



HAL
open science

Annotazione di Documenti nel Web Semantico Document annotation in the Semantic Web

Mario Lezoche

► **To cite this version:**

Mario Lezoche. Annotazione di Documenti nel Web Semantico Document annotation in the Semantic Web. 2004, pp.196. hal-00656684

HAL Id: hal-00656684

<https://hal.science/hal-00656684>

Submitted on 5 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÀ DEGLI STUDI ROMA TRE



Facoltà di Ingegneria
Corso di Laurea in Ingegneria Informatica

Tesi di Laurea

Annotazione di Documenti nel Web Semantico

Candidato

Mario Iezzoche

Matricola: 60471/88

Relatore

Prof.

Paolo Atzeni

Anno Accademico 2003-2004

*A Mamma e Papà che lo hanno permesso.
A Diego e Debora che lo hanno sopportato.
A Misha che non può vederlo.*

Ringraziamenti

Ringrazio immensamente il mio relatore, prof. Paolo Atzeni, per tutto quello che ha fatto per me, per ciò che mi ha insegnato, per quello che mi ha mostrato con il suo modo di agire. É raro avere la fortuna di incontrare una persona che ha una percezione globale di ciò che accade attorno a sé così spiccata ed è un piacere potervi collaborare.

Ringrazio di cuore la mia grande famiglia, Mamma che è sempre, completamente, dedita a noi e colma d'amore in ogni suo gesto, Papà che come un faro inamovibile nella tempesta è sempre accanto a noi a proteggerci, Zia Rossana che mi ha sempre trattato come un secondo figlio, Zio Gino che attraverso la sua calma e razionalità mi ha fatto comprendere modi differenti di affrontare i problemi, Zia RosaMaria che mi è stata accanto dall'infanzia e anche se siamo distanti fisicamente è sempre con me, Zio Pietro che mi ha seguito nei primi passi della vita e con cui mi confronto per qualunque cosa, Zia Velia che con i suoi meravigliosi modi è ciò che più mi dà la carica per andare avanti, Zio Franco con cui mi diverto tantissimo e che mi ha introdotto al piacere del buon vino, Nonna che mi è sempre vicina, Zia Enza che sempre mi è accanto in tutte le scelte di vita, Zio Italo con cui amo discorrere per la sua stupenda visione dell'esistenza e tutti gli altri miei parenti che come nostra tradizione sono sempre disponibili quando qualcuno è in difficoltà. Con una famiglia del genere difficilmente ci si sente soli e si perde la retta via. Grazie.

Fuori della sfera lavorativa e dalla stretta sfera di parentela di sangue mi fa piacere ringraziare un'altra famiglia. La famiglia che si ingrandisce con la vita. Gli Amici. In primis Debora per tutto quello che abbiamo condiviso assieme e tutto quello che condivideremo nel futuro, per la forza che esprime con lo sguardo, per le parole che accordano sempre i fili dell'anima, per il rispetto estremo che sempre dimostra per le scelte altrui, per il fatto d'esistere. Raffaele, un Amico di tal fatta è difficile trovarlo anche nello scorrere di più vite. Sto vivendo proprio la mia vita fortunata per questo! Simone, il mio Guf, colui con cui amo condividere passioni e sogni. Claudiuccio aka Soth, è bello essere con lui perché comunque splende il sole della felicità dove si trova. Diego, mio fratello non solo di sangue! Più andiamo avanti e più siamo stretti da un'intesa che solo chi ha un fratello può sperimentare. Moira (o anagrammino), sempre pronta a qualunque sacrificio, sempre attenta ad ogni esigenza spesso dimenticandosi di sé stessa a 'diamond in the rough' che si rivela solo a chi sa cercare bene. Fabrizio, il mio 'piccolo' fratellino che ha un animo tanto caro quanto d'effetto è la sua presenza. SimySimy, con la quale condivido la passione per la lettura e con cui mi fa sempre un piacere immenso scambiare idee ed emozioni. Marcello, l'anima artistica della famiglia, tutto ciò che tocca diviene un'opera d'arte. Un vero bohémien d'altri tempi. Andrea, l'anima del giurista gaudente, con persone come lui continuerò sempre ad aver fiducia nella giustizia. Diego, che con la sua calma e tranquillità affronta tutta la vita come una splendida nave su di un mare piatto, è bello stargli vicino. Federica, la mia piccola e bella sorellina, spero che

nonostante le avversità della vita rimarremo sempre uniti. Stefania, la mia sorellona, colei che tanto mi conosce e possiede le chiavi dei cassetti nascosti. Cristian, una persona trasmette energia vitale e quando si fa conoscere meglio trovi un tesoro in lui. Stefy ed Ory, che con il loro amore e la loro simpatia mi fanno vedere tutto più sereno. Giulia, che mi manca tanto a Kung Fu ma so che con lo spirito è sempre con me. Alessandro, con cui tante avventure sono state affrontate e tante ne affronteremo in futuro. Andrew, colui che ha per primo varcato tante soglie e che ha in se lo spirito di un artista. Stefano, con la sua amicizia incondizionata e il suo amore per il fantasy è il mio Master preferito. Alessandra, che con la volontà che la contraddistingue è riuscita a superare prove inaudite. Claudio, con cui ricordo il primo incontro nei parchi di un liceo scientifico. Cristina, la mia psicologa preferita, e già prenotata! Federica, uno spirito libero che affronta spavalidamente la vita. Riccardo, il mio meteorologo guida, colui con cui è un piacere sciare, parlare, volare e godersi la vita. Ivan, il mio cuginetto dolce e timido che rende la nostra fantastica città ancora più artistica di quello che già è. Milena, che da Catania mi ha sempre mostrato, con il suo immenso affetto per me e per la sua meravigliosa famiglia, che l'amicizia non è una questione di quotidianità e vicinanza fisica. Peppole, il mio amico, colui che nelle più dure avversità degli anni novanta è stata la voce della mia coscienza, l'amico con cui ho gioito di fantastiche giornate di allegria e pianto di tristezza, una vita assieme che continuerà ovunque il destino ci invierà. Iancelo (o al secolo Angelo), che con la sua passione calcistica e il suo sorriso hanno sempre portato la luce nei nostri incontri. Carlo, che è sempre nel mio cuore come il nostro amico comune a quattro zampe KenCiccio che mi ha accompagnato per tanti anni. Igor, che con il suo essere naturalmente speciale era, ed è, una guida in tutto ciò che faccio. PierGiorgio, una persona che affronta la vita come solo un guerriero di pace potrebbe fare. GianMario, che con il suo amore per le arti marziali e la sua gioia di vivere rallegra tutti i giorni! Tutti i miei amici con cui condivido la pratica delle arti marziali antiche, una cammino, una ricerca interiore che è bello fare in compagnia di tante anime presenti. Gabriella ed Arianna che hanno tifato sempre per me, due persone che hanno sempre avuto una parola di lode e di sprone.

E tutti gli altri che mi sono dimenticato di citare, ben pochi visto la lunghezza dei ringraziamenti, e che, in un modo o nell'altro, mi sono stati vicini in questi anni di studi. Per finire ringrazio rapidamente: Raffaele e le sue vicissitudini con MS Word che mi hanno spinto ad entrare nel meraviglioso mondo L^AT_EX.

Roma, 14 Dicembre 2004

Mario Lezoche

Indice

Introduzione	6
Obiettivo e Motivazioni Generali	6
Struttura della Tesi	9
1 Le premesse del Semantic Web	11
1.1 Introduzione	11
1.2 Meta Informazioni	12
1.3 Le caratteristiche del Web	13
2 Ontologie	16
2.1 Introduzione	16
2.2 Definizione di Ontologia	16
2.3 Ingegneria delle ontologie	19
2.3.1 Metodi di costruzione delle ontologie	19
3 Annotazioni Semantiche	21
3.1 Introduzione	21
3.2 Criteri di classificazione delle Annotazioni Semantiche	22
3.2.1 Livello di formalità del linguaggio usato	22
3.2.2 Posizionamento dell'annotazione	23
3.2.3 Destinatario dell'annotazione	23

3.2.4	Tipo di risorsa annotata	23
3.2.5	Livello di restrizione imposto al linguaggio	24
3.2.6	Modalità di Annotazione per le annotazioni semantiche su Ontologie	24
3.2.7	Classificazioni di alcuni Tool per annotazioni	24
4	I linguaggi del Semantic Web	26
4.1	Introduzione	26
4.2	La base	27
4.2.1	XML	27
4.3	Schematizzazione	33
4.3.1	RDF	33
4.3.2	Topic Maps - XTM	39
4.3.3	Differenze con RDF	47
4.4	Oltre lo schema	48
4.4.1	Il linguaggio DAML+OIL e OWL	49
5	MOSES: Un progetto di ricerca	53
5.1	Introduzione	53
5.2	Descrizione del progetto	53
5.2.1	Punti cruciali	54
5.2.2	Lista dei partecipanti	55
5.3	Innovazione	55
5.3.1	Approccio di dimensionamento dinamico al Web Semantico . .	56
5.3.2	Sistema ontologico di dimensione variabile	57
5.3.3	Integrazione di Ontologie	59
5.3.4	Agenti d'Apprendimento	60

5.3.5	Apprendimento da nuovi contenuti	60
5.4	Motore di ricerca Semantico	60
5.4.1	Aspetti di multilinguismo	61
5.5	Piano di lavoro	62
5.6	Sviluppo di una linea di lavoro	62
5.6.1	WP2 - Analisi dei requisiti e del Dominio	62
5.6.2	WP4 - Gestione della conoscenza e software semantico	63
5.6.3	WP7 - Incremento della conoscenza	63
6	MOSES: Analisi dei requisiti	64
6.1	Analisi dei requisiti	64
6.2	Sito Centrale	65
6.2.1	Organizzazione generale	65
6.3	Siti Satelliti	78
6.4	Requisiti	83
6.5	Domini e Applicazioni	88
6.6	Criteri di classificazione delle domande per MOSES	90
6.6.1	Formato della risposta	91
6.6.2	Siti coinvolti	92
6.6.3	Struttura della sorgente di informazione	93
6.6.4	Processo di estrazione	93
6.7	Classificazione	94
6.8	Struttura schede domande	98
6.9	Esempi schede domande	100
7	MOSES: Rappresentazione della conoscenza e linguaggi del web semantico	103

7.1	Rappresentazione della conoscenza e linguaggi del web semantico . . .	103
7.2	XTM e OWL	103
7.2.1	Il livello Tassonomico: Classi, Sottoclassi ed Istanze	104
7.2.2	Distinzioni tra classi ed istanze	105
7.2.3	Coerenza della Tassonomia	105
7.2.4	Supporto per la possibilità di avere più sottoclassi	106
7.2.5	Proprietà OWL e vincoli delle Topic Map	106
7.2.6	Identificazione dei soggetti, di classi e proprietà	107
7.3	Ontologia Italiana	107
7.4	Topic e strumento ITM Mondeca	114
7.4.1	Descrizione di due generici topic	115
7.4.2	Inserimento istanze delle ontologie in sistema ITM	119
7.4.3	Inserimento istanze delle ontologie in sistema Mondeca ITM attraverso script	123
8	MOSES: Annotazione in XML e OWL RDF	130
8.1	Annotazione in XML e OWL RDF	130
8.1.1	Linguaggi in Topic Map	131
8.2	Annotazione in Gate	134
8.2.1	Selezione del sottoinsieme di pagine da Annotare e delle On- tologie scelte	136
8.2.2	Basename introdotti	136
8.2.3	Tuning della configurazione del software GATE	137
9	Conclusioni	141
9.1	Semantic Web	141
9.2	Progetto svolto	142

Appendici	143
Appendice A: Ontologies	144
Appendice B: Question Card	147
Appendice C: Script PHP	162
Appendice D: File per la configurazione di GATE	184
Bibliografia	190

Introduzione

Obiettivo e Motivazioni Generali

La maggior parte dei contenuti del Web sono, al giorno d'oggi, espressamente rappresentati affinché possano essere letti da utenti umani e non sono progettati per essere elaborati da software. Un browser può, secondo ben precise istruzioni, disporre e visualizzare le informazioni in una certa maniera, ma non ha modo di riconoscerle o attribuire loro un particolare significato per elaborarle. Il 'Semantic Web'¹ o 'Web Semantico' nasce proprio in questo ambiente e permette di introdurre una struttura semantica nei contenuti del Web. Non separato dal Web 'tradizionale', ma, introducendo il significato dei dati, ne è un'estensione, che aggiunge una nuova funzionalità alle macchine, che diventano in grado di elaborare i dati, che fino ad ora semplicemente visualizzavano. Quindi, se fino ad oggi il Web è stato un mezzo di trasmissione di documenti per le persone, in questo modo diventa una fonte di dati e informazioni che possono essere elaborate automaticamente. Per realizzare tutto questo, un software deve poter accedere ad un insieme strutturato di informazioni e ad un insieme di regole d'inferenza, che gli permettano di fare ragionamenti automatizzati: questo comporta uno stretto legame tra le problematiche del Semantic Web ciò che concerne i sistemi di rappresentazione della conoscenza. Tali sistemi dispongono di un piccolo insieme di regole d'inferenza, caratteristiche nell'ambito di un dominio e limitano le interrogazioni cui possono rispondere per garantire delle

¹Che verrà introdotto nel Capitolo 2

risposte sicure, o almeno attendibili. Nel Semantic Web, invece, si vuole ottenere maggiore flessibilità, a scapito della completezza delle risposte che si possono ricavare, esattamente come nel Web tradizionale la mancanza di struttura e di controllo centralizzato il prezzo da pagare per avere ampia disponibilità delle informazioni. Aggiungere la semantica ai contenuti del Web richiede la creazione di un linguaggio che permetta di esprimere dati e regole per i ragionamenti: tale linguaggio deve permettere che le regole proprie di un sistema di rappresentazione della conoscenza possano essere utilizzate sul Web. Naturalmente, per la natura stessa del Web, è possibile che dati concettualmente diversi, siano rappresentati con lo stesso nome: un programma dovrebbe essere in grado di distinguere la situazione e quindi elaborare le informazioni in maniera appropriata. Perché questo possa avvenire, si devono creare delle ‘ontologie’ ed annotare, manualmente o automaticamente, le informazioni che le pagine contengono.

Il termine ontologia, preso in prestito dal linguaggio filosofico², indica un documento condiviso, che contiene la descrizione formale dei concetti di un dato dominio; identifica le classi più importanti, le organizza in una gerarchia, specifica le loro proprietà (che caratterizzano anche gli oggetti appartenenti alla classe) e descrive anche le relazioni più significative, che legano queste classi. Le convenzioni usate per presentare queste descrizioni, vanno dal linguaggio naturale a formalismi logici, ma è chiaro che la regolarità e una specifica formale facilitano la comprensione via software.

Le annotazioni, che rappresentano un sistema per descrivere il contenuto di una risorsa web, sono in grado di stabilire delle relazioni tra un oggetto esistente e la classe di un’ontologia, attraverso linguaggi ad hoc. Attuali sistemi come Yawas e Annotea permettono di creare e condividere annotazioni di documenti web come

²Il termine ontologia, in filosofia, indica la teoria che studia quali tipi di cose esistono.

commenti, note, spiegazioni, o altre informazioni utili al lettore. Queste annotazioni possono essere separate dai documenti, come nel caso di Annotea, e possono riferirsi all'intero documento o ad una parte di esso. Tuttavia, questi due sistemi non utilizzano un'ontologia per associare le risorse alle annotazioni. Altri sistemi, invece, sfruttano le ontologie oppure cercano di costruire metodi con i quali è possibile realizzare una marcatura semiautomatica servendosi di regole di 'estrazione della conoscenza'.

In definitiva, la struttura dei dati e la semantica introdotta dalle ontologie migliorano le potenzialità del Web: i programmi di ricerca, basandosi su un preciso concetto, cercano e trovano le pagine che effettivamente si riferiscono a quel concetto, anziché quelle che contengono parole-chiave ambigue o generiche; in questo modo la ricerca è più accurata: i programmi di ricerca, che nel Web tradizionale riportano una serie di pagine tra le quali l'utente deve ulteriormente cercare quelle d'interesse, con queste nuove funzionalità restituiscono le pagine che si riferiscono a un preciso concetto; inoltre le interrogazioni possono riguardare informazioni che non risiedono sulla stessa pagina Web: il programma di ricerca, inferendo sulle regole specificate, può individuare il dato richiesto e rispondere all'interrogazione. Le ontologie consentono anche di migliorare le prestazioni di altre applicazioni basate sul Semantic Web: ad esempio, visitando i siti di commercio elettronico, il confronto di informazioni si ottiene solo visitando 'negozi' diversi, analizzando quelle informazioni che compaiono nel layout della pagina e trascurando quelle di più difficile reperimento; l'introduzione di una o più ontologie, per uniformare le informazioni disponibili, permette un confronto più immediato dei cataloghi e anche un'analisi via software dei dati.

È chiaro che HTML, il linguaggio standard per la creazione di pagine Web, non ha le caratteristiche per strutturare i dati in base alla loro semantica, poiché specifica

come devono apparire le informazioni, e pertanto deve essere affiancato da un altro strumento per la strutturazione semantica: tutto questo comporta la separazione tra il contenuto e il layout delle pagine e permette l'indipendenza tra i due.

Le potenzialità del Semantic Web sarebbero comunque inutili se non ci fossero dei software in grado di raccogliere le informazioni dalle diverse sorgenti, elaborarle e scambiare i risultati con altri programmi. La potenza di questi software sarà tanto più grande quanto più i contenuti del Web e i servizi che li usano sono resi disponibili. L'oggetto di questa tesi è quello di capire cosa sia il Web Semantico attraverso la comprensione degli elementi costitutivi quali le ontologie, le annotazione e i linguaggi in cui lo si sta sviluppando. Altro obiettivo è quello di scegliere ed utilizzare alcune delle tecnologie vagliate per sperimentarne la validità sul campo focalizzando l'attenzione sulle metodologie di annotazione semantica.

Struttura della Tesi

Partendo dalla nozione di Semantic Web che si è introdotta, si è cercato di individuare gli strumenti, che, affiancando HTML, potenziano le funzionalità del Web: la ricerca ha portato ad individuare da una parte gli strumenti standard, sviluppati con il patrocinio del W3C³ e dall'altra dei prodotti che sono nati indipendentemente, pur essendo ad essi legati dagli scopi comuni e da alcune analogie di implementazione. La tesi è strutturata nel seguente modo:

Il *primo capitolo* presenta quelle che sono le premesse al Semantic Web.

Nel *secondo capitolo* si introduce il concetto di ontologia e si studiano i metodi per la costruzione delle ontologie e alcuni tool per il loro sviluppo.

Nel *terzo capitolo* si presentano le annotazioni ed una loro classificazione formale.

³Il World Wide Web Consortium è un'organizzazione che produce tecnologie (specifiche, software e tools) per portare il Web alla sua piena capacità, realizzando i protocolli comuni che ne promuovono lo sviluppo e ne garantiscono l'interoperabilità come fonte di informazioni e come mezzo per il commercio e la comunicazione.

Si analizzeranno anche alcuni tool utilizzati per la creazione delle annotazioni.

Nel *quarto capitolo* si introducono i linguaggi che permettono l'implementazione del Semantic Web. Si inizierà con quelli di più basso livello come l'XML, continuando poi con i linguaggi utili alla schematizzazione, ad esempio RDF e Topics Map. Si procederà poi con l'analizzare due linguaggi che vanno al di là della semplice schematizzazione, DAML + OIL (OWL) e SHOE.

Nel *quinto capitolo* si illustrerà il progetto di ricerca MOSES, applicazione pratica di ciò di cui si è trattato nei precedenti capitoli. Nel *sesto capitolo* si analizzerà la scelta delle ontologie utilizzate. Nel *settimo ed ottavo capitolo* si illustreremo le metodologie utilizzate per annotare le informazioni all'interno delle pagine.

Infine nel *nono capitolo*, si riporteranno i risultati ottenuti.

Capitolo 1

Le premesse del Semantic Web

1.1 Introduzione

Il World Wide Web è stato reso possibile dalla creazione di un insieme di standard che garantiscono l'interoperabilità a vari livelli; il protocollo di trasporto TCP/IP consente la comunicazione fra processi distribuiti sulla rete, HTTP e HTML consentono il recupero, il collegamento e la presentazione di documenti ipertestuali.

L'espansione del WWW, fenomeno di dimensioni senza precedenti nel campo della comunicazione umana, ha visto al momento il succedersi di due generazioni del web:

- Una prima generazione di pagine HTML statiche create a mano dalle persone.
- Una seconda generazione di pagine dinamiche create in maniera automatica dalle macchine grazie all'adozione di linguaggi come PHP, Perl, Java ecc...

Entrambe queste fasi della storia del Web vedono comunque i contenuti della rete creati per elaborazioni fatte dagli uomini, che sono al momento gli unici in grado di interpretarne i contenuti.

1.2 Meta Informazioni

Al momento molti degli sforzi del WWW hanno uno scopo: generare informazioni che non siano soltanto destinati alla lettura, ma che possano essere riutilizzati per applicazioni automatiche. Non c'è niente in un documento HTML che indichi l'argomento trattato o la fonte delle informazioni. L'unico tipo di ricerca che si può fare su un documento è la ricerca sul contenuto. Questo non è sufficiente nella maggior parte delle volte: usando un motore di ricerca si ottiene un qualche migliaio di hit, la maggior parte dei quali non serve assolutamente a niente.

La terza generazione del Web mira a rendere le risorse 'comprensibili' ai processi automatici utilizzando meta-informazioni sui contenuti. Le meta informazioni permettono agli autori di specificare informazioni sui loro documenti (informazioni su informazioni) che siano non soltanto leggibili, ma anche interpretabili in maniera intelligente dalle applicazioni di rielaborazione, (es. dai motori di ricerca). L'utilizzo sistematico di meta-informazioni ci porterà alla prossima generazione di Web: il Web Semantico.

Il W3C considera l'ideale, futura evoluzione del Web quella dal *machine-representable* al *machine-understandable*. L'idea è di generare documenti che possano al tempo stesso essere letti ed apprezzati da esseri umani, ma anche acceduti ed interpretati da software che potranno agire come agenti di ricerca, filtri di informazioni o intermediari (brokers) nelle comunicazioni.

Il Web si deve dunque dotare di una sovrastruttura per l'*interoperabilità semantica* tra le applicazioni, in modo da poter svolgere quelle funzioni che oggi debbono essere fatte a mano o codificate nei programmi.

Questo porta al web semantico, in cui non si esprimono *testi* (all'interno dei quali le informazioni si trovano nascoste e richiedono un umano per la loro comprensione), ma *affermazioni* (informazioni non ambigue, che esprimono relazioni tra oggetti,

risorse, esseri umani, fatti del mondo reale, e che possono essere utilizzate anche da applicazioni automatiche). Il Semantic Web NON risulta separato e distinto del Web odierno, ma è piuttosto una sua estensione.

1.3 Le caratteristiche del Web

Una base di conoscenza è l'insieme delle 'credenze' di un sistema e può essere rappresentata in modi diversi, secondo il linguaggio di rappresentazione che si utilizza. Il Web può essere inteso come una base di conoscenza: una fonte di informazioni facilmente reperibili e su cui si possono effettuare interrogazioni.

Questo accostamento ha fatto emergere, però, alcune caratteristiche del Web, che sono in contrasto con la concezione tradizionale di knowledge base:

- è un immenso contenitore d'informazioni: le pagine indirizzate crescono a dismisura ed ognuna contiene una parte infinitesima di tutto quello che può essere 'raccolto' da una singola applicazione; il database che si otterrebbe sarebbe così grande da essere ingestibile dai sistemi di rappresentazione esistenti.
- è un mondo 'aperto' e l'informazione che è raccolta da un software è sempre una piccola parte di quella disponibile; un sistema di rappresentazione della conoscenza, invece, per l'efficienza dei ragionamenti, considera il dominio come un mondo 'chiuso' e ciò che non vi è compreso è considerato 'falso': questa completezza non può essere garantita nella gestione del Web.
- è dinamico, cioè cambia molto rapidamente, intervenendo su di esso con:
 - Aggiunta di nuove pagine;
 - Modifica delle pagine esistenti: tali variazioni possono avvenire ad intervalli regolari o in qualunque momento, e riguardare parti non inerenti

- al contenuto informativo (come la correzione di errori di sintassi o il cambiamento di formattazione) oppure coinvolgere dati significativi;
- Rimozione di pagine (o di interi siti). Questa instabilità si deve riflettere in sistemi estremamente flessibili, in grado di riconoscere dati non disponibili, ma non per questo meno significativi.

Il Web, quindi, è un mondo in continua evoluzione: i dati disponibili cambiano in maniera imprevedibile, la pagine possono essere inserite, cancellate o modificate senza avvertimenti di sorta; tutto questo, in aggiunta alla mancanza di un controllo centralizzato, fa sì che i dati possano essere

- Inconsistenti, per la mancanza di vincoli di qualunque tipo;
- Inaffidabili, perchè non sono controllati in alcun modo;
- Non disponibili, se la pagina in cui sono contenuti è rimossa.

L'introduzione dei contenuti semantici deve cercare di mantenere le caratteristiche peculiari del Web (soprattutto la quantità d'informazioni largamente disponibili), ma deve cercare di sfruttarlo meglio, consentendo una ricerca più mirata e precisa e uno scambio di informazioni tra le applicazioni, che le trovano nel Web e successivamente le elaborano in base al loro significato.

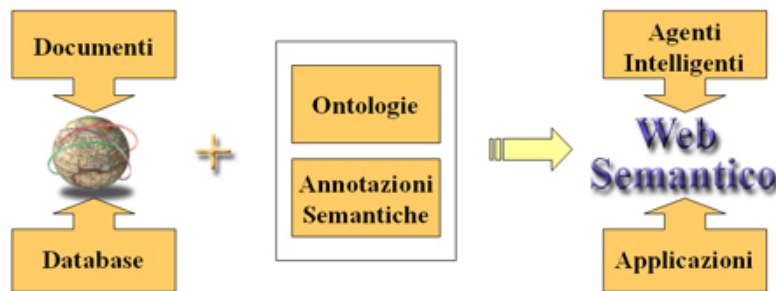


Figura 1.1: Struttura del Web Semantico

Concettualmente, la creazione del Semantic Web si basa su due elementi fondamentali, come mostrato in figura 1.1, che vaglieremo nei prossimi capitoli:

- Le ontologie di riferimento.
- Documenti sul web ‘annotati’ con informazioni semantiche, ovvero che abbiano al loro interno dei contenuti semantici.

Capitolo 2

Ontologie

2.1 Introduzione

Il termine ontologia ha origini nella filosofia e designa *lo studio e ricerca dell'essere e dell'esistenza in quanto tale, principio che condiziona le sue determinazioni particolari*.

Le ontologie hanno un ruolo di primo piano per l'elaborazione e condivisione della conoscenza sul Web. In pratica forniscono una visione comune di un determinato insieme di concetti e/o argomenti, che possono essere scambiati fra uomini o applicazioni.

2.2 Definizione di Ontologia

Che cos'è un'ontologia? Sono state date molte definizioni

- dalla filosofia "a systematic explanation of being"
- "...an explicit specification of a conceptualization" (Gruber 1993)
- "...a logical theory which gives an explicit, partial account of a conceptualization" (Guarino 1995)
- "...a formal specification of a shared conceptualization" (Borst 1997)

Dove i termini *explicit*, *formal*, *shared* e *conceptualization* hanno i seguenti significati:

- *explicit*: i tipi di concetti utilizzati e i vincoli su di essi sono definiti in modo esplicito.
- *formal*: definita in un linguaggio formale.
- *shared*: la conoscenza rappresentata in un'ontologia non è scaturita dalla decisione di un singolo individuo, ma è riconosciuta da una collettività.
- *conceptualization*: perchè un'ontologia identifica i concetti rilevanti di un dominio applicativo stabilito.

Un'ontologia contiene generalmente una gerarchia di concetti di un determinato dominio, descrive ogni concetto con attributi a cui sono assegnati dei valori, esprime relazioni fra i concetti mediante espressioni logiche e mette le costanti in relazione con uno o più concetti.

Il contenuto di un'ontologia è quindi rappresentato da:

- Un insieme di classi (concetti e astrazioni del mondo reale).
- Un insieme di relazioni (tra 2 classi); un esempio è la relazione ISA.
- Un insieme di proprietà, che sono attributi di classi o relazioni.

Come già detto, un'ontologia è un documento che indica in modo formale il significato e i legami fra termini. Più precisamente, introdotta in una specifica realtà, individua gli aspetti che sono ritenuti rilevanti nel contesto e quelli che possono essere ignorati: specifica sia i concetti sia come sono legati tra loro, quali sono le proprietà di ciascuno e come queste proprietà sono connesse, mediante regole d'inferenza e

logiche. L'introduzione delle ontologie nel Web ne consente la strutturazione delle informazioni e permette di superare alcuni aspetti critici del Web tradizionale.

In primo luogo, le varie sorgenti *'producono'* informazioni in diversi formati e la creazione di indici per localizzare queste sorgenti è piuttosto complessa, poichè è molto difficile ottenere indicazioni da sorgenti audio o video: un'ontologia può facilitare questa operazione, descrivendo in modo formale i contenuti di ogni sorgente e aiutando un utente nella ricerca di quelle sorgenti che generano un particolare tipo di informazione.

Di conseguenza, si risolve anche il problema della mancanza di struttura del Web, dovuta al fatto che HTML è un linguaggio di formattazione e non si occupa di gestire le informazioni per facilitare il loro reperimento: un'ontologia descrive il dominio dal punto di vista strutturale, definendone i componenti e i loro legami.

Il terzo problema affrontato è quello della dipendenza dal contesto: definendo mediante un'ontologia il contesto in cui è presentato un certo documento, i termini utilizzati sono propri di quell'ambito e quindi il loro significato non è ambiguo. Esattamente come il Web tradizionale si è sviluppato senza controlli e senza regole, è difficile immaginare che il Semantic Web si evolva come un insieme ordinato di ontologie, costruite da esperti.

Un esempio per comprendere meglio l'utilizzo delle ontologie è la classificazione. Un insieme di entità organizzate seguendo un criterio di accesso alla risorsa come le immagini delle proprie vacanze divise per stato in cui si sono svolte, per tipologie, mare, montagna città. Si possono avere però dei problemi per le classificazioni multiple, ad esempio si possono avere diversi domini, differente terminologia, delle proprietà interessanti per alcuni oggetti lo sono molto meno per altri, il significato degli attributi può essere differente, solo per fare alcuni esempi.

La differenza fondamentale tra le classificazioni e le ontologie è che le classificazioni si focalizzano sull'accesso basato su dei criteri predefiniti (codificati su delle chiavi sintattiche), mentre le ontologie si basano sul significato dei termini, sulla natura e la struttura del dominio cui si riferiscono.

Nei primi tempi il web semantico si estenderà con un grande numero di ontologie di dimensioni ridotte, che faranno riferimento l'una all'altra (per riutilizzare o modificare termini comuni) e saranno sviluppate dagli utenti, analogamente a quanto avviene per i contenuti del Web attualmente creati. Un altro aspetto molto importante per la creazione del Semantic Web consiste nel rendere possibile per utenti differenti, che non sono esperti, non solo la creazione di ontologie, ma anche la realizzazione di contenuti del Web che siano *'leggibili'* dalle macchine.

è importante che gli strumenti per fare tutto questo siano liberamente utilizzabili (un costo, anche basso, potrebbe frenare alcuni utenti) e, soprattutto, che diventino di uso comune.

Per far fronte a tutte queste problematiche il World Wide Web Consortium ha progettato una serie di linguaggi, che nel corso del tempo sono diventati degli standard, che consentono proprio la strutturazione semantica delle informazioni e la condivisione dei significati dei concetti.

Nel corso dei prossimi capitoli [RIFERIMENTO AI CAPITOLI INTERESSANTI], saranno analizzati questi strumenti, mettendone in luce caratteristiche e modi d'impiego.

2.3 Ingegneria delle ontologie

2.3.1 Metodi di costruzione delle ontologie

Un'Ontologia, come già detto, rappresenta una concettualizzazione condivisa di un certo dominio. Contiene, quindi, l'insieme dei concetti (entità, attributi,

processi), le definizioni e le relazioni fra concetti. Le relazioni possono essere di vario tipo: *tassonomico*(IS-A), *meronimico* (PART-OF), *telico* (PURPOSE-OF..) ecc. Un'Ontologia può avere vari livelli di formalizzazione, ma, necessariamente, deve includere un vocabolario di termini (*concept names*) con associate definizioni (*assiomi*), e (*almeno*) relazioni tassonomiche.

Capitolo 3

Annotazioni Semantiche

3.1 Introduzione

Le annotazioni sono informazioni semantiche associate a risorse web. Si usano per arricchire il contenuto informativo dei documenti e per esprimere in maniera formale, il significato di una porzione di testo in un documento, dei Web service, delle strutture dati coinvolte in un processo di collaborazione (interoperabilità tra applicazioni software).

Le annotazioni sono delle meta risorse di conoscenza, in Figura 3.1 si può avere una rappresentazione della struttura sistema di risorse di conoscenza con le annotazioni poste come livello intermedio.

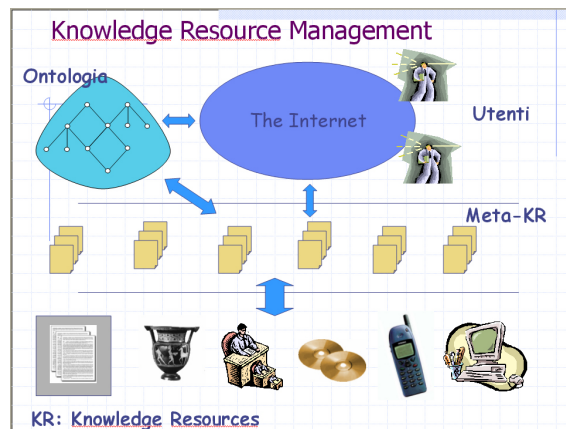


Figura 3.1: Struttura web semantico

3.2 Criteri di classificazione delle Annotazioni Semantiche

Vi sono molti criteri per classificare le annotazioni. Si può analizzare il livello di formalità del linguaggio usato, il posizionamento dell'annotazione, il destinatario dell'annotazione, il tipo di risorsa annotata, il livello di restrizione imposto sul linguaggio e la modalità di annotazione.

3.2.1 Livello di formalità del linguaggio usato

Il livello di formalità può andare dallo strettamente formale all'informale. Nel livello formale l'annotazione è espressa in un linguaggio formale di rappresentazione della conoscenza come per esempio RDF e RDFS: un framework per esprimere i metadati nel web, DAML, DAML+OIL, OWL usati per la definizione di ontologie. Il livello informale esprime l'annotazione in linguaggio naturale o in un linguaggio controllato e permette di aggiungere informazioni su documenti o risorse, fruibili da un utente umano.

Un esempio Annotazione Formale in RDF:

```
<Professor rdf:about=''file:///C:\...\PagineHtml\PaoloAtzeni.htm#PaoloAtzeni">  
<first_Name> Paolo </first_Name>  
<last_Name> Atzeni </last_Name>  
<has_affiliation rdf:resource=''file:///C:\...\PaoloAtzeni.htm#UniRomaTRE"/>  
</Professor>
```

Un esempio di Annotazione Informale in linguaggio naturale:

Paolo Atzeni è un professore dell'università Roma TRE.

3.2.2 Posizionamento dell'annotazione

Un'annotazione può essere *Embedded*, ovvero inserita all'interno dello stesso documento annotato oppure *Attached*, ovvero memorizzata separatamente e collegata al documento da un link. Nella Figura 3.2 si schematizzano le due tipologie di posizionamento.

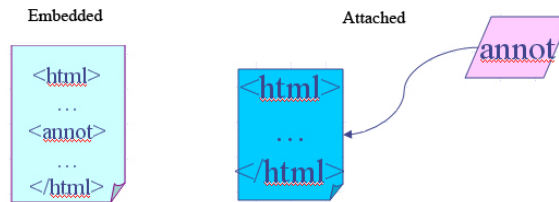


Figura 3.2: Posizionamento delle Annotazioni

3.2.3 Destinatario dell'annotazione

Il destinatario di un'annotazione può essere un utente umano o il computer. Nel primo caso l'annotazione ha lo scopo di fornire informazioni aggiuntive su una risorsa, in questo caso l'annotazione in genere sarà poco formale probabilmente realizzata in linguaggio naturale, nel secondo caso l'annotazione ha lo scopo di esplicitare il significato di una risorsa, in tal caso l'annotazione deve essere di tipo formale per poi essere processabile da una macchina.

3.2.4 Tipo di risorsa annotata

Vari tipi di risorsa possono essere annotati. *Annotazione di documenti*, ad esempio frammento di testo, pagina HTML, immagini, ecc., oppure *Annotazione di Web Services*, esprimere cosa fa un servizio, descrivere i parametri di input e l'output (questa solitamente è di tipo formale), *Annotazione di strutture dati*, processi, coinvolti nello scambio di informazioni tra applicazioni software che cooperano anche questa solitamente è di tipo formale, ed è basata sulle ontologie.

3.2.5 Livello di restrizione imposto al linguaggio

Le restrizioni imposte dal linguaggio possono essere: *Absence* in cui i termini del linguaggio possono essere usati senza vincoli, *Advised* (consigliato), in cui si è liberi di scegliere se utilizzare o meno un glossario, un ontologia, (o anche parole chiavi) in alternativa al linguaggio naturale o *Mandatory* (vincolante) in cui c'è l'obbligo di usare un glossario di riferimento o un ontologia (Annotazione Ontology-based).

3.2.6 Modalità di Annotazione per le annotazioni semantiche su Ontologie

Questo tipo di annotazione può avvenire attraverso istanze o concetti. Nell'annotazione attraverso istanze di concetti l'annotazione consiste in un'associazione di una istanza all'elemento annotato in modo da valorizzare le proprietà che descrivono l'istanza. Nell'annotazione attraverso concetti l'annotazione consiste in un'associazione di un concetto dell'ontologia all'elemento annotato è un'associazione di una composizione di concetti dell'ontologia (attraverso opportuni operatori) all'elemento annotato. In Figura 3.3 vi è un esempio di entrambe le modalità di annotazione.

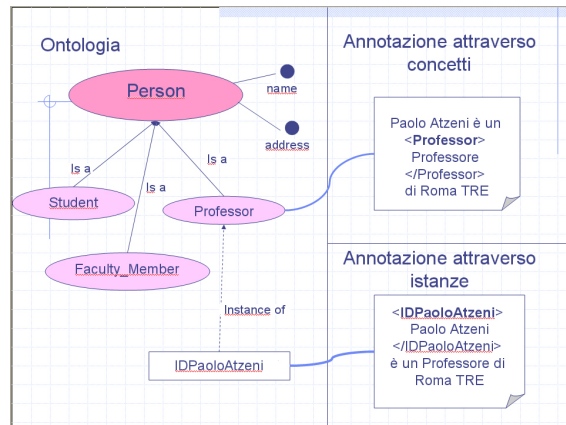


Figura 3.3: Annotazioni attraverso concetti o attraverso istanze

3.2.7 Classificazioni di alcuni Tool per annotazioni

	Formalità	Posiz.	Destinatar	Risorsa	Restriz.	ElemAnn
Ontomat	Formale (DAML+ OIL)	Embedded (nell'hea- der)	Computer	Pg.html	Ontology- based	Istanze
GATE	Formale (RDF, OWL)	Embedded / Attached	Computer	Pg.html	Ontology- based	Istanze / Concetti
Mnm	Formale (DAML+ OIL, RDF)	Attached	Computer	Pg.html	Ontology- based	Istanze
Smore	Formale (RDF, DAML+ OIL, OWL)	Embedded / Attached	Computer	Foto, mail, Pg.html	Advised	??
Annotea	Infomale	—	Human user	Pg.html	Absence	Non O-B
Cohse	Formale	—	Computer	Doc/ Pg.html	Ontology- based	Concetti
Melita	Formale	—	Computer	Doc	Ontology- based	Istanze

Tabella 3.1: Relazioni inserite in GATE

Capitolo 4

I linguaggi del Semantic Web

4.1 Introduzione

Originariamente, il Web è cresciuto intorno ad HTML, che rappresenta lo standard per strutturare i documenti, in modo che i browser possano interpretarli e tradurli uniformemente: le informazioni per la formattazione sono racchiuse nei markup di formattazione, cioè il testo da visualizzare è racchiuso tra istruzioni (i tag) che appunto indicano l'aspetto che l'informazione deve assumere a video. HTML è un linguaggio semplice, ma, se da una parte la sua semplicità ha permesso un incremento molto rapido del Web, dall'altra ha impedito l'utilizzo di alcune applicazioni avanzate. Il mondo delle informazioni tecnologiche è un ambiente in continua evoluzione e per un utente umano può essere difficile trovare ciò di cui ha bisogno in un universo così eterogeneo: un agente software potrebbe potenzialmente essergli d'aiuto, ma, sfortunatamente, non è in grado di comprendere o interagire con dati, che non sono elaborabili o compaiono in linguaggio naturale. La soluzione per risolvere questa incapacità di capire la semantica dei concetti contenuti in una pagina Web, consiste di strumenti che permettano di associare ad ogni concetto il suo significato, affinché esso sia disponibile per gli agenti software. XML (eXtension Markup Language) è il primo linguaggio che separa il markup dei contenuti da quello di presentazione: è uno standard elaborato per fare in modo che le informazioni non

siano semplicemente formattate per facilitarne il reperimento per un utente umano, ma anche facilmente elaborate da agenti software; consente la creazione di documenti strutturati, ma sfortunatamente, non gestisce la semantica dei contenuti: un programma riconosce i contenuti, ma non è in grado di attribuire loro un significato. Inoltre non fornisce meccanismi di classificazione o ragionamento e quindi deve essere affiancato ad altri linguaggi più potenti. A supporto di questo strumento, è nato RDF (Resource Description Framework): è stato sviluppato dal W3C come uno standard per la gestione dei metadati, con lo scopo di aggiungere semantica formale al Web. È definito sopra XML, che fornisce un modello dei dati e una convenzione sintattica, per rappresentare i significati dei dati in modo interoperabile. RDF è un mezzo per descrivere le relazioni tra le risorse in termini di proprietà e valori. Con RDF è stato sviluppato anche lo Schema RDF (RDFS), un sistema di tipo object-oriented, che può essere effettivamente pensato come un linguaggio ontologico e/o di rappresentazione della conoscenza. Vediamo ora più in dettaglio questi linguaggi, tenendo comunque presente che da soli non potrebbero realizzare le piene funzionalità del Web Semantico.

4.2 La base

4.2.1 XML

XML è utilizzato per creare una struttura ‘digitale’ dei documenti, cioè i documenti sono rappresentati in maniera tale da essere interpretati da un computer, affinché le informazioni contenute possano essere elaborate, memorizzate, ricercate o stampate. Prima di tutto, si deve definire una struttura logica del documento (che è comune a tutti i documenti di quel tipo) e individuare i suoi componenti: ognuno di questi è un elemento XML ed è legato agli altri da vincoli di dipendenza; tutti gli elementi che compongono un documento assumono una struttura ad albe-

ro: l'elemento principale è la radice, gli altri sono i suoi sottoelementi, che sono a loro volta rami o foglie a seconda che abbiano o no dei discendenti. Ad esempio, si consideri come tipo di documento un messaggio di posta elettronica: partendo dall'elemento principale, e-message, si considerano le sue componenti, che a loro volta sono scomposte, ottenendo la struttura rappresentata in figura 4.1.

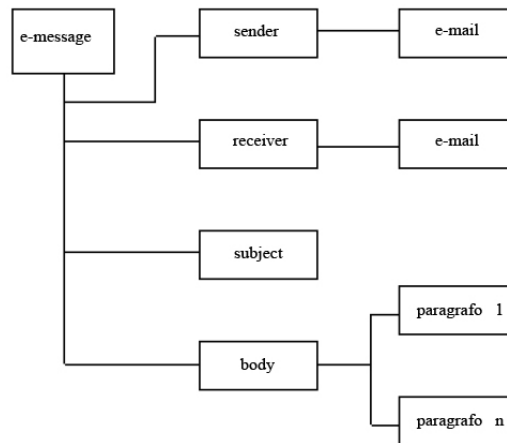


Figura 4.1: Struttura del tipo di documento e-message

Ogni elemento può avere degli attributi, che ne descrivono le proprietà. Si noti che la strutturazione di un documento in questi termini non è immediata: alcuni aspetti possono essere rappresentati sia come elementi sia come attributi e da alcuni tipi di documento non è possibile ricavare una struttura perfettamente ad albero: anche alla presenza di collegamenti fra un elemento dell'albero ed un altro, XML non dà problemi. La struttura fisica di un documento XML è formata da entità: un'entità è un costrutto che rappresenta una parte del testo, ha una dimensione variabile (da un carattere a tutto il documento) ed ha un nome, per fare i riferimenti ad essa (con un funzionamento analogo a quello delle macro). Come gli elementi indicano i componenti della struttura logica del documento, le entità specificano la posizione di insiemi di byte: si deve fare in modo che questi due aspetti siano correlati tra

loro. L'insieme di comandi che stabiliscono questo legame è il 'markup', detto anche 'tag': questi comandi sono distinti dal testo mediante dei delimitatori e in pratica ciò che è racchiuso fra i simboli di minore e maggiore costituisce il markup, che è la parte che è compresa dagli elaboratori. La forma è analoga a quella dei tag HTML, ma, se in questo caso danno le indicazioni per la formattazione delle informazioni, in un documento XML non ci dovrebbero essere istruzioni per la loro visualizzazione. Quello che non è racchiuso fra i tag, è interpretato come testo e non è generalmente analizzato.

Il documento XML dell'esempio precedente può essere il seguente:

```
<?XML version="1.0"?>
<e-message>
<sender e-mail=lezoche@dia.uniroma3.it>
Mario Lezoche
</sender>
<receiver e-mail=atzeni@dia.uniroma3.it>
Paolo Atzeni
</receiver>
<subject>
Semantic Web
</subject>
<body>
<primo-paragrafo>
... testo primo paragrafo
</primo-paragrafo>
<secondo-paragrafo>
... testo secondo paragrafo
```

```
</secondo-paragrafo>
```

```
</body>
```

```
</e-message>
```

Si noti come l'indirizzo di mittente e destinatario siano stati considerati come attributi: non ci sono regole precise per scegliere una soluzione piuttosto che un'altra, ma genericamente si considera come attributo ciò che non contiene degli elementi e dà informazioni semplici e non strutturate. Generalmente, gli attributi sono delle stringhe che indicano una proprietà dell'elemento a cui sono associati. Come già detto, i documenti di un certo tipo hanno la stessa struttura; a questo proposito la tecnologia XML consente di creare la dichiarazione di tipo di documento o *DTD* (Data Type Definition), che rappresenta la definizione formale di ogni tipo (analogamente allo schema, che, nell'ambito delle basi di dati, stabilisce i vincoli per la strutturazione dei dati). Una DTD specifica le caratteristiche di un certo tipo di documento, cioè gli elementi, gli attributi e le entità: un documento che si conforma ad alcuna DTD si dice valido; se non è valido, potrebbe essere ben formato (cioè contenente un markup comprensibile), ma non conforme ad una DTD. Un documento XML deve dichiarare a quale DTD si conforma, per verificarne la validità, ma tale dichiarazione non è obbligatoria e il documento è considerato non valido, non valido.

L'esempio precedente è conforme a una DTD del tipo:

```
<!DOCTYPE e-message [
```

```
<!ELEMENT e-message (sender, receiver, subject, body)>
```

```
<!ELEMENT sender (#PCDATA)>
```

```
<!ATTLIST sender e-mail CDATA #REQUIRED>
```

```
<!ELEMENT receiver (#PCDATA)>
```

```
<!ATTLIST receiver e-mail CDATA #REQUIRED>
```

```
<!ELEMENT subject (#PCDATA)>
<!ELEMENT body (paragrafo+)>
<!ELEMENT paragrafo (#PCDATA)>
]> <e-message>
```

dove PCDATA è un tipo predefinito e indica una sequenza di caratteri parsificata; CDATA è una sequenza di caratteri qualsiasi, non parsificata. Il simbolo ‘+’ indica che l’elemento ‘body’ contiene uno o più sottoelementi ‘paragrafo’. L’elemento ‘e-message’ è costituito da una sequenza di sottoelementi, separati da una virgola: in alternativa alla sequenza si può avere una scelta; ad esempio:

```
<!ELEMENT recapito (e-mail | telefono)>
```

in un documento XML si traduce con

```
<recapito>
<e-mail>
claudio@educ.di.unito.it
</e-mail>
</recapito>
```

oppure

```
<recapito>
<telefono>
999999999
</telefono>
</recapito>
```

Le entità sono usate come abbreviazioni per riutilizzare le informazioni o condividerle fra documenti diversi. Nel primo caso, si definiscono le entità interne, del tipo:

```
<!ENTITY nome "Nome lungo dell'entità">
```

che è utilizzata nel seguente modo:

```
<abbreviazione>
```

```
Questo il &nome che viene abbreviato
```

```
</abbreviazione>
```

e il parsificatore sostituisce la stringa ‘nome’, preceduta dal carattere ‘&’, con la stringa ‘Nome lungo dell’entità’. Nel secondo caso, si definisce un’entità per fare riferimento ad una definizione esterna alla DTD in cui è dichiarata. Si ha:

```
<!ENTITY esterna SYSTEM "nome-del-file.estensione">
```

In questo modo si dichiara che l’entità ‘esterna’ è definita nel file indicato tra apici e individuato dalla parola-chiave SYSTEM; come nel caso precedente, il parsificatore sostituisce ogni occorrenza della stringa ‘&esterna’ con il contenuto del file indicato. XML, da solo, non è in grado di realizzare una delle caratteristiche fondamentali del Web: i collegamenti ipertestuali. Per questo motivo, è affiancato da uno standard, *XLink*, che permette di inserire elementi in un documento in modo da creare e descrivere collegamenti tra risorse: i link che si possono ottenere non sono solo di tipo unidirezionale e uno a uno come nel caso di HTML, ma possono essere bidirezionali, uno a molti, molti a molti. Come detto all’inizio, in un documento XML non devono comparire informazioni o elementi che riguardano il modo in cui i dati vengono visualizzati. Accanto ad XML esiste, quindi, una seconda specifica, XSL (eXtensible Style Language), che permette di creare fogli di stile per specificare come il contenuto di un documento XML debba essere visualizzato sul supporto scelto (browser, stampante, ...). Con questo strumento si realizzano gli scopi primari di XML: mantenere separati gli elementi di contenuto dagli elementi di formattazione.

É chiaro che i contenuti di XML sono molto più ampi: questa breve panoramica permette solo di inquadrare il linguaggio nel contesto del Semantic Web.

4.3 Schematizzazione

4.3.1 RDF

Nel Web esiste un altro tipo di dati: i metadati. Questi, genericamente, sono informazioni riguardanti le informazioni o, più precisamente, sono i dati che descrivono le risorse del Web. Il primo passo per introdurre i metadati in Internet consiste in un'aggiunta standardizzata ad XML, che si chiama RDF (Resource Description Framework): si può vedere questa struttura ad un livello direttamente superiore rispetto ad XML e, se XML è un'estensione del documento, RDF è un'estensione dei dati introdotti da XML. Non è propriamente un linguaggio, ma un modello per la rappresentazione dei metadati. I documenti RDF permettono di stabilire nuove relazioni fra i documenti, immagini e qualunque altro tipo di risorsa Web, consentendo l'interoperabilità tra le applicazioni che si scambiano via Web informazioni a loro comprensibili: uno dei suoi scopi è di rendere possibile la specificazione della semantica dei dati basati su XML, in modo standardizzato. Inoltre, esso permette di definire un meccanismo per descrivere le risorse, indipendentemente dal dominio di applicazione e senza definire tale dominio a priori. RDF descrive non solo ciò che è presente nel Web, come le pagine o parti di esse, ma anche tutto quello a cui può essere associato un URI (Uniform Resource Identifier)¹: ad esempio, se si può associare un URI ad una persona, tutte le informazioni ad essa relative potranno essere rappresentate mediante RDF. RDF ha un sistema a classi, paragonabile a quelli della programmazione ad oggetti: una collezione di classi è detta schema e le

¹Un URI definisce o specifica un'entità, non necessariamente indicando la sua posizione nel Web. Un tipo particolare di URI è l'URL (Uniform Resource Locator), che è la base dei link ipertestuali.

classi formano una gerarchia, che, con il meccanismo delle sottoclassi, consente la specializzazione dei concetti e la condivisione delle definizioni dei metadati. Inoltre, è permessa l'ereditarietà multipla, che consente la vista multipla dei dati e lo sfruttamento di definizioni create da altri.

Il modello dei dati è rappresentato con un grafo, che schematizza la tripla *risorsa-proprietà-valore*: i nodi del grafo sono le risorse descritte da RDF (pagine Web, insieme di pagine, oggetti non direttamente accessibili via Web), gli archi sono etichettati con il nome della proprietà relativa alla risorsa e sono direzionati dal nodo che rappresenta la risorsa verso quello indica il valore (che può essere di tipo atomico oppure un'altra risorsa, ottenendo in questo caso delle relazioni fra le risorse). La risorsa (identificata univocamente da un URI) insieme con una proprietà ad essa associata e al corrispondente valore forma uno *statement*, in cui i tre elementi prendono il nome di soggetto, predicato e oggetto. Ad esempio, nella figura 4.2, la risorsa 'http://www.dia.uniroma3.it/atzeni.htm' ha la proprietà 'creatore' il cui valore è 'Paolo Atzeni' (gli ovali vengono utilizzati per rappresentare le risorse e i rettangoli i valori atomici).



Figura 4.2: Statement RDF

Sintatticamente, i concetti espressi con RDF vengono serializzati mediante XML; per specificare se un elemento appartiene alla sintassi RDF, si utilizza il prefisso 'rdf:'. Ad esempio, lo statement precedente è rappresentato nel seguente modo:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://www.dia.uniroma3.it/~prova/TestRDF/">
```

```
<rdf:Description about="http://www.dia.uniroma3.it/~atzeni/
index.html">
<Creatore>
Paolo Atzeni
</Creatore>
</rdf:Description>
</rdf:RDF>
```

I primi due elementi specificano i namespace, a cui i nomi delle proprietà sono associati, in modo che la loro semantica non sia ambigua: pi precisamente, il primo indica lo spazio dei nomi relativo a RDF, cioè i concetti definiti dalla sintassi di RDF; il secondo è il namespace in cui sono definiti tutti gli altri nomi (in particolare quelli delle proprietà, la cui definizione è indipendente da RDF). Nell'esempio, quindi, lo schema `http://www.dia.uniroma3.it/~prova/TestRDF/` contiene la definizione di 'creatore'. Lo statement può anche essere specificato mettendo in evidenza i suoi tre componenti:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
<rdf:Description>
<rdf:subject resource="http://www.dia.uniroma3.it/~atzeni/
index.html">
<rdf:predicate resource="http://www.dia.uniroma3.it/
~prova/TestRDF/Creatore"/>
<rdf:object>
Paolo Atzeni
</rdf:object>
</rdf:Description>
```



```
</rdf:RDF>
```

Il significato dei termini utilizzati in uno statement RDF è molto importante per stabilire il tipo di elaborazione che si deve effettuare su una particolare informazione: chi scrive e chi legge lo statement devono capire lo stesso concetto per i termini usati. Questo significato comune è espresso mediante uno schema, che definisce i termini che vengono usati e il loro contenuto semantico: lo schema RDF (RDFS) si trova a livello superiore rispetto ad RDF. Lo schema è identificabile con un'ontologia e consente la definizione di classi e di risorse, delle loro proprietà, dei vincoli di dominio e di range sulle proprietà, delle relazioni (comprese quelli di sottoclasse di una risorsa e sottotipo di una proprietà), i vincoli e la documentazione. Ad esempio, in corrispondenza dell'URL <http://www.dia.uniroma3.it/prova/TestRDF/>, può essere contenuto il seguente schema, che definisce la classe delle 'Persone' e la loro proprietà 'Afferenza', impostando anche i namespace:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:xml="http://www.w3.org/xml#">
<rdf:Description ID="ClassOfPerson">
<rdfs:comment>
```

La classe delle persone

```
</rdfs:Comment>
<rdfs:label xml:lang="en">
Person
</rdfs:label>
<rdfs:label xml:lang="it">
```

Persona

```
</rdfs:label>
<rdfs:type resource="http://www.w3.org/2000/01/
rdf-schema#Class"/>
<rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/
rdf-schema#Resource"/>
</rdf:Description>
<rdf:Description ID="PropertyAffiliated">
<rdfs:comment>
```

Proprieta' delle Persone

```
</rdfs:Comment>
<rdfs:type resource="http://www.w3.org/2000/01/
rdf-schema#Property"/>
<rdfs:label xml:lang="en">
```

Affiliated

```
</rdfs:label>
<rdfs:label xml:lang="it">
```

Affiliato

```
</rdfs:label>
<rdfs:domain rdf:resource="ClassOfPerson"/>
<rdfs:range rdf:resource="ClassOfOrganization"/>
</rdf:Description>
</rdf:RDF>
```

Questo schema definisce:

- Una nuova risorsa, il cui identificativo è 'classOfPerson': è definita come 'Class' ed è sottotipo di 'Resource' (questa dipendenza è di default). A questa classe

sono associati un commento e due label (che indicano la forma, secondo la lingua, in cui può comparire in un documento).

- Una proprietà della classe delle persone, il cui identificativo è ‘PropertyAffiliated’: è definita come ‘Property’ e, ad essa sono associati un commento, due label, il dominio (cioè la risorsa a cui fa riferimento, in questo caso ‘ClassOfPerson’) e il range (cioè l’insieme di valori che la proprietà può assumere, in questo caso una risorsa ‘ClassOfOrganization’, che si suppone definita in un altro schema ‘http://www.dia.uniroma3.it/schema1’).

In base a questo schema, un documento può contenere un elemento del tipo:

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://www.dia.uniroma3.it/~prova/TestRDF">
<Person>
<Affiliated>
Dipartimento di Informatica
</Affiliated>
</Person>
</rdf:RDF>
```

o alternativamente

```
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://www.dia.uniroma3.it/~prova/TestRDF">
<Persona>
<Affiliato>
```

Dipartimento di Informatica

</Affiliato>

</Persona>

</rdf:RDF>

La specifica del namespace lega l'uso di un termine in un dominio allo schema (dizionario) in cui si trova la sua definizione: nell'esempio precedente, un motore di ricerca, facendo riferimento allo schema, è in grado di interpretare correttamente i tag in entrambe le versioni, poiché è definito il legame tra i due concetti. RDF, come base per l'elaborazione dei metadati, può essere utilizzato in diverse aree: nella ricerca di risorse, nella catalogazione, nei software di intelligenza artificiale, nel commercio elettronico, in cui è affiancato a tecniche di firma digitale. Sebbene RDF e lo schema rappresentino un buon strumento per la definizione di un linguaggio di markup per il Semantic Web (ad esempio, per determinare le relazioni semantiche tra termini differenti), non hanno sufficiente potere espressivo: non consentono di specificare le proprietà delle proprietà, le condizioni necessarie e sufficienti per l'appartenenza alle classi e gli unici vincoli che si possono definire sono quelli di dominio e range delle proprietà. Il loro utilizzo nei sistemi di rappresentazione della conoscenza è limitato, inoltre, da un'altra caratteristica: non permettono di specificare meccanismi di ragionamento, ma rappresentano semplicemente un sistema a frame. I meccanismi di ragionamento devono essere costruiti ad un livello superiore.

4.3.2 Topic Maps - XTM

Topic Maps è definito dallo standard ISO 13250 e rappresenta un meccanismo per la rappresentazione di metainformazioni, con l'intento di facilitare l'organizzazione, il mantenimento e la navigazione di risorse informative. Un'istanza di documento Topic Maps (d'ora in poi mappa) crea una sovrastruttura semantica alla base

informativa cui si riferisce. Il paradigma Topic Maps può essere utilizzato in più contesti diversi, non necessariamente per l'indicizzazione e la caratterizzazione di risorse web; il web attraverso gli URI offre un meccanismo di indicizzazione che ne facilita l'utilizzo. Permette la creazione di ontologie attraverso meccanismi di tipizzazione gerarchica e relazioni di classe-istanza e superclasse-sottclasse tra concetti descritti.

Nascono con l'obiettivo di creare un meccanismo di indicizzazione che si adatti ad ogni supporto informativo. In genere, gli indici sono conformi alla struttura dell'informazione cui si riferiscono, quindi i loro modelli di meta-informazioni sono impliciti e diversi caso per caso. Creando un modello formale generico per la descrizione di indici si ottiene l'interoperabilità necessaria ad automatizzare la loro gestione e la fusione di documenti di meta-informazioni aventi intersezione di contenuti. Topic Maps definisce una serie di costrutti che creano una classificazione a livello di linguaggio delle meta-informazioni: i concetti che rappresentano i mattoni con cui costruire delle strutture semantiche. Si possono distinguere tre costrutti fondamentali:

- Argomenti (d'ora in poi Topics)
- Occorrenze
- Associazioni

XTM: XML Topic Maps (1.0). Il modello concettuale definito dallo standard ISO 13250 è indipendente da qualsiasi meccanismo di serializzazione. Per favorire l'interscambio di mappe e l'automazione della loro gestione, un consorzio indipendente formato dalle principali aziende sostenitrici della tecnologia Topic Map ha creato XTM: grammatica formale per la rappresentazione di mappe, basata su XML e su XLink come meccanismo di referenziazione. Esistono diversi formati di rappresen-

tazione delle mappe. XTM, oltre ad essere il più diffuso, è recentemente diventato parte dello standard ISO.

I Topics sono il concetto fondamentale dello standard. Qualsiasi cosa che debba essere descritta viene fatta corrispondere ad un topic, cui eventualmente si associano ulteriori caratteristiche. Topic è un concetto simile a risorsa in RDF. Tuttavia un topic può anche essere completamente scorrelato da qualsiasi riferimento a risorse esterne, e descrivere quindi un concetto astratto. Le caratteristiche che possono essergli associate sono:

- Un soggetto (subject identity).
- Uno o più nomi (base-name), di cui uno indicabile come preferito nella visualizzazione .
- Uno o più tipi (relativamente ad una ontologia prestabilita).
- una o più occorrenze.

I Topic Names sono l'associazione ai topic di uno o più nomi per la visualizzazione. Il doppio livello di tag è dovuto alla possibilità di contestualizzare il nome, ossia associare degli attributi di validità o di significatività alla stringa che segue.

```
<topic id="mario_rossi">
<baseName>
<baseNameString>Rossi, Mario</baseNameString>
</baseName>
<baseName>
<baseNameString>Mario Rossi</baseNameString>
</baseName>
</topic>
```

La tipizzazione dei topic non è imposta, è il meccanismo essenziale mediante il quale utilizzare una ontologia preesistente. Il tipo di un topic è un altro topic o un insieme di altri topic: è prevista la tipizzazione multipla. Non esiste una separazione a livello di modello tra la parte dichiarativa e quella descrittiva di una mappa: non esistono vincoli che impongono ad un topic il ruolo di tipo o di istanza. In assenza di un linguaggio di constraint che ponga dei vincoli, sta al creatore della mappa mantenere la coerenza interna. Mediante il tag `instanceOfi` di un `topici` è possibile specificare una lista di topic-types, che lo tipizzano.

Tipizzazione: Esempi

```
<topic id="tt-person">
  <baseName>
    <baseNameString>Person</baseNameString>
  </baseName>
</topic>
<topic id="tt-artist">
  <instanceOf>
    <topicRef xlink:href="#tt-person"/>
  </instanceOf>
  <baseName>
    <baseNameString>Artist</baseNameString>
  </baseName>
</topic>
<topic id="tt-musician">
  <instanceOf>
    <topicRef xlink:href="#tt-artist"/>
  </instanceOf>
```

```
<baseName>  
<baseNameString>Musician</baseNameString>  
</baseName>  
</topic>
```

Il Soggetto e le Occorrenze sono i costrutti mediante i quali si creano correlazioni tra un topic di una mappa e le reali risorse informative. Il soggetto indica ciò che un topic rappresenta. Può essere qualsiasi risorsa indirizzabile, in particolare qualsiasi URI oppure un'altro topic, nel qual caso si attua una reificazione. Le occorrenze indicano una o più risorse indirizzabili (o definite online nella mappa stessa) nelle quali il concetto che il topic rappresenta 'ricorre' in una qualsiasi accezione del termine ricorrere. Distinguono i ruoli delle risorse referenziate all'interno di una mappa: una cosa è associare delle meta-informazioni ad una risorsa, un'altra è descrivere un concetto e associarlo ad una risorsa in cui tale concetto viene analizzato, espresso o semplicemente citato. Un'occorrenza è una sottostruttura di un topic, intrinsecamente legata ad esso, caratterizzata da un ruolo (occurrence role type). Un esempio di Soggetti e Occorrenze: una home page viene modellata con un topic che ha per soggetto il suo URL, e viene ulteriormente caratterizzata da un'occorrenza sulla pagina in cui è linkata, opportunamente tipizzata in modo da indicare il ruolo dell'occorrenza.

```
<topic id="pagina_di_link"/>  
<topic id="la_mia_pagina_web">  
<subjectIdentity>  
<resourceRef xlink:href="http://www.dia.uniroma3.it/~atzeni"/>  
</subjectIdentity>  
<occurrence>  
<instanceOf>
```



```
<topicRef xlink:href="#pagina_di_link"/>
</instanceOf>
<resourceRef link:href="http://cs.unibo.it/students.html"/>
</occurrence>
</topic>
```

Le Associazioni sono il secondo costrutto cardine delle mappe. Un documento XTM è in effetti una collezione di elementi topic e association. Servono (non sorprendentemente) a realizzare una relazione tra due o più topic, ma anche tra topic e risorse. Si noti che già le occorrenze ed il soggetto realizzano un particolare tipo di associazione tra un topic e una risorsa: associazioni esplicite sono necessarie per tutti gli altri casi.

Esempio:

```
<association>
<member>
<topicRef xlink:href="#mario_rossi"/>
</member>
<member>
<resourceRef xlink:href="http://www.mariorossi.com/index.html"/>
</member>
</association>
```

Anche le associazioni sono soggette a tipizzazione per evidenziare il tipo di relazione che sussiste fra i membri. Le associazioni in TM sono inerentemente bidirezionali: $A \tilde{B} = B \tilde{A}$

Ciascun membro `memberi` può essere caratterizzato mediante un attributo che ne determini il ‘ruolo’ nella relazione, mediante il tag `roleSpeci`. Si può aumentare il livello di dettaglio dell’esempio precedente:

```
<association>
<instanceOf> <topicRef xlink:href="#creatore"/> </instanceOf>
<member>
<roleSpec> <topicRef xlink:href="#autore"/> </roleSpec>
<topicRef xlink:href="#mario_rossi"/>
</member>
<member>
<roleSpec> <topicRef xlink:href="#pagina_web"/> </roleSpec>
<resourceRef xlink:href="http://www.mariorossi.com/index.html"/>
</member>
</association>
```

Un unico costrutto member può contenere più ‘players’, ossia più riferimenti a risorse o ad altri topic che stanno nella stessa relazione e nello stesso ruolo. Esempio: Se Mario Rossi è il creatore di altre pagine web:

```
<association>
<instanceOf> <topicRef xlink:href="#creatore"/> </instanceOf>
<member>
<roleSpec> <topicRef xlink:href="#autore"/> </roleSpec>
<topicRef xlink:href="#mario_rossi"/>
</member>
<member>
<roleSpec> <topicRef xlink:href="#pagina_web"/> </roleSpec>
<resourceRef xlink:href="http://www.mariorossi.com/index.html"/>
<resourceRef xlink:href="http://www.myhsot.org/mrossi.html"/>
</member>
</association>
```

Nella definizione di ontologie, per evitare problemi di ambiguità e di sinonimia dei concetti descritti è utile ricorrere, quando possibile, a soggetti definiti pubblicamente (Published Subject Indicators). I Published Subject Indicators (PSI) sono topic che rappresentano concetti basilari, utili nella caratterizzazione di argomenti. Sono definiti su namespace noti e permettono di:

- facilitare l'interscambio di mappe e la loro fusione;
- fornire un meccanismo di base per la differenziazione tra parte dichiarativa e parte descrittiva delle mappe.

Ne sono un esempio quelli definiti all'url: <http://www.topicmaps.org/xtm/1.0/core.xtm>. Sono essenzialmente dei topic con ruolo di tipizzazione. Il modello Topic Maps prevede la possibilità di contestualizzare le associazioni, le occorrenze ed i nomi assegnabili ai topic col meccanismo degli scope. Gli scope descrivono un contesto nel quale le caratteristiche descritte hanno significato: un'applicazione di navigazione potrà considerare tale informazione per filtrare le meta-informazioni rilevanti in ogni momento. L'assenza di uno scope indica validità generale. Ad esempio, nel seguente topic il basename 'Rossi, Mario' deve essere usato per l'applicazione definita con topic 'sorting', ovvero per l'ordinamento dei nomi.

```
<topic id="mario-rossi">
  <baseName>
    <scope>
      <topicRef xlink:href="#sorting"/>
    </scope>
    <baseNameString>Rossi, Mario</baseNameString>
  </baseName>
</topic>
```

Un topic usato per definire uno scope viene detto tema (theme). In generale uno scope può essere costituito da un numero arbitrario di temi.

Esempio:

```
<association>
<instanceOf> <topicRef xlink:href="#creatore"/> </instanceOf>
<scope>
<topicRef xlink:href="#web_authoring"/>
<topicRef xlink:href="#completed_works"/>
</scope>
<member> <topicRef xlink:href="#mario_rossi"/> </member>
<member>
<resourceRef xink:href="http://www.mariorossi.com/index.html"/>
</member>
</association>
```

4.3.3 Differenze con RDF

Topic Maps è semanticamente più ricco, tanto che un vero e proprio confronto non può essere fatto senza considerare i livelli superiori di RDF (RDF Schema, modelli di II ordine come DAML+OIL, etc.).

Similarità:

- Permettono di creare strutture di metadati complesse
- Permettono la definizione di ontologie
- Possono essere utilizzati come 'building block' per il Semantic Web
- Si prestano entrambi ad una rappresentazione a grafo.

Differenze:

- Topic Maps consente delle relazioni n-arie multidirezionali invece di richiedere la creazione di $n(n-1)$ predicati monodirezionali di RDF.
- TM fornisce regole formali per il merging di mappe, ovvero la creazione di mappe unione di due o più mappe più semplici.
- Non esiste ancora uno standard per definire schemi di mappe (TMCL)

4.4 Oltre lo schema

Il livello superiore è quello che fornisce i costrutti per la creazione di ontologie, il cui potere espressivo è necessario per descrivere le risorse in modo sufficientemente dettagliato. Un'ontologia è una collezione di asserzioni, che definiscono le relazioni tra i concetti e specificano le regole logiche per ragionare su di essi; gli agenti software comprendono il significato di un dato riferendosi all'ontologia indicata. Proseguendo sulla strada degli standard, il livello superiore è rappresentato dal linguaggio DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer), il cui inserimento nel contesto del Semantic Web porta ad avere la struttura rappresentata in figura 4.3.

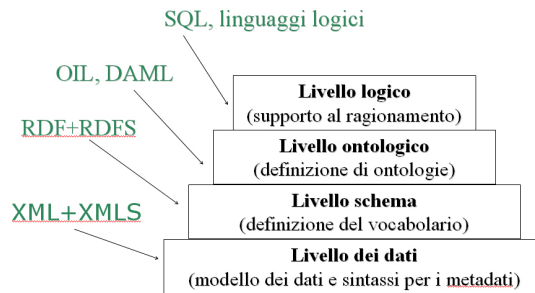


Figura 4.3: Oltre lo schema

4.4.1 Il linguaggio DAML+OIL e OWL

È un linguaggio standard che consente la rappresentazione delle informazioni del Web in modo che il loro significato sia comprensibile alle macchine. Originariamente erano due linguaggi separati: DAML permetteva di inserire il contenuto semantico dei dati, basandosi sulle ontologie definite con RDFS (e favorendo la nascita di un nuovo linguaggio per la definizione di ontologie, DAML-ONT, che estendeva RDF con costrutti propri dei linguaggi object-oriented e di rappresentazione della conoscenza basati su frame); OIL è nato come un altro linguaggio per la creazione di ontologie, basato anch'esso su RDFS. Ad un certo punto, è stato evidente che i due prodotti potevano essere uniti e il risultato è un linguaggio ontologico, che consente di descrivere la struttura di un dominio. Per realizzare ciò per cui è stato progettato, si serve del frutto di altre discipline; in particolare si basa su:

- Linguaggi basati su frame: modellano la realtà in classi (o *textitframe*), ognuna delle quali ha delle proprietà (o *textitslot*); le classi sono legate tra di loro da relazioni di sottoclasse o superclasse. DAML+OIL si basa su questi concetti, consentendo la definizione di classi, delle loro sottoclassi e delle loro proprietà;
- Description Logic: è un linguaggio, elaborato nella ricerca sulle rappresentazioni della conoscenza, che descrive la conoscenza in termini di concetti (paragonabili ai frame) e ruoli (paragonabili agli slot); è importante il fatto che le espressioni in DL possono essere descritte in forma matematica, in modo che si possano fare ragionamenti, basati sulla descrizione dei concetti, e classificazioni automatiche. DAML+OIL sfrutta proprio questo supporto per i ragionamenti;
- XML e RDF: la sintassi di DAML+OIL è quella di XML ed RDF; inoltre, lo schema RDF fornisce un insieme di primitive (come la relazione di 'sottoclas-

se') e le regole sintattiche per definire le gerarchie (che non sono previste in XML).

Il linguaggio OWL deriva direttamente da DAML+OIL come si può vedere dalla Figura 4.4.

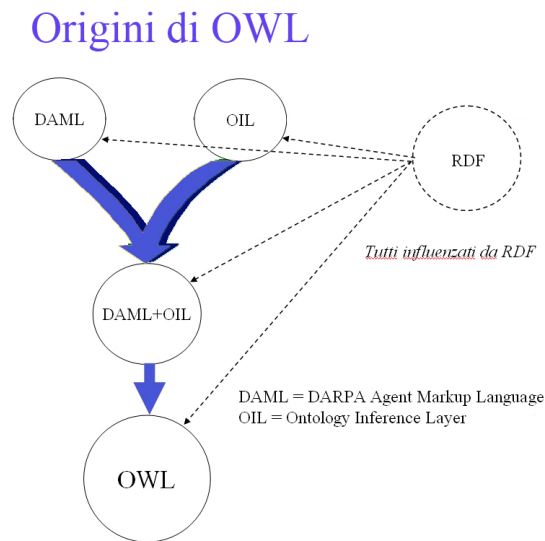


Figura 4.4: Origini di OWL

OWL usa `rdfs:range`, `rdfs:domain`, e `rdfs:subPropertyOf` ed introduce due classi di Property: `owl:ObjectProperty` che è la classe di proprietà che hanno come valore un oggetto istanza di una classe (`Resource2Resource`) e `owl:DatatypeProperty` che è la classe di proprietà che hanno come valore un dato di tipo semplice o strutturato (`rdfs:Literal` o XML Schema built-in datatype). Un esempio di uso di OWL si può avere sulla tassonomia dei corsi d'acqua presentata in Figura 4.5. Le classi di OWL sono molto più espressive delle classi di RDF Schema classes. Un esempio è mostrato nella Figura 4.6.

L'ontologia universitaria utilizzata come riferimento per il progetto MOSES è un'ontologia DAML+OIL definita a: <http://www.cs.umd.edu/projects/plus/DAML/onts/univ1.0.da>. Questa ontologia importa tre ontologie più generali:

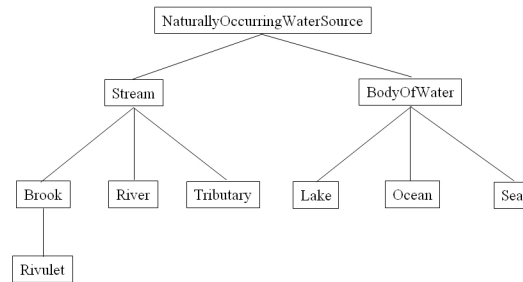


Figura 4.5: Tassonomia corsi d'acqua

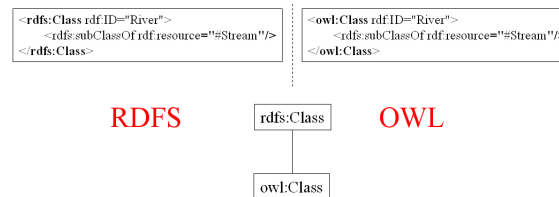


Figura 4.6: Confronto generazione classi RDF ed OWL

- <http://www.cs.umd.edu/projects/plus/DAML/onts/docmnt1.0.daml>
- <http://www.cs.umd.edu/projects/plus/DAML/onts/general1.0.daml>
- <http://www.cs.umd.edu/projects/plus/DAML/onts/base1.0.daml>

Siccome i riferimenti alle ontologie DAML contengono delle caratteristiche base del linguaggio DAML, la loro trasformazione in OWL è stata abbastanza semplice. L'unico problema è stata la mancanza di distinzione tra DatatypeProperty ed ObjectProperty nei file originali. Si è assunto che le proprietà DAML senza una specifica fossero implicitamente DatatypeProperty, ma senza specifica del Range. Per esempio in <http://www.cs.umd.edu/projects/plus/DAML/onts/docmnt1.0.daml>

```

<Property ID="publishDate">
  <label>was published on</label>
  <domain resource="#Document"/>
</Property>
    
```


Questa proprietà è stata tradotta in http://www.mondeca.com/owl/document1_1.xml

```
<owl:DatatypeProperty rdf:ID="publishDate">
  <rdfs:label xml:lang="en">was published on</rdfs:label>
  <rdfs:domain rdf:resource="#Document"/>
  <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#date"/>
</owl:DatatypeProperty>
```

Questo è un caso dove il Datatype è stato interpretato in maniera da seguire il senso comune: `publishDate` è ovviamente una data... Ma in molti casi, nessun aiuto è dato dall'ontologia DAML originale circa il tipo di dato che si ci deve aspettare ed il Range è stato impostato al generico valore stringa. D'altro canto, la trasformazione di Classi e Proprietà di un oggetto è stato semplicemente una questione di cambiare i tag DAML negli equivalenti tag OWL più discorsivi. Esempio nelle due sintassi, DAML prima e OWL successivamente:

```
<Property ID="teachingAssistantOf">
  <label>is a teaching assistant for</label>
  <domain resource="#TeachingAssistant" />
  <range resource="#Course" />
</Property>
```

```
<owl:ObjectProperty rdf:ID="teachingAssistantOf">
  <rdfs:label xml:lang="en">is a teaching assistant for</rdfs:label>
  <rdfs:domain rdf:resource="#TeachingAssistant"/>
  <rdfs:range rdf:resource="#Course"/>
</owl:ObjectProperty>
```

Capitolo 5

MOSES: Un progetto di ricerca

5.1 Introduzione

L'opinione espressa all'interno del progetto *MOSES* (MOdular and Scalable Environment for the Semantic web) è che l'unico modo per far affermare e divenire un successo il Semantic Web è procedere con un approccio '*bottom up*' che gli permetta di emergere dall'aggregazione di frammenti organizzati di conoscenza di varie dimensioni.

Per dimostrare questo approccio il progetto mira alla creazione di un piccolo, ma scalabile, sistema di gestione della conoscenza e di un motore di ricerca basato sulle ontologie che accetterà interrogazioni e produrrà risposte in linguaggio naturale.

L'obiettivo principale dell'ambiente scalabile sarà quello di dimostrare che è possibile migliorare i sistemi ontologici attraverso nuovi contenuti, sia creando nuovi domini di conoscenza sia incorporando nuova conoscenza in domini preesistenti. Far ciò richiede l'estrazione di conoscenza strutturata dal semplice testo.

5.2 Descrizione del progetto

Il lavoro di MOSES inizia dal fissare il dominio del dimostratore basandosi sulle necessità degli utenti del progetto. Successivamente si disegna la griglia d'architettura.

tura sia per quanto riguarda l'aspetto computazionale che di conoscenza.

L'implementazione del primo dimostratore è basata su tre 'Work Packages' (macro suddivisioni tematiche del progetto) ognuno dei quali relativi ad una specifica tecnologia. Ogni work package è sotto la diretta responsabilità di uno dei partner, colui che possiede maggiori requisiti di conoscenza in quel campo. Questi work package saranno accoppiati con il WP8 (Work Package 8) che ha l'obiettivo principale di amalgamare le tecnologie che vengono utilizzate in modo da evitare incompatibilità tecnologiche. Il lavoro si focalizza prima sulla struttura della conoscenza e sul motore di ricerca ontologico. Questo è diviso in due parti quella relativa all'immissione delle richieste e alle risposte.

Il lavoro sul dimostratore dinamico si sviluppa dopo il completamento della struttura di conoscenza. Il WP8 è ugualmente posto come coordinatore tecnologico.

5.2.1 Punti cruciali

Si può dividere il progetto MOSES in quattro punti cruciali ('milestone') che caratterizzano dei significativi punti d'arrivo:

1. *Campo d'azione del dimostratore:* il primo punto cruciale si raggiunge nel momento in cui si fissano i domini di contenuto e conoscenza del dimostratore.
2. *Struttura di conoscenza:* il secondo quando si ottiene una struttura di conoscenza capace di descrivere contenuti e semantica degli esempi identificati nei web site che sono rappresentativi dei domini di conoscenza scelti.
3. *Dimostratore statico:* il motore di ricerca ontologico è capace di valutare domande e generare risposte sulla struttura di conoscenza su cui si basa.
4. *Dimostratore dinamico:* in questo ultimo punto cruciale la Struttura di Conoscenza è capace di generare nuova conoscenza da nuovi contenuti e il motore

di ricerca ontologico è capace di estrarre informazioni significative da questa struttura di conoscenza arricchita.

5.2.2 Lista dei partecipanti

Una lista dei partecipanti al progetto MOSES si può vedere nella Tabella 5.1.

Nome Partecipante	Nome corto	Nazione
FINSA Consulting Società a Responsabilità Limitata - Futuro ed Innovazione - Strategie Aziendali Ricerca e Consulenza	FINSA	Italia
Mondeca	Mondeca	Francia
Center for Sprogteknologi	CTS	Danimarca
Koebenhavns Universitet - Facoltà Umanistica	UKBH	Danimarca
Università Degli Studi Roma TRE - Dipartimento di Informatica e Automazione	DIA	Italia
Università Degli Studi Roma 'Tor Vergata' - Dipartimento di Informatica, Sistemi e Produzione	RTV	Italia
ParaBotS	ParaBotS	Olanda

Tabella 5.1: Lista dei soggetti partecipanti al progetto europeo MOSES

5.3 Innovazione

Avvantaggiandosi di molte tecnologie differenti MOSES propone un approccio al web semantico innovativo per molti aspetti. Uno dei più importanti è la possibilità di aumentare le risorse semantiche prese in considerazione, il suo uso intelligente della Griglia di Conoscenza e lo sviluppo di informazioni attraverso contenuti web in varie lingue.

5.3.1 Approccio di dimensionamento dinamico al Web Semantico

Solo un approccio dinamico rispetto la dimensione dei dati alla fine potrà:

- Identificare e gestire il gran numero, necessario, di oggetti relativi alla conoscenza che formano un'adeguata ontologia e le relazioni complesse tra i singoli termini dell'ontologia stessa.
- Gestire i contenuti sconosciuti (non ancora categorizzati ed annotati), e farlo per diverse lingue.
- Processare un gran numero di informazioni abbastanza velocemente da essere di uso pratico stando sempre adeguatamente attenti agli aspetti di sicurezza.

La soluzione proposta in MOSES è un ambiente capace di adattarsi alle reali necessità di un utente anche in termini di performance. L'ambiente è disegnato in modo da poter:

- Agire come un nodo cooperativo all'interno di una Griglia di Conoscenza. Ciò significa la possibilità di poter scambiare conoscenza con sistemi simili creati appositamente per particolari situazioni e quindi aventi conoscenza differente.
- Evolversi con l'uso, aumentare la conoscenza in accordo con le necessità che si vengono a creare con l'uso reale.

Entrambi questi aspetti sono innestati nell'approccio di dimensionamento dinamico del web semantico. Il primo combina, infatti, la modularità e il dimensionamento nella cooperazione tra i nodi della griglia di conoscenza che costituiscono il modo più importante di sviluppare la conoscenza. La soluzione adottata in MOSES è quella di un *sistema ontologico di dimensione variabile* in ogni nodo della Griglia

di conoscenza che mimerà lo scenario del web semantico e sull'abilità della Griglia di interconnettere, in maniera dinamica ed efficiente, i nodi di conoscenza.

5.3.2 Sistema ontologico di dimensione variabile

Nei vari tentativi effettuati dai ricercatori e dai gruppi industriali di definire ontologie che potessero gestire terminologia e relazioni complesse tra i termini si può vedere una dicotomia tra coloro che tentano di risolvere il problema di un dominio di conoscenza specifico e altri che cercano di creare delle ontologie generali.

I due approcci hanno differenti svantaggi. Le ontologie generiche corrono il rischio di essere troppo grandi per essere gestite in maniera efficiente. Infatti, sebbene sistemi di gestione per ontologie e basi di conoscenza sono stati già sviluppati (ad esempio protégé¹), rimane comunque non risolto il problema di come costruire e gestire un ambiente che renda possibile la creazione di un gran numero di ontologie formali mantenendo i tempi di analisi al minimo.

In generale le ontologie per uno specifico dominio sono usate per applicazioni pratiche. Queste sono adeguate per risolvere il problema per cui sono nate ma difficilmente possono essere utilizzate in altri contesti o unificarle con altre ontologie. Infatti spesso può verificarsi il caso in cui due ontologie si sovrappongono mentre la loro concettualizzazione no (ad esempio due ontologie utilizzano lo stesso termine con un differente significato).

Questo significa che un approccio dal basso verso l'alto 'bottom-up' all'integrazione dei sistemi basati sulla semplice integrazione di varie ontologie locali potrebbe non funzionare. Questo è specialmente vero se le ontologie locali sono focalizzate solo sulle relazioni concettuali rilevanti ad uno specifico contesto senza relazionarsi ad un'ontologia più generale. MOSES non si propone di risolvere il problema definendo

¹protégé è un editor di ontologie e di basi di conoscenza. Informazioni più precise si possono trovare al seguente URL: <http://protege.stanford.edu/>

una nuova ontologia universale, ma provando ad adottare una nuova prospettiva dalla quale lavorare con il problema della mappatura tra le ontologie esistenti².

Un esempio di conoscenza a dimensione variabile per un nodo della griglia di conoscenza, in accordo con le necessità e la conoscenza disponibile, potrebbe essere poter rispondere ad una domanda in modi progressivamente sempre meno specifici. Ad esempio alla domanda: *'Chi è Paolo Atzeni?'* si potrebbe rispondere (i termini ontologici sono sottolineati):

1. *É un professore dell'Università di Roma TRE dove insegna i corsi di Basi di Dati.*
2. *É un professore dell'Università di Roma TRE che si occupa di Informatica.*
3. *É un professore dell'Università di Roma TRE a cui interessano le scienze.*
4. *Insegna all'Università di Roma TRE.*
5. *Lavora all'Università di Roma TRE.*
6. *É un uomo e ha un lavoro.*

Questo comportamento significa la possibilità di generalizzare un contenuto facendo riferimento a termini più generali in una data ontologia o ad una differente e meno specifica ontologia. Per arrivare a questo obiettivo bisogna assicurarsi che un *applicativo ontologico* disegnato per un particolare dominio di conoscenza è connesso con ontologie che definiscono diversi livelli di generalizzazione. Alla fine di questa scala c'è un'ontologia che descrive in maniera molto generale i concetti che, per definizione, sono indipendenti da qualunque particolare problema o dominio ma

²Per coprire il dominio che si andrà a considerare nel progetto si è fatto uso di ontologie esistenti, ove possibile, definendone di nuove dove necessario

ancora capaci di garantire un minimo livello di condivisione di conoscenza. La combinazione di conoscenza specifica e generale si è rivelata utile in altri progetti (come ad esempio OntoQuery³). Recentemente sono stati sviluppati sistemi di gestione per ontologie e basi di conoscenza (come protégé) che permettono una progettazione *'object oriented'* di un sistema ontologico il che è molto utile per il proposito di combinare ontologie di gradi di specializzazione differenti.

Ma il solo riferimento ad una ontologia comune di livello superiore non garantisce la possibilità di integrare differenti ontologie. Il problema viene affrontato, in MOSES, attraverso altre soluzioni innovative: queste sono l'integrazione (o mapping) di ontologie, l'uso di agenti d'apprendimento, i sistemi intelligenti di gestione di contenuto e l'apprendimento da nuovi contenuti.

5.3.3 Integrazione di Ontologie

Una *'Mapping Ontology'* è un'ontologia disegnata per descrivere le relazioni tra i concetti di due diverse ontologie piuttosto che tra i concetti di un particolare dominio.

In realtà la problematica di integrare ontologie ha una sua propria semantica ed è per questo che si parla di *'Mapping Ontologies'* invece che di algoritmi o tool di integrazione. Così il problema di integrazione non è equivalente alla *fusione di ontologie*: gli elementi nell'ontologia traduttrice devono contenere delle formule che servono alla traduzione tra concetti nelle ontologie tradotte. Come abbiamo visto vi sono vari linguaggi formali che specificano questo tipo di ontologie e le loro formule. Questo progetto ha utilizzato **RDF-schemata** in quanto compatibile con XML.

³Per avere più informazioni sul progetto OntoQuery: www.ontoquery.dk

5.3.4 Agenti d'Apprendimento

Nello scenario MOSES, ogni agente conosce la propria ontologia e tutti vogliono condividere la conoscenza integrando le potenzialità delle altre ontologie, approfittando del fatto che sono tutte basate su di una ontologia comune. In verità, più è generale l'ontologia comune, più grande sarà il numero dei sistemi che potranno sfruttarla per le proprie necessità. Per ogni paio di agenti che vogliono condividere la conoscenza il sistema genera un'*intersezione* contenente solo l'informazione necessaria a comporre la conoscenza dalle due sorgenti e rispondere alle domande poste dalla particolare applicazione.

5.3.5 Apprendimento da nuovi contenuti

Un altro modo in cui il dimensionamento dinamico è utilizzato è quello di permettere agli agenti di imparare non solo termini e relazioni tra di loro ma anche da nuovi contenuti web. Si sono utilizzati i metodi correntemente adottati nei sistemi di classificazione testuale arricchiti con tecniche NLP. Il proposito di questa sezione del progetto è quello di provvedere ad una base per la classificazione semantica e l'annotazione di nuovi testi. Ciò che viene estratto dalle pagine web viene confrontato con parole chiavi di differenti nodi in un'ontologia rilevante.

5.4 Motore di ricerca Semantico

I motori di ricerca tradizionali dipendono, più o meno esclusivamente, dal riconoscimento di parole chiave o da campioni (pattern) di parole chiave nel materiale testuale. Comunque l'uso di dizionari dei sinonimi e ontologie, per formare la spina dorsale semantica che guida il processo di ricerca dei termini, sta riscuotendo sempre più attenzione sia nella comunità di ricerca che in quella industriale. Tra i progetti di ricerca che lavorano in questa direzione si può menzionare OntoSeek[2] ed

OntoQuery[3], entrambi gli strumenti di ricerca incorporano al loro interno un'ontologia. In OntoSeek l'ontologia è utilizzata per aiutare interattivamente gli utenti nella costruzione di precise descrizioni di risorse testuali e nella formulazione di query non ambigue, che possono essere successivamente generalizzate o specializzate. In OntoQuery l'ontologia è utilizzata nell'analisi sintattica e concettuale delle query e dei testi ed anche nella guida dell'algoritmo che confronto: il motore di ricerca ricava del testo il cui contenuto è lo stesso o più specifico di quello che è stato richiesto dall'utente. Per esempio, un testo che ha a che fare con 'Gandalf' sarà trovato interessante ad una query relativa a 'maghi'.

Le interrogazioni basate sul contenuto in MOSES sono simili all'approccio trovato in OntoSeek nel quale le ontologie sono utilizzate per strutturare i contenuti web prima dell'interrogazione. Comunque MOSES è anche ispirato da OntoQuery nell'utilizzo di tecniche NLP usate nel processare le query e i testi. Un'altra importante similarità con l'approccio di OntoQuery è il fatto che il confronto del contenuto in MOSES è mediato dalla conoscenza concettuale.

5.4.1 Aspetti di multilinguismo

Il progetto si rivolge verso la tematica del multilinguismo in vari modi. Il primo di tutti è che i due insiemi di contenuti web utilizzati sono scritti in due lingue differenti (Italiano e Danese) e descrivono due differenti realtà culturali, sebbene il dominio coperto può essere visto come il medesimo.

Invece di provare a risolvere il problema attraverso l'imposizione di contenuti utilizzando una strutturazione del dominio indipendente dal linguaggio, come primo passo si compara la struttura del contenuto esistente nei linguaggi presi in considerazione in modo da esaminare in che modo divergono. In questo modo l'integrazione tra frammenti di ontologia multilingua è il più grande contributo del progetto al

problema del multilinguismo.

5.5 Piano di lavoro

Il progetto MOSES è suddiviso in varie sottosezioni chiamate work package. Ogni work package è inerente ad un aspetto differente del progetto. Gli output dei work package iniziali sono utilizzati come input di quelli successivi. Nella Tabella 5.2 vi è la suddivisione del progetto MOSES con i relativi nomi dei singoli work package.

Work Package	Titolo Work Package
WP1	Coordinazione Progetto
WP2	Analisi dei requisiti e del Dominio
WP3	Progettazione Architettura
WP4	Gestione della conoscenza e software semantico
WP5	Implementazione del sistema multi agente
WP6	Query basate sulle ontologie
WP7	Incremento della conoscenza
WP8	Integrazione e dimostrazione
WP9	Valutazione e Validazione
WP10	Diffusione e Utilizzo

Tabella 5.2: Suddivisione in sottosezioni (work package) del progetto europeo MOSES

5.6 Sviluppo di una linea di lavoro

I work package approfonditi relativamente all'argomento della presente tesi sono stati il WP2, il WP4 e il WP7.

5.6.1 WP2 - Analisi dei requisiti e del Dominio

Il maggior obiettivo di questo work package è stato quello di definire l'ambito di utilizzo del dimostratore. È stato trovato un equilibrio tra il significato rispetto i

contenuti trattati, e le aspettative degli utenti, verso la fattibilità rispetto ai problemi trattati, considerato lo stato dell'arte delle tecnologie e le risorse disponibili.

5.6.2 WP4 - Gestione della conoscenza e software semantico

Il work package 4 ha implementato il sistema di gestione della conoscenza richiesto in accordo ai requisiti derivanti dal WP2. L'obiettivo è stato quello di far evolvere un sistema di gestione dei contenuti in un sistema di gestione della conoscenza basato sulle ontologie.

5.6.3 WP7 - Incremento della conoscenza

Questo work package ha implementato la tecnologia che era necessaria affinché la conoscenza potesse crescere. Prendendo come punto di partenza la struttura di conoscenza creata sulla base di un sottoinsieme dei contenuti web universitari, ha indagato il modo in cui i nuovi contenuti (nuovi testi), come le nuove strutture di conoscenza (concetti e relazioni) possano essere aggiunti alla struttura di conoscenza e ha approfondito le tematiche delle annotazioni semantiche.

Capitolo 6

MOSES: Analisi dei requisiti

6.1 Analisi dei requisiti

La prima operazione che si è effettuata è la descrizione dei contenuti del sito web dell'Università di Roma TRE, questo è stato il primo passo verso la definizione del raggio d'azione dei dimostratori del progetto MOSES. Data la complessità della struttura del sito web, che ha un componente centrale e molti siti satelliti eterogenei (ed autonomi), l'analisi è proceduta fino ad un livello di raffinamento che ha permesso l'identificazione di *'domini di applicazione'* che diverranno il reale raggio d'azione del dimostratore del progetto.

La descrizione del sito web centrale sarà dettagliata con una discussione della struttura generale ed una indicazione delle principali categorie di utenti (con associate le necessità ed aspettative). Per i siti satelliti vi sarà una descrizione generale con qualche esempio, una discussione dei requisiti per alcuni domini di applicazioni potenziali che possono essere di interesse per il dimostratore. La discussione dei requisiti è l'aspetto più interessante, si riferisce ad un primo tentativo di mettere insieme le aspettative degli utenti con le sorgenti d'informazione.

Questi requisiti definiranno i reali contenuti di interesse per il dimostratore, il dominio semantico e le ontologie.

6.2 Sito Centrale

La missione di un sito di Ateneo è quella di fornire informazioni e servizi ai propri utenti con due obiettivi principali:

- Supporto nei confronti dei cosiddetti utenti interni, ovvero studenti immatricolati e personale interno (docente e non).
- Promozione della propria immagine nei confronti degli utenti esterni, ovvero studenti non ancora immatricolati e visitatori più o meno occasionali quali Aziende, Istituzioni Pubbliche e Media.

6.2.1 Organizzazione generale

Per quanto riguarda l'impostazione dell'organizzazione generale (di seguito IA: Information Architecture) qui documentata è necessario fare alcune considerazioni preliminari:

- L'impostazione data alla nuova IA del sito privilegia in particolare la categoria degli studenti (attuali e potenziali), rivolgendo principalmente ad essi i propri servizi e le proprie informazioni. Offrendo quindi, per esempio, in modo ampio e articolato informazioni relative a Didattica, Facoltà, Corsi di Studio, Insegnamenti ecc... Ma anche alla dislocazione delle varie strutture quali Facoltà stesse, Biblioteche, Centri di ateneo, Dipartimenti e Segreteria studenti. Non vengono trascurate, però, le altre categorie di utenti, fornendo un insieme di informazioni e servizi di ampio interesse, con la presenza quindi di aree e sezioni non direttamente rivolte agli studenti ma ad una platea più vasta.
- La IA pensata e qui documentata prevede, come detto, una offerta di informazioni globale e centralizzata che copre l'intera struttura dell'Ateneo lasciando

spazio, per approfondimenti ed informazioni più dettagliate, alla presenza di siti satellite gestiti autonomamente. Viene quindi a crearsi una struttura centralizzata con l'aggiunta di una sorta di decentramento governato costituita dai vari siti satellite, ospitati e/o supportati centralmente.

La struttura organizzativa del sito di Ateneo è prevalentemente di tipo gerarchico, pur non mancando componenti ipertestuali che permettano di 'saltare' da una sezione all'altra e componenti di tipo entità relazionali che evidenziano le differenze tra pagine 'singole' e pagine ad 'istanza multipla'. Vengono quindi analizzate a partire dalla 'root' le varie sezioni della struttura gerarchica scendendo lungo di essa, con una definizione più o meno dettagliata delle sezioni stesse e delle pagine in funzione della centralità della gestione delle stesse. Il formalismo utilizzato per la descrizione grafica della struttura organizzativa è abbastanza intuitivo. E' comunque utile consultare la figura 6.1 per comprendere meglio i vari elementi del formalismo utilizzato.

HOME PAGE

URL : <http://www.uniroma3.it>

Tecnologia utilizzata: Active Server Page.

Questa in figura 6.2 è la pagina principale del sito centrale dell'università. È il punto di partenza di navigazione e ricerca. Da questa pagina è possibile cercare velocemente tutti i documenti e le sezioni. Ci sono anche connessioni con i vari siti satelliti. Sono presentate anche le informazioni importanti in modo da catturare l'attenzione dell'utente.

Le seguenti sono le cinque sezioni principali del sito:

- **ATENEO**: contiene informazioni di carattere generale relative alla struttura di governo e amministrativa dell'Ateneo, ai regolamenti, alle strutture didattiche e di ricerca, ai servizi e agli enti (associazioni, comitati, ecc..) di rilievo.

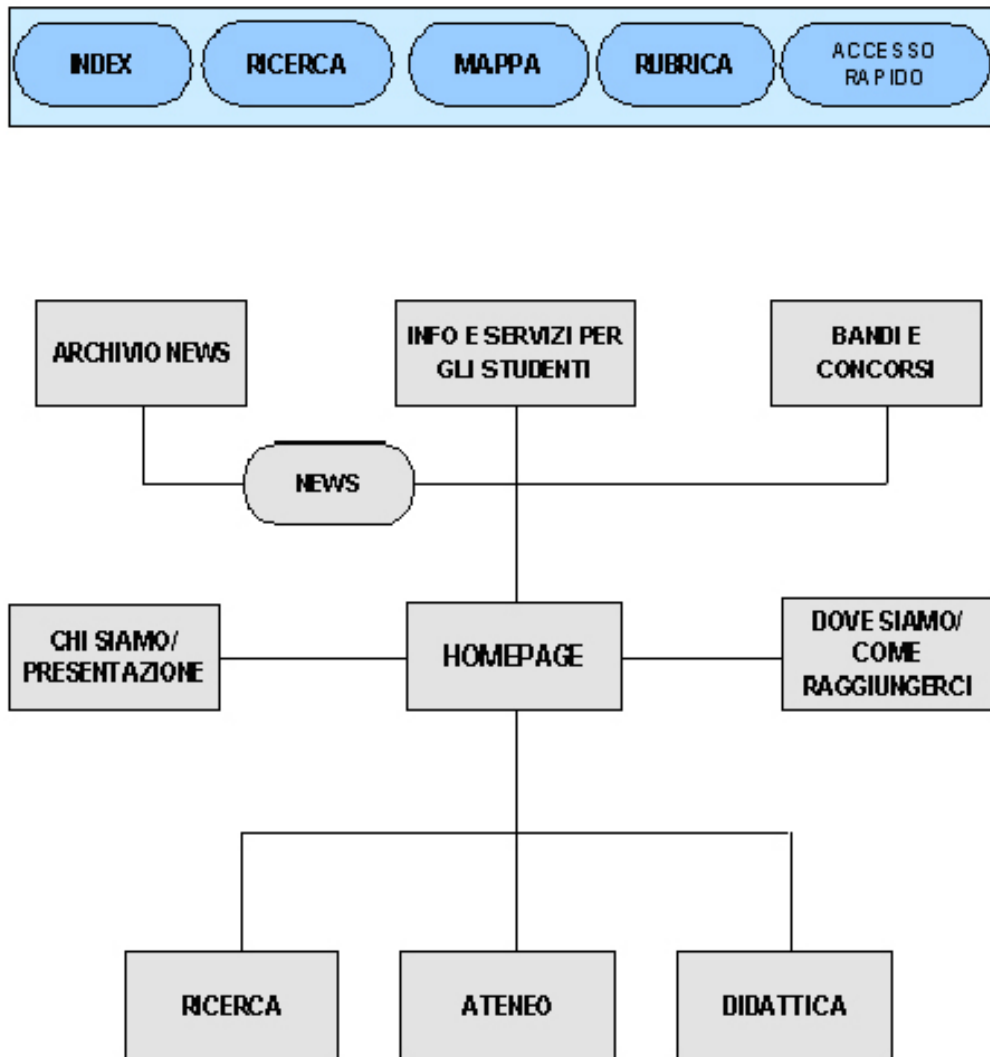


Figura 6.1: Organizzazione del sito centrale dell'Università Roma TRE



Figura 6.2: Home Page del sito centrale dell'Università Roma TRE

- **DIDATTICA:** In questa sezione si vuole offrire la più ampia offerta possibile di informazioni relativa all'area didattica dell'Ateneo. Rispetto alla situazione attuale, in cui compaiono la lista delle facoltà, la lista delle tipologie di corsi di studio e il collegamento al Centro di Ateneo, l'offerta informativa risulta più articolata e completa. Si rivolge prevalentemente agli studenti, anche post-lauream, fornendo informazioni relative soprattutto all'offerta didattica dell'Ateneo.
- **RICERCA:** offre informazioni relative in particolare alle attività e ai progetti di ricerca che coinvolgono RomaTre, e alle strutture di Ateneo (Dipartimenti, Centri, ecc..) collegabili a queste attività.
- **INFORMAZIONE E SERVIZI PER GLI STUDENTI:** si rivolge ancora una volta agli studenti, fornendo informazioni a trecentosessanta gradi su iniziative,

problematiche, strutture, servizi ecc... Che possano essere di interesse per lo studente. In questa sezione, ma è prassi comune nell'architettura proposta, vengono fornite in forma ridotta informazioni anche ridondanti quali quelle relative alla didattica o alla ricerca, per esempio.

- BANDI E CONCORSI: Questa sezione è stata pensata per ospitare tutto ciò che riguardi qualsiasi tipo di bando e/o concorso che coinvolga direttamente o indirettamente, attraverso i propri dipendenti e *utenti*, l'Ateneo.
- CHI SIAMO: contiene alcune informazioni di presentazione dell'Ateneo. Si tratta di una ridondanza degli elementi di presentazione che si troveranno anche all'interno della suddetta sezione ATENEO.
- DOVE SIAMO: informazioni relative alla posizione geografica e alla raggiungibilità delle strutture di Ateneo.
- ARCHIVIO NEWS: rappresenta un archivio delle cosiddette news, le più *fresh* delle quali (quattro o cinque) saranno presenti, in forma ridotta, direttamente nella Homepage.

ATENEO

URL : <http://www.uniroma3.it/ateneo/indexateneo.asp>

Technology used : Active Server Page

Come visualizzato nella figura 6.3, questa sezione contiene informazioni generali relative alla struttura di governo ed amministrativa dell'Ateneo.

La sezione ATENEO contiene quattro sottosezioni:

- PRESENTAZIONE: Questa sezione presenta l'università attraverso vari documenti. A sua volta è composta da sottosezioni:

The screenshot shows the website for the University of Rome TRE, specifically the 'ATENEO' section. The header includes the university logo, name, and website URL (www.uniroma3.it). A navigation bar contains links for 'Cerca', 'NEWS', 'MAPPA', 'RUBRICA', and 'ACCESSO RAPIDO'. The main content is organized into several sections:

- CHI SIAMO / DOVE SIAMO**: A vertical menu on the left with links to 'CHI SIAMO' and 'DOVE SIAMO'.
- PRESENTAZIONE**: A central section with a list of links including 'Ateneo Roma Tre', 'Interventi del Rettore', 'Personaggi celebri incontrano Roma Tre', 'Pubblicazioni', 'Rassegna stampa di Ateneo', and 'Rappresentanti studenti'.
- ORGANI CENTRALI DI ATENEO**: A section listing various university bodies such as 'Rettore', 'Senato Accademico', 'Consiglio d'Amministrazione', 'Prorettore', 'Direttore Amministrativo', 'Collegio dei Direttori di Dipartimento', 'Consiglio degli Studenti', and 'Nucleo di Valutazione'.
- IN PRIMO PIANO**: A section with links to 'Riforma didattica' and 'Inaugurazione A.A.'.
- SEMPRE UTILI**: A section with links to 'Facoltà', 'Dipartimenti', 'Segreterie Studenti', 'Biblioteche', 'Attività internazionali', 'Centri di Ateneo', 'URP', 'Servizi Informatici', 'ADISU', and 'Host'.
- STATUTO E REGOLAMENTI**: A section with links to 'Statuto', 'Reg. generale di Ateneo', 'Reg. didattico di Ateneo', 'Reg. amministrativo contabile', 'Reg. reclutamento docenti', 'Reg. Consiglio Studenti', and 'Tutti i regolamenti'.
- BILANCI**: A section with links to various financial reports, including 'Bilancio 2002/04 Relazione politica', 'Relazione tecnica al Bilancio di Previsione per l'Esercizio Finanziario 2002', 'Situazione finanziaria presunta al 31.12.2001', 'Relazione Collegio dei Revisori dei Conti al Bilancio di Previsione 2002', 'Bilancio di Previsione Anno 2002', 'Bilancio Triennale Previsione Anno 2002/04', and 'Anni precedenti'.

Figura 6.3: Sezione ATENEO del sito centrale dell'Università Roma TRE

- Ateneo Roma Tre: Contiene documenti inerenti la storia e l'evoluzione di ROMA TRE.
- Pubblicazioni: Contiene tutte le guide alle varie facoltà e informazioni utili per comprenderne lo sviluppo.
- Rassegna stampa di Ateneo: permette l'accesso ad articoli di giornali di interesse per l'università.
- Rappresentanti studenti: lista di tutti i rappresentanti degli studenti.

C'è anche una lista di link agli uffici satelliti che informano gli utenti delle attività universitarie.

- BILANCI: In questa sezione ci sono link a documenti in formato PDF di tutti i bilanci universitari.
- ORGANI CENTRALI DI ATENEO: In questa sezione vi sono i link alle liste dei principali organi di governo universitario. Le liste sono semplici pagine html.
- STATUTO E REGOLAMENTI: Qui sono presentati lo statuto dell'università e tutte le regole che governano l'ateneo.

DIDATTICA

URL : <http://www.uniroma3.it/ateneo/indexateneo.asp>

Technology used : Active Server Page

Questa sezione cerca di offrire il maggior numero di informazioni sulla didattica universitaria. C'è una lista delle facoltà, una lista degli insegnamenti, servizi per gli studenti.

La pagina della sezione didattica, come si vede dalla figura 6.4, include nove sottosezioni:

Università degli Studi ROMA TRE www.uniroma3.it

Q Cerca | NEWS | MAPPA | RUBRICA | ACCESSO RAPIDO

CHI SIAMO
DOVE SIAMO

ATENEEO
DIDATTICA E STUDENTI
RICERCA SCIENTIFICA
INFORMAZIONI E SERVIZI
BANDI E CONCORSI

IN PRIMO PIANO

- Iscrizioni on line
- Scadenze Erasmus
- Riforma didattica

SEMPRE UTILI

- Facoltà
- Dipartimenti
- Segreterie Studenti
- Biblioteche
- Attività internazionali
- Centri di Ateneo
- URP
- Servizi Informatici
- ADISU
- Host

DIDATTICA E STUDENTI

FACOLTA'	CORSI DI STUDIO
<ul style="list-style-type: none"> · Architettura · Economia "Federico Caffè" · Giurisprudenza · Ingegneria · Lettere e Filosofia · Scienze della Formazione · Scienze Matematiche, Fisiche e Naturali · Scienze Politiche 	<p>NUOVO ORDINAMENTO</p> <ul style="list-style-type: none"> · Corsi di Laurea · Corsi di Laurea Specialistica · Master · Dottorati · Scuole di Specializzazione · Corsi di Perfezionamento <p>VECCHIO ORDINAMENTO</p> <ul style="list-style-type: none"> · Corsi di Laurea · Corsi di Diploma Universitario · Prenotazione esami ed esoneri
SUPPORTO ALLA DIDATTICA	INSEGNAMENTI
<ul style="list-style-type: none"> · Dipartimenti · Biblioteche · Centro Linguistico di Ateneo (CLA) 	<ul style="list-style-type: none"> · Elenco completo
ISCRIVERSI ALL'ATENEO	SERVIZI PER GLI STUDENTI
<ul style="list-style-type: none"> · Riforma didattica · Segreterie studenti <p>Iscrizioni</p> <ul style="list-style-type: none"> · Immatricolazioni a.a. 2002/2003 · Lauree e Diplomi · Master · Scuola di Spec. all'Insegnamento Secondario · Scuola di Spec. per le Professioni Legali · Corsi di Perfezionamento · Corsi Singoli <hr/> <ul style="list-style-type: none"> · Ammissione studenti con titolo estero · Pagamento delle tasse on-line 	<ul style="list-style-type: none"> · Centro Accoglienza · Ascolto psicologico · Accoglienza disabili · Servizi ADISU · Servizi Informatici · Licenze Microsoft Campus
GUIDE	AREA STUDENTI
<ul style="list-style-type: none"> · Guida dello Studente · Guide brevi di Facoltà · Ordini degli Studi · GiGa(Guida a iscrizione e autocertificazione) · Guida ai corsi di studio 	<ul style="list-style-type: none"> · Risultati elezioni marzo 2002 · Consiglio degli Studenti · LUG - Pinguini in ascolto · Archi 1999 · UnioneStudenti.it
SERVIZIO DI LEVA	
<ul style="list-style-type: none"> · Rinvio del servizio militare · Servizio Civile a Roma Tre 	

Figura 6.4: Sezione DIDATTICA del sito centrale dell'Università Roma TRE

- **FACOLTÁ:** contiene dei link a delle schede con le informazioni più importanti sulla facoltà. Ci sono l'indirizzo, il telefono, il fax, la posta elettronica, il sito web (se esistente), la presentazione della facoltà, l'offerta didattica, informazioni sugli uffici amministrativi, come immatricolarsi.
- **CORSI DI STUDIO:** Qui si trovano le informazioni sui singoli corsi di studio (vecchio e nuovo ordinamento), Master e corsi di Dottorato. Tutti i link aprono una pagina dove sono presentati i vari oggetti. C'è anche un link per la prenotazione degli esami online.
- **SUPPORTO ALLA DIDATTICA:** In questa sottosezione si possono trovare degli aiuti agli studenti, un link al centro linguistico d'ateneo, una lista con tutti i dipartimenti dell'università. C'è anche un link alle biblioteche di Roma TRE che sono organizzate in un sistema di biblioteche strutturato con una rete composta da singoli nodi interconnessi.
- **INSEGNAMENTI:** Ci sono link a tutti i corsi offerti dall'università organizzati per facoltà. Ogni link presenta una breve descrizione del corso.
- **SERVIZI PER GLI STUDENTI:** Qui si possono trovare una serie di servizi per studenti. Si ci trova il centro di benvenuto per nuovi studenti, per gli studenti Erasmus ecc... C'è un centro per l'aiuto psicologico, un centro per i portatori di disabilità, servizi informatici e molto altro. Tutti i servizi sono creati per aiutare gli studenti a vivere bene gli anni trascorsi all'università.
- **ISCRIVERSI ALL'ATENEO:** Sono descritte, in questa sezione, le informazioni inerenti alla riforma didattica e come immatricolarsi ai vari tipi di Laurea, Diplomi, Master e Dottorati. Ci sono informazioni per chi possiede una laurea straniera. Molto utile sono le istruzioni di come pagare le tasse via internet.

- AREA STUDENTI: Iniziative per studenti come il LUG, Linux University Group (Gruppo Universitario per Linux), o i risultati delle elezioni dei rappresentanti degli studenti.
- GUIDE: Si trovano le guide su vari argomenti. Ad esempio la *Guida degli studenti* che aiuta gli studenti del primo anno a capire la vita universitaria.
- SERVIZIO DI LEVA: Istruzioni e regole su come posporre la chiamata alle armi per motivi di studio.

RICERCA

URL : <http://www.uniroma3.it/ricerca/indexricerca.asp>

Tecnologia utilizzata : Active Server Page

Questa sezione offre informazioni relative alle attività di ricerca che fanno capo all'università di Roma TRE e alle strutture universitarie come dipartimenti, centri ecc... che sono coinvolti in queste attività.

Come si vede dalla figura 6.5 la sezione Ricerca include quattro sottosezioni:

- DIPARTIMENTI: Lista esaustiva di tutti i dipartimenti di Roma TRE. Da questa lista è possibile raggiungere le home page dei dipartimenti che offrono informazioni interessanti sui modi per contattarli, sugli edifici, sugli orari di apertura e così via.
- CENTRI DI ATENEO: I centri sono unità organizzative che hanno degli specifici obiettivi, e quindi sono distinti dai dipartimenti. Il modo in cui sono presentati è simile a quello in cui si presentano i dipartimenti.
- BIBLIOTECHE: Lista completa di tutte le biblioteche di Roma TRE. Questa lista permette all'utente di navigare il sito e reperire le informazioni offerte

The screenshot shows the 'RICERCA SCIENTIFICA' section of the University of Rome TRE website. The header includes the university logo, name, and website URL (www.uniroma3.it). A navigation bar contains 'Cerca', 'NEWS', 'MAPPA', 'RUBRICA', and 'ACCESSO RAPIDO'. The main content is organized into several sections:

- Left Navigation Menu:**
 - CHI SIAMO
 - DOVE SIAMO
 - ATENEEO
 - DIDATTICA E STUDENTI
 - RICERCA SCIENTIFICA
 - INFORMAZIONI E SERVIZI
 - BANDI E CONCORSI
 - IN PRIMO PIANO**
 - Cofinanziamento MIUR
 - Assegni di ricerca
 - SEMPRE UTILI**
 - Facoltà
 - Dipartimenti
 - Segreterie Studenti
 - Biblioteche
 - Attività internazionali
 - Centri di Ateneo
 - URP
 - Servizi Informatici
 - ADISU
 - Host
- Central Column: DIPARTIMENTI**
 - Biologia
 - Comunicazione Letteraria e Spettacolo
 - Economia
 - Elettronica Applicata
 - Filosofia
 - Fisica
 - Informatica e Automazione
 - Ingegneria Elettronica
 - Ingegneria Meccanica e Industriale
 - Istituzioni politiche e Scienze sociali
 - Italianistica
 - Letterature Comparete
 - Linguistica
 - Matematica
 - Progettazione e Scienze dell'Architettura
 - Progettazione e Studio dell'Architettura e Studi Urbani
 - Ricerche Aziendali
 - Scienze dell'Educazione
 - Scienze dell'Ingegneria Civile
 - Scienze Geologiche
 - Studi Americani
 - Studi Giuridici
 - Studi Storici Geografici Antropologici
 - Studi Storico-Artistici, Archeologici e sulla Conservazione
 - Studi sul Mondo Antico
- Right Column: CENTRI DI ATENEEO**
 - Centro Linguistico di Ateneo (CLA)
 - Centro di Studi Italo-Francesi
 - Centro per lo studio di Roma (CROMA)
- Right Column: BIBLIOTECHE**
 - Presentazione Sistema Bibliotecario di Ateneo
 - Area delle Arti
 - Area Giuridico-Economico-Politica
 - Area Scientifico-Tecnologica
 - Area Umanistica
 - Biblioteca Centro Studi Italo-Francesi
 - Catalogo Informatizzato
- Right Column: ALTRE RISORSE**
 - Dottorati
 - Banca Dati Trasferimento Tecnologico
 - Cofinanziamento MIUR
 - Assistenza per progetti comunitari
 - Relazione sulla Ricerca a Roma Tre (dalla Giornata della Ricerca, 31/10/2001)

Figura 6.5: Sezione RICERCA del sito centrale dell'Università Roma TRE

dal sistema di biblioteche. Presenta inoltre un catalogo informatico per la consultazione online.

- ALTRE RISORSE: Informazioni varie.

INFORMAZIONI E SERVIZI

URL : <http://www.uniroma3.it/infoservizi/indexinfoservizi.asp>

Tecnologia usata: Active Server Page

Questa sezione si rivolge, ancora una volta, agli studenti, dando tutte le informazioni circa i problemi, le strutture, i servizi ecc... Come è stato già notato vi sono varie sovrapposizioni con le altre sezioni ma l'idea è quella di presentare gli stessi argomenti ma dall'ottica degli studenti e non dell'organizzazione.

- ISCRIZIONI: Questa parte della pagina contiene informazioni utili sulla procedura di registrazione dello studente. Ogni link contenuto in questa zona spiega un tipo differente di registrazione ad esempio, registrazione universitaria, registrazioni degli esami, e così via.
- GUIDE: In questa sezione sono riunite delle guide dell'università. Queste guide sono scritte per aiutare gli studenti ad orientarsi nella vita universitaria. È possibile consultare le guide sull'università, sulla procedura di registrazione e così via.
- STRUTTURE DIDATTICHE E DI RICERCA: Breve lista di tutte le unità organizzative di Roma TRE, Facoltà, Dipartimenti, Centri, Biblioteche.
- BORSE FINANZIAMENTI E STAGES: Con i link contenuti in questa parte della pagina gli studenti possono accedere alle informazioni inerenti i concorsi banditi da Roma TRE. Ci sono anche link per il programma Erasmus e altri.

The image shows a screenshot of the University of Rome TRE website. At the top, there is a navigation bar with the university logo, the name 'Università degli Studi ROMA TRE', and the website address 'www.uniroma3.it'. Below this is a search bar and several menu items: 'NEWS', 'MAPPA', 'RUBRICA', and 'ACCESSO RAPIDO'. The main content area is titled 'INFORMAZIONI E SERVIZI' and is organized into several columns and sections. On the left, there is a vertical menu with categories like 'CHI SIAMO', 'DOVE SIAMO', 'ATENE0', 'DIDATTICA E STUDENTI', 'RICERCA SCIENTIFICA', 'INFORMAZIONI E SERVIZI', 'BANDI E CONCORSI', 'IN PRIMO PIANO', and 'SEMPRE UTILI'. The main content area is divided into sections: 'ISCRIZIONI' (with sub-sections like 'Segreterie studenti', 'Iscriversi all'Ateneo', 'Lauree e Diplomi', etc.), 'GUIDE' (with 'Guida dello Studente', 'Guide brevi di Facoltà', etc.), 'SERVIZI' (with 'Centro Accoglienza', 'URP - Ufficio Relazioni con il Pubblico', etc.), 'STRUTTURE DIDATTICHE E DI RICERCA' (with 'Facoltà', 'Dipartimenti', etc.), 'SERVIZIO DI LEVA' (with 'Rinvio del servizio militare', etc.), 'AREA STUDENTI' (with 'Rappresentanti Studenti', etc.), 'BORSE, FINANZIAMENTI & STAGES' (with 'Borse di collaborazione', etc.), and 'STUDENTI CON TITOLO ESTERO' (with 'Certificato IT', etc.).

Figura 6.6: Sezione INFORMAZIONE E SERVIZI del sito centrale dell'Università Roma TRE

- SERVIZIO DI LEVA: Ci sono due sezioni principali raggiungibili da questa parte della pagina. La prima permette agli studenti di posporre il servizio militare fino alla fine degli studi o fino ad una particolare età. Con il secondo link gli studenti possono sapere come lavorare con Roma TRE invece di espletare il servizio militare.
- STUDENTI CON TITOLO ESTERO: Questi due link sono dedicati agli studenti stranieri. Il secondo, in particolare, permette agli studenti stranieri di accedere all'ufficio relazioni internazionali con molte informazioni utili per il proprio soggiorno in Italia.
- AREA STUDENTI: Questa parte contiene link interessanti sui gruppi politici studenteschi a Roma TRE. Ci sono informazioni sulle persone coinvolte in queste attività, il Concilio degli studenti e così via. C'è anche il link al LUG (Il Gruppo Universitario per Linux)
- ENTI E INIZIATIVE: In questa sezione vi sono informazioni sulle iniziative di Roma TRE.

ARCHIVIO NEWS

URL : <http://www.uniroma3.it/news/archivionews.asp>

Tecnologia usata: Active Server Page

Questa sezione contiene, come si può vedere dalla figura 6.7, tutte le notizie importanti e gli archivi di tutte le notizie mai presentate.

6.3 Siti Satelliti

La scelta dei Siti satelliti è stata fatta sulla base dei seguenti criteri:

- Quantità di informazione

Università degli Studi ROMA TRE www.uniroma3.it

Cerca NEWS MAPPA RUBRICA ACCESSO RAPIDO

CHI SIAMO DOVE SIAMO

ATENEEO
DIDATTICA E STUDENTI
RICERCA SCIENTIFICA
INFORMAZIONI E SERVIZI
BANDI E CONCORSI

IN PRIMO PIANO

- Riforma didattica
- Inaugurazione A.A.

SEMPRE UTILI

- Facoltà
- Dipartimenti
- Segreterie Studenti
- Biblioteche
- Attività internazionali
- Centri di Ateneo
- URP
- Servizi Informatici
- ADISU
- Host

DETTAGLI NEWS

Turismo: opportunità di carriera per laureati

3 dicembre 2002 – ore 12.00 Aula 23, Facoltà di Lettere e Filosofia - via Ostiense 234

Il seminario metterà in luce il settore turistico ruotando su tre temi principali:

- Chi sono le aziende che operano nel settore e quale è il loro ruolo.
- Quali sono i profili professionali di maggior interesse per le aziende, ora e nel prossimo futuro.
- Come compilare un curriculum e come proporlo alle aziende. Opportunità di stage in aziende del settore.

Il seminario è rivolto agli studenti di tutte le facoltà.

Le date successive degli incontri sono :

- 06 dicembre 2002, 12.00-14.00
- 10 dicembre 2002, 12.00-14.00
- 13 dicembre 2002, 12.00-14.00 - Aula 23
- 17 dicembre 2002, 12.00-14.00
- 20 dicembre 2002 : 12.00-14.00

Per informazioni
Prof.ssa Marinella Rocca Longo
Corso di laurea in lingue e culture straniere

Data Evento: **12/3/02**
Tipo Evento: **Seminario**
Data Emissione: 11/29/02

ARCHIVIO NEWS

Figura 6.7: Sezione NEWS del sito centrale dell'Università Roma TRE

- Aggiornamento dell'informazione esistente
- Importanza dell'informazione

Le seguenti tabelle presentano, tematicamente, i siti che sono stati vagliati con il relativo risultato. Ad ogni singolo sito è stato dato un codice che ne rappresenta l'interesse per il progetto. I valori possibili per tale codice sono:

- 0: Il sito web è in costruzione o non disponibile
- 1: Sito interessante per il progetto MOSES
- 2: Sito non interessante per il progetto MOSES

Un esempio, per la sola categoria dei dipartimenti, si può vedere nelle Tabelle 6.1 e 6.2.

La stessa procedura di controllo è stata effettuata per le altre categorie di siti. Si sono vagliate le seguenti categorie:

- Facoltà: 8 siti vagliati
- Dipartimenti: 29 siti vagliati
- Uffici: 14 siti vagliati
- Eventi: 3 siti vagliati
- Laboratori: 8 siti vagliati
- Master: 6 siti vagliati
- Progetti: 15 siti vagliati
- Magazine: 3 Siti vagliati

Dipartimenti	URL	Codice
Biologia	http://www.bio.uniroma3.it/biologia/	1
Comunicazione Letteraria e Spettacolo	http://luna.uniroma3.it/presen_spett.htm	1
Economia	Non disponibile	0
Filosofia	http://host.uniroma3.it/dipartimenti/filosofia/	1
Fisica	http://www.fis.uniroma3.it/	1
Informatica e Automazione	http://www.dia.uniroma3.it/	1
Ingegneria Elettronica	http://www.ele.uniroma3.it/	1
Ingegneria Meccanica e Industriale	http://www.dimi.uniroma3.it/	1
Istituzioni politiche e Scienze sociali	http://host.uniroma3.it/dipartimenti/scienzepolitiche/	1
Italianistica	http://host.uniroma3.it/dipartimenti/italianistica	1
Letterature Compare	http://host.uniroma3.it/dipartimenti/lettcomp/	1
Linguistica	http://w3.uniroma3.it/didattica/dipartimenti/ling/linguistica/	1
Matematica	http://www.mat.uniroma3.it/	1
Progettazione e Scienze dell'Architettura	http://host.uniroma3.it/dipartimenti/dipsa-cos/	1
Scienze dell'Educazione	http://host.uniroma3.it/dipartimenti/educazione/	1

Tabella 6.1: Siti Satelliti vagliati. Categoria Dipartimenti

Dipartimenti	URL	Codice
Scienze dell'Ingegneria Civile	http://w3.uniroma3.it/didattica/dipartimenti/scieing/	1
Scienze Geologiche	Non disponibile	0
Studi Americani	http://host.uniroma3.it/dipartimenti/studiamericani/	1
Studi Giuridici	Non disponibile	0
Studi Storici Geografici Antropologici	http://host.uniroma3.it/dipartimenti/stusto/index.htm	1
Studi Storico-Artistici, Archeologici e sulla Conservazione	Non disponibile	0
Studi sul Mondo Antico	Non Disponibile	0
CdS Lingue e culture straniere	Non Disponibile	0
CdS in Meccanica	Non Disponibile	0
CdL Scienze Storiche del Territorio - Fac Lettere	http://host.uniroma3.it/cds/ScienzeStoriche/	1
CdL Storia e Conservazione patrimonio Artistico - Fac Lettere	http://host.uniroma3.it/cds/scpa/	1
Dipartimento Ingegneria Meccanica e Industriale	Non Disponibile	0
Dipartimento di Geologia	Non Disponibile	0
Dipartimento Studi Giuridici	Non Disponibile	0

Tabella 6.2: Siti Satelliti vagliati. Categoria Dipartimenti

Alla fine dell'analisi si sono scelti i 25 siti riportati nelle Tabelle 6.3 e 6.4 di cui si è data una descrizione particolareggiata. Si sono poi confrontati i siti selezionati, all'interno di ogni categoria, utilizzando dei criteri di aggiornamento ed affidabilità dell'informazione specifici per le singole categorie. Un esempio è riportato nella Tabella 6.5 e nella relativa Tabella 6.6 che ne è la leggenda.

Siti	URL
Facoltà Architettura	http://w3.uniroma3.it/didattica/facolta/archit/
Facoltà Giurisprudenza	http://host.uniroma3.it/facolta/giur/
Facoltà Scienze Matematiche, Fisiche e Naturali	http://www.smfn.uniroma3.it
Facoltà Scienze Politiche	http://host.uniroma3.it/facolta/politiche/
Dipartimento Biologia	http://www.bio.uniroma3.it/biologia/
Dipartimento Filosofia	http://host.uniroma3.it/dipartimenti/filosofia/
Dipartimento Fisica	http://www.fis.uniroma3.it/
Dipartimento Informatica e Automazione	http://www.dia.uniroma3.it/
Dipartimento Ingegneria Elettronica	http://www.ele.uniroma3.it/
Dipartimento Ingegneria Meccanica e Industriale	http://www.dimi.uniroma3.it/
Dipartimento Istituzioni politiche e Scienze sociali	http://host.uniroma3.it/dipartimenti/scienzepolitiche/
Dipartimento Italianistica	http://host.uniroma3.it/dipartimenti/italianistica

Tabella 6.3: Siti Satelliti scelti

6.4 Requisiti

Si sono stabilite delle *classi di domande*, presentate di seguito con alcuni esempi, per ogni argomento che può essere utilizzato nell'applicazione. Ogni domanda ha i seguenti attributi:

- Sotto applicazione

Siti	URL
Dipartimento Letterature Comparate	http://host.uniroma3.it /dipartimenti/lettcomp/
Dipartimento Linguistica	http://w3.uniroma3.it /didattica/dipartimenti/ling/linguistica/
Dipartimento Matematica	http://www.mat.uniroma3.it/
Dipartimento Progettazione e Scienze dell'Architettura	http://host.uniroma3.it /dipartimenti/dipsa-cos/
Dipartimento Scienze dell'Educazione	http://host.uniroma3.it /dipartimenti/educazione/
Dipartimento Scienze dell'Ingegneria Civile	http://w3.uniroma3.it /didattica/dipartimenti/scieing/
Dipartimento CdL Scienze Storiche del Territorio - Fac Lettere	http://host.uniroma3.it /cds/ScienzeStoriche/
Dipartimento CdL Storia e Conservazione patrimonio Artistico	http://host.uniroma3.it /cds/scpa/
Ufficio Relazioni Internazionali	http://europa.uniroma3.it /relint_nuovo/sito/index_new.asp
Ufficio Centro Linguistico d'Ateneo	http://www.cla.uniroma3.it/
Ufficio servizi generali - Obiettori di Coscienza	http://host.uniroma3.it /uffici/servizigenerali/
Ufficio Relazioni con il Pubblico	http://host.uniroma3.it /uffici/urp/homeurp.htm
Tutorato per studenti con disabilità	http://host.uniroma3.it /progetti/handicap/

Tabella 6.4: Siti Satelliti scelti

Nome	Legge	Architettura	Scienze politiche
Prenotazione esami	1	2	2
Presentazione	5	5	2
Lista degli insegnanti	4d	5d	5d
Lista dei dipartimenti	5	0	0
Informazioni logistiche	5	4	5
Indirizzi email dello staff docente	2	2	2c
Lista dei dottorati	4	0	0
Informazioni Amministrative	4	4	4
Pagamento delle tasse online	5	0	0
Materiale Didattico	4c	0	0
Tabella degli insegnamenti	5	0	5
Congressi	4c	5	0
Ordine degli Studi	5	5	5
Lista delle Tesi	2d	5d	0
Informazione per il test d'ammissione	5	0	0
Programmi di scambio internazionale	5	4	4
Informazioni sui Master	5	5	5
Lista degli studenti	2c	0	0
Date degli esami	5b	4a	0
Ricevimento dei professori	5	0	0
Seminari	2d	0	0
Riviste Online	0	0	3
Corsi di Perfezionamento Post-Lauream	0	5	5
Attività Culturali	0	4	0
News	5	4	5
Motore di ricerca interno	0	5	0

Tabella 6.5: Confronti tra siti satelliti scelti all'interno di una singola categoria

Valore	Stato d'aggiornamento
0	Non Presente
1	Presente ma non funzionante
2	Presente ma non verificato
3	Presente ed aggiornato negli ultimi 12 mesi
4	Presente ed aggiornato negli ultimi 6 mesi
5	Presente ed aggiornato negli ultimi 1 mesi
a	Servizio offerto in maniera ottimale - > 90%
b	Servizio offerto in maniera sufficiente - >40% AND <90%
c	Servizio offerto in maniera insufficiente - <40%
d	Non verificabile

Tabella 6.6: Leggenda del confronto tra siti satelliti della medesima categoria

- Utenti coinvolti
- Formato della domanda
- Formato della risposta
- Aspetti multi linguistici

Questa lista è un possibile dominio di applicazione per l'utilizzo con MOSES. Si è definita una categorizzazione gerarchica dell'informazione disponibile nel sito dell'università di Roma TRE e le possibili domande sulle foglie di questa struttura. Il modello utilizzato per la descrizione dei requisiti è il seguente:

1. Dominio

(a) Classi di domanda

i. Applicazione

A. Sotto applicazione

- Utenti coinvolti

- Formato della domanda
- Formato della risposta

Sono stati riconosciuti quattro domini di interesse:

- Argomento inerente alle persone che lavorano all'interno della struttura universitaria.
- Argomento inerente i corsi che vengono offerti.
- Argomento inerente le ricerche svolte.
- Notizie offerte dai vari siti universitari.

Un duplice esempio del modello utilizzato è offerto di seguito. Il primo è relativo all'argomento Persone e il secondo all'argomento Corsi:

1. Persone

(a) Corpo docente

i. Accessibilità

A. A che ora riceve il professor X?

- Questa informazione è di interesse maggiormente per gli studenti, ma probabilmente è una notizia utile anche per coloro che vogliono incontrare il professore e in questo modo saranno sicuri di trovarlo nello studio nella data e nell'orario trovato.
- Parole Chiavi: 'riceve, professor X'.
- Una pagina contenente una lista di possibili orari e luoghi.

B. Qual'è il sito web del professor X?

- Tutti gli utenti sono coinvolti

- Parole Chiavi: 'website, URL, professor X'.
- Una pagina contenente un link al sito del professore.

1. Corsi

(a) Didattica

i. Corso

A. Quali corsi sono offerti dal dipartimento X?

- Questa informazione è di interesse maggiormente per coloro che hanno intenzione di divenire studenti.
- Parole Chiavi: 'corsi, lista di corsi, dipartimento X'.
- Una pagina contenente una lista di corsi che a loro volta dei link agli URL delle pagine dei corsi stessi.

B. Chi insegna il corso/argomento X?

- Tutti gli utenti sono coinvolti
- Parole Chiavi: 'insegna, chi, lezione, insegnamento, corso, argomento'.
- Una pagina di un corso o di una persona o parte di una pagina.

Il lavoro di definizione dei requisiti si è svolto su tutti e quattro i principali argomenti. Si sono identificate 40 sotto applicazioni che sono reputate di estremo interesse per la definizione delle necessità ed aspettative dell'utente. Si sono indicati pochi domini di applicazione ognuno con un insieme di requisiti che formano la base per un'ulteriore analisi per selezionare informazione e servizi di interesse.

6.5 Domini e Applicazioni

Si è raffinato ulteriormente il concetto di domini di interesse che sono stati riconosciuti in: Persone, Didattica, Organizzazione e Ricerca. Queste saranno le aree di

conoscenza in cui MOSES proverà le proprie funzionalità. L'attenzione, quindi, si è focalizzata su questi domini e sulle loro sotto categorie e le applicazioni in modo da ottenere una classificazione per discernere quale parte di conoscenza è rilevante per il progetto. La struttura gerarchica realizzata è la seguente :

1. Persone

(a) Corpo Docente

- i. Informazioni per il contatto
- ii. Profilo
- iii. Attività

(b) Corpo non docente

- i. Informazioni per il contatto

2. Didattica

(a) Corsi

- i. Organizzazione
- ii. Contenuti
- iii. Lezioni
- iv. Materiale
- v. Prerequisiti

(b) Master

- i. Argomento
- ii. Requisiti

(c) Laboratori

- i. Strumentazione
 - ii. Requisiti
3. Organizzazione
- (a) Accessibilità
 - i. Luogo
 - ii. Informazioni per un contatto
 - iii. Trasporto
 - (b) Amministrazione
 - i. Documenti
 - ii. Scadenze
4. Ricerca
- (a) Informazioni
 - (b) Attività

La gerarchia presentata è alla base della classificazione delle domande a cui MOSES potrà offrire una risposta.

6.6 Criteri di classificazione delle domande per MOSES

MOSES mira ad essere un sistema che offre servizi ai suoi utenti essendo capace di rispondere a domande che non sono limitate alle ricerche standard che sono soddisfatte dagli attuali motori di ricerca. In relazione a ciò è importante comprendere quali sono le caratteristiche delle domande di interesse. Gli attuali motori di ricerca si comportano, più o meno, accettando delle parole chiave e restituendo una lista

ordinata di link (con degli estratti della pagina a cui punta il link). Producono delle risposte interessanti ma vi sono delle caratteristiche che non sono completamente soddisfacenti:

- Non sono capaci di produrre degli insiemi di dati strutturati, come ad esempio tabelle.
- Non sono capaci di collegare tra loro dati provenienti da differenti sorgenti (pagine, siti, database).
- Non supportano le domande basate su frasi.

Di conseguenza, il processo che implementa le risposte alle domande ha una difficoltà che dipende sia dalla domanda in se stesso (e dal tipo di risposta che richiede) sia nel modo in cui l'informazione è conservata nella sorgente (o sorgenti). Devono essere utilizzati dei criteri di classificazione per descrivere queste caratteristiche.

6.6.1 Formato della risposta

Il più semplice formato per una risposta è *un frammento di dato*, un oggetto semplice (atomico o no). Questo è il caso in cui si sta cercando il nome di una persona (identificato in qualche modo, ma questo origina altre coordinate per la classificazione che verranno considerate più avanti) o il suo numero di telefono; Questa tipologia di domanda può essere evasa facilmente da una interrogazione ad un database. Un punto rilevante addizionale in questa situazione è l'utilità di offrire dettagli per il tipo di oggetto, come l'indirizzo email o l'URL. Variazioni dell'oggetto possono sembrare minime ma significanti da molti punti di vista:

- Una pagina web (o un documento) è spesso ciò che viene cercato; questo è ciò che viene restituito dai motori di ricerca; un caso più generale è un frammento (sezione, paragrafo...) da una pagina web (o un documento).

- Una tupla di dati è un insieme di oggetti (generalmente eterogenei ma tipizzati: telefono e indirizzo di una persona).
- Un concetto è il tipo di risposta che un sistema sofisticato può offrire in qualche caso. Questa è una nozione preliminare, in quanto è necessaria una definizione di ciò che un concetto è: un paragrafo in un documento o un nodo in un'ontologia.

Per ognuno dei precedenti punti, è spesso significativo ottenere non semplicemente un elemento ma un insieme (o una lista di elementi omogenei):

- Una lista di oggetti (per esempio, una lista di tipi di corsi); l'ordinamento (esatto o approssimato) potrebbe essere rilevante in alcuni casi e non importante in altri.
- Una lista di pagine; questo è esattamente ciò che i motori di ricerca offrono, ed in genere c'è un ordinamento (ma anche in questo caso potrebbe non essere necessario come nel caso di una ricerca di home page di professori o dipartimenti).
- Una lista di tuple (essenzialmente una tabella): una lista di professori, ognuno con i corsi che il professore insegna e un link alla home page; quest'ultima osservazione suggerisce una generalizzazione dei casi precedenti.
- Lista di concetti

6.6.2 Siti coinvolti

Un aspetto generale è la conoscenza di se si ha uno o più siti, più o meno autonomi: per esempio, si può avere solo un sito per un dipartimento, o il sito centrale dell'università più i siti satellite dei dipartimenti dell'università o i siti di due o più università indipendenti tra loro.

6.6.3 Struttura della sorgente di informazione

Questo ha a che fare con come l'informazione che è necessaria per la risposta alla domanda è organizzata nel sito che deve essere utilizzato. Si possono avere, usualmente, tre casistiche:

- L'informazione è direttamente offerta dal sito (siti) come un insieme unico: questo accade quando l'informazione è già tutta su di una pagina (o su un insieme di pagine, se si sta cercando un insieme di risposte) e si deve solo estrarla dalle pagine.
- L'informazione è disponibile nei siti, è correlata ma non aggregata: in questo caso, l'informazione di interesse è sparpagliata su di un insieme di pagine collegate dai link che devono essere seguiti per raggiungerla.
- L'informazione è disponibile nel sito in una maniera non collegata: questo caso si ha quando l'informazione è sparpagliata su pagine che sono *implicitamente* correlate (per esempio la correlazione potrebbe essere stabilita da valori comuni che appaiono nella pagina ma non sono supporti dai link).

6.6.4 Processo di estrazione

In questo caso si modella come l'utente deve procedere nel sito in cui si trova, in modo da raggiungere la sorgente dell'informazione che gli è necessaria per rispondere alla domanda.

- Servizio esplicito: questo è il caso, per esempio, in cui il sito offre una rubrica o un tool per la ricerca per trovare i professori.
- Navigazione a senso unico: qui l'utente deve navigare, ma senza la necessità di andare avanti ed indietro.

- Navigazione con 'backtracking': in questo caso la navigazione richiede più tentativi (ad esempio, si sta cercando per un determinato professore, ma non si conosce il dipartimento al quale afferisce e non vi è un indice).

Riassumendo, per ogni coppia domanda-risposta si ha una classificazione basata sui seguenti criteri:

- Formato della risposta.
- Numero (e relazione) dei siti coinvolti (uno o più, della stessa università o meno).
- Processo di estrazione

6.7 Classificazione

Sulla base dei criteri discussi nella sezione precedente, è possibile generare una classificazione per le varie domande. Si riportano alcuni esempi di classificazione nella Tabella 6.7 e la relativa legenda. Nell'intestazione della Tabella 6.7 App ha il significato di Applicazione; FdR: Formato della Risposta; EvO: Esatta vs Ordinata; CoF: Centrale o Federato; SSI: Struttura della Sorgente dell'Informazione; PE: Processo d'Estrazione;

Leggenda della Tabella 6.7

Applicazione

- PTC: (Persone - Personale Docente - Informazioni per Contatto)
- PTA: (Persone - Personale Docente - Attività)
- PTP: (Persone - Personale Docente - Profili)

App	Domanda	FdR	EvO	CoF	SSI	PE
PTC	Qual'è la mail del professore X Y ?	LOE	E	C+S	S2	E2
PTC	Qual'è il numero di telefono del professore X Y ?	LOPN	E	C+S	S1	E1
DCC	All'interno del programma del corso X c'è l'argomento Y?	P	E	C+S	S2	E3
DCC	Quali corsi contengono l'argomento X ?	LOI	E	C+S	S2	E3
PTA DCC	Chi è il professore che insegna un corso in cui c'è l'argomento X?	T	E	C+S	S3	E3
PTP	Qual'è il sito web del professore X?	P	E	C+S	S1	E1
OAE	Quale tassa scade il mese prossimo?	D	E	C+S	S1	E3
All	Ci sono delle notizie recenti sull'argomento X?	NL	E	C+S	S1	E1
PTC	Quando riceve il professore X?	I	E	C+S	S1	E2
PTP	Che bandi di concorso sono offerti da dipartimento X?	LOI	E	S	S3	E3

Tabella 6.7: Classificazione delle domande

- PNC: (Persone - Personale Non Docente - Informazioni per Contatto)
- DCO: (Didattica - Corsi - Organizzazione)
- DCC: (Didattica - Corsi - Contenuto)
- DCL: (Didattica - Corsi - Lezioni)
- DCM: (Didattica - Corsi - Materiale)
- DCF: (Didattica - Corsi - Prerequisiti)
- DMS: (Didattica - Master - Soggetto)
- DMR: (Didattica - Master - Requisiti)
- DLE: (Didattica - Laboratori - Strumentazione)
- DLR: (Didattica - Laboratori - Requisiti)
- OAL: (Organizzazione - Accessibilità - Luogo)
- OAC: (Organizzazione - Accessibilità - Informazioni per Contatto)
- OAT: (Organizzazione - Accessibilità - Trasporti)
- OAD: (Organizzazione - Accessibilità - Documenti)
- OAE: (Organizzazione - Accessibilità - Scadenza)
- RIN: (Ricerca - Informazione)
- ROA: (Ricerca - Organizzazione - Attività)
- ROS: (Ricerca - Organizzazione - Argomento)
- RAC: (Ricerca - Administration - Informazioni per Contatto)

Formato della risposta

- LOx: Elemento x della lista
- I: Oggetto
- E: E-mail
- PN: Numero di telefono
- C: Concetto
- D: Data
- P: Pagina
- NL: Linguaggio naturale
- U: URL
- T: Tabella
- TD: Tupla di dati

Esatto vs Ordinata

- E : La risposta è un singolo oggetto esatto.
- R : La risposta è una lista ordinata di oggetti.

Centrale o Federato

- C: L'informazione risiede all'interno del sito centrale.
- S: L'informazione risiede all'interno di uno dei siti satelliti.
- C+S: L'informazione risiede nel sito centrale o in uno (o più) dei siti satelliti.

- C*S: L'informazione deve essere composta utilizzando come sorgente il sito centrale e uno o più dei siti satelliti.

Struttura della sorgente d'informazione

- S1 : L'informazione è direttamente offerta dal sito (siti) come un insieme unico.
- S2 : L'informazione è disponibile nei siti, è correlata ma non aggregata.
- S3 : L'informazione è disponibile nel sito in una maniera non collegata.

Processo d'estrazione

- E1 : Servizio esplicito
- E2 : Navigazione unidirezionale
- E3 : Navigazione con 'backtracking'

6.8 Struttura schede domande

Una descrizione più dettagliata delle domande può essere fornita dalle *schede*, definite nel seguente modo:

Una prima parte che descrive la domanda

Dominio:	Persone / Didattica / Ricerca / Notizie
Applicazione:	Xx
Sotto applicazione:	Yy
Domanda:	Bla Bla
Formato della risposta:	Oggetto / Pagina / Tupla di dati / Concetto / Lista di ... / Tabella / Linguaggio Naturale / Altro

Tabella 6.8: Prima parte della struttura della scheda Domanda/Risposta

La prima parte della Tabella 6.8 indica la classificazione della domanda all'interno dell'organizzazione discussa nella sezione 6.5. La seconda parte presenta una

descrizione basilare di una domanda, con le sue specifiche testuali (nelle tre lingue, Inglese, Italiano e Danese), gli utenti interessati e la struttura del risultato. Il formato della risposta riporta la classificazione presentata nella sezione 6.6.

La seconda parte della scheda è strutturata nel seguente modo:

Nome del sito:	Nome del sito
Tipo di sito:	Centrale o Satellite
Tipo di organizzazione:	Facoltà / Dipartimento / Ufficio / Progetto / Sito universitario
URL:	http//...
Risposta:	
Struttura della sorgente d'informazione - Classificazione:	Informazione offerta direttamente dal sito / Informazione disponibile e correlata ma non aggregata / Informazione disponibile ma non correlata
Struttura della sorgente d'informazione - Descrizione:	Una descrizione di come l'informazione necessaria sia organizzata nella sorgente, includendo la pagina, o le pagine, dove compare e la struttura in cui è inserita nelle pagine (tabella, lista, testo,...)
Struttura della sorgente d'informazione - Percorso d'accesso:	Una descrizione del percorso necessario per raggiungere l'informazione richiesta
Tecnologia utilizzata:	ASP / PHP / JSP / HTML statico
Processo d'estrazione - Classificazione:	Servizio Esplicito / Navigazione unidirezionale / Navigazione con backtracking
Processo d'estrazione - Descrizione:	Una breve descrizione del processo seguito dall'utente (interagendo con il sistema) per ottenere la risposta.

Tabella 6.9: Seconda parte della struttura della scheda Domanda/Risposta

La prima parte della Tabella 6.9 contiene le informazioni basilari sul sito (o i siti) che si stanno considerando per ottenere la risposta alla domanda posta. La seconda parte della tabella è probabilmente la più importante, in quanto illustra come la domanda può essere risposta utilizzando il sito (e quali sono i modi diversi di

accedere allo stesso sito in relazione alle domande poste). Per una singola domanda possono esserci più risposte e quindi la Tabella 6.9 può essere ripetuta più volte per una singola domanda.

6.9 Esempi schede domande

Si riportano ora nelle Tabelle 6.10 e 6.11 degli esempi delle schede per la descrizione delle domande e delle risposte. Una più ampia casistica di schede (Question card) sviluppate può essere trovata nell'Appendice 9.2.

Dominio:	Persone
Applicazione:	Personale Docente
Sotto applicazione:	Informazione per contatto
Domanda:	Qual'è la mail del Professor X Y ? / What is the e-mail of professor X Y ?
Utente:	Studente
Formato della risposta:	Una lista di oggetti (indirizzi email). Più precisamente, potrebbe essere una tabella che associa i nomi con gli indirizzi.
Nome del sito:	Roma TRE
Tipo di sito:	Centrale
Tipo di organizzazione:	Sito universitario
URL:	http://www.uniroma3.it
Risposta:	
Struttura della sorgente d'informazione - Classificazione:	Informazione offerta direttamente dal sito.
Struttura della sorgente d'informazione - Descrizione:	L'informazione è nella pagina con altri oggetti.
Struttura della sorgente d'informazione - Percorso d'accesso:	Un link (Rubrica) sulla home page conduce direttamente ad una form. La risposta della form è una tabella con delle informazioni (incluso il nome e i numeri di telefono) di tutti i professori selezionati. Un link in questa pagina conduce direttamente ad un'altra pagina con la mail del professore
Tecnologia utilizzata:	ASP
Processo d'estrazione - Classificazione:	Servizio Esplicito
Processo d'estrazione - Descrizione:	L'utente seleziona un link sulla home page del sito che conduce ad una form che permette la ricerca di professori (ed anche di personale non docente) fornendo il nome (o parte del nome). La risposta è una tabella (ogni professore su una riga) con nome, affiliazione e numero di telefono. L'utente deve selezionare il professore desiderato per accedere ad una nuova pagina con più informazioni. In questa pagina troverà l'indirizzo email.

Tabella 6.10: Esempio di una scheda Domanda/Risposta sul dominio Persone

Dominio:	Didattica
Applicazione:	Corsi
Sotto applicazione:	Contenuti
Domanda:	All'interno del programma del corso X c'è l'argomento Y ? / Within a course X program there is argument Y?
Utente:	Studente
Formato della risposta:	Potrebbe essere un URL al corso X cercato.
Nome del sito:	Roma TRE
Tipo di sito:	Centrale
Tipo di organizzazione:	Sito universitario
URL:	http://www.uniroma3.it
Risposta:	
Struttura della sorgente d'informazione - Classificazione:	Informazione offerta direttamente dal sito
Struttura della sorgente d'informazione - Descrizione:	L'informazione si trova in un file PDF
Struttura della sorgente d'informazione - Percorso d'accesso:	Un link (DIDATTICA E STUDENTI) dalla home page conduce alla pagina dove il link (ELENCO COMPLETO) mostra tutti i corsi disponibili a Roma TRE. Selezionando il corso desiderato l'utente accede alla pagina della facoltà e ad un documento in formato PDF degli Ordini degli studi che contiene i programmi di tutti i corsi della facoltà.
Tecnologia utilizzata:	ASP
Processo d'estrazione - Classificazione:	Navigazione unidirezionale
Processo d'estrazione - Descrizione:	L'utente seguendo il link (DIDATTICA E STUDENTI) dalla home page raggiunge una pagina in cui vi è un link (ELENCO COMPLETO) che presenta tutti i corsi disponibili nell'università Roma TRE. Selezionando il corso desiderato l'utente accede alla pagina della facoltà dove trova varie informazioni su di essa compreso un link che porta all'Ordine di Studi in formato PDF. Questo documento contiene i programmi di tutti i corsi della facoltà. L'utente dovrà scorrere tale file fino a trovare il programma del corso X e verificare se vi è l'argomento Y.

Tabella 6.11: Esempio di una scheda Domanda/Risposta sul dominio Didattica

Capitolo 7

MOSES: Rappresentazione della conoscenza e linguaggi del web semantico

7.1 Rappresentazione della conoscenza e linguaggi del web semantico

I linguaggi che erano stati proposti nel progetto MOSES per la rappresentazione della conoscenza erano originariamente DAML per l'espressione delle ontologie, XTM per le espressioni delle relazioni nella base di conoscenza ed RDF per l'annotazione dei contenuti web. La specifica DAML+OIL, che inizialmente era stata utilizzata, è stata rimpiazzata da OWL durante lo sviluppo del progetto, le ontologie inizialmente proposte in DAML sono state tutte convertite nella sintassi OWL. Più generalmente OWL è proposto come un unico formato per la rappresentazione di tutto il livello ontologico, inclusi gli elementi originariamente definiti utilizzando i concetti delle Topic Maps.

7.2 XTM e OWL

Questa sezione dettaglia l'approccio sviluppato per integrare gli elementi XTM ed OWL nella stessa struttura semantica. Le soluzioni proposte non sono quelle

standard, in quanto lo stato dell'arte non dà regole precise per un'integrazione di questo tipo. Ma questo approccio può essere considerato un input per i gruppi di standardizzazione, ISO/JTC1 SC34/WG3 (per le Topic Maps), W3C Semantic Web Activity (per OWL), e OASIS Published Subjects Technical Committee. Lo stato corrente dei tre standard può essere trovato ai seguenti indirizzi:

- Guide to OWL specification: <http://www.w3.org/TR/owl-guide/>
- Topic Maps family of standards, including the XTM syntax: <http://www.isotopicmaps.org>
- Published Subjects Technical Committee: http://www.oasis-open.org/committees/tc_home.php?pubsubj

Siccome le ontologie universitarie usano sia definizioni OWL che Topic Map, e lo strumento di Mondeca, ITM, usa come formato di input XTM, questa sezione esplorerà come XTM e le caratteristiche base di OWL sono interoperabili, e come entrambi i linguaggi possono essere integrati come input in un'unica base di conoscenza.

7.2.1 Il livello Tassonomico: Classi, Sottoclassi ed Istanze

Il livello tassonomico sembra offrire una semplice interoperabilità tra le ontologie e le Topic Map. Classi, sottoclassi ed istanze, che sono le caratteristiche base nelle ontologie, sono facilmente esprimibili mediante la sintassi Topic Map ed XTM. Infatti classe-istanza è l'unica associazione generica definita dallo standard ISO 13250, e la proprietà di sottoclasse è generalmente definita come un tipo particolare di associazione binaria, con i tipi per i ruoli *classe* e *sottoclasse*. Queste caratteristiche sono gestite in maniera naturale in ITM, e qualunque tassonomia (un file strutturato che rappresenta un insieme di classi, sottoclassi ed istanze) può essere importato nel software qualunque sia il linguaggio in cui è espressa.

7.2.2 Distinzioni tra classi ed istanze

Nelle Topic Map sia le classi che le istanze sono rappresentate da *topic*, e nulla impedisce allo stesso topic di essere utilizzato come un'istanza ad un livello semantico e come una classe ad un altro livello. Infatti quando è creato un topic, non è necessariamente definito come classe o istanza, sebbene sia abbastanza semplice farlo. ITM supporta la stessa flessibilità, ma ha al suo interno un modo per distinguere i topic utilizzati come classi dagli altri.

Le specifiche OWL, in questo caso, possono supportare due opzioni, dipende dalla versione del linguaggio. In OWL-Lite e OWL-DL, classi ed istanze sono due oggetti distinti, mentre in OWL-Full, lo stesso oggetto può essere considerato classe ad un livello ed istanza ad un altro livello, offrendo la stessa flessibilità di XTM (ed RDF).

7.2.3 Coerenza della Tassonomia

Questo punto è più critico. A differenza delle ontologie formali, le Topic Map non offrono alcun modo per controllare la coerenza di una tassonomia, come ad esempio la ricerca di cicli classe-sottoclasse, errori di accoppiamento tra classi ed istanze e altre strutture generalmente proibite. Infatti nelle specifiche delle Topic Map vi è il chiaro intento di supportare qualunque tipo di inconsistenza nel caso l'autore decida che vada bene. Questa particolarità ha attratto molte critiche sulle Topic Map dalla comunità di Logica Formale.

Una soluzione per il problema della coerenza della tassonomia dovrebbe essere trovata nella creazione del futuro linguaggio Topic Map Constraint Language che però è ancora allo stadio embrionale. Nel frattempo il miglior metodo per un uso di tassonomie formali in Topic Map è validarle con un validatore OWL oppure costruirle con un software dedicato (come Protégé-2000). Delle tassonomie costruite in questo modo possono facilmente essere espresse in sintassi XYM e proposte come input ad

un motore di Topic Map. Ovviamente qualunque successiva modifica al file XTM potrebbe introdurre inconsistenze.

7.2.4 Supporto per la possibilità di avere più sottoclassi

Sia le Topic Map che OWL supportano la possibilità di avere molte sottoclassi, e le ontologie che sono utilizzate in MOSES ne contengono un certo numero. Sebbene la struttura generica ITM può supportare l'uso di sottoclassi multiple è stata fatta una scelta di non introdurre questa caratteristica nell'editor ITM. I Topic, in ITM, sono organizzati sotto poche superclassi disgiunte tra loro: *concept*, *individual* e *document*. Queste definizioni di superclassi è la causa di una scelta di fogli di stili specifici per la pubblicazione e la creazione di topic, adattati alle proprietà generiche di questa classe. Le sottoclassi multiple trovate nelle ontologie importate saranno accettate da ITM sempre che non vadano in conflitto con le superclassi gestite all'interno di ITM.

7.2.5 Proprietà OWL e vincoli delle Topic Map

Le due classi di proprietà in OWL, *owl:DatatypeProperty* e *owl:ObjectProperty*, possono facilmente essere fatte corrispondere alle occorrenze di un topic e al ruolo delle associazioni binarie con vincoli che permettono ad una classe tipi di occorrenze e tipi di ruolo, includendo la possibilità di restrizioni di cardinalità.

Sfortunatamente, lo standard ISO 13250 delle Topic Map, attualmente, non supporta questo tipo di vincoli né nel modello né nella sintassi. Questo tipo di espressioni saranno possibili quando il linguaggio Topic Maps Constraint Language (TMCL) sarà sviluppato.

7.2.6 Identificazione dei soggetti, di classi e proprietà

L'identificazione dei soggetti, rappresentati da Topic, utilizzeranno URI, e di fatto maggiormente URL, per conformarsi al meccanismo generale raccomandato in un ambiente web (semantico), la generica nozione di risorsa in RDF/OWL e la nozione di Indicatore del soggetto pubblicato come è definita in Topic Map.

OWL offre il meccanismo generico RDF per l'identificazione di classe e proprietà attraverso gli URI. Come mostrato nella sezione precedente i meccanismi di rappresentazione offrono una identificazione diretta per le occorrenze, per i tipi di ruolo e associazione, rispettivamente identificati da DatatypeProperty, ObjectProperty e Class.

Questi identificatori non devono essere necessariamente univoci in quanto l'equivalenza è dichiarata nelle ontologie utilizzando la proprietà *sameAs*. L'ontologia italiana usa traduzioni di molte classi dell'Ontologia dell'Università in italiano, infatti nel seguente esempio i due identificatori rappresenteranno la stessa classe:

<http://www.mondeca.com/owl/univ1.1.xml#Faculty>

http://www.mondeca.com/owl/univ_ita1.1.xml#Facoltà

7.3 Ontologia Italiana

In questa sezione, utilizzando degli sforzi preesistenti nelle modellazione di domini di conoscenza, viene data una concettualizzazione del dominio di conoscenza richiesto con le peculiarità del sistema universitario italiano.

Il mondo universitario è un dominio ben studiato e per questo si trovano alcune risorse preesistenti già disponibili. La concettualizzazione contenuta nella biblioteca ontologica di DAML-OIL è un'interessante rappresentazione alla luce degli scenari di applicazione previsti in MOSES (ad esempio persone/Corsi/Ricerca). Nel documento univ1.0.daml pubblicato nel sito www.daml.org vi sono delle classi e le relazioni

che coprono molti dei concetti rilevanti per gli scenari considerati nel progetto. Si riportano, nel seguito, sia la gerarchia delle classi, per i concetti, che la gerarchia delle proprietà, per le relazioni. Nella gerarchia dei concetti le porzioni delle ontologie generali (*general1.0.daml*) e delle ontologie dei documenti (*docmnt1.0.daml*) attivate dai concetti nell'ontologie dell'università (*univ1.0.daml*) sono indicate nella gerarchia delle classi (i concetti definiti localmente sono scritti in maniera normale mentre i concetti esterni sono in *italico*). Tutte le ontologie coinvolte sono descritte nell'Appendice 9.2.

Ontologia universitaria.

Risorsa ufficiale: <http://www.cs.umd.edu/projects/plus/DAML/onts/univ1.0.daml>

Gerarchia delle Classi

- *Communication (communicator, recipient)*
 - *Document (author, authorOrg, containedIn, publishDate, publisher, subject, title)*
 - * Publication ()
 - Article ()
 - + BookArticle ()
 - + ConferencePaper ()
 - + JournalArticle ()
 - + WorkshopPaper ()
 - Book ()
 - Periodical (volume)
 - + Journal ()
 - + Magazine ()

- Proceedings ()
- NewsItem ()
- Thesis ()
 - + DoctoralThesis ()
 - + MastersThesis ()
- Schedule (listedCourse) is a base1.0.daml#SHOEEntity
- *Activity* ()
 - Work ()
 - * Course ()
 - * Research ()
 - * MASTER
 - * LaureaDiPrimoLivello
 - * LaureaDiSecondoLivello
 - * Laurea
- *Agent* (*engagesIn*, *performs*)
 - Person (emailAddress, homeAddress, homePhone, workAddress, workPhone)
 - * Employee (worksFor)
 - Assistant ()
 - + ResearchAssistant ()
 - + TeachingAssistant (teachingAssistantOf)

- AdministrativeStaff ()
 - + Chair ()
 - + ClericalStaff ()
 - + Dean ()
 - + Director ()
 - + SystemsStaff ()
- Faculty (teacherOf)
 - + Professore
 - ++ ProfessoreOrdinario
 - +++ ProfessoreEmerito
 - +++ Decano
 - ++ ProfessoreStraordinario
 - ++ ProfessoreAssociato
 - ++ ProfessoreAContratto
 - ++ ProfessoreVisitatore
 - + TitolareCorsoIntegrativo
 - + Ricercatore
 - + RicercatoreAContratto
 - + AssegnistaDiRicerca
- * Student (advisor, takesCourse)
 - GraduateStudent ()
 - UnderGraduateStudent ()
- *SocialGroup (member)*
 - * Organization (affiliateOf, affiliatedOrganization, alumnus, head, isLocated, orgAddress, orgPhone, subOrganizationOf)

- EducationOrganization ()
 - + Department ()
 - + Institute ()
 - + Program ()
 - + ResearchGroup (researchProject)
 - + School ()
 - + University (offers)
- ResearchOrganization ()
 - + Laboratory

- AdministrativeOrganization ()
 - + SupportOffice

- *Event (eventEnd, eventLocation, eventStart, participant)*
 - Conference ()

- *PhysicalObject ()*
 - *Thing ()*
 - * *Organism ()*
 - Person (emailAddress, homeAddress, homePhone, workAddress, workPhone)

- SUBJECT

- EMAILADDRESS

- PHONENUMBER

- TIME

- TimeInterval
- RepeatedTimeInterval
- Instant
- * DATE

- CREDIT OpenPosition

Da questa gerarchizzazione si evince facilmente come una domande vada ad interessare alcune classi ed alcune relazioni. Ad esempio la domanda:

- Qual è l'indirizzo email dei professori che insegnano informatica nell'università Roma TRE?

Concetti:

- Faculty
- Course: Informatics
- University: RomaTRE
- emailAddress

Relazioni binarie (proprietà in DAML+OIL:

- teacherOf(Faculty, Course)
- offers(University, Course)
- emailAddress(Person, emailAddress)

Nome della Relazione	Argomento 1	Argomento 2
affiliateOf	Organization	Person
affiliatedOrganization	Organization	Organization
alumnus	Organization	Person
containedIn	Document	Document
emailAddress	Person	EMAILADDRESS
head	Organization	Person
member	SocialGroup	Person
publicationAuthor	Document	general1.0.daml#Person
publicationDate	Document	DATE
publicationOrg	Document	general1.0.daml#Organization
publisher	Document	general1.0.daml#Organization
subOrganizationOf	Organization	Organization
subject	Document	SUBJECT
subject	Work	SUBJECT
advisor	Student	Professor
doctoralDegreeFrom	general1.0.daml#Person	University
listedCourse	Schedule	Course
mastersDegreeFrom	general1.0.daml#Person	University
offers	University	Course
publicationResearch	docmnt1.0.daml#Publication	Research
researchInterest	general1.0.daml#Person	Research
researchProject	ResearchGroup	Research
takesCourse	Student	Course
teacherOf	Faculty	Course
teachingAssistantOf	TeachingAssistant	Course
tenured	Professor	
undergraduateDegreeFrom	general1.0.daml#Person	University
activityStart	Activity	DATE
activityEnd	Activity	DATE
hasProgram	Course	Document
assignCredits	Course	CREDIT
hasQuestionTime	Faculty	RepeatedTimeInterval
hasOpeningTime	SupportOffice	RepeatedTimeInterval
emailAddress	Organization	EMAILADDRESS
hasMachinery	Laboratory	Artifact
offers	Organization	OpenPosition
typeOfRole	OpenPosition	Employee
prerequisite	Work	SHOE#Entity

Tabella 7.1: Gerarchia delle relazioni (proprietà)

7.4 Topic e strumento ITM Mondeca

Gerarchia Topic

Tutti i topic hanno origine dai topic primitivi *Document* e *Subject* che sono definiti in un file esterno di mondeca in questo modo:

```
<owl:Class rdf:ID="Document">
<rdfs:label xml:lang="en">Document</rdfs:label>
<rdfs:subClassOf rdf:resource="#Topic"/>
</owl:Class>
```

```
<owl:Class rdf:ID="Subject">
<rdfs:label xml:lang="en">Subject</rdfs:label>
<rdfs:subClassOf rdf:resource="#Topic"/>
</owl:Class>
```

La classe Topic è definita in questo modo:

```
<owl:Class rdf:ID="Topic">
<rdfs:label xml:lang="en">Topic</rdfs:label>
</owl:Class>
```

In più nello stesso file vi è la definizione di associazione e di ruolo dell'associazione:

```
<owl:Class rdf:ID="Association">
<rdfs:label xml:lang="en">Association</rdfs:label>
</owl:Class>
```

```
<owl:ObjectProperty rdf:ID="associationRole">
<rdfs:label xml:lang="en">Association Role</rdfs:label>
<rdfs:domain rdf:resource="#Association"/>
<rdfs:range rdf:resource="#Topic"/>
</owl:ObjectProperty>
```

Da queste classi discendono tutti i *topic* che sono stati utilizzati per la costruzione delle specifiche associazioni e topic. Un esempio di gerarchia si trova nella Figura 7.1.

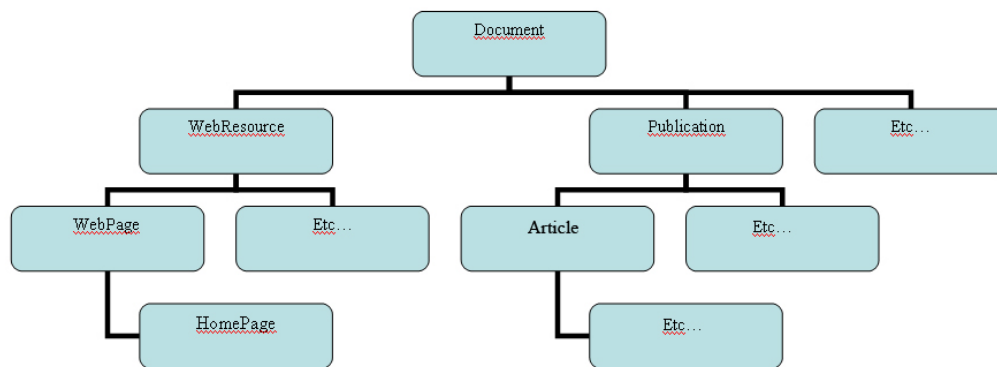


Figura 7.1: Esempio di gerarchia delle classi

7.4.1 Descrizione di due generici topic

Definizione di Professore Associato:

```
<owl:Class rdf:ID=" ProfessoreAssociato ">
<rdfs:subClassOf>
<owl:Class rdf:about="#Professore"/>
</rdfs:subClassOf>
</owl:Class>
```



```
<owl:Class rdf:ID="Address">
<rdfs:subClassOf rdf:resource="#Individual"/>
</owl:Class>
```

Esempio di istanziazione topic:

```
<!-- topic: Riccardo Torlone, -->
<topic id="personid_13">
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#ProfessoreAssociato"/>
</instanceOf>
<occurrence>
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#personalEmailAddress"/>
</instanceOf>
<resourceRef xlink:href="torlone@dia.uniroma3.it"/>
</occurrence>
<occurrence>
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#workPhone"/>
</instanceOf>
<resourceData> 06 5517 3377 </resourceData>
</occurrence>
<baseName>
```

```
<baseNameString>Riccardo Torlone</baseNameString>
</baseName>
<baseName>
<baseNameString>Torlone Riccardo </baseNameString>
</baseName>
</topic>

<!-- topic:Riccardo Torlone office address -->
<topic id="address_personid_13">
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#Address"/>
</instanceOf>
<occurrence>
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#addressStreet"/>
</instanceOf>
<resourceData> Via della Vasca Navale, 79 00146 Roma </resourceData>
</occurrence>
<baseName>
<baseNameString>Riccardo Torlone office address</baseNameString>
</baseName>
</topic>
```

Descrizione della generica association:

```
<owl:Class rdf:ID="workAddress">
```

```
<rdfs:subClassOf rdf:resource="#Association"/>
<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty>
<owl:ObjectProperty rdf:about="#workAddressSubject"/>
</owl:onProperty>
<owl:allValuesFrom>
<owl:Class rdf:about="#Person"/>
</owl:allValuesFrom>
</owl:Restriction>
</rdfs:subClassOf>

<rdfs:subClassOf>
<owl:Restriction>
<owl:onProperty rdf:resource=
"#workAddressObject" rdf:type="http://www.w3.org/2002/07/owl#ObjectProperty"/>
</owl:onProperty>
<owl:allValuesFrom>
<owl:Class rdf:about="#Address"/>
</owl:allValuesFrom>
</owl:Restriction>
</rdfs:subClassOf>
</owl:Class>
```

Esempio di istanziazione association:

```
<!-- association: Person - Address -->
<association id="person_13_workaddress">
```

```
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#workAddress"/>
</instanceOf>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#workAddressSubject"/>
</roleSpec>
<topicRef xlink:href="#personid_13"/>
</member>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#workAddressObject"/>
</roleSpec>
<topicRef xlink:href="#address_personid_13"/>
</member>
</association>
```

7.4.2 Inserimento istanze delle ontologie in sistema ITM

Inserimento istanze delle ontologie in sistema ITM attraverso sito Mon-deca

La descrizione della metodologia per l'inserimento di istanze delle ontologie attraverso l'uso di un'interfaccia web si attua inizialmente attraverso le immissioni delle istanze nel sito: <http://mondeca-publishing.com/moses/audit>. utilizzando l'account danese e navigando nella Knowledge Base. Alcuni termini utili alla comprensione

delle pagine sono: Hjemmeside = Home Page e Bevis = Argomento. Una volta acceduti nella home page si viene dirottati alla pagina mostrata nella Figura 7.2.



Figura 7.2: Home page ITM editor

Sulla barra sinistra vi è un albero di cartelle che contengono le ontologie generiche, un esempio è mostrato in Figura 7.3. Se si espandono si naviga attraverso la gerarchia presentata precedentemente.

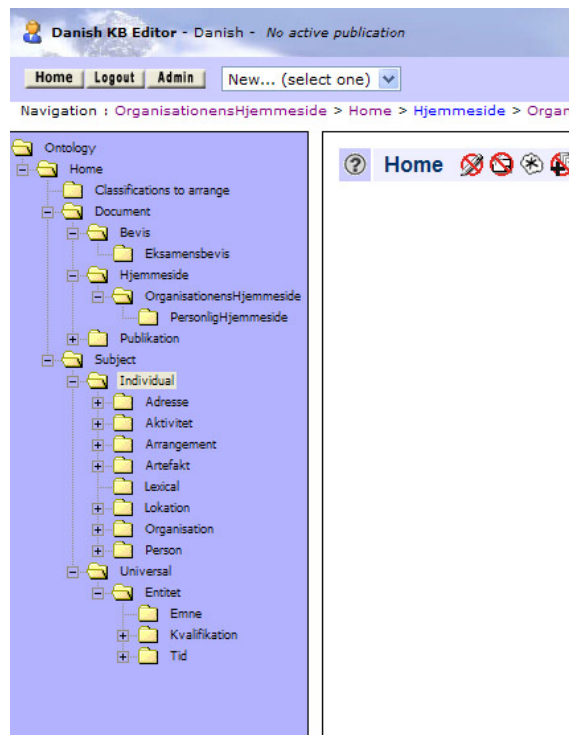


Figura 7.3: Indice delle Ontologie in ITM editor

Selezionando i nomi delle singole classi si naviga attraverso le proprietà di appartenenza delle ontologie come si può vedere nella Figura 7.4.



Figura 7.4: Particolare delle ontologie in ITM editor

Per operare delle ricerche sulle istanze inserite nello strumento ITM si va sul campo search e si inserisce un nominativo. Si sceglie in che set cercare (all, individuals, universals o documents). La ricerca è possibile solo sulle *topic* inserite. Se l'istanza è inserita allora si ottiene come risultato una pagina che riassume le proprietà possedute dall'oggetto ricercato e se fa parte o meno di associazioni.



Figura 7.5: Possibilità di modifica delle istanze delle ontologie in ITM editor

Accanto al nome dell'istanza cercata vi sono quattro simboli, si possono vedere nella Figura 7.5. Ognuno ha una funzione specifica:

- Una matita: per editare l'istanza stessa
- Una letterina: per inserire commenti
- Un grafo: per visualizzare un grafo dei ruoli che svolge l'istanza.

Lo strumento di visualizzazione grafo è quello più interessante in quanto pone, in forma grafica, tutte le relazioni che intercorrono tra l'istanza cercata e vari altri oggetti. Un esempio è visibile nella Figura 7.6.

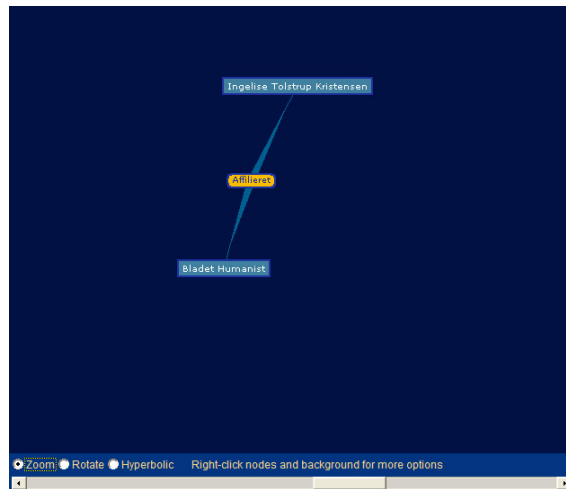


Figura 7.6: Strumento di visualizzazione delle relazioni delle istanze cercate

Lo strumento è molto semplice da utilizzare. Presenta dei grafi in cui le istanze sono dei rettangoli azzurri e i ruoli sono delle ellissi gialle. Se si seleziona una istanza, passandoci sopra con la freccia del mouse, si ha poi la possibilità, con il tasto destro, di eseguire alcune operazioni:

- Center Topic: carica la pagina delle proprietà dell'istanza.
- Expand Node: espande tutte le varie relazioni esistenti tra quella istanza ed altri oggetti.
- Collapse Node: fa collassate tutte le relazioni tranne la prima.
- Hide Node: nasconde il nodo.

La Figura 7.7 è il grafo con l'espansione dell'istanza *Bladet Humanist*

Chiudendo lo strumento si ritorna alla pagina che descrive le proprietà dell'istanza cercata. Sulla destra dello schermo si trova la classe di appartenenza dell'istanza.

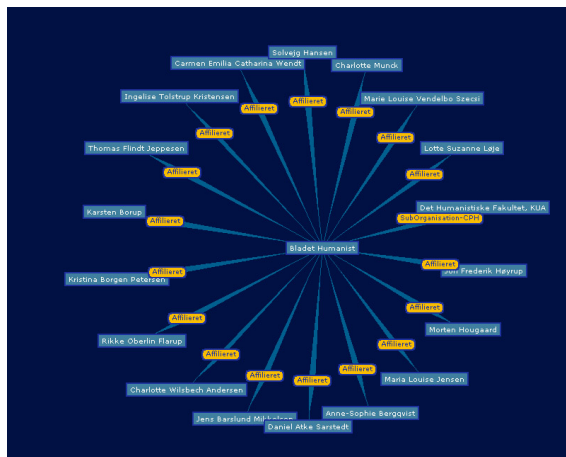


Figura 7.7: Visualizzazione delle relazioni dell'istanza cercata con altre istanze

Se con il mouse si seleziona il link viene presentata una lista di tutte le istanze della classe in esame e anche varie proprietà della classe stessa. La navigazione procede nello stesso modo per tutte le istanze dei *topic* inseriti.

7.4.3 Inserimento istanze delle ontologie in sistema Mondeca ITM attraverso script

Per la generazione del file XML contenente le informazioni provenienti dal database del Dipartimento di Informatica ed Automazione e dal sito dell'Università RomaTRE e che sarebbe andato a creare le istanze delle ontologie definite nel sistema ITM, sviluppato da Mondeca, si è scelto di utilizzare il linguaggio di scripting PHP. La scelta è derivata dal fatto che l'immissione manuale presentata nella precedente sezione è risultata troppo laboriosa e lenta rispetto la mole di istanze da inserire.

Si è proceduto nel seguente modo :

- Analisi delle schede presentate nella sezione 6.8.
- Studio del linguaggio XTM

- Analisi dei vari file presenti sul sito di Mondeca contenenti le classi e le associazioni utilizzabili.
- Analisi del dump del Database del Dipartimento di Informatica ed Automazione (DIA)
- Installazione e configurazione dei software necessari per lo sviluppo:
 - Apache Server
 - PhP
 - PostgreSQL
 - Oxygen
- Creazione del file di script php che avrebbe generato il file xml a partire dalla base di dati del DIA con successive integrazioni manuali.

Elenco Istanze generate:

1. Istanze Università RomaTRE e Dipartimenti

- Definizione Università RomaTre
- Definizione Dipartimenti
- Affiliazione Dipartimento Università
- Homepage Dipartimenti
- Email Dipartimenti
- Telefono Dipartimenti
- Stato Geografico in cui si trova il Dipartimento
- Città in cui si trova il Dipartimento

- Indirizzo del Dipartimento
- Cap del Dipartimento
- Definizione di Persona con qualifica di Preside
- Associazione Preside Dipartimento

2. Istanze Dipartimento Informatica e Automazione

- Tipologia Argomento Persona
- Schede Persona
- Associazione Lexicalization Persona Nome
- Lexical Persona
- Associazione affiliateOf Dipartimento Persona
- Homepage Persona
- Email Persona
- Lexical email Persona
- Associazione Lexicalization Persona email
- Telefono Persona
- Lexical Telefono Persona
- Associazione Lexicalization Persona Telefono
- Indirizzo Persona
- Lexical Address Persona
- Associazione Lexicalization Persona Indirizzo
- Ricevimento Persona
- Lexical Ricevimento Persona

- Associazione Lexicalization Persona Indirizzo
- Associazione questionTime Persona Time

3. Tipologia Argomento Corsi

- Definizione Corso
- Scheda Corso
- Lexical Nome Corso
- Associazione Lexicalization Corso
- Associazione Università offre Corso
- HomePage Corso
- Lexical HomePage Corso
- Associazione Lexicalization Homepage Corso
- Associazione Professore Insegna Corso
- Definizione Programma del Corso
- Associazione Corso ha il Programma

4. Tipologia Argomento Ricerca

- Definizione Ricerca
- Associazione Persona interessata ad un campo di Ricerca
- Definizione CFU
- Associazione Corso assegna Numero Crediti
- Definizione Articoli di Ricerca
- Associazione Autore di quale articolo

- Associazione Campo di Ricerca dell'Articolo

5. Tipologia Argomento Tempo

- Definizione anni
- Giorni Settimana
- Definizione orario

Descrizione Istanze generate:

relazione sul DIA xml file.doc pag 3

Nome: Definizione Università RomaTre

Istanza della classe: University

ID Topic: RomaTRE_University

Utilizzo: Definisce RomaTRE come istanza della classe University

Dominio: Istanziata RomaTRE

Codice XTM generato:

```
<!-- Definizione Istanza University RomaTRE -->
<!-- topic: University RomaTRE -->
<topic id="RomaTRE_University">
<instanceOf>
<subjectIndicatorRef xlink:href=
"http://www.mondeca.com/owl/moses/ita.owl#University"/>
</instanceOf>
<subjectIdentity>
<subjectIndicatorRef xlink:href="www.uniroma3.it"/>
</subjectIdentity>
```

```
<baseName>  
<baseNameString>Universita degli Studi di Roma RomaTRE</baseNameString>  
</baseName>  
<baseName>  
<baseNameString>RomaTRE</baseNameString>  
</baseName>  
</topic>
```

Problemi affrontati.

Si sono affrontate diverse problematiche durante la creazione del file xml. La più rilevante è stata la gestione dei dati contenuti nel database DIA che in formato HTML destavano problemi con la codifica UTF-8 del file xml generato.

Il database del DIA possiede i dati in formato html, in modo da non doverli trattare durante la materializzazione delle pagine dinamiche del sito. Questo accorgimento, molto utile durante la creazione delle pagine del sito web, ha posto in essere la necessità di creare delle funzioni di conversione dati così da poter utilizzare senza problemi le informazioni provenienti dal database.

Modifica della struttura del Database

Si è lavorato su due versioni differenti del Database del DIA. La prima versione meno completa della seconda. Questo ha fatto sì che si sono dovute modificare alcune delle funzioni che si utilizzavano per estrarre i dati.

Modifica dei file a cui fare riferimento sul sito Mondeca

Dopo un incontro tra i partner che lavorano sul progetto MOSES si è deciso di modificare i file che erano utilizzati come base di riferimento per la definizione di classi e associazioni. Si è quindi dovuto modificare lo script in modo che generasse

i nuovi riferimenti in maniera corretta. Si sono anche modificate le istanze create manualmente.

Inserimento classe Lexical e associazione Lexicalization per annotazione pagine

Per annotare le pagine web i partner hanno deciso di creare una classe ed un'associazione. Si sono quindi apportati dei cambiamenti nello script PHP per gestire i due nuovi oggetti.

Capitolo 8

MOSES: Annotazione in XML e OWL RDF

8.1 Annotazione in XML e OWL RDF

In questa sezione, si esaminano i dettagli tecnici di come gli standard del web semantico si comportano con le delle situazioni in cui vi è multilinguismo, focalizzandosi in particolare con gli standard XML utilizzati in MOSES: OWL-RDF e XTM Topic Map. Risorse di linguaggi nella sintassi XML sono indicate attraverso un attributo interno

`xml:lang`.

In OWL, come in RDF, questo attributo è usato per indicare il linguaggio di un elemento etichetta. Il seguente esempio mostra l'etichetta inglese 'Book' unita alla classe 'Libro' nell'ontologia italiana.

```
<owl:Class rdf:about="http://www.mondeca.com/owl/moses/ita.owl#Libro">  
<rdfs:label xml:lang="en">Book</rdfs:label>  
</owl:Class>
```

Le etichette sono semplici proprietà di annotazione e non trasmettono alcun tipo di semantica. Classi diverse possono portare la stessa etichetta nello stesso linguaggio con un impatto semantico nullo. Infatti non c'è alcun modo, in OWL,

di catturare un'etichetta linguistica come una caratteristica semantica, sebbene le applicazioni sono libere di trarre qualunque interpretazione dalle annotazioni. Un modo per aggirare l'ostacolo è di utilizzare delle etichette preferenziali come degli identificatori univoci in un *namespace*, avere tanti namespace quante sono le lingue e dichiarare delle equivalenze semantiche tra le classi definite nei differenti namespace. Nell'esempio seguente è dichiarata una equivalenza semantica tra la classe 'Libro' nel namespace dell'ontologia italiana e la classe 'Book' nel namespace riferito all'ontologia dell'università .

```
<owl:Class rdf:about="http://www.mondeca.com/owl/moses/ita.owl#Libro">
<owl:equivalentClass
rdf:resource="http://www.mondeca.com/owl/moses/univ.owl#Book"/>
</owl:Class>
```

In quest'ultimo caso, una query semantica che sta cercando delle istanze del ita:Libro dovrebbe anche trovare, se disponibile nel database, istanze di univ:Book. E questo è completamente indipendente da qualsiasi etichetta unita a qualunque classe in qualunque linguaggio.

Questa gestione dei linguaggi in XML, RDF ed OWL causerà serie difficoltà se i gestori di ontologie e gli utenti non saranno accorti alla distinzione tra la mappatura semantica e linguistica. Vedremo nella prossima sezione che la specifica delle topic map ha un chiaro e nativo sistema per la gestione dei linguaggi che sono considerati semplicemente degli attributi di annotazione.

8.1.1 Linguaggi in Topic Map

Sebbene ci sia nella sintassi XTM, XTM Topic Map non fa uso dell'attributo `xml:lang`, ed offre uno specifico, e in qualche modo controverso, modo di rapportarsi con i le lingue. I linguaggi sono uniti a delle caratteristiche dei topic (nomi ed occor-

renze) utilizzando un elemento *scope*. *Scope* (Vista) si suppone definisca un contesto in cui l'unione della proprietà, che è visibile, è valido, ed è formalmente definito in XTM come un insieme di Topic, che può contenere altri elementi di contesto oltre che quelli relativi ai linguaggi, ma la semantica di *scope* in generale è rimasta oscura, in particolare quando vari elementi *scope* sono utilizzati insieme.

Rimanendo in un uso dell'elemento *scope* che sia consensuale, si può inserire il Topic 'italiano' nello *scope* di un topic 'baseName'. In questo modo il nome dato è valido in italiano. Questo è l'utilizzo che se ne fa nel database ITM, sia per topic che rappresentano classe nell'ontologia sia per topic rappresentanti istanze nella base di conoscenza.

Per esempio, il topic rappresentante la classe 'Libro' nell'ontologia italiana apparirà nel file XTM come segue. Si noti che due elementi *subjectIndicatorRef* che contengono gli identificatori del topic come riferimento ai file originali OWL sia nell'*ontologia italiana* che nell'*ontologia generale* sono distinti dall'elemento *baseName* che definisce il nome del topic in italiano.

```
<topic id="Libro">
<subjectIdentity>
<subjectIndicatorRef
xlink:href="http://www.mondeca.com/owl/moses/univ.owl#Book"/>
<subjectIndicatorRef
xlink:href="http://www.mondeca.com/owl/moses/ita.owl#Libro"/>
</subjectIdentity>
<baseName>
<scope>
<subjectIndicatorRef xlink:href=
"http://psi.oasis-open.org/iso/640/#ita"/>
```

```
</scope>  
<baseNameString>Libro</baseNameString>  
</baseName>  
</topic>
```

Anche qui l'identificazione e il nominare sono due distinti meccanismi. Ma al baseName XTM è permesso portare più informazione semantica che alle etichette RDF. Se scelta, la cosiddetta Topic Naming Constraint (TNC) può essere rinforzata, ciò significa che in un dato ambito di visibilità, due distinti topic non possono avere lo stesso nome. In altre parole, due topic con lo stesso baseName e nello stesso campo di visibilità dovrebbero essere dichiarati identici e per questo fusi insieme. Questo vincolo, rinforzato da XTM 1.0, è stato alleggerito nelle versioni successive ma le applicazioni possono comunque sempre rinforzarlo o meno. In MOSES il TNC è applicato ad ogni nodo ed a ogni ambito di visibilità di ogni linguaggio. In altre parole, in ogni nodo MOSES l'ontologia più la base di conoscenza usa l'*assunzione del nome unico*. Due topic con lo stesso nome nello stesso ambito di visibilità di un linguaggio in un dato nodo saranno considerati identici. La ragione di rinforzare la TNC in ogni nodo è legata al fatto dell'uso di tool linguistici che trasformano domande degli utenti in richieste semantiche. Questi tool lavorano su di un singolo linguaggio e non riescono a gestire la polisemia in questo modo deve esserci una corrispondenza uno a uno al livello a cui loro operano tra nomi e concetti.

Ma la TNC non verrà utilizzata attraverso i nodi. Questo significa che i concetti nei differenti nodi di MOSES usando lo stesso nome nello stesso linguaggio saranno considerati distinti, a meno che non si riferiscano esplicitamente allo stesso identificatore URI. Questo principio viene mantenuto sia a livello delle ontologie che a livello della base di conoscenza. Tecnicamente significa che le topic map definite per ogni nodo non saranno fuse tra loro, ogni nodo offrirà il supporto per un singolo

tool linguistico indipendente.

8.2 Annotazione in Gate

GATE è uno strumento di annotazione che contiene la funzionalità per l'interfaciamento con tutti i software di linguaggio naturale. Lavora con le risorse. Identifica due tipi di risorse: le Risorse per il Linguaggio e le Risorse per il Processamento.

Risorse per il Linguaggio (LR): si riferiscono a risorse contenenti solo dati come i lessici, i corpora, i thesauri o le ontologie.

Risorse per il Processamento (PR): si riferiscono a risorse la cui caratteristica primaria è quella programmatica o algoritmica, come i generatori di lemmi, traduttori, parser, riconoscitori di fasi. Queste risorse possono essere caricate in GATE e salvate insieme in Data Store. Le PR possono essere accorpate in una pipeline che forma un'applicazione in GATE.

Per annotare un file in GATE deve per prima cosa caricarsi il file da annotare e caricare successivamente le risorse. Per caricare le risorse vi sono due modi, sia selezionando le appropriate risorse dal menù file sia cliccando con il pulsante destro su risorse nell'interfaccia di GATE e selezionando l'appropriata risorsa dal menù a scomparsa che appare. Per aprire un file si deve selezionare 'New Language Resource' e poi 'GATE Document'. Apparirà una finestra in cui possono essere immessi i parametri per il documento. L'unico parametro essenziale è il luogo dove si trova il file chiamato, utilizzando la terminologia GATE, `SourceUrl`. Selezionando un file presente questo verrà caricato come LR in GATE. Cliccando due volte sul file questo verrà caricato nella finestra principale di GATE e si aprirà l'Editor per le annotazioni. Per vedere le annotazioni si deve cliccare sulle Annotazioni e sugli insiemi di annotazione nella barra dell'editor di annotazione. Si apre un frame in cui sono presentate tutte le annotazioni inserite nel file. Esiste un insieme di annotazione

chiamato 'Original markup annotations' che contiene tutti i tag che sono trovati nel file.

La fase di annotazione si è così svolta:

1. Si sono selezionati n annotatori.
2. Si sono selezionati i documenti che dovevano essere annotati che sono andati a formare il *corpus*.
3. Valutazione dei concetti e delle relazioni che dovranno essere utilizzate
 - (a) Discussione sulla parte di ontologia che si sarebbe utilizzata nell'operazione di annotazione.
 - (b) Selezione casuale di 10 documenti (non vuoti) dal corpus.
 - (c) Selezione di uno dei 10 documenti e annotazione collaborativa tra gli annotatori di questo documento.
 - (d) Annotazione degli altri 9 documenti da parte di ogni annotatore.
 - (e) Calcolo, con *annotation diff* di GATE, delle differenze tra le annotazioni.
 - (f) Discussione delle differenze.
4. Distribuzione del carico di lavoro
 - (a) Rimuovere il 10% del corpus
 - (b) Dividere il rimanente 90% in 'n' porzioni.
 - (c) Inviare ad ogni annotatore $1/n$ del 90% e del 10% lasciato a parte. L'annotatore deve lavorare in primo luogo sul $1/n$ del 90% e secondariamente sul 10% comune.
5. Valutazione del lavoro fatto.

8.2.1 Selezione del sottoinsieme di pagine da Annotare e delle Ontologie scelte

Sono state scelte le pagine dei siti selezionati in base alla loro ricchezza di contenuto e alla loro varietà di struttura. Sono state preferite le pagine personali del corpo docente e le pagine dei corsi in quanto differenti fra loro e che offrivano una gran quantità di dati da annotare. In modo da rendere il processo di annotazione fattibile oltre a selezionare un sottoinsieme di pagine dai siti selezionati si è scelto anche di utilizzare un sottoinsieme delle ontologie inserite. L'ontologia finale annotata contiene:

- person / student - professor - ...
- course
- time measures
- research interests/projects
- publications
- affiliations
- contact info

8.2.2 Basename introdotti

Si sono introdotti dei baseName uniformi per identificare univocamente un'istanza di un concetto che ha più modi di essere nominato. Se ne sono introdotti per le università, per le facoltà la cui regola di creazione è: f(prime 3 lettere del nome della facoltà)[...][prime 3 lettere ultimo nome della facoltà]. Esempio: Facoltà Scienze della Formazione = fscifor. Si sono create regole analoghe per i nomi dei corsi, delle persone, dei gruppi di ricerca, dei dipartimenti,

8.2.3 Tuning della configurazione del software GATE

Durante l'utilizzo del software GATE si sono evidenziati numerose mancanze. Inizialmente si è notata la mancanza di varie relazioni. La comprensione della modalità di inserimento di nuove relazioni è stata tutt'altro che banale. Alla fine si è arrivati ad una procedura per l'inserimento delle relazioni mancanti. Le relazioni inserite sono descritte nella Tabella 8.1

Nome in GATE	Nome della Relazione	Primo argomento	Secondo Argomento
relTeachingOf	teacherOf	Faculty	Course
relTeachingAssistantOf	teachingAssistantOf	TeachingAssistant	Course
relActivityStart	activityStart	Activity	DATE
relActivityEnd	activityEnd	Activity	DATE
relHasProgram	hasProgram	Course	Document
relAssignCredits	assignCredits	Course	CREDIT
relHasQuestionTime	hasQuestionTime	Faculty	RepeatedTimeInterval
relHasOpeningTime	hasOpeningTime	SupportOffice	RepeatedTimeInterval
relCourseBook	CourseBook	Course	Book
relOrgEmailAddress	emailAddress	Organization	EMAILADDRESS

Tabella 8.1: Relazioni inserite in GATE

Per inserire le relazioni si sono prima dovute creare, un esempio è il seguente:

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relTeacherOf" >
<complexType>
<attribute name="Faculty" type="string" use="optional" />
<attribute name="Course" type="string" use="optional" />
</complexType>
</element>
</schema>
```

Ogni relazione creata è stata inserita in un file XML con il proprio nome e poi si è dovuto aggiornare il file principale di GATE che gestisce le relazioni il file creole.xml. Dopo l'inserimento di codice nel file creole si sono dovute far apparire le nuove relazioni nella tendina di annotazione. Quindi i file xml creati sono stati inseriti in due cartelle:

- Gate.jar/Gate/resources/moses/schema
- File://cartella installazione Gate/moses/schema

Successivamente di è modificato il file creole.xml che si trova nella cartella 'Gate.jar/Gate/resources/creole' nel quale è stata inserita la seguente istruzione nella sezione 'moses schemata':

```
<AUTOINSTANCE>  
<PARAM NAME ="xmlFileUrl" VALUE =  
"gate:/moses/schema/relHasProgram.xml"/>  
</AUTOINSTANCE>
```

Infine si è modificato il file Italian_ontology.daml nelle seguenti cartelle:

- Gate.jar/Gate/resources/moses/ontology
- File://cartella installazione Gate/moses/ontology

inserendo in coda la seguente istruzione :

```
<rdfs:Class rdf:about='file:/D:/Users/Matthijs/Projects/Moses  
/Gate/v2.0/Installatie/italian/moses/ontology/  
italian_ontology.daml#relHasProgram'  
rdfs:label='relHasProgram'  
rdfs:comment=''>
```

```
<rdfs:subClassOf>file:/D:/Users/Matthijs/Projects/Moses  
/Gate/v2.0/Installatie/italian/moses/ontology  
/italian_ontology.daml#relations</rdfs:subClassOf>  
</rdfs:Class>
```

A questo punto si è inserita la nuova relazione nel menù a tendina che viene usato per creare una nuova annotazione. Si è trovato un modo per annotare in maniera automatica alcuni concetti. Per far ciò è stato necessario inserire in GATE delle regole attraverso i suoi file di configurazione. Si sono eseguiti i seguenti passi:

- caricare ontology daml file nel language resource
- Creare un ontogazetteer
- creare un jape transducer
- creare (in language resource) un nuovo corpus
- popolarlo scegliendo la dir dove sono i file html da annotare (ricordarsi di mettere l'estensione perché in caso contrario carica tutte le tipologie di file).
- andare su corpus pipeline e caricare a destra (ovvero premendo l'icona a forma di freccia che fa passare dal frame di sinistra a quello di destra) l'ontogazetteer
- fare la stessa cosa per il jape transducer.
- selezionare il corpus creato nel punto d
- premere su 'Run'
- dopo la fine dell'operazione aprire uno dei file e vedere se ha inserito ciò che si desiderava nelle default annotation

Si devono inserire nel file mappings.def tutti i nomi dei file .lst che contengono le liste di istanze di concetti. Si deve anche verificare che il tutto (i concetti interessati) sia in italian.jape se i concetti mancano si devono creare le regole in jape appropriate. Nel file mappings.def la sintassi di riconoscimento è la seguente:

```
university.lst:gate:/moses/ontology/italian_ontology.daml:University
```

Con university.lst che è il nome del file contenente la lista delle istanze del concetto una per riga e University che è la label che si è data nel file .daml al concetto university. I file mappings.def e gli eventuali lists.def e .lst modificati vanno inseriti nel gate.jar nella cartella moses/ontogazetteer e nella cartella moses/ontogazetteer.

Capitolo 9

Conclusioni

9.1 Semantic Web

Lo sviluppo del Semantic Web ha interessato diverse aree del web tradizionale: i siti di commercio elettronico, in cui le ontologie facilitano le comunicazioni tra venditore e cliente, permettono l'integrazione dei mercati, consentendo che le descrizioni siano utilizzate in diversi 'punti vendita';

i motori di ricerca, a cui le ontologie permettono di effettuare una ricerca basata non solo su parole chiave, ma anche sui contenuti semantici inclusi nella pagina;

i Web Services, che sono rappresentati dalle ontologie, in maniera tale che la loro descrizione semantica sia interpretata da un agente intelligente per il reperimento immediato del servizio.

In ogni caso, gli sviluppi e le innovazioni possono nascere in due tipi di situazioni diverse: da una parte, la sicurezza di un gruppo piccolo, che agisce indipendentemente dagli altri, in cui i cambiamenti possono avvenire in modo rapido; dall'altra, la necessità di essere parte di una comunità più ampia, che nonostante la lentezza delle evoluzioni e la gran quantità di comunicazione necessaria per il coordinamento, garantisce la creazione di materiale facilmente condivisibile.

Per lo sviluppo di un progetto si parte con un'idea, la si sviluppa 'in piccolo', per poi inserirsi in un contesto più ampio. L'espansione del Semantic Web potrebbe seguire

questa strada.

Nel prossimo futuro, si prevede che ‘ogni compagnia, università, agenzia di governo vorrà che le proprie risorse siano collegate ad un contesto ontologico, poiché molti potenti tools saranno disponibili per usarli’ (J. Hendler). Le applicazioni potranno scambiarsi le informazioni e i programmi saranno in grado di raccogliere ed elaborare i contenuti del Web. Il Semantic Web, facendo riferimento ad ogni entità con un URI, permetterà a ciascuno di descrivere nuovi concetti, eventualmente basandosi su altri già esistenti e, l’utilizzo di linguaggi standard, consentirà alle informazioni individuali di essere collegate al resto del Web. La struttura che si verrà a formare, imperniata sulla conoscenza, fornirà una serie di strumenti nuovi in grado di comunicare, lavorare ed imparare nel medesimo istante.

9.2 Progetto svolto

La partecipazione al progetto MOSES, oltre ad aver materializzato tutto quello che ho studiato riguardo il web semantico, è stata anche un’esperienza formativa di notevole impatto in quanto mi ha permesso di collaborare in un progetto europeo e comprendere le dinamiche che regolano un lavoro così imponente. Proprio l’aver partecipato ad un tale progetto mi ha fatto comprendere che la strada verso un web semantico nel senso di Berners Lee è ancora lunga, ma non impossibile da raggiungere. Strumenti formidabili quali le ontologie, le annotazioni e tutti i linguaggi formali che sono stati appositamente creati hanno dato, e daranno, un impulso sempre più forte per la realizzazione di una struttura basata sulla conoscenza e non più sulla mera rappresentazione.

Appendici

Appendice A: Ontologies

This appendix lists the two kinds of ontologies that have been integrated in the Semantic Structure: + Reference DAML ontologies and their transcription in OWL + Extensions for Italian University and Danish University Given their size, the ontologies themselves are not included in the document, but are available on-line from: <http://www.mondeca.com/owl/>

A very practical way to browse the ontologies is to use the 'hyperdaml navigation', making all RDF elements fully addressable. For example, to navigate the Danish Ontology, use:

http://www.daml.org/cgi-bin/hyperdaml?http://www.mondeca.com/owl/univ_dan1_1.xml

GENERAL ONTOLOGY The General Ontology contains elements useful across many domains: classes such as Person, Organization, Activity, Event ... properties like Event Location, Affiliation ... Those elements are further imported and refined in more specific ontologies (Document, University).

Sources:

- <http://www.cs.umd.edu/projects/plus/DAML/onts/general1.0.daml>

Imports:

- <http://www.cs.umd.edu/projects/plus/DAML/onts/base1.0.daml>

OWL Transcription:

- http://www.mondeca.com/owl/general1_1.xml

Imports:

- http://www.mondeca.com/owl/base1_1.xml

DOCUMENT ONTOLOGY The Document Ontology contains classes and properties relevant to the domain of documentation, publication and general communication. Sources:

- <http://www.cs.umd.edu/projects/plus/DAML/onts/docmnt1.0.daml>

Imports: <http://www.cs.umd.edu/projects/plus/DAML/onts/general1.0.daml>

- <http://www.cs.umd.edu/projects/plus/DAML/onts/base1.0.daml>

OWL Transcription:

- http://www.mondeca.com/owl/document1_1.xml

Imports:

- http://www.mondeca.com/owl/general1_1.xml

- http://www.mondeca.com/owl/base1_1.xml

UNIVERSITY ONTOLOGY The University Ontology is an extension of both General and Document ontologies including classes and properties specific to the academic world.

Sources:

- <http://www.cs.umd.edu/projects/plus/DAML/onts/univ1.0.daml>

Imports:

- <http://www.cs.umd.edu/projects/plus/DAML/onts/docmnt1.0.daml>

- <http://www.cs.umd.edu/projects/plus/DAML/onts/general1.0.daml>

- <http://www.cs.umd.edu/projects/plus/DAML/onts/base1.0.daml>

OWL Transcription:

- http://www.mondeca.com/owl/univ1_1.xml

Imports:

- http://www.mondeca.com/owl/document1_1.xml
- http://www.mondeca.com/owl/general1_1.xml
- http://www.mondeca.com/owl/base1_1.xml

ITALIAN UNIVERSITY ONTOLOGY The Italian University ontology is an extension of University Ontology, including classes and properties specific to Italian academic organization.

Sources: This ontology has been developed specifically for MOSES project, and defined in Deliverable 2.3 of Moses Project.

OWL Transcription:

- http://www.mondeca.com/owl/univ_ita1_1.xml

Imports:

- http://www.mondeca.com/owl/univ1_1.xml

DANISH UNIVERSITY ONTOLOGY The Danish University ontology is an extension of University Ontology, including classes and properties specific to Danish academic organization.

Sources: This ontology has been developed specifically for MOSES project, and defined in Deliverable 2.3 of MOSES Project. OWL Transcription:

- http://www.mondeca.com/owl/univ_dan1_1.xml

Imports:

- http://www.mondeca.com/owl/univ1_1.xml

Appendice B: Question Card

Domain:	Didactics
Application:	Courses
Sub-application:	Content
Question:	Quali sono i corsi che contengono l'argomento X ? / Which are the courses that contain argument X ?
User:	Student
Format of the response:	List of items (course name). More precisely, it could be a table associating names of the course with URLs of the courses pages.
Site title(s):	Roma TRE
Site type(s):	Central site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole.
Structure of the source information - Description:	The information is in various PDF documents.
Structure of the source information - Access Path(s):	A link ('DIDATTICA e STUDENTI') on the home page leads to a page where a table show all the faculty of RomaTRE. Clicking on a faculty name user can access the faculty page and the 'Ordine degli studi' PDF document that contains the programs of all the courses of the faculty.
Technology used:	ASP
Extraction process - Classification:	Navigation with backtracking
Extraction process - Description:	The user follow the link ('DIDATTICA e STUDENTI') on the home page to reach a new page where a table show all the faculty of RomaTRE. By clicking on a faculty name the user can access a faculty page where there are different information related to the faculty. The link 'Ordine degli studi' leads to a PDF document that contains the programs of all the courses of the faculty. In this document the user has to find the argument X and verify the course in which it is taught. This operation has to be repeated for every 'Ordine degli studi' in the various faculty pages.

Tabella 9.1: Question card sul dominio Didactics

Domain:	Didactics
Application:	Courses
Sub-application:	Content
Question:	Qual'è il professore che insegna un corso in cui c'è l'argomento X ? / What's the professor who teaches a course in which there's argument X?
User:	Student
Format of the response:	List of items (professor name). More precisely, it could be a table associating names of the professor and his email and the course name.
Site title(s):	Roma TRE
Site type(s):	Central site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole
Structure of the source information - Description:	The information is in various PDF documents.
Structure of the source information - Access Path(s):	A link ('DIDATTICA e STUDENTI') on the home page leads to a page where a table show all the faculty of RomaTRE. Clicking on a faculty name user can access the faculty page and the 'Ordine degli studi' PDF document that contains the programs of all the courses of the faculty.
Technology used:	ASP
Extraction process - Classification:	Navigation with backtracking
Extraction process - Description:	The user follow the link ('DIDATTICA e STUDENTI') on the home page to reach a new page where a table show all the faculty of RomaTRE. By clicking on a faculty name the user can access a faculty page where there are different information related to the faculty. The link 'Ordine degli studi' leads to a PDF document that contains the programs of all the courses of the faculty. In this document the user has to find the argument X and retrieve the course name in which it is taught At the beginning of the course section can be found the professor name. This operation has to be repeated for every 'Ordine degli studi' in the various faculty pages.

Tabella 9.2: Question card sul dominio Didactics

Domain:	People
Application:	Teaching Staff
Sub-application:	Contact Info
Question:	A che ora riceve il professore X ? / At what time are professor X's meeting hours?
User:	Student
Format of the response:	List of Tuple of Data. (day of week, hour).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Rubrica') on the home page leads to a form interface. The response to the form is a table with some information (including names and phone numbers) of the selected professors. A link in this page leads on a more detailed page with the url of the professor' web site.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a form that allows to search for professors (and staff members) given the name (or part of it) and provides a table (one row for each professor, with name, affiliation and phone number) as a response. The user have to click the desired professor's name to access a new page where more information, like the personal page url, can be found. Usually the meetings hours information can be found in the professor personal page.

Tabella 9.3: Question card sul dominio People

Domain:	People
Application:	Teaching Staff
Sub-application:	Contact Info
Question:	Quali risorse (informazioni aggiuntive) sono disponibili per il corso del professor X? /What resources are available for professor X's course?
User:	Student
Format of the response:	List of URL (about pages or external documents).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Rubrica') on the home page leads to a form interface. The response to the form is a table with some information (including names and phone numbers) of the selected professors. A link in this page leads on a more detailed page with the url of the professor' web site. Inside professor web site can be found the researched resources.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a form that allows to search for professors (and staff members) given the name (or part of it) and provides a table (one row for each professor, with name, affiliation and phone number) as a response. The user have to click the desired professor's name to access a new page where more information, like the personal page url, can be found. Usually the resources are linked from the professor personal page.

Tabella 9.4: Question card sul dominio People

Domain:	People
Application:	Teaching Staff
Sub-application:	Activities
Question:	Quali corsi insegna il Professore X? / Which courses does professor X teach?
User:	Student
Format of the response:	List of Page (URL about courses pages or external documents).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link (Rubrica) on the home page leads to a form interface. The response to the form is a table with some information (including names and phone numbers) of the selected professors. A link in this page leads on a more detailed page with the professor's department name. In the department site, usually, a link to the 'Didattica' section allow to know the courses who have been held by the professor X. E.g. for the DIA: in the home page a link 'Didattica' leads in a new page where is present a link 'Informazioni generali sulla didattica'. A new link http://www.dia.uniroma3.it/didattica/ lead into a page where is available the 'CORSI, DOCENTI, ESAMI' section. The link 'docent' of this section leads to an ASP page where the professor X name can be clicked to access the list of course who have been held by him.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a form that allows to search for professors (and staff members) given the name (or part of it) and provides a table (one row for each professor, with name, affiliation and phone number) as a response. The user have to click the desired professor's name to access a new page where more information, like the personal page url, can be found. The next step is to find the department name where the professor X teaches. This information is also available in the previously page. Once opened the department site a certain number of links must be follow to reach a complete list of professors of that

Domain:	People
Application:	Teaching Staff
Sub-application:	Activities
Question:	A quali ricerche il Professore X è interessato? / What type of research is professor X involved in?
User:	Student
Format of the response:	List of item (research topic title and a NL description).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Ricerca Scientifica') on the home page leads to page with the list of all department. Clicking on a department name the user can access a page containing a list of the research performed in those department. Then, clicking on each research title, the user can check if the professor X is, or not, involved in those research. Otherwise the user can search this information in the professor personal web page.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a list of department. For each department name the user has to check the research topic and the name of the professor involved in those.

Tabella 9.6: Question card sul dominio People

Domain:	People
Application:	Non Teaching Staff
Sub-application:	Contact info
Question:	Qual'è la stanza della persona X? / Where is person X's room?
User:	Student
Format of the response:	List of items (room numbers). More precisely, it could be a table associating names and affiliations with room numbers (useful in case of multiple professors satisfying the selection condition).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information available in the site in a unrelated way.
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Rubrica') on the home page leads to a form interface. The response to the form is a table with some information (including names and phone numbers) of the selected professors. A link in this page leads on a more detailed page with the professor's department name. In the department site, usually, a link 'Persone' ('People') allow to know the room number of the professor X. E.g. for the DIA: in the home page a link 'Persone' leads in a new page where is present table that contains name, email, room number and telephone number of professor X.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a form that allows to search for professors (and staff members) given the name (or part of it) and provides a table (one row for each professor, with name, affiliation and phone number) as a response. The user have to click the desired professor's name to access a new page where more information, like the personal page url, can be found. The next step is to find the department name where the professor X teaches. This information is also available in the previously page. Once opened the department site a certain number of links must be follow to reach a complete list of people related to the department. For each person is available its room number.

Domain:	Didactics
Application:	Courses
Sub-application:	Organization
Question:	Quale facoltà/dipartimento/Istituto offre il corso X? / Which faculty/department/institute offers the course X?
User:	Student
Format of the response:	List of item (names of the faculty/department/institute(s) that offers course X).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole.
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Didattica e Studenti') on the home page leads to a page that contains various information. The link 'Elenco Completo', under the section 'INSEGNAMENTI', allow to access the complete list of courses divided into faculty.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	The user selects a link on the home page of the site, that leads, through an intermediate page, to a list of courses available for each faculty.

Tabella 9.8: Question card sul dominio Didactics

Domain:	Didactics
Application:	Courses
Sub-application:	Organization
Question:	Quali corsi sono offerti al Dipartimento X? / What courses are offered at department X?
User:	Student
Format of the response:	List of item (names of the courses available at department X).
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole.
Structure of the source information - Description:	The information is in a page with other items.
Structure of the source information - Access Path(s):	A link ('Didattica e Studenti') on the home page leads to a page that contains various information. The link 'Dipartimenti', under the section 'SUPPORTO ALLA DIDATTICA', allow to access the complete list of department of RomaTRE. Clicking on every department name a brief card report the department URL. Typically, in the departments site, a link 'DIDATTICA' presents the list of all courses available in the department.
Technology used:	ASP
Extraction process - Classification:	Navigation with backtracking
Extraction process - Description:	The user selects a link on the home page of the site, that leads, through an intermediate page, to a list of department of RomaTRE. The user have to access the wished department site to examine the list of offered courses.

Tabella 9.9: Question card sul dominio Didactics

Domain:	Didactics
Application:	Master
Sub-application:	Subject
Question:	Quali master sono disponibili sull'argomento X? / Which masters are available on subject X?
User:	Student
Format of the response:	List of items (master title) with link to the related page.
Site title(s):	Roma TRE
Site type(s):	Central site + Satellite Site
Organization type(s):	University site
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information available in the site and correlated but not aggregated.
Structure of the source information - Description:	The information is in a free text page.
Structure of the source information - Access Path(s):	A link ('Master') on the home page leads to a page containing different links. Information about the Master available in the university can be found in some of the page linked from this page.
Technology used:	HTML
Extraction process - Classification:	Navigation with backtracking
Extraction process - Description:	The user selects a link on the home page of the site, that leads to a page that allows to check if a master on subject X is available.

Tabella 9.10: Question card sul dominio Didactics

Domain:	Didactics
Application:	Courses
Sub-application:	Lesson
Question:	Quando comincia il corso X? / When does course X begin?
User:	Student
Format of the response:	Item (Date).
Site title(s):	DIA
Site type(s):	Satellite
Organization type(s):	Department
Url(s):	http://www.dia.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole.
Structure of the source information - Description:	The information is in a table.
Structure of the source information - Access Path(s):	Following various links from the department home page the user will find the response embedded in a table. The links to be followed from the home page are : ‘Didattica’: http://web.dia.uniroma3.it/teaching/ ; ‘informazioni sugli insegnamenti del dipartimento’: http://www.dia.uniroma3.it/CCSIngInf/ ‘Orario Lezioni’: http://www.dia.uniroma3.it/CCSIngInf/orariCalendari/orario2001.html ; ‘Orario Lezioni’: http://www.dia.uniroma3.it/CCSIngInf/orariCalendari/calendario-20030214.pdf In the page there is a textual description for every program course.
Technology used:	PDF
Extraction process - Classification:	One-way navigation
Extraction process - Description:	A link (‘Didattica’) on the home page leads to a page where are listed links to various didactic services and from there selecting (‘Informazioni sugli insegnamenti del dipartimento’). the user, will find, under the section (‘Informazioni Generali’) a link, from which he’ll arrive to a page containing a textual description for the information he was looking for.

Tabella 9.11: Question card sul dominio Didactics

Domain:	Didactics
Application:	Courses
Sub-application:	Organization
Question:	Chi insegna l'argomento X all'interno di un corso? / Who teaches subject X within a course?
User:	Student
Format of the response:	Item (Professor's name).
Site title(s):	DIA
Site type(s):	Satellite
Organization type(s):	Department
Url(s):	http://www.dia.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information available in the site(s) in a unrelated way.
Structure of the source information - Description:	The information is in a description page.
Structure of the source information - Access Path(s):	Following various links from the department home page the user will find the response embedded in a table. The links to be followed from the home page are : 'Didattica': http://web.dia.uniroma3.it/teaching/ ; 'Programmi': http://www.dia.uniroma3.it/CCSIngInf/programmi.html ; There user finds what courses concerns subject X. 'Didattica': http://web.dia.uniroma3.it/teaching/ ; 'docenti': http://www.dia.uniroma3.it/CCSIngInf/ There the user finds who teaches the interested course.
Technology used:	ASP
Extraction process - Classification:	Navigation with backtracking
Extraction process - Description:	First the user has to discover what course concerns subject X and then has to find who teaches that course. To do this he has to start from a link ('Didattica') on the home page leads to a page where are listed links to various didactic services and from there selecting ('Informazioni sugli insegnamenti del dipartimento'). the user, will find a link ('programmi') where he has to search for the subject X. After he found it he has to return to the previous page and then select the link ('docenti') where a link list present all the professors and the user has to look at all the description pages to find the interested course.

Tabella 9.12: Question card sul dominio Didactics

Domain:	Organization
Application:	Accessibility
Sub-application:	Contact Info
Question:	Quando è aperto l'ufficio X? / When is the office X open?
User:	Student
Format of the response:	Item (Date). More precisely, it could be a tuple of data associating opening hours and how to reach the office.
Site title(s):	Roma Tre
Site type(s):	Central site
Organization type(s):	University
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information directly provided by the site(s) as a whole.
Structure of the source information - Description:	The information is in various pages all reachable from a single page where are listed all the office links.
Structure of the source information - Access Path(s):	Following various links from the home page the user will find the response embedded in a describing page. The links to be followed from the home page are : 'Ateneo': http://www.uniroma3.it/ateneo/index_ateneo.asp ; In the page there are various links to the presentation of all the offices course.
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	A link ('Ateneo') on the home page leads to a page where are listed the links to the presentation page of every single office. The user, selecting the right link, will find the textual description and embedded in he will locate the information he was looking for.

Tabella 9.13: Question card sul dominio Organization

Domain:	Organization
Application:	Administration
Sub-application:	Deadline
Question:	Quando scade la presentazione della domanda di iscrizione? / When does the presentation of the application for the scholarship expire?
User:	Student
Format of the response:	Item (Date). More precisely, it could be a tuple of data associating period with others information like what type of scholarship.
Site title(s):	Roma Tre
Site type(s):	Central site
Organization type(s):	University
Url(s):	http://www.uniroma3.it
Answer:	
Structure of the source information - Classification:	Information available in the sites and correlated but not aggregated .
Structure of the source information - Description:	The information is in a page where are listed all the link to the active scholarships. In that page there are also lot of other information and are all merged.
Structure of the source information - Access Path(s):	Following various links from the home page the user will find the response.. The links to be followed from the home page are : ‘... archivio’: http://www.uniroma3.it/inevidenza/archivio_inevidenza.asp ; From here you the user has to select the desiderated argument
Technology used:	ASP
Extraction process - Classification:	One-way navigation
Extraction process - Description:	A link (‘... archivio’) inside the box named (‘in evidenza’) on the home page leads to a page where are listed all the important events offered by the University. From this page following the various link about scholarship the user reaches the page where he has to specify, choosing the right link, the type of inscription and eventually he finds the information searched. There’s another way to collect the information required. From the home page following the link (‘Informazioni e Servizi’) and then selecting one of the link under the section (‘Borse, Finanziamenti & Stages’) the user will find the page where are link to documents about the scholarship selected.

Tabella 9.14: Question card sul dominio Organization

Appendice C: Creazione file XTM tramite script

Si riportano sezioni fondamentali dello script creato per estrarre i dati dal database del Dipartimento di Informatica ed Automazione e trasformarli in istanze delle ontologie formattate in file xml con sintassi XTM per poter essere inserite nello strumento di Mondeca ITM.

```
<?PHP
```

```
    // Creazione dell'intestazione del file xml
```

```
/* echo '
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<topicMap xmlns:xlink="http://www.w3.org/1999/xlink">
```

```
<!-- topics: -->
```

```
';
```

```
*/
```

Connessione al database PostGres. Con specifica della prima query.

```
$connessione = pg_connect("host=localhost
```

```
dbname=DIA
```

```
user=mario
```

```
password=mario")
```

```
or die("Impossibile stabilire una connessione.");
```

```
$sql = "SELECT codicepersona as person_id, nome, cognome, descrizionequalifica,
itmurl, homepage, email, telefono, ricevimento, indirizzo
FROM persone, qualifica
WHERE qualifica.codicequalifica = persone.codicequalifica
ORDER BY persone.codicequalifica, cognome, nome";
```

```
$sql_risultato = pg_exec($connessione, $sql)
or die("Impossibile eseguire la query.");
```

```
$num = pg_numrows($sql_risultato);
```

```
$i = 0;
```

```
echo " ";
```

Fetching del risultato.

```
while ($i < $num)
```

```
{
```

```
$riga = pg_fetch_array($sql_risultato, $i);
```

```
$person_id = $riga["person_id"];

```

```
$nome = $riga["nome"];

```

```
$cognome = $riga["cognome"];

```

```
$descrizionequalifica = $riga["descrizionequalifica"];

```



```

$itmurl = $riga["itmurl"];
$homepage = $riga["homepage"];
$email = $riga["email"];
$telefono = $riga["telefono"];
$ricevimento = $riga["ricevimento"];
$indirizzo = $riga["indirizzo"];
echo " ";

```

Creazione dell'istanza dell'ontologia.

```

echo "
<!-- topic: $nome $cognome, $descrizionequalifica -->
<topic id='personid_{$person_id}'>
<instanceOf>
<subjectIndicatorRef xlink:href='{$itmurl}'/>
</instanceOf>
<subjectIdentity>
<subjectIndicatorRef xlink:href='http://ricerca.dia.uniroma3.it
/persona/schedaPersona.php?id={$person_id}'/>
</subjectIdentity>
<baseName>
<baseNameString>{$nome $cognome}</baseNameString>
</baseName>
<baseName>
<baseNameString>{$cognome $nome}</baseNameString>
</baseName>

```

</topic>

<!-- topic: \$nome \$cognome personal homepage -->

<topic id='personid_\$person_id'."_homepage'>

<instanceOf>

<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#WebPage' />

</instanceOf>

<subjectIdentity>

<resourceRef xlink:href='\$homepage' />

</subjectIdentity>

<baseName>

<baseNameString>\$nome \$cognome homepage</baseNameString>

</baseName>

</topic>

<!-- topic: \$nome \$cognome personal email -->

<topic id='personid_\$person_id'."_email'>

<instanceOf>

<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_ita1_1.xml#personalEmailAddress' />

</instanceOf>

<subjectIdentity>

<resourceRef xlink:href='\$email' />

```

</subjectIdentity>

<baseName>

<baseNameString>$nome $cognome email</baseNameString>

</baseName>

</topic>

<!-- topic: $nome $cognome work phone -->
<topic id='personid_$person_id"."_phone'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#workPhone' />
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='$telefono' />
</subjectIdentity>
<baseName>
<baseNameString>$nome $cognome phone</baseNameString>
</baseName>
</topic>

<!-- topic: $nome $cognome office address -->
<topic id='personid_$person_id"."_address'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#workAddress' />
</instanceOf>

```

```

<subjectIdentity>
<resourceRef xlink:href='$indirizzo' />
</subjectIdentity>
<baseName>
<baseNameString>$nome $cognome office address</baseNameString>
</baseName>
</topic>

<!-- topic: $nome $cognome meeting hours -->
<topic id='personid_$person_id"."_ meeting_hours'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#Time' />
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='$ricevimento' />
</subjectIdentity>
<baseName>
<baseNameString>$nome $cognome meeting hours</baseNameString>
</baseName>
</topic>

";
$i++;
}

```

Esecuzione della seconda query per estrarre le informazioni sui Corsi e relativo fetching del risultato.

```
$sql = "SELECT DISTINCT canali.codicecorso, aa, denominazione
FROM canali, corsi
WHERE canali.codicecorso = corsi.codicecorso
ORDER BY codicecorso";
```

```
$sql_risultato = pg_exec($connessione, $sql)
or die("Impossibile eseguire la query.");
```

```
$num = pg_numrows($sql_risultato);
```

```
$i = 0;
```

```
while ($i < $num)
```

```
{
```

```
$riga = pg_fetch_array($sql_risultato, $i);
```

```
$course_id = $riga["codicecorso"];

```

```
$denominazione = $riga["denominazione"];

```

```
$aa = $riga["aa"];

```

```
echo "
```

```

<!-- topic: Course $denominazione -->
<topic id='courseid_$course_id'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ1_1.xml#Course' />
</instanceOf>
<subjectIdentity>
<subjectIndicatorRef xlink:href='http://didattica.dia.uniroma3.it
/corsi/schedaEdizione.php?cc=$course_id&AA_=$aa/' />
</subjectIdentity>
<baseName>
<baseNameString>$denominazione</baseNameString>
</baseName>
</topic>

";
$i++;

}

```

Creazione delle istanze per la definizione del tempo.

```

for ($anno=2000;$anno<2011;$anno++)
{
echo "
<!-- topic: the year $anno expressed as a 'date' -->

```

```

<topic id='anno_$(anno)'\>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#dato'/>
</instanceOf>
<baseName>
<baseNameString>$(anno)</baseNameString>
</baseName>
</topic>

```

```

";
}

```

```

echo "

```

```

<!-- topic: the weekday lunedì (monday) -->
<topic id='Giorno_settimana_lunedì'\>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Ugedag'/>
</instanceOf>
<baseName>
<baseNameString>Lunedì</baseNameString>
</baseName>
</topic>

```

```

<!-- topic: the weekday martedì (tuesday) -->
<topic id='Giorno_settimana_martedì'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Ugedag' />
</instanceOf>
<baseName>
<baseNameString>Martedì</baseNameString>
</baseName>
</topic>
<!-- topic: the weekday mercoledì (wednesday) -->
<topic id='Giorno_settimana_mercoledì'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Ugedag' />
</instanceOf>
<baseName>
<baseNameString>Mercoledì</baseNameString>
</baseName>
</topic>
<!-- topic: the weekday giovedì (thursday) -->
<topic id='Giorno_settimana_giovedì'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Ugedag' />
</instanceOf>

```



```

<baseName>
<baseNameString>Giovedi</baseNameString>
</baseName>
</topic>
<!-- topic: the weekday venerdi (friday) -->
<topic id='Giorno_settimana_venerdi'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Ugedag' />
</instanceOf>
<baseName>
<baseNameString>Venerdi</baseNameString>
</baseName>
</topic>
";

for ($orario=8;$orario<20;$orario++)
{
if ($orario<12) $fase_giornata = "am";
else $fase_giornata = "pm";

echo "
<!-- topic: the hour $orario ($orario $fase_giornata) -->
<topic id='orario_$orario'>
<instanceOf>

```

```

<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan1_1.xml#Klokkesl?/>
</instanceOf>
<baseName>
<baseNameString>$orario</baseNameString>
</baseName>
</topic>

";
}

```

Query per prelevare nome cognome ed homepage

```

$sql = "SELECT codicepersona as person_id, nome, cognome, homepage
FROM persone
WHERE persone.homepage <> ''
ORDER BY cognome, nome";

$sql_risultato = pg_exec($connessione, $sql)
or die("Impossibile eseguire la query.");

$num = pg_numrows($sql_risultato);

$i = 0;

```

Definizione di istanze delle relazioni.

```

echo "
<!-- associations-->
<!-- association: Persona con propria personal homepage
(links the two topics 'Persona' and 'personal homepage')-->
";

while ($i < $num)
{
$riga = pg_fetch_array($sql_risultato, $i);

$person_id = $riga["person_id"];
$nome = $riga["nome"];
$cognome = $riga["cognome"];
$homepage = $riga["homepage"];
echo " ";

//Abbiamo usato il person id invece che il nome perché
//alcuni record il nome possiede uno spazio bianco alla
//fine che non ci permetteva di concatenare il nome con il cognome.
echo "

<association id='ass_{$cognome"."_"."$person_id"."_webpage}'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com

```

```

/owl/univ_dan_asowl_1_0.xml#PersonHjemmeside' />
</instanceOf>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan_asowl_1_0.xml#hjemmesideRole' />
</roleSpec>
<topicRef xlink:href='#personid_$person_id'."_homepage' />
</member>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href='http://www.mondeca.com/owl
/univ_dan_asowl_1_0.xml#genstandForRole' />
</roleSpec>
<topicRef xlink:href='#personid_$person_id' />
</member>
</association>

";

$i++;
}

```

Query per prelevare nome cognome e corso insegnato e relazioni associate

```

$sql = "SELECT DISTINCT persone.codicepersona as person_id, persone.nome,
persone.cognome, corsi.denominazione, corsi.codicecorso AS course_id

```

```
FROM persone, canali, corsi
WHERE canali.codicepersona=persone.codicepersona AND canali.codicecorso=
corsi.codicecorso
ORDER BY cognome, nome";
```

```
$sql_risultato = pg_exec($connessione, $sql)
or die("Impossibile eseguire la query.");
```

```
$num = pg_numrows($sql_risultato);
```

```
$i = 0;
```

```
echo "
```

```
<!-- associations-->
```

```
<!-- association: Persona con propria personal homepage
(links the two topics 'Persona' and 'personal homepage')-->
";
```

```
while ($i < $num)
```

```
{
```

```
$riga = pg_fetch_array($sql_risultato, $i);
```

```
$person_id = $riga["person_id"];

```

```
$nome = $riga["nome"];

```

```
$cognome = $riga["cognome"];

```

```

$denominazione = $riga["denominazione"];
$course_id = $riga["course_id"];

echo "

<!-- association: $nome $cognome teaches $denominazione course -->
<association id='personid_".$person_id."_teaches_courseid_".$course_id."'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com/owl
/univ_dan_asowl_1_0.xml#UnderviserP?>
</instanceOf>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href='http://www.mondeca.com/owl
/univ_dan_asowl_1_0.xml#underviserRole' />
</roleSpec>
<topicRef xlink:href='#personid_".$person_id."' />
</member>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href='http://www.mondeca.com/owl
/univ_dan_asowl_1_0.xml#kursusRole' />
</roleSpec>
<topicRef xlink:href='#courseid_".$course_id."' />
</member>

```

```
</association>
```

```
";
```

```
$i++;
```

```
}
```

```
/*
```

```
echo "
```

Chiusura del file

```
</topicMap>
```

```
";
```

```
*/
```

```
pg_freeresult($sql_risultato);
```

```
pg_close($connessione);
```

```
?>
```

Sezione del file xml creato.

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<topicMap xmlns:xlink="http://www.w3.org/1999/xlink">
```

```
<!-- topics: -->
```

```
<!-- topic: Paolo Atzeni, Professore ordinario -->
```

```
<topic id='personid_1'>
```

```
<instanceOf>
```

```
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_ita1_1.xml#ProfessoreOrdinario' />
```

```
</instanceOf>
```

```
<subjectIdentity>
```

```
<subjectIndicatorRef xlink:href='http://ricerca.dia.uniroma3.it
/persona/schedaPersona.php?id=1' />
```

```
</subjectIdentity>
```

```
<baseName>
```

```
<baseNameString>Paolo Atzeni</baseNameString>
```

```
</baseName>
```

```
<baseName>
```

```
<baseNameString>Atzeni Paolo</baseNameString>
```

```
</baseName>
```

```
</topic>
```

```
<!-- topic: Paolo Atzeni personal homepage -->
```

```
<topic id='personid_1_homepage'>
```

```
<instanceOf>
```

```
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#WebPage' />
```



```

</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='http://atzeni.dia.uniroma3.it'/>
</subjectIdentity>
<baseName>
<baseNameString>Paolo Atzeni homepage</baseNameString>
</baseName>
</topic>

```

```

<!-- topic: Paolo Atzeni personal email -->
<topic id='personid_1_email'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_ita1_1.xml#personalEmailAddress'/>
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='atzeni@dia.uniroma3.it'/>
</subjectIdentity>
<baseName>
<baseNameString>Paolo Atzeni email</baseNameString>
</baseName>
</topic>

```

```

<!-- topic: Paolo Atzeni work phone -->
<topic id='personid_1_phone'>

```

```

<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#workPhone' />
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='06 5517 3213' />
</subjectIdentity>
<baseName>
<baseNameString>Paolo Atzeni phone</baseNameString>
</baseName>
</topic>

<!-- topic: Paolo Atzeni office address -->
<topic id='personid_1_address'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#workAddress' />
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='' />
</subjectIdentity>
<baseName>
<baseNameString>Paolo Atzeni office address</baseNameString>
</baseName>
</topic>

```

```

<!-- topic: Paolo Atzeni meeting hours -->
<topic id='personid_1_meeting_hours'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/general1_1.xml#Time' />
</instanceOf>
<subjectIdentity>
<resourceRef xlink:href='' />
</subjectIdentity>
<baseName>
<baseNameString>Paolo Atzeni meeting hours</baseNameString>
</baseName>
</topic>

<association id='ass_Atzeni_1_webpage'>
<instanceOf>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan_asowl_1_0.xml#PersonHjemmeside' />
</instanceOf>
<member>
<roleSpec>
<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan_asowl_1_0.xml#hjemmesideRole' />
</roleSpec>
<topicRef xlink:href='#personid_1_homepage' />
</member>

```

<member>

<roleSpec>

<subjectIndicatorRef xlink:href='http://www.mondeca.com
/owl/univ_dan_asowl_1_0.xml#genstandForRole'/>

</roleSpec>

<topicRef xlink:href='#personid_1'/>

</member>

</association>

Appendice D: File configurazione GATE

Per introdurre le relazioni che sono state usate in GATE è stato necessario creare dei file xml che rispettassero una determinata sintassi utile al programma GATE per la generazione, all'interno delle sue librerie delle regole per l'annotazione di quelle relazioni.

Il file *relTeacherOf.xml* è relativo alla relazione di insegnamento.

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relTeacherOf" >
<complexType>
<attribute name="Faculty" type="string" use="optional" />
<attribute name="Course" type="string" use="optional" />
</complexType>
</element>
</schema>
```

Il file *relTeachingAssistantOf.xml* è relativo alla relazione di insegnamento di un Assistente.

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relTeachingAssistantOf" >
<complexType>
<attribute name="TeachingAssistant" type="string" use="optional" />
<attribute name="Course" type="string" use="optional" />
</complexType>
```

```
</element>
```

```
</schema>
```

Il file *relOrgEmailAddress.xml* è relativo alla relazione di indirizzo di posta elettronica con un'organizzazione.

```
<?xml version="1.0" ?>
```

```
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
```

```
<element name="relOrgEmailAddress" >
```

```
<complexType>
```

```
<attribute name="Organization" type="string" use="optional" />
```

```
<attribute name="EMAILADDRESS" type="string" use="optional" />
```

```
</complexType>
```

```
</element>
```

```
</schema>
```

Il file *relHasQuestionTime.xml* è relativo alla relazione di orario di ricevimento del personale docente.

```
<?xml version="1.0" ?>
```

```
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
```

```
<element name="relHasQuestionTime" >
```

```
<complexType>
```

```
<attribute name="Faculty" type="string" use="optional" />
```

```
<attribute name="RepeatedTimeInterval" type="string" use="optional" />
```

```
</complexType>
```

```
</element>
```

```
</schema>
```

Il file *relHasProgram.xml* è relativo alla relazione tra programma e corso.

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relHasProgram" >
<complexType>
<attribute name="Course" type="string" use="optional" />
<attribute name="Document" type="string" use="optional" />
</complexType>
</element>
</schema>
```

Il file *relHasOpeningTime.xml* è relativo alla relazione tra orario di apertura ed ufficio.

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relHasOpeningTime" >
<complexType>
<attribute name="SupportOffice" type="string" use="optional" />
<attribute name="RepeatedTimeInterval" type="string" use="optional" />
</complexType>
</element>
</schema>
```

Il file *relAssignCredits.xml* è relativo alla relazione tra crediti attribuiti e corsi.

```
<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relAssignCredits" >
<complexType>
```

```

<attribute name="Course" type="string" use="optional" />
<attribute name="CREDIT" type="string" use="optional" />
</complexType>
</element>
</schema>

```

Il file *relActivityStart.xml* è relativo alla relazione tra attività e data di inizio.

```

<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relActivityStart" >
<complexType>
<attribute name="Activity" type="string" use="optional" />
<attribute name="DATE" type="string" use="optional" />
</complexType>
</element>
</schema>

```

Il file *relActivityEnd.xml* è relativo alla relazione tra attività e data di fine.

```

<?xml version="1.0" ?>
<schema xmlns="http://www.w3.org/2000/10/XMLSchema">
<element name="relActivityEnd" >
<complexType>
<attribute name="Activity" type="string" use="optional" />
<attribute name="DATE" type="string" use="optional" />
</complexType>
</element>
</schema>

```


Una volta sviluppati i file contenenti le relazioni si è andato a modificare uno dei file principali per la configurazione di GATE, ovvero il file

`creole.xml`

. In questo file si è inserito il seguente codice:

```
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relActivityEnd.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relActivityStart.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relAssignCredits.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relCourseBook.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relOrgEmailAddress.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relHasOpeningTime.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relHasProgram.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relHasQuestionTime.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relTeacherOf.xml"/></AUTOINSTANCE>
<AUTOINSTANCE><PARAM NAME ="xmlFileUrl"
  VALUE ="gate:/moses/schema/relTeachingAssistantOf.xml"/></AUTOINSTANCE>
```

In modo da rendere possibili annotare utilizzando le relazioni inserite con i nuovi file xml.

Bibliografia

- [1] T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. Scientific American, May 2001.
- [2] Guarino et al (1999) OntoSeek: Content-based Access to the Web. IEEE Intelligent System, 14 (3), pp.70-80.
- [3] Andreasen et al (2000) Ontology-Based Querying. In Larsen et al (eds) Flexible Query Answering system, Recent Advances, Springer, pp.15-26.
- [4] Living with Topic Maps and RDF: <http://www.ontopia.net/topicmaps/materials/tmrdf.html>
- [5] S. Bechhofer, C. Goble. Towards Annotation using DAML+OIL K-CAP 2001 workshop on Knowledge Markup and Semantic Annotation, Victoria B.C, October 2001. (COHSE)
- [6] Ontology-based question answering in a federation of university sites: the MOSES case study Paolo Atzeni, Roberto Basili, Dorte H. Hansen, Paolo Missier, Patrizia Paggio, Maria Teresa Pazienza, Fabio Massimo Zanzotto, Giugno 2004.
- [7] M. Denny. Ontology Building: A Survey of Editing Tools November 06, 2002 <http://www.xml.com/pub/a/2002/11/06/ontologies.html?pages=2>
- [8] S. Handschuh, S. Staab, A. Mädche . CREAM - Creating relational metadata with a component-based, ontology driven annotation framework . K-CAP 2001

- [9] S. Handschuh, S. Staab, A. Maedche. An AnnotationFramework for the Semantic Web Karlsruhe, Germany 2001
- [10] M. Missikoff, F. Schiappelli, F.Taglino A Controlled Language for Semantic Annoatation and Interoperability in e-Business Application IASI-CNR (Italy) 2003
- [11] M. Missikoff F. Schiappelli. Semantic Annotation for Enterprises Interoperability IASI-CNR (Italy) 2003
- [12] www.semanticweb.org
- [13] DAML + OIL (March 2001). <http://www.daml.org>.
- [14] Resource Description Framework (RDF) Model and Syntax Specification. <http://www.w3.org/TR/1999/REC-rdf-suntax-19990222>, February, 22 1999.
- [15] OWL Guide <http://www.w3.org/TR/owl-guide/>
- [16] OWL Reference <http://www.w3.org/TR/owl-ref/>
- [17] Steve Pepper, Graham Moore, XML Topic Maps (XTM) 1.0, <http://www.topicmaps.org/xtm/1.0/xtm1-20010806.html>, International Organization for Standardization, ISO/IEC13250:2000 Document description and processing languages. In Topic Maps, Ginevra, CH, 2000.
- [18] S. Pepper. The TAO of Topic Maps, finding the way in the age of infoglut. XML Europe 2000, 2000.
- [19] Kal Ahmed, Danny Ayers, Mark Birbeckand, Jay Cousinsand, David Dodd-sand, Joshua Lubelland, Miloslav Nicand, Daniel Rivers-Mooreand, Andrew Wattand, Rob Worden, and Ann Wrightson. Professional XML Meta Data. Wrox, 2001.

- [20] Anitta Altenburger. Topic Maps Tutorial, 2001
<http://topicmaps.bond.edu.au/tutorial1>.