7.000*	*0*	*0*	90.000
Unità nº 4	*14.000*	*0*	*2.500*	*0*	*40.500*	*2.500*	*20.000*	*4.000*	*0*	*0*	83.500
TOTALE	**68.000**	**0**	**12.500**	**0**	**130.500**	**2.500**	**140.000**	**35.000**	**5.000**	**0**	**393.500**

3.2 Costo complessivo del Programma di Ricerca

Unità di Ricerca	Voce di spesa					
	RD	RA	RD+RA	Cofinanziamento di altre amministrazioni	Cofinanziamento richiesto al MIUR	Costo totale del programma
Unità n. 1	*11.500*	*23.000*	*34.500*	*0*	*80.500*	*115.000*
Unità n. 2	*0*	*31.500*	*31.500*	*0*	*73.500*	*105.000*
Unità n. 3	*16.600*	*10.400*	*27.000*	*0*	*63.000*	*90.000*
Unità n. 4	*25.100*	*0*	*25.100*	*0*	*58.400*	*83.500*
TOTALE	**53.200**	**64.900**	**118.100**	**0**	**275.400**	**393.500**

	Euro
Costo complessivo del Programma	*393.500*
Fondi disponibili (RD)	*53.200*
Fondi acquisibili (RA)	*64.900*
Cofinanziamento di altre amministrazioni	*0*
Cofinanziamento richiesto al MIUR	*275.400*

(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; legge del 31.12.96 n° 675 sulla "Tutela dei dati personali")

Firma _____ Data 30/03/2004 ore 19:45