

**UNIVERSITA' DEGLI STUDI DI MODENA
E REGGIO EMILIA**

Facoltà di Ingegneria - Sede di Modena

Corso di Laurea in Ingegneria Informatica

Annotazione Semantica di Portali Web

Relatore
Prof. Sonia Bergamaschi

Tesi di Laurea di
Luigi Martino

Correlatore
Ing. Alain Fergnani

Anno Accademico 2002 – 2003

RINGRAZIAMENTI

Desidero ringraziare la Professoressa Sonia Bergamaschi e l'Ing. Alain Fergnani per l'aiuto fornito durante la realizzazione della presente tesi.

Un ringraziamento speciale, quello che ha più valore, va poi alla mia famiglia e ai miei nonni che mi hanno dato la possibilità di raggiungere questo importante obiettivo, sognato da sempre.

Non posso non ricordare e ringraziare i miei amici tra cui, Leonardo, Antonio e Piero per tutte le giornate in cui non mi hanno fatto pensare allo studio. Ringrazio anche gli amici di studio della BSI e tutti quelli che hanno diviso con me gioie e fatiche in questi anni di studio.

E infine ringrazio Antonella che mi ha sempre sostenuto, supportato, e anche sopportato, facendomi acquisire quella fiducia e tranquillità nel raggiungere l'obiettivo che mi serviva negli ultimi periodi.

INDICE

Introduzione

1 Semantic Web e annotazione

- 1.1 Semantic Web
- 1.2 Ontologie
- 1.3 Annotazioni
- 1.4 RDF
- 1.5 XML
- 1.6 Il sito di esempio : il portale DBGROUP

2 Annotea

- 2.1 Il progetto Annotea
- 2.2 Struttura di Annotea
 - 2.2.1 L'architettura di base
 - 2.2.2 Il protocollo di Annotea
- 2.3 Amaya
 - 2.3.1 Installazione e requisiti
 - 2.3.2 Funzioni
 - 2.3.2.1 Configurazione del server
 - 2.3.2.2 Creazione di un'annotazione
 - 2.3.2.3 Salvataggio di un'annotazione
 - 2.3.2.4 Caricamento, navigazione, spostamento di un'annotazione
 - 2.3.2.5 Risposta ad un'annotazione
 - 2.3.3 Test su sito Web
- 2.4 Archiviazione delle annotazioni

3 OntoMat-Annotizer

- 3.1 CREAM il framework per l'annotazione
- 3.2 S-CREAM Semi-automatic CREATION of Metadata
- 3.3 Ontomat
 - 3.3.1 Installazione e requisiti
 - 3.3.2 Funzioni
 - 3.3.2.1 Operazioni da eseguire all'avvio
 - 3.3.2.2 Creazione di un'annotazione

3.3.2.3 Creazione di relazioni fra istanze
3.3.2.4 Salvataggio di annotazioni e della pagina
HTML

3.3.3 Test su sito Web

3.4 Archiviazione delle annotazioni

4 Confronto fra i due tools

5 Gli altri tools

5.1 Smore

5.2 MnM

6 Conclusioni

INTRODUZIONE

Il World Wide Web (WWW) contiene una enorme quantità di documenti e di informazioni che si stanno espandendo rapidamente. Per questo motivo la ricerca di informazioni sta diventando sempre più onerosa sia in termini di tempo, che di risorse. All'utente moderno, occorre ricercare le informazioni in modo facile e veloce, così negli ultimissimi anni sono nate delle tecnologie che permettono all'utente di ottimizzare queste ricerche.

Scopo di questa tesi è quello di provare e di confrontare due strumenti per effettuare l'*annotazione* di pagine web, cercando di capire che tipo di approccio utilizzano, qual è il loro raggio d'azione ed i loro possibili sviluppi. Nel mio caso l'annotazione è stata fatta su alcune pagine del portale del *DBGROUP* (www.dbgroup.unimo.it).

La fase di annotazione è essenziale per dare inizio a quel processo di ottimizzazione delle risorse web che permetterà di migliorare la ricerca di informazioni sul web.

Infatti da qui nasce l'idea del *Semantic Web*, una nuova visione del Web, dove si cerca di costruire una infrastruttura che raccolga il significato, il senso di risorse web dal punto di vista dei vari utenti (*Ontologie*).

Quindi mettendo insieme tutte queste ontologie, ordinate per classi di interesse, si potrebbe avere una banca di dati (database) che raccolga il significato di ogni classe e creare così degli opportuni motori di ricerca che sfruttino queste informazioni per ottimizzare le ricerche sul web.

I due tools che ho provato sono *Annotea* [1] e *OntoMat* [2]; il primo nasce come standard di riferimento della W3C [3] mentre il secondo è uno sviluppo dell'Università di Karlsruhe.

Sono due tools sviluppati con lo scopo di permettere ad un qualsiasi utente di poter annotare i propri documenti Web, ma adottano degli approcci sicuramente diversi specie nella gestione delle Ontologie.

Verranno messi a confronto i due tools cercando di capire i vantaggi o gli svantaggi dell'uno rispetto all'altro, come ad esempio la facilità d'uso, la portabilità, l'aggiornamento, lo sviluppo, l'archiviazione delle annotazioni.

Nella tesi verrà effettuato l'approfondimento di entrambi i tools, verranno descritte le pagine annotate, ci sarà la parte dei confronti, una breve citazione di altri tools e ci sarà anche una spiegazione delle parole chiavi come ad esempio ontologie, annotazioni, semantic web.

CAPITOLO 1

Semantic Web e annotazione

1.1 Semantic Web

Definizione: The Semantic Web is the representation of data on the World Wide Web. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URIs for naming.

"Il Semantic Web è un'estensione del Web attuale in cui alle informazioni sono date un senso, un significato ben definito, migliorando in questo modo la cooperazione tra i computer e le persone." -- Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001 [4]

In questi ultimi anni avere a disposizione macchine che siano in grado di comprendere i dati sul Web sta diventando una priorità per molte aziende. Il Web potrebbe fare un grande passo in avanti se diventasse un luogo dove i dati possono essere condivisi ed elaborati da strumenti automatizzati e permettere così alle persone di trarne beneficio.

Il Semantic Web è un'idea: l'idea di avere i dati sul Web definiti e collegati in un modo che possono essere usati non solo dalle macchine per scopi di visualizzazione, ma anche per l'automazione, l'integrazione e il riuso dei dati attraverso varie applicazioni.

Il successo del Semantic Web dipende in modo cruciale dalla facile creazione, l'integrazione e l'uso dei dati semantici. Infatti una delle sfide del Semantic Web è la creazione di metadata con la collaborazione di utenti web, cioè combinando il contenuto semantico creato da un grande numero di persone. Per ottenere questo obiettivo diversi approcci sono stati concepiti (ad esempio, CREAM, MnMv, o Mindswap) che gestiscono in modo manuale e/o semiautomatico la creazione di metadati da informazioni già esistenti.

Questi strumenti, annotazione solo le pagine statiche, che hanno però il difetto che possono con il tempo dare vita a metadati vecchi non aggiornati.

Oggi, comunque, una alta percentuale delle pagine del Web sono dinamiche. Per le pagine Web dinamiche (ad esempio quelle che sono generate da database) non sembra sia utile annotarle manualmente. Piuttosto si vuole annotare il database per riusarlo ogni volta per scopi riguardanti il Semantic Web. Per questo obiettivo, gli approcci sono stati concepiti per permettere la costruzione di wrappers da definizione di interrogazioni HTML o XML.

Due importanti tecnologie per lo sviluppo del Semantic Web sono già in fase avanzata: eXtensible Markup Language (XML) e il Resource Description Framework (RDF). XML permette di creare dei propri tag e permette agli utenti di aggiungere delle strutture arbitrarie ai loro documenti, ma non dice niente circa il significato delle strutture. Per questo motivo il significato è espresso dal linguaggio RDF, che utilizza il codice in set di triple, dove ogni tripla contiene il soggetto, il verbo e l'oggetto di una frase elementare. Queste triple sono scritte usando tags XML.

1.2 Ontologie

Le ontologie sono spesso considerate come le componenti di base per il Web Semantico. Esse forniscono una parte riutilizzabile della conoscenza su un dominio specifico. Comunque, queste conoscenze non sono statiche, ma evolvono nel tempo. I cambiamenti del dominio, gli adattamenti ai compiti differenti, o i cambiamenti nella concettualizzazione richiedono modifiche dell'ontologia. L'evoluzione delle ontologie causa dei problemi di operabilità, che crea forti dubbi sul loro uso efficace. Un meccanismo che può aiutare a ridurre questi problemi, è quello di costruire delle relazioni attraverso le revisioni differenti di un'ontologia esplicita.

L'idea del Web semantico ha creato un gran interesse nell'uso delle ontologie, le descrizioni di una parte del mondo per descrivere il significato delle informazioni sul Web. Comunque, quando le ontologie sono state usate sul Web, ci furono diversi problemi : come possono essere combinate ontologie differenti, che cosa succede quando sono modificate, e come devono essere adattate per nuovi compiti? Queste domande richiedono la gestione delle ontologie, il loro uso e la loro evoluzione.

Le ontologie stabiliscono una terminologia tra i membri di una comunità di interesse. Questi membri possono essere umani o agenti automatizzati. Per rappresentare una concettualizzazione è necessario un linguaggio di rappresentazione.

Le ontologie giocano anche un ruolo importante in nuove aree di interesse, il Semantic Web, come confermato da progetti come OntoWeb (<http://www.ontoweb.org>) e DAML (<http://www.daml.org>)

La parola Ontologia assume diversi significati nella Artificial Intelligence, dove indica un *“engineering artifact”* che viene compreso da uno specifico vocabolario e da un set di esplicite ipotesi che riguardano il significato che si vuole dare ad una parola che compone il vocabolario.

La definizione di Ontologia secondo Gruber [5]

La definizione più citata di ontologia fu data da Tom Gruber nel 1993. Un'ontologia è ***una esplicita specificazione di una concettualizzazione***. La definizione di Gruber costruisce l'idea che la formalizzazione del dominio di conoscenza inizia con la concettualizzazione del dominio [Genesereth e Nilsson 1987], cioè è l'identificazione degli oggetti che si è ipotizzato esistano nel mondo e delle relazioni tra loro.

Secondo Gruber un'ontologia è una quintupla composta di classi, istanze, funzioni, relazioni, assiomi. Le classi corrispondono alle entità del dominio, le istanze sono gli oggetti che stanno nel dominio, le funzioni e le relazioni collegano le entità al dominio, ed in fine gli assiomi che limitano il senso e l'uso delle classi, istanze, funzioni e relazioni.

La definizione di Ontologia secondo Guarino [6]

Un'ontologia secondo Guarino, è invece: ***un set di assiomi logici progettati per considerare il senso di un vocabolario***.

Da questa definizione nasce la relazione tra una concettualizzazione e l'ontologia che la specifica e la formalizza.

La definizione di Gruber è così stata estesa prendendo non solo la concettualizzazione ma anche il linguaggio usato per descriverla ed un insieme di assiomi associati con essa. In questo senso un'ontologia è un linguaggio dipendente mentre una concettualizzazione è un linguaggio indipendente.

1.3 Annotazioni

Le annotazioni sono i commenti, le note, le spiegazioni, o altri tipi di note esterne che possono essere allegate a un documento Web o ad una parte selezionata del documento. Dato che le annotazioni sono esterne, è possibile annotare ogni documento Web indipendentemente, senza aver bisogno di editare quel documento. Dal punto di vista tecnico, le annotazioni sono viste di solito come metadata, e danno informazioni supplementari su un documento. Per descrivere le annotazioni viene usato uno schema di annotazione speciale: RDF. Le annotazioni possono essere memorizzate localmente o in uno o più server per l'annotazione. L'annotazione di documenti Web è uno delle maggiori tecniche per creare metadata sul Web. Possono essere memorizzate nella workstation dell'utente per il suo uso privato o in uno o più server comuni. Le annotazioni memorizzate nei server pubblici possono essere disponibili per chiunque ma con una autenticazione propria.

Un'annotazione ha molte proprietà [7]:

- La posizione fisica: l'annotazione è memorizzata in un file o in un server per l'annotazione
- Scopo: l'annotazione è associata ad un intero documento o solamente a un frammento di questo documento
- Il tipo dell'annotazione: "Annotazione", "commento", "interrogazione", ...

1.4 RDF



RDF- Resource Description Framework [8] è un modo standard per descrivere semplicemente ciò che è stato costruito. Quello che XML è per la sintassi, RDF è per la semantica un set completo di regole per fornire semplici descrizioni delle informazioni. RDF-Schema fornisce poi un modo per queste descrizioni per essere unite in un vocabolario solo. Poi c'è bisogno di una strada che permetta di sviluppare una materia o un dominio per vocabolari. Questo è il ruolo di un'ontologia.

Un'ontologia definisce i termini usati per descrivere e rappresentare un dominio di conoscenza. Le ontologie sono usate dalle persone, da database, e per le applicazioni che hanno bisogno di condividere materie specifiche (dominio) di alcune informazioni, come ad esempio la medicina, riparazione di automobili, gestioni finanziarie, eccetera. Il Resource Description Framework (RDF) è uno standard per rappresentare le informazioni nel Web.

L'RDFs il linguaggio della descrizione del vocabolario, detto anche RDF schema, è un'estensione semantica dell' RDF. Fornisce dei meccanismi per descrivere i gruppi delle risorse collegate e le relazioni tra queste risorse.

Con l'estensione e lo sviluppo di RDF e di RDF-Schema nasce [OIL \(Ontology Interchange Language\)](#). Un successore di OIL è [DAML+OIL](#), creato da un gruppo di sviluppatori europei e americani.

Questa tabella descrive il vocabolario RDF, mettendo insieme il vocabolario originale definito nel modello RDF con le classi e proprietà generate con RDF schema.

RDF classes

Class name	comment
rdfs:Resource	The class resource, everything.
rdfs:Literal	The class of literal values, e.g. textual strings and integers.
rdf:XMLLiteral	The class of XML literals values.
rdfs:Class	The class of classes.
rdf:Property	The class of RDF properties.
rdfs:Datatype	The class of RDF datatypes.
rdf:Statement	The class of RDF statements.
rdf:Bag	The class of unordered containers.
rdf:Seq	The class of ordered containers.
rdf:Alt	The class of containers of alternatives.
rdfs:Container	The class of RDF containers.
rdfs:ContainerMembershipProperty	The class of container membership properties, rdf:_1, rdf:_2, ..., all of which are sub-properties of 'member'.
rdf:List	The class of RDF Lists.

RDF properties

Property name	Comment	domain	range
rdf:type	The subject is an instance of a class.	rdfs:Resource	rdfs:Class
rdfs:subClassOf	The subject is a subclass of a class.	rdfs:Class	rdfs:Class
rdfs:subPropertyOf	The subject is a subproperty of a property.	rdf:Property	rdf:Property
rdfs:domain	A domain of the subject property.	rdf:Property	rdfs:Class
rdfs:range	A range of the subject property.	rdf:Property	rdfs:Class
rdfs:label	A human-readable name for the subject.	rdfs:Resource	rdfs:Literal
rdfs:comment	A description of the subject resource.	rdfs:Resource	rdfs:Literal
rdfs:member	A member of the subject resource.	rdfs:Resource	rdfs:Resource
rdf:first	The first item in the subject RDF list.	rdf:List	rdfs:Resource
rdf:rest	The rest of the subject RDF list after the first item.	rdf:List	rdf:List
rdfs:seeAlso	Further information about the subject resource.	rdfs:Resource	rdfs:Resource
rdfs:isDefinedBy	The definition of the subject resource.	rdfs:Resource	rdfs:Resource
rdf:value	Idiomatic property used for structured values (see the RDF Primer for an example of its usage).	rdfs:Resource	rdfs:Resource

<code>rdf:subject</code>	The subject of the subject RDF statement.	<code>rdf:Statement</code>	<code>rdfs:Resource</code>
<code>rdf:predicate</code>	The predicate of the subject RDF statement.	<code>rdf:Statement</code>	<code>rdf:Property</code>
<code>rdf:object</code>	The object of the subject RDF statement.	<code>rdf:Statement</code>	<code>rdfs:Resource</code>

1.5 XML

Quasi tutte le informazioni a disposizione sul Web, finora sono state rappresentate usando Hypertext Markup Language (HTML), che è stato progettato per permettere agli sviluppatori del Web di rappresentare le informazioni in un modo che sia accessibile agli utenti tramite web browsers. Mentre l' HTML ci permette di visualizzare le informazioni sul Web, non fornisce però molte potenzialità per descrivere le informazioni, in modo da facilitare l'uso di software per la sua interpretazione e traduzione. Il [World Wide Web Consortium \(W3C\)](#) ha sviluppato l'[Extensible Markup Language \(XML\)](#) [9] che permette alle informazioni di essere descritte usando dei tags.

L'eXtended Markup Language è accettato come lo standard emergente per lo scambio dei dati sul Web. È semplice, il formato di testo è molto flessibile e deriva da SGML (l'ISO 8879). Originariamente progettato per far incontrare le varie sfide dell'editoria elettronica, XML sta giocando anche un ruolo molto importante e che è in grande aumento, nello scambio di dati sul Web. L'uso delle ontologie fornisce un mezzo molto potente per descrivere gli oggetti e le loro relazioni con altri oggetti. Affinché XML diventi un linguaggio standard per lo scambio dei dati sul Web, è necessario scambiare le ontologie usando una sintassi XML. Comunque è recente la nascita di un nuovo linguaggio, il DAML che si sta sviluppando come un'estensione dell' XML e del Resource Description Framework (RDF).

1.6 Il sito di esempio:il portale DBGROUP

Il portale presenta una home-page(vedi figura 1) da dove si può accedere alle altre pagine. E' formato sia da pagine statiche che dinamiche.

E' facile capire che la parte che riguarda "notizie ed eventi" è dinamica e viene aggiornata periodicamente.

C'è poi tutta una parte che riguarda il personale che fa parte del DBGROUP e relativi progetti, i corsi di Basi di Dati che presentano ulteriori pagine tra appelli, esami, iscrizioni, materiale didattico, tesi ed altro ancora.

The screenshot shows the DataBase Group home page in Microsoft Internet Explorer. The header includes the university's logo and the text "DataBase Group", "Dipartimento di Ingegneria dell'Informazione", "Facoltà di Ingegneria - sede di Modena", and "Università di Modena e Reggio Emilia". The sidebar on the left is titled "Members" and lists "Research" (with links to "Funded projects" and "Publications"), "Courses 02/03" (with links to "Basi di Dati", "Basi di Dati A", "Lab. di Basi di Dati", and "Altri Corsi"), and "Courses 98-02" (with links to "Thesis" and "Other" resources like "dbgroup mail reader" and "sw/hw resources"). The main content area features four project logos: "sewasie" (with the tagline "Semantic Web and Agents in integrated economies"), ";-)", "MOMIS Project", and "MIKS Project". To the right, a sidebar titled "Notizie ed Eventi" lists news items: "09/09/2003 Risultati appello 05/09/2003" (with a note about the written exam results for Basi di Dati VOD and Basi di Dati A NOD), "28/08/2003 Prova orale - Basi di Dati e Basi di Dati A" (with a note about the oral exam for Basi di Dati VOD and Basi di Dati A NOD), and "31/07/2003 Risultati appello 28/07/2003" (with a note about the written exam results for Basi di Dati VOD and Basi di Dati A NOD). A link "Le altre notizie..." is also present. The bottom of the page shows the Windows taskbar with the Internet icon and the system tray.

Figura 1 : Home page del portale del DBGROUP

CAPITOLO 2

ANNOtea

2.1 Il Progetto Annotea

Annotea[1] è un progetto LEAD (Live Early Adoption and Demonstration) nato per migliorare ed accrescere la collaborazione di W3C con le annotazioni divise. Secondo il W3C le annotazioni sono i commenti, le note, le spiegazioni, o altri tipi di informazioni esterne che possono essere associate ai documenti Web o anche ad una parte di essi senza aver bisogno di manipolare o cambiare il documento. Quando l'utente apre il documento, può caricare le annotazioni ad esso associate da un server o da più server.

Per esempio, Annotea usa uno schema per l'annotazione (RDF) per descrivere le annotazioni come metadata ed usa XPointer per posizionare o localizzare le annotazioni nel documento annotato.

Le annotazioni sono memorizzate in server per l'annotazione come metadata e presentati all'utente con un client capace della comprensione di questi metadata e capace di interagire con un server mediante il protocollo HTTP.

Il primo client realizzato da Annotea è W3C's Amaya editor/browser.[12]

Annotea è un sistema basato su un'infrastruttura RDF, dove le annotazioni sono modellate come una classe di metadata. Le annotazioni sono viste come una specificazione da parte di un autore di un documento Web. Le annotazioni sono esterne ai documenti e possono essere memorizzate in uno o più server per l'annotazione. Uno dei traguardi di questo progetto è stato quello di poter riutilizzare molte tecnologie W3C già esistenti per quanto possibile. Infatti si è cercato il più possibile di combinare RDF con XPointer, XLink, e HTTP. In questa implementazione, la fusione delle annotazioni con i documenti ha luogo nel client.

La prima applicazione della infrastruttura Annotea fu uno schema per Annotazioni; oggetti con varie proprietà con lo scopo di comunicare commenti tra i lettori di un documento web. Lo schema di Annotea Annotation definisce le proprietà per l'identificazione del documento esistente annotato, uno specifico contesto per ogni risorsa facente parte di quel documento, al quale il corpo della annotazione fa riferimento, l'autore della annotazione, e altri tipi di informazioni. Annotea client può provvedere a una varietà di rappresentazioni di queste annotazioni. Dato che Annotea framework è progettato al top di RDF, esso può facilmente essere esteso ad altre azioni per applicazioni metadata.

Un'annotazione ha molte proprietà [10] :

- La posizione fisica: è l'annotazione memorizzata in locale o in un server per l'annotazione
- Scopo: è l'annotazione associata un intero documento o ad un frammento del documento
- Il tipo dell'annotazione: "Annotazione", "Commento", "Interrogazione", ...

2.2 Struttura di Annotea

2.2.1 L'architettura di Base

La schema di Annotea Bookmark (BookmarkNS) fornisce i concetti di base istituiti nelle comuni implementazioni browser bookmark. Questi concetti di base sono catturati anche nel XML Bookmark Exchange Language [XBEL]. L'uso di RDF in Annotea permette ai bookmarks di esprimere semantiche supplementari. XBEL può essere rappresentato facilmente in questa schema [11].

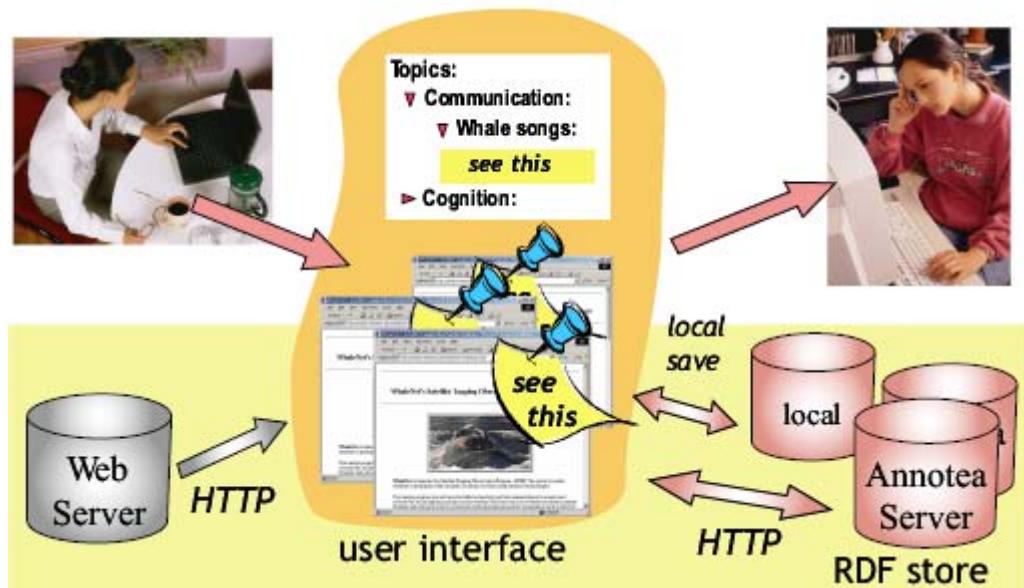


Figura 2 : L'architettura bookmark di Annotea

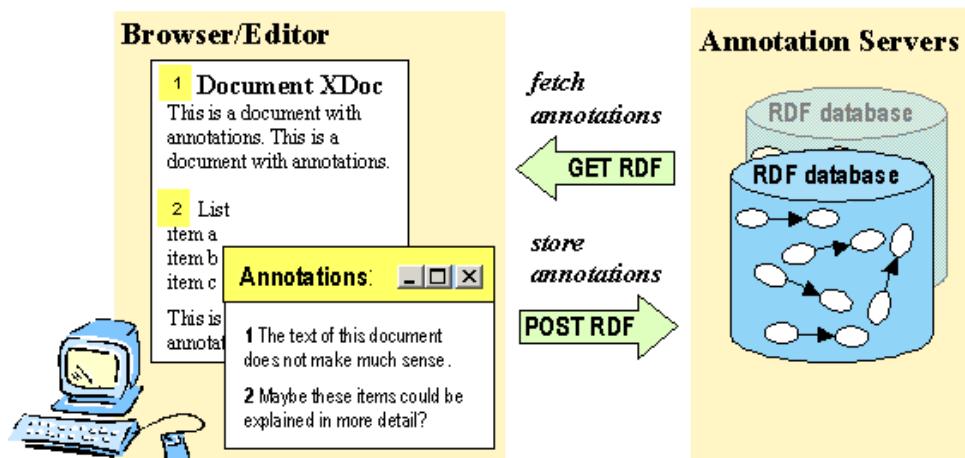


Figura 3: L'architettura di base di Annotea.

L'applicazione bookmark usa la stessa architettura di Annotea così come usato per le annotazioni (vedi la figura 3). Ogni singolo bookmark è rappresentato come una risorsa

RDF proprio di tipo bookmark. Le proprietà dei bookmark e dei Topics sono espresse come espressioni RDF e sono memorizzate in file locali o in uno o più server bookmark via HTTP.

2.2.2 Il protocollo di Annotea

Annotea è un sistema per la creazione e la pubblicazione di annotazioni di documenti web. Realizzato in HTTP, RDF, e XML, Annotea fornisce un protocollo adatto per l'implementazione con un web browser per permettere agli utenti di aggiungere dati alle pagine web. Questi dati aggiunti possono comunque essere consultati da altri utenti in qualsiasi momento. Il protocollo di Annotea lavora senza modificare il documento originale. Esso è adatto per tutti i dati che sono utilizzati per essere consultati dall'utente sotto forma di annotazioni, sia per dati annotati utilizzati da altre applicazioni, come strumenti automatici, di ricerca, e applicazioni di flusso.

Il protocollo di Annotea è comunque adatto per l'uso di applicazioni che scambiano dati per usi non-visuali da allegare a pagine web.

Un servizio di annotazione è un'istanza di un http server che risponde al protocollo Un client Annotea, come Amaya, è capace di interagire con uno o più servizi annotati usando un set di protocolli che sono costruiti al top del protocollo HTTP come descritto in seguito.

Annotea attualmente specifica 3 tipi di protocolli [10] :

Annotation protocols

Describe le interazioni client-server collegate al caricamento, scaricamento, aggiornamento e cancellazione di annotazioni.

Reply-to-annotation protocols

Describe le interazioni client-server che supportano risposte alle annotazioni e così promuovere discussioni.

Optimizing query requests

Definisce i parametri di interrogazione di speciali URI per combinare le interrogazioni in una singola richiesta HTTP.

2.3 AMAYA

2.3.1 Installazione e requisiti



Amaya è un Web Editor, uno strumento per creare ed aggiornare documenti direttamente sul Web.

Amaya include un'applicazione per l'annotazione, basata su Resource Description Framework (RDF), XLink, e XPointer.

La versione attuale di Amaya è la 8.1a e supporta: HTML 4.01, XHTML 1.0, XHTML Basic, XHTML 1.1, HTTP 1.1, MathML 2.0, molte caratteristiche CSS 2, un supporto SVG (trasformazione, trasparenza, e animazioni SMIL su piattaforme OpenGL). Si può visualizzare e parzialmente editare il documento XML.

Amaya include Annotea e Annotea Bookmarks, progetti di W3C Semantic Web Advanced Development che forniscono una struttura per la comunicazione sulle pagine del Web attraverso RDF metadata. Annotea Bookmarks è attualmente disponibile solo sulle piattaforme Linux.

Le versioni future

La prossima versione 8.2 è prevista per la fine di ottobre del 2003

- Download Amaya 8.1b pubblic release (4 September 2003) [13]

La binary distribution di Amaya è disponibile per PC Linux (GTK/OpenGL), Windows (NT, XP, 2000, 98).

La binary distribution è disponibile per un set di piattaforme di Windows e Unix.

Quando si installa Amaya per la prima volta sulle piattaforme di Windows NT/2000/XP, Amaya proverà a creare la directory di utente di Amaya nell'ordine seguente: 1)Guardando i valori delle variabili d'ambiente HOMEDRIVE e HOMEPATH, 2)Document and settings\login-name\WINNT\PROFILES\LOGIN-NAME.

Piattaforma	Lunghezza	http download	ftp Europe o US
Windows NT/2000/XP	about 4.752, 000 bytes**	amaya-WinXP-8.1b.exe	amaya-WinXP-8.1b.exe
Windows 98/ME	about 4, 789, 000 bytes**	amaya-Win98-8.1b.exe	amaya-Win98-8.1b.exe
Linux Debian 3.0: GTK version	About 3, 453, 000 bytes**	amaya-GTK-8.1b.tgz	amaya-GTK-8.1b.tgz

Debian 3.0 (woody) package: GTK version	About 5, 412, 000 bytes	[amaya]_gtk-8.1b-1_i386.deb	[amaya]_gtk-8.1b-1_i386.deb
RedHat 8 RPM i386 packages: GTK version (should work with Mandrake and SuZE)	about 4, 660, 000 bytes**	[amaya]_gtk-8.1b-1.i386.rpm	[amaya]_gtk-8.1b-1.i386.rpm
Windows NT/2000/XP: OpenGL version	about 4, 969, 000 bytes**	amaya-WinGL-8.1b.exe	amaya-WinGL-8.1b.exe
Linux Debian 3.0: OpenGL version	about 3, 485, 000 bytes**	amaya-GL-8.1b.tgz	amaya-GL-8.1b.tgz
RedHat 8 RPM i386 packages: OpenGL version (should work with Mandrake and SuZE)	about 4, 222, 000 bytes**	[amaya]_gl-8.1b-1.i386.rpm	[amaya]_gl-8.1b-1.i386.rpm

**Questi distribuzioni includono i dizionari inglesi. L'extra file amayadoc ha in più i dizionari francesi.

Come installare la versione per Windows di Amaya

Dopo il download, il processo di installazione di Amaya è automatico. Basta fare un doppio click sul programma di installazione di Amaya e seguire le istruzioni del processo di installazione.

Se si sceglie di installare Amaya nella directory X (di default C:\Program Files), il programma di installazione creerà un nuovo directory chiamato X\Amaya con alcuni file e directory. Il file amaya.exe è posizionato nella directory X\Amaya\Windows\bin.

Come installare la versione per Unix di Amaya

Scegliere una directory per Amaya

- /Usr/local per un'installazione completa
- la user home directory per un'installazione personale

Estrazione dei file

Bisogna copiare i file nella directory, poi controllare che i diritti su questa directory siano tali da permetterti di creare una directory ed estrarre i file.

Su una piattaforma Linux bisogna invocare:

tar xvzf amaya-**Platform-Release**.tgz

o più generalmente:

gunzip -c amaya-**Platform-Release**.tgz | tar xvf -

Set up dell'applicazione ed avvio di Amaya

L'Amaya binary è il file Amaya/Platform/bin/amaya. La cosa migliore è aggiungere il percorso Amaya/Platform/bin al PATH della macchina utilizzata dall'utente.

Ad esempio su un sistema GTK con una struttura csh o tcsh:

```
setenv PATH $PATH:/usr/local/Amaya/GTK/bin
```

Oppure se si usa sh, bash o un'altra variante della struttura Bourne:

```
PATH=$PATH:/usr/local/Amaya/GTK/bin ; export PATH
```

Una volta eseguite queste operazioni si può lanciare Amaya semplicemente digitando "amaya". La finestra principale di Amaya dovrebbe aprirsi e mostrare la prima pagina della documentazione di Amaya.

2.3.2 Funzioni

2.3.2.1 Configurazione del server

Il menù di configurazione per le annotazioni è situato sotto il menù "Annotations/Configure". La figura seguente mostra la versione di Windows di questo menù. La versione di Unix ha un'interfaccia utente leggermente differente, ma con le stesse funzionalità.

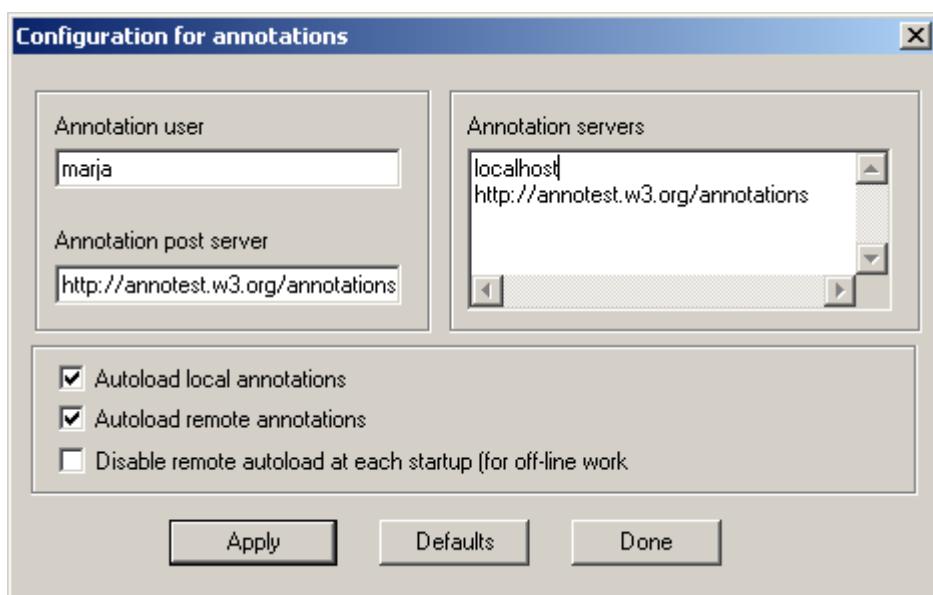


Figura 4:il menù di configurazione per le annotazioni

-Annotation user

Un campo che è associato ad ogni nuova annotazione e rappresenta l'autore dell'annotazione. Di default, Amaya usa il login dell'utente quando deve inserire l'autore del metadato. La annotation user setting permette a un utente di cambiare questo nome con uno più significativo.

-Annotation servers

Questa funzione dice ad Amaya quali server dovrebbe contattare quando cerca le annotazioni. Si può specificare uno o più server. Viene utilizzato il nome del server riservato “localhost” per dire ad Amaya che si vuole ricercare le annotazioni in locale. Sotto Unix, i server dell'annotazione sono specificati con uno spazio in una lista separata. Sotto Windows, sono aggiunti uno dopo l'altro digitando i loro nomi nella finestra di dialogo, usando il tasto di return per aggiungerne uno nuovo. Di default, il campo dei server dell'annotazione è inizializzato come localhost.

CONSIGLIO: Se temporaneamente si vuole disattivare un server dell'annotazione, basta aggiungere un carattere “-” prima del suo URL, Amaya lo ignorerà.

-Autoload local annotations/Autoload remote annotations

Queste funzioni dicono ad Amaya se si vuole richiedere le annotazioni automaticamente, ogni volta che un URL è stato selezionato. Se questo campo non è stato selezionato, allora l'utente manualmente deve invocare la funzione Load annotations dal menù delle annotazioni.

-Disable remote autoload at each startup

Se questa opzione è scelta, Amaya ripristinerà l'opzione autoload remote annotations ad ogni caricamento. Questa opzione può essere utile se si lavora senza essere sempre collegati, ma si vuole di tanto in tanto anche l'autoload delle local e remote annotations.

-Annotation post server

Questa funzione definisce il server nel quale saranno salvate le annotazioni. Le annotazioni locali sono salvate sempre in una directory locale. Di default, questa funzione non è selezionata.

2.3.2.2 Creazione di un'annotazione

Questa versione di Amaya supporta due tipi di annotazioni:

- annotazioni che si applicano ad un intero documento
- annotazioni che si applicano ad una parte selezionata del documento.

Per annotare un documento intero, selezionare **Annotations/Annotate document** dal menù. Per annotare una selezione bisogna prima selezionare la parte del documento con il puntatore del mouse e poi dal menù scegliere l'opzione **Annotations/Annotate selection**. In tutti i casi comunque, una finestra dell'annotazione apparirà sullo schermo (vedi la figura seguente). In un riquadro di queste finestre c'è il metadato dell'annotazione (autore, titolo del documento annotato, tipo dell'annotazione, data di creazione e ultima modifica), mentre nel resto della finestra c'è il corpo vero e proprio dell'annotazione.

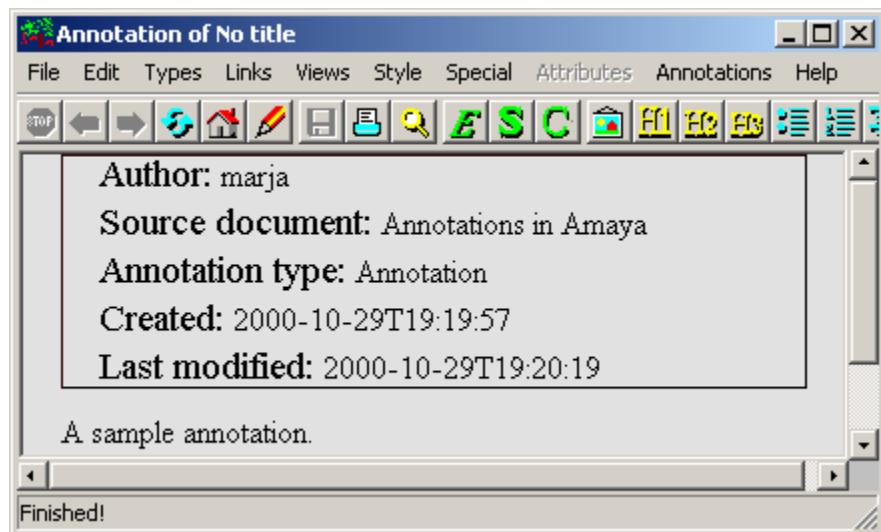


Figura 5:il metadato di una annotazione

Attualmente, il metadato è costituito da, nome dell'autore, il titolo del documento annotato, il tipo dell'annotazione, la data della creazione, e la data dell'ultima modifica. Alcuni dei campi del metadato hanno delle particolari proprietà:

- Il campo del titolo del documento annotato è anche un collegamento per il testo annotato. Se l'utente fa doppio clic sul campo, il documento annotato sarà mostrato ed il testo annotato sarà illuminato.
- Il campo del tipo di annotazione permette all'utente di classificare l'annotazione e comunque di cambiare il suo tipo. Facendo doppio clic sul testo "annotation type" si può vedere la lista dei tipi disponibili. Si può anche definire un tipo proprio per l'annotazione.
- Il campo ultima modifica è aggiornato automaticamente ogni volta che un'annotazione è salvata.

2.3.2.3 Salvataggio di un'annotazione

Salvare un'annotazione con Amaya è molto semplice. L'utente ha solamente bisogno di selezionare il comando "File/Save" (o usare il pulsante equivalente). Le annotazioni locali sono memorizzate in una directory, mentre le annotazioni remote sono memorizzate nel post server se l'utente ha l'accesso in scrittura. Per portare un'annotazione locale in un'altro posto, l'utente deve usare il comando di Annotations/Post annotation e l'annotazione sarà salvata nel post server come definito nel menù di configurazione. Se questa operazione ha successo, l'annotazione locale sarà spostata ed inoltre le operazioni successive andranno memorizzate direttamente in quel server per l'annotazione. La versione release di Amaya non supporta l'operazione di salvataggio di una copia di un'annotazione remota nella directory delle annotazioni locali.

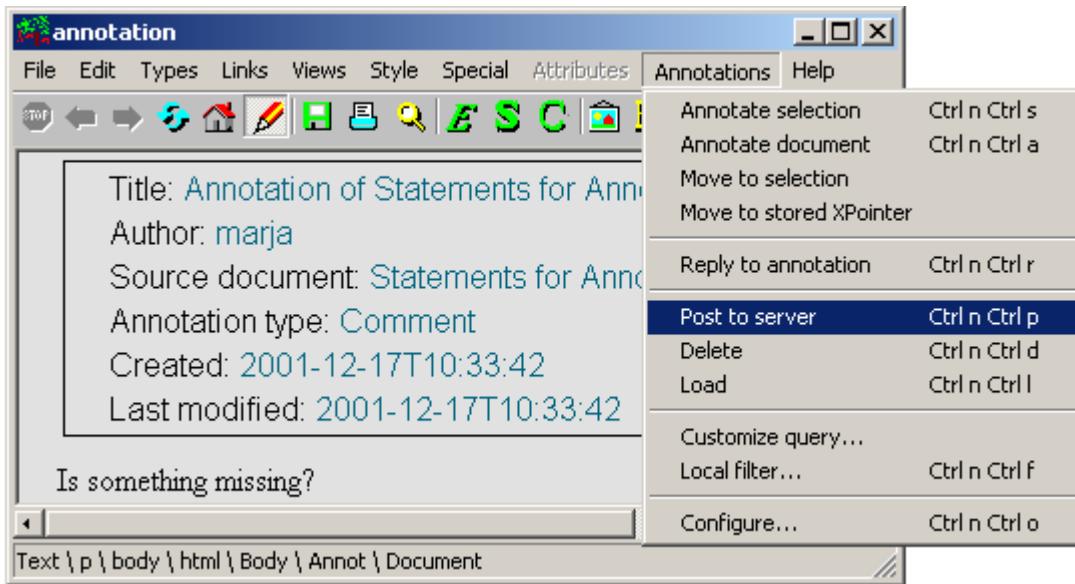


Figura 6: l'operazione di post to server

2.3.2.4 Caricamento, navigazione, spostamento di annotazioni

Amaya indica le annotazioni con un'icona che varia in base al tipo di annotazione; ad esempio nel caso si tratti di un commento l'icona è quella del pennello (☞) vedi la figura sotto. Se l'utente effettua un clic su un'icona per l'annotazione, il testo che è annotato viene illuminato. Mentre se l'utente fa doppio clic sull'icona, il testo dell'annotazione ed il metadato sono visualizzati in una finestra a parte.

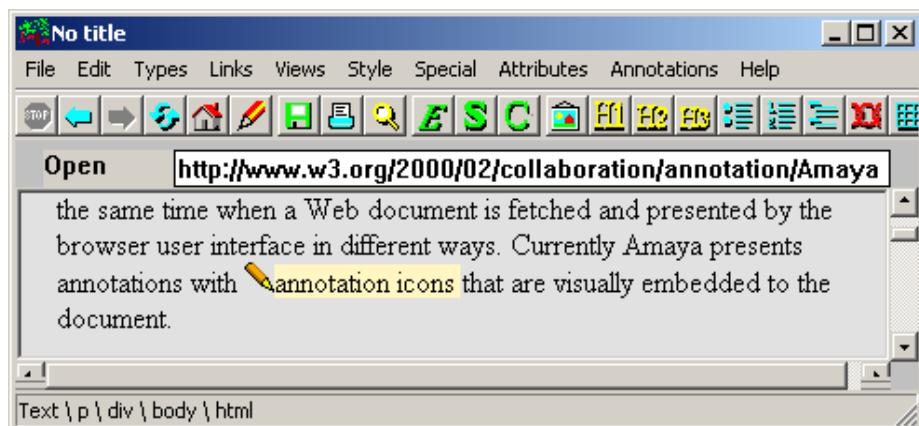


Figura 7: la parte del documento annotato viene evidenziata

Loading and Presenting Annotations

Il comando Load annotations carica le annotazioni che sono associate all'URL del documento. Amaya interrogherà i server per le annotazioni, come settato nella Annotations/Configure, per ogni relativa annotazione.

Le annotazioni possono anche essere caricate automaticamente ogni qualvolta una nuova pagina è caricata selezionando Autoload annotations checkbox nel Annotations/Configure dialogue.

Il menù Annotations/Local filter permette ad un utente di mostrare o selezionare le icone dell'annotazione dalla finestra del documento, ad esempio per concentrare quello che realmente interessa, o per facilitare la lettura di un documento pesantemente annotato. L'utente può mostrare o selezionare le annotazioni in base a tre tipi di metadato: per l'autore dell'annotazione, per il tipo dell'annotazione e per il nome del server dell'annotazione. Per applicare qualsiasi filtro, si deve fare clic sulla parte del testo e selezionare un tipo di annotazione, poi cliccare sul pulsante dell'azione corrispondente.

Navigating Annotations

Le annotazioni appaiono come collegamenti nella “Link window” che può essere attivata tramite il comando Views/ Show links. Le annotazioni in questa finestra sono segnate con la stessa icona come le annotazioni nella finestra del documento. La “Link window” mostra tutte le annotazioni, senza tener conto se sono state nascoste con il menù di Annotation Local Filter. Come con la finestra del documento, un solo-clic sull'annotazione selezionerà il testo annotato nella finestra del documento ed un doppio clic aprirà l'annotazione.

Moving annotations

Amaya permette anche di muovere un'annotazione all'interno dello stesso documento. Si può muovere un'annotazione in base alla selezione attuale o in base al valore di un memorizzato XPointer. Per il momento in questa versione è solo possibile muovere le annotazioni nello stesso documento dove sono state create.

Per spostare un'annotazione nella selezione attuale, per primo bisogna aprire la finestra dell'annotazione, poi cliccare sul documento annotato e selezionare la parte interessata, sulla finestra dell'annotazione, poi invocare il menù “Annotations/Move to selection”. Amaya sposterà l'icona dell'annotazione al testo selezionato e marcherà il documento dell'annotazione come modificato. Per rendere questo cambiamento efficace, si dovrà salvare l'annotazione, altrimenti si perderà il cambiamento. Similmente alla creazione delle annotazioni, si può anche muovere un'annotazione alla posizione attuale del puntatore senza dover fare una selezione.

E' anche possibile memorizzare la posizione dove muovere l'annotazione. Questa funzione è utile, ad esempio, se si vuole spostare molte annotazioni alla stessa posizione. Per fare questo, si può evidenziare la posizione dove si vuole muovere l'annotazione. A questo punto si usa il menù Link/Store link as XPointer per memorizzare un XPointer che rappresenta quella selezione. Alla fine, nella finestra delle annotazioni, usare il menù Annotations/Move to stored XPointer per muovere l'annotazione alla sua nuova posizione. Come prima si deve salvare l'annotazione per rendere questo cambiamento definitivo.

2.3.2.5 Risposta di un'annotazione

Le annotazioni possono essere viste come un commento alle pagine. La proprietà Annotation / Replies accresce la collaborazione tra workspace permettendo agli utenti di rispondere alle annotazioni.

Il comando Annotations / Reply to annotation permettono di creare una risposta a un'annotazione esistente. Si può chiamare questo comando da un'annotazione aperta o dalla finestra della risposta. Come risultato si ha una nuova finestra della risposta. I campi nella finestra della risposta possono essere compilati proprio come in una finestra dell'annotazione come se si dovesse creare un'annotazione.

Quando la risposta è pronta, è possibile memorizzarla in un server con il comando Annotations / Post to Server o memorizzarla localmente usando il comando File / Save dal menù. Per cancellare una risposta invece, si può usare il comando Annotations / Delete.

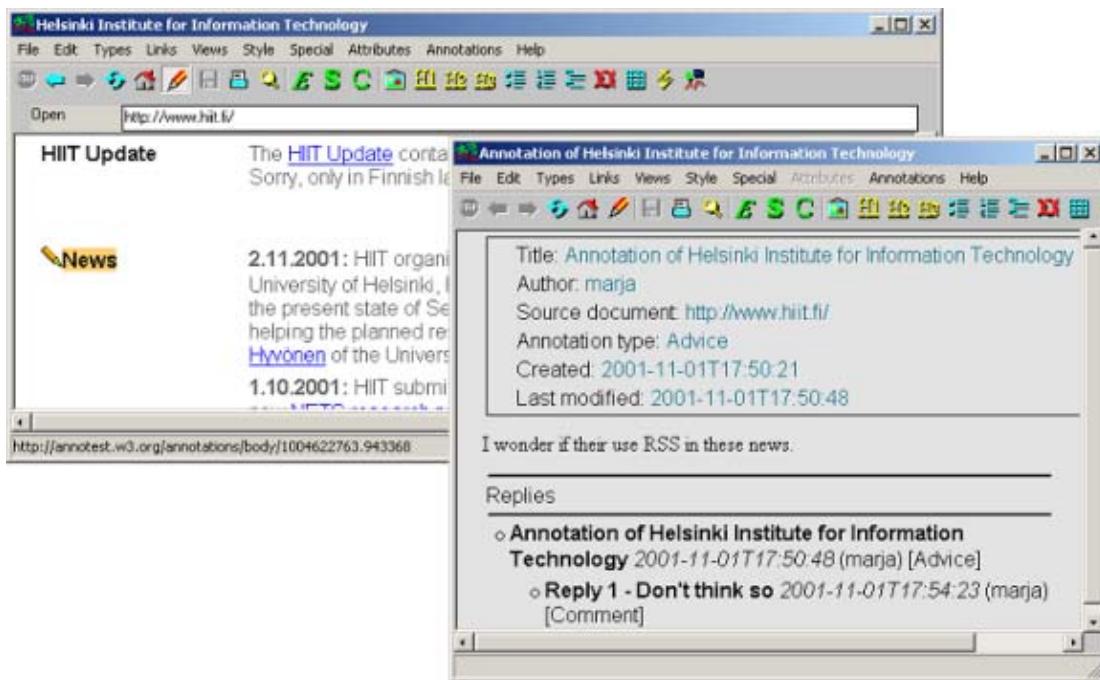


Figura 8: l'operazione di risposta ad un'annotazione

2.3.4 Test su sito Web

Sono state annotate varie pagine del sito www.dbgroup.unimo.it.

Qui presento alcune immagini di queste pagine:

Nella figura 9 è visualizzata l'operazione di annotazione della pagina web <http://dbgroup.unimo.it/bdati/bdati.htm> che rappresenta la pagina principale del corso di Basi di Dati e contiene tutto il materiale che riguarda il corso (informazioni, date degli appelli, risultati, tesine, ecc..)

In questo caso nella pagina sono presenti varie annotazioni. Qui è evidenziata l'annotazione che riguarda gli appelli d'esame, le modalità d'esame, i risultati dei vari corsi di Basi di Dati NOD, VOD e Diploma (come nel riquadro) e si può osservare nella finestra dell'annotazione il Metadato relativo con autore, risorsa annotata, data, ecc.

Notare che in base al tipo di annotazione effettuata, Amaya assegna un'icona particolare al testo selezionato nel documento, in questo caso una spiegazione.

Nella figura 10 è visualizzato tutto il codice RDF della pagina annotata, e nella parte evidenziata c'è l'RDF della parte selezionata in figura 9 a cui corrisponde il metadato in figura, si può notare il tipo di annotazione fatta, il titolo, l'autore, la data di creazione e modifica, la posizione nel testo della parte annotata.

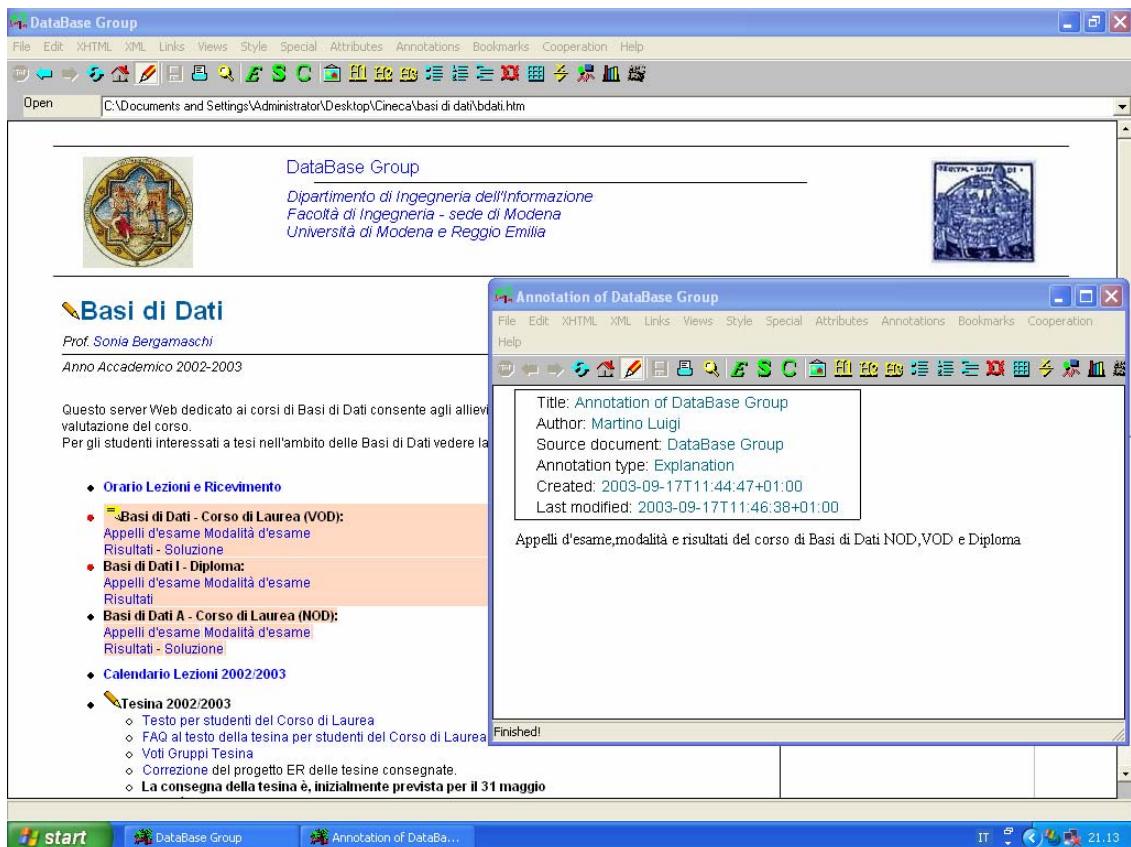


Figura 9: l'annotazione della pagina web <http://dbgroup.unimo.it/bdati/bdati.htm>

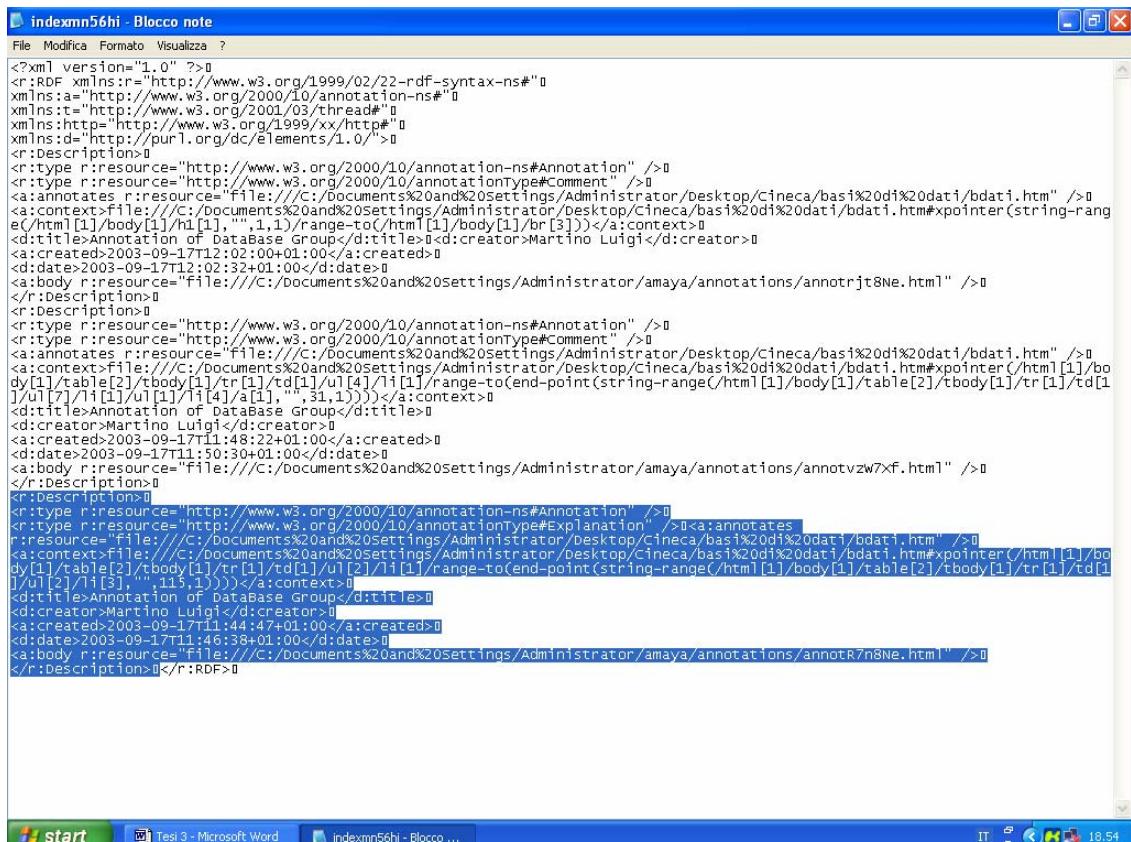


Figura10. Codice RDF della pagina annotata.

-Altre immagini di alcune pagine Web del portale del DBGROUP annotate con Amaya:

The screenshot shows a dual-pane Amaya interface. The left pane displays the homepage of the University of Modena and Reggio Emilia, featuring a navigation menu on the left and content on the right. The right pane shows an annotation for a specific page, with the following details:

Title: Annotation of University of Modena e Reggio Emilia - Facolta di Ingegneria Modena
Author: Martino Luigi
Source document: University of Modena e Reggio Emilia - Facolta di Ingegneria Modena
Annotation type: SeeAlso
Created: 2003-09-17T12:25:04+01:00
Last modified: 2003-09-17T12:26:54+01:00

Link dei servizi offerti on-line dal sito di Ingegneria di Modena www.ing.unimo.it

Annotation content (from left to right):

- Basi di Dati A**
- Docente:** Sonia Bergamaschi
- telefono:** +39 059 2056132
- fax:** +39 059 2056129
- e-mail:** bergamaschi.sonia@unimo.it
- ricevimento:** Giovedì 17-19 Ufficio
- ufficio:** Dipartimento di Ingegneria dell'Informazione, Primo Piano
- Propedeuticità**
- Consigliate:**
- Materiale Didattico:** <http://www.dbgroup.unimo.it/bdati/index.html>
- Appelli Attivi:**
- II RECUPERO PER TUTTI I CORSI DI LAUREA NOD**
- Prova Orale del 12/09/03 ore 15.00 Aula FA-1-A**
- Prova Scritta del 05/09/03 ore 10.00 Aula FA-1-B**
- Prova Scritta del 24/09/03 ore 10.00 Aula FA-1-B**

At the bottom of the annotation pane, it says "Finished!"

Figura 11.

The screenshot shows a dual-pane Amaya interface. The left pane displays the homepage of the DataBase Group, featuring a navigation menu on the left and content on the right. The right pane shows an annotation for a specific page, with the following details:

Title: Annotation of DataBase Group
Author: Martino Luigi
Source document: DataBase Group
Annotation type: Explanation
Created: 2003-09-17T12:39:24+01:00
Last modified: 2003-09-17T12:40:43+01:00

Tutte le informazioni e il materiale relativo al Seminario sulla Metodologia di progetto di applicazioni WebDB.

Annotation content (from left to right):

- Basi di Dati**
- Prof. Sonia Bergamaschi**
- Seminari relativi al corso**
- Seminario sulla Metodologia di progetto di applicazioni WebDB**
- Prof. Stefano Ceri - Politecnico di Milano**
- Venerdì 6 giugno 2003, ore 17.00**
- Aula FA-1B - Facoltà di Ingegneria - Sede di Modena, Via Vignolese 905**
- Presentazione: Prima Parte - Seconda Parte**
- Ciclo di seminari su Tecnologia XML e Corba per Sistemi di Gestione di Basi di Dati**
- Prof. Omar Boucelma - Université de Provence (Marseille)**
- Aula FA-2A - Facoltà di Ingegneria - Sede di Modena, Via Vignolese 905**
- Lucidi del Seminario**

Client-Server	Corba Applicazioni Corba	XML
● Martedì 29 maggio, ore 9-11 Introduzione alle Architetture Client/Server ed ai Sistemi Informativi in rete	○ Protocolli di rete Internet Protocolli di Alto livello Port number Telnet, ftp, SMTP, HTTP Internet, Intranet	

At the bottom of the annotation pane, it says "Finished!"

Figura 12.

2.4 Archiviazione delle annotazioni

Amaya può memorizzare i dati dell'annotazione in un file (chiamato "local annotations") oppure in un file remoto (chiamato "remote annotations"). Le annotazioni remote sono memorizzate in server dedicati per le annotazioni e possono essere scaricate e archiviate da chiunque abbia i diritti di accesso, come nel caso di altri documenti HTML. Se si vuole installare un proprio server per le annotazioni basta seguire le indicazioni come spiegato in questo sito: [Annotation-Server-HOWTO](#) [14].

Mentre per fare delle annotazioni in locale, l'utente non ha bisogno di un server. Le annotazioni in locale sono memorizzate in una speciale directory chiamata annotations, e possono essere consultate solo dall'utente stesso. Sono previsti tre tipi di files:

- annot.index

associa URLs con i files dove i metadati dell'annotazione sono archiviati.

- index + suffisso casuale

è un file che archivia il metadato dell'annotazione relativa ad un dato URL. Il metadato è espresso usando RDF.

- annot + random suffix.html

contiene il corpo dell'annotazione, archiviato come XHTML.

In qualunque momento un utente può convertire un'annotazione locale in un altro tipo, semplicemente usando il menù **Annotations/Post annotation**. Se questa operazione ha successo, l'annotazione locale sarà cancellata e sarà spostata in un server per l'annotazione.

CAPITOLO 3

OntoMat-Annotizer

3.1 CREAM il framework per l'annotazione

Il Semantic Web è fondato su metadati che descrivono i contenuti delle pagine del Web. In particolare, il Semantic Web richiede il metadato relativo per ogni tipo di specificazione, un metadato che rappresenta la descrizione di classi istanziate e descrive come sono semanticamente collegate le proprietà. Così nasce CREAM (CREAting Metadata) [15], una struttura per l'annotazione di pagine web utilizzata per permettere la creazione di metadati, e contiene lo strumento per l'annotazione che realizza questa struttura, OntoMat appunto.

Nel sistema CREAM per ora, è stato costruito un processo per l'annotazione fondato su SQL e database relazionali. Gli schemi futuri utilizzeranno XQuery o un'ontologia basata su linguaggi di interrogazione.

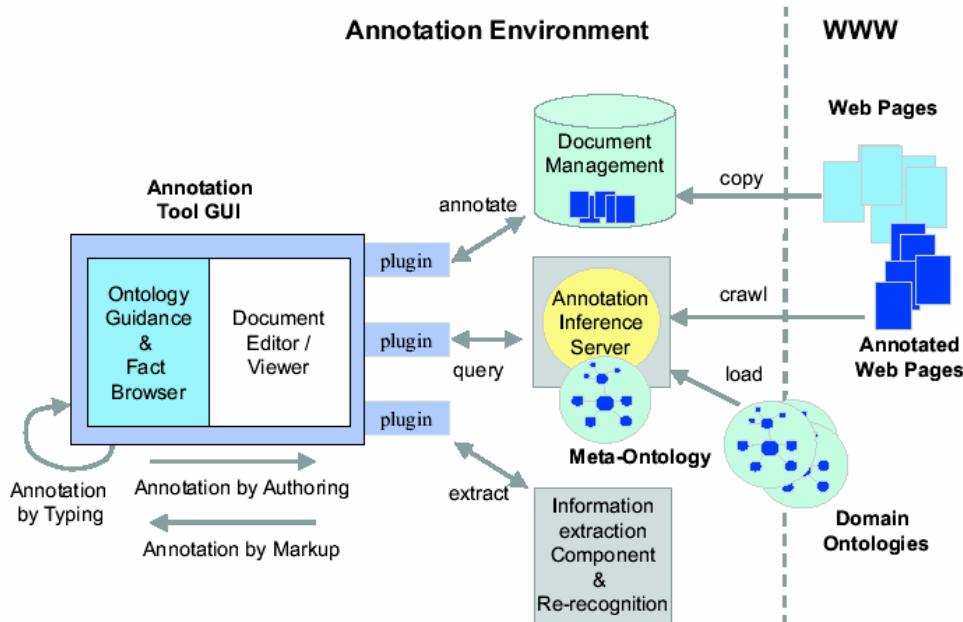


Figura 12: L'Architettura di CREAM

CREAM è una struttura per creare annotazioni. CREAM include Crowler, sistemi per il management dei documenti, guide per browser di ontologie, editors/viewers di documenti e meta-ontologie.

L'obiettivo di CREAM è quello di permettere una semplice generazione di rappresentazioni di oggetti. Questo obiettivo è ottenuto in base al tipo di interazione. Nella versione più recente di CREAM ne esistono tre diversi tipi:

1. Annotazione in base al tipo di esposizione, richiede di lavorare quasi esclusivamente con un browser per ontologie.
 2. Annotazione di markup richiede il riuso dei dati, dall'editor della pagina al browser per l'ontologia, da prima evidenziando le parti della pagina e poi facendo un drag'n'drop di questi nell'ontologia.
 3. Annotazione di authoring Web pages, che richiedono l'uso dei dati dall'ontologia e scegliendo nell'editor della pagina facendo il solito drag'n'drop.
- OntMat di solito memorizza il derivato dell'annotazione in un file HTML sotto forma di codice RDF, è anche possibile memorizzare il tutto in un file separato o in un database.
- Gli autori di CREAM dicono che d'ora in poi le annotazioni potranno perfino collegarsi l'una con le altre, creando una interconnessione dinamica per il SemanticWeb che permetterà traslazioni tra server differenti.

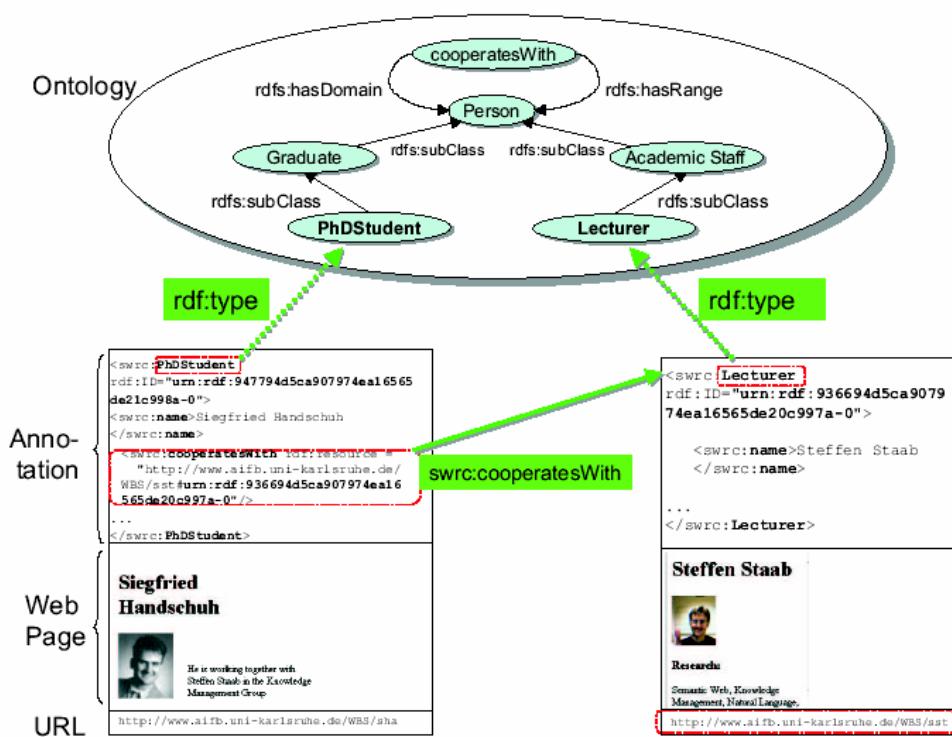


Figura 13: un esempio di annotazione

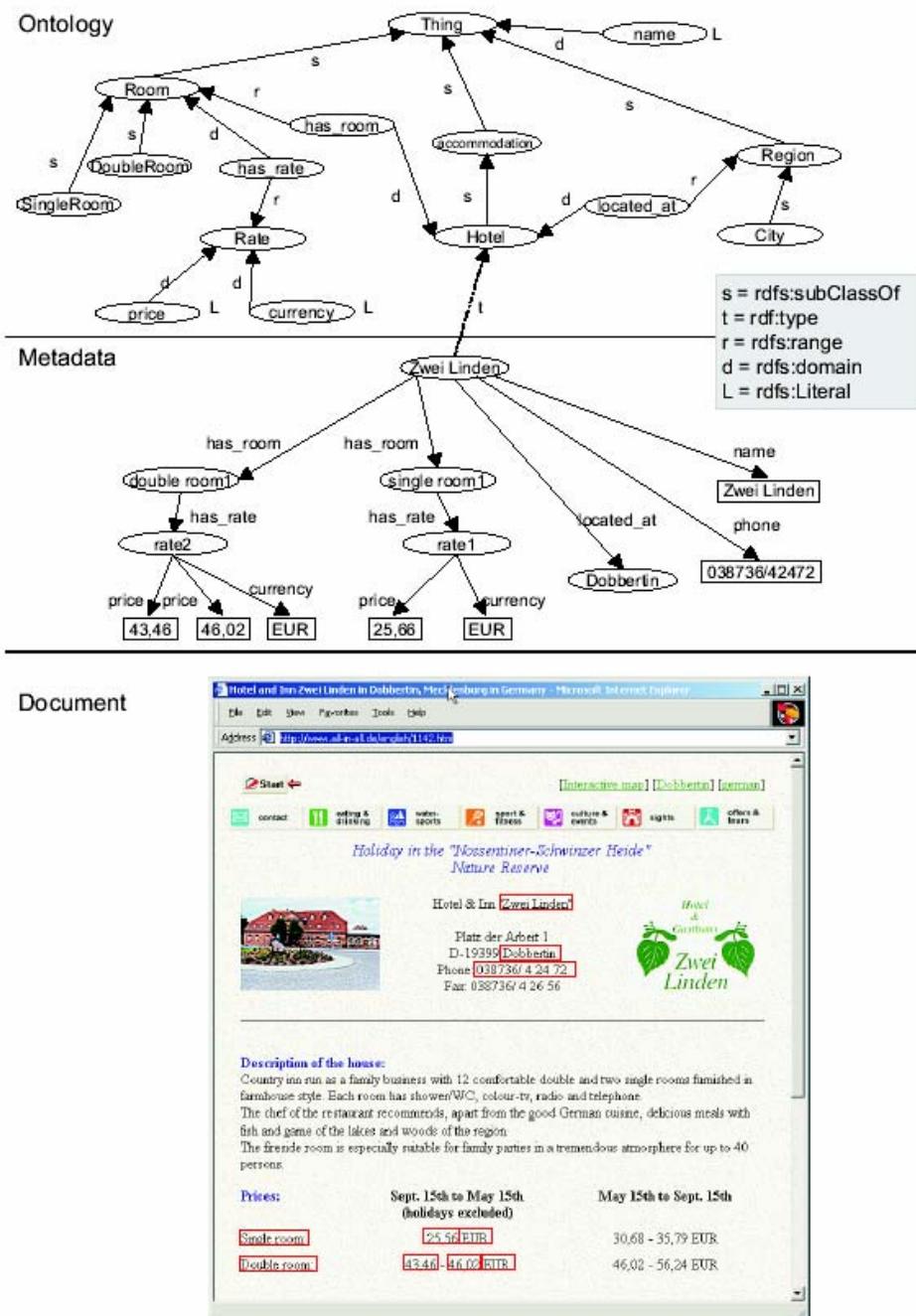


Figura 14: esempio di un'Ontologia di una pagina web e relativo metadato

Gli sviluppatori di OntoMat hanno elaborato una terminologia propria che usano per la loro struttura, perché molti dei termini che sono usati a proposito della creazione di metadata portano a connotazioni ambigue che implicano decisioni importanti per la configurazione di CREAM :

-Ontologia: Un'ontologia è un'esplicita, formale, specificazione di una concettualizzazione di un dominio di interesse. Nel nostro caso è costituita da frasi, esprimendo definizioni di classi e proprietà di DAML+OIL.

-Annotazioni: Un'annotazione nel nostro contesto è un set di istanze allegate a un documento HTML. Distinguiamo:

(i) istanze di classi DAML+OIL, (ii) proprietà di istanze da un'istanza della classe a un tipo di dato chiamato attributo dell'istanza, (iii) proprietà di un'istanza da una classe di istanza a un'altra classe d'istanza chiamata istanza della relazione.

-Metadata: Metadata sono dati che rappresentano dati. Nel nostro contesto le annotazioni sono metadati sulle pagine HTML.

-Metadata Relazionali: Usiamo il termine metadata relazionale per denotare le annotazioni che contengono istanze relazionali.

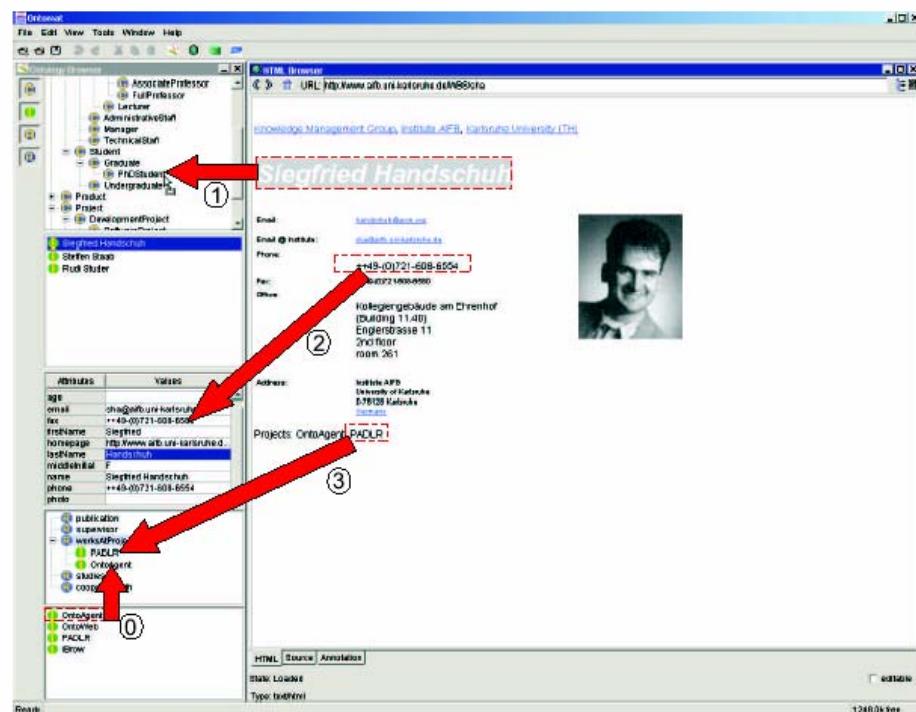


Figura 15: creazione di relazioni in un'annotazione

3.2 S-CREAM Semi-automatic CREAtion of Metadata

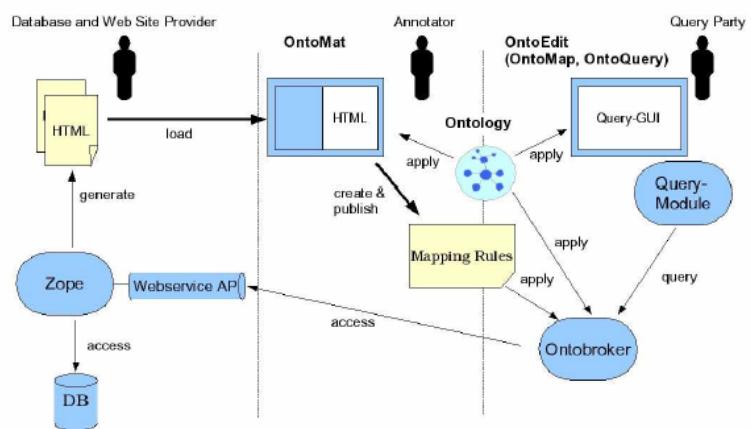
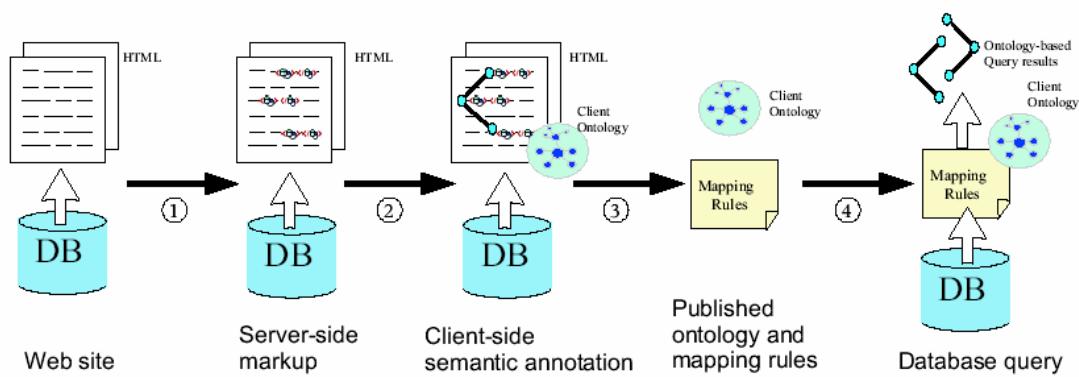
OntoMat è l'implementazione di riferimento di CREAM. CREAM è il framework per l'annotazione studiato all'università di Karlsruhe. C'è una versione di CREAM – S-CREAM – [16] che consente di estrarre le informazioni in modo automatico. Per estrarre automaticamente le informazioni occorre prima eseguire una fase di 'addestramento' del tool che esegue l'estrazione. L'utente deve fornire un set di documenti (simili) e annotarli in modo opportuno prima di poter iniziare la fase di "learning". L'output di questa fase è una serie di regole di inferenza che verranno poi usate dal tool per estrarre la conoscenza dai documenti sui quali verrà invocato. Le informazioni estratte possono essere rifinite.

Per estrarre le informazioni vengono usati due strumenti: Annie e Amilcare, per trovare i token dei documenti (parole, frasi, etc.) e per indurre le regole di estrazione. L'annotazione che può anche essere fatta manualmente, fa riferimento ad una o più ontologie.

Nella implementazione di SCREAM, OntoMat supporta ora l'annotazione semi-automatica delle pagine del Web. Questa annotazione semi-automatica è basata su un componente per estrarre informazioni, Amilcare. OntoMat estrae con l'aiuto delle strutture di conoscenza di Amilcare da pagine Web e attraverso l'uso di regole di estrazione già note. Queste regole sono il risultato di una fase di "learning" basato su pagine già annotate.

Amilcare è un sistema che apprende le regole per l'estrazione delle informazioni.

Amilcare è uno strumento adatto per l'estrazione di informazioni da testi ed è progettato per sostenere l'annotazione di documenti per Knowledge Management (KM). Esso esegue IE arricchendo i testi con annotazioni XML; ad esempio il sistema segna le informazioni estratte con annotazioni XML.



3.3 ONTOMAT

3.3.1 Installazione e requisiti

OntoMat-Annotizer finora è stato testato solo per **Windows** e **Linux**, quindi non si conoscono ancora gli sviluppi che si potrebbero avere su altre piattaforme [17].

- Istruzioni per Windows :
 - Dopo aver scaricato il file, fai doppio-click su **install.exe**
 - E' richiesta l'installazione di Java 1.4.0 (o superiore) virtual machine. E' possibile scaricarla da: Sun's Java web site.
- Istruzioni per Linux :
 - Dopo aver scaricato il file apri una shell e cd alla directory dove hai scaricato il file.
 - Nel prompt digita: **sh ./install.bin**
 - E' richiesta l'installazione di Java 1.4.0 (o superiore) virtual machine. E' possibile scaricarla da: Sun's Java web site

3.3.2 Funzioni

3.3.2.1 Operazioni da eseguire all'avvio

All'avvio di OntoMat bisogna compiere delle operazioni di caricamento dei vari browser utilizzati per effettuare l'annotazione, cioè bisogna aprire l'Ontology browser e il Web browser. L'apertura dell'Ontology browser consiste nel caricare la DAML-OIL ontology, e questo è possibile farlo sia collegandosi via URI, scaricando il file <http://annotation.semanticweb.org/iswc/iswc.daml> oppure in locale una volta salvato questo file sul proprio hard disk. Ovviamente questa ultima possibilità permette di lavorare senza dover necessariamente essere collegati al web.

3.3.2.2 Creazione di un'annotazione

Per creare un'annotazione, una volta aperto il documento HTML, basta scegliere nell'Ontology browser quale ontologia tra quelle presenti è quella che si addice di più al tipo di annotazione da fare, quindi selezionare il testo che farà parte dell'annotazione. Dopo di che bisognerà creare l'istanza della classe che conterrà l'annotazione e questo è possibile farlo semplicemente con un click (tasto destro del mouse) sull'ontologia prima selezionata. Ad ogni istanza sono assegnati una serie di attributi con cui completare l'annotazione.

3.3.2.3 Creazione di relazioni tra istanze

L'operazione di annotazione può essere arricchita perché con OntoMat è possibile creare delle relazioni tra istanze, visto che in base alla scelta fatta (classe), OntoMat già preseleziona le istanze della classe con cui fare delle relazioni (vedi figura 6).

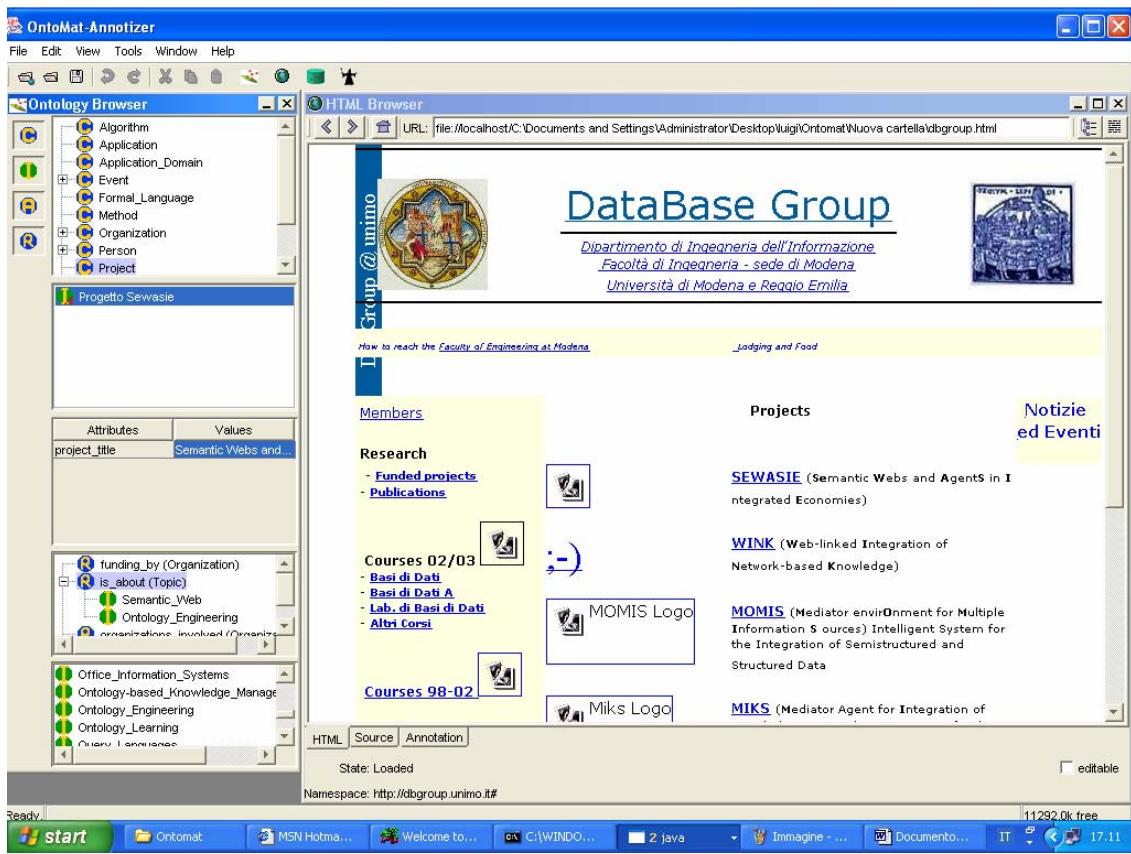


Figura 17:Annotazione di una pagina web

3.3.2.4 Salvataggio di un'annotazione

Per salvare l'annotazione c'è il comando “write annotation into document” dal menù “tools”, che memorizzerà le istanze in un tag del codice HTML del documento web. Mentre per salvare la pagina HTML annotata, sempre dallo stesso menù c'è il comando “save HTML file”.

3.3.3 Test su sito web

Sono state annotate varie pagine del sito www.dbgroup.unimo.it. Qui presento alcune immagini di queste pagine:

-In questa immagine è visualizzata l'operazione di annotazione della pagina web <http://dbgroup.unimo.it/bdati/MatDid02-03A.html> che rappresenta la pagina che contiene le informazioni e i file del materiale didattico utilizzati nel corso di Basi di Dati.

Qui si può notare l'annotazione che riguarda il libro di testo utilizzato, e viene creata la classe “Book” indicata da “Libri di testo”. Vengono creati gli attributi della classe cioè il titolo e l'anno del libro (ma se ne potevano creare anche degli altri).

Inoltre ho creato una relazione con questa classe ed è “formal-language” che ho completato con l’istanza “SQL” relativa al libro in questione.

Bisogna dire che dopo l’annotazione non rimane nessuna traccia visibile dell’annotazione fatta, a differenza di Amaya che utilizza dei simboli.

Il contenuto dell’annotazione è sotto forma di codice RDF nell’HTML della pagina.

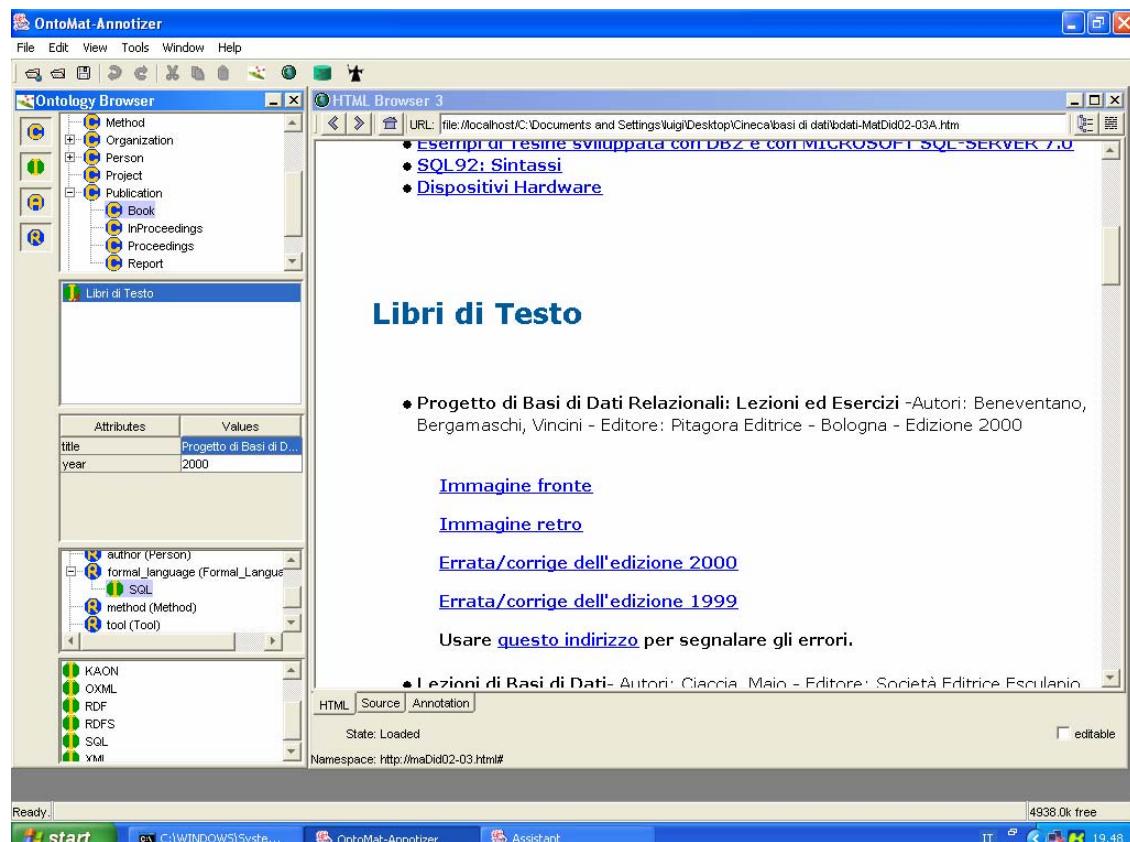


Figura 18: annotazione della pagina web
<http://dbgroup.unimo.it/bdati/MatDid02-03A.html>

Presento qui una immagine che visualizza il codice RDF della pagina prima annotata, e nel codice si può osservare il riferimento alla classe creata, all'indirizzo dell'pagina(che però non è quello reale ma solo un riferimento), alle relazioni e agli attributi creati.

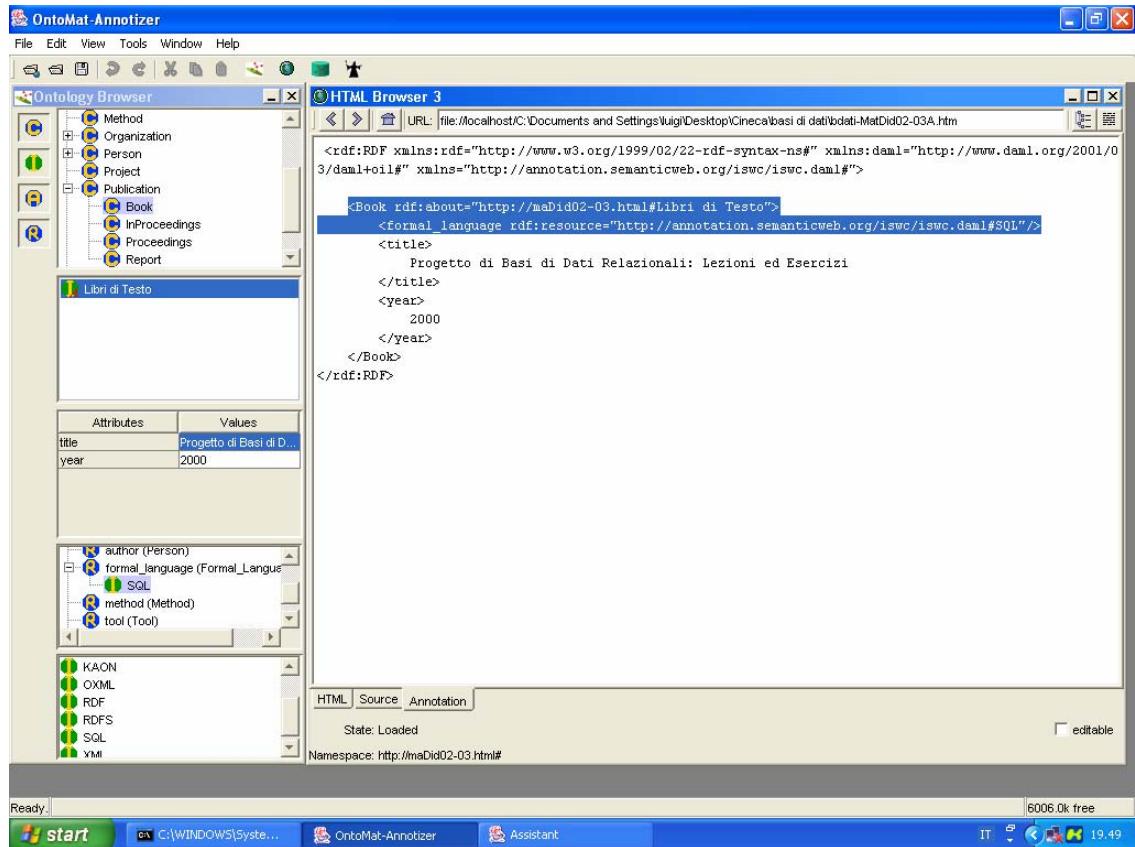


Figura 19. codice RDF della pagina prima annotata.

Altri esempi di pagine annotate con OntoMat:

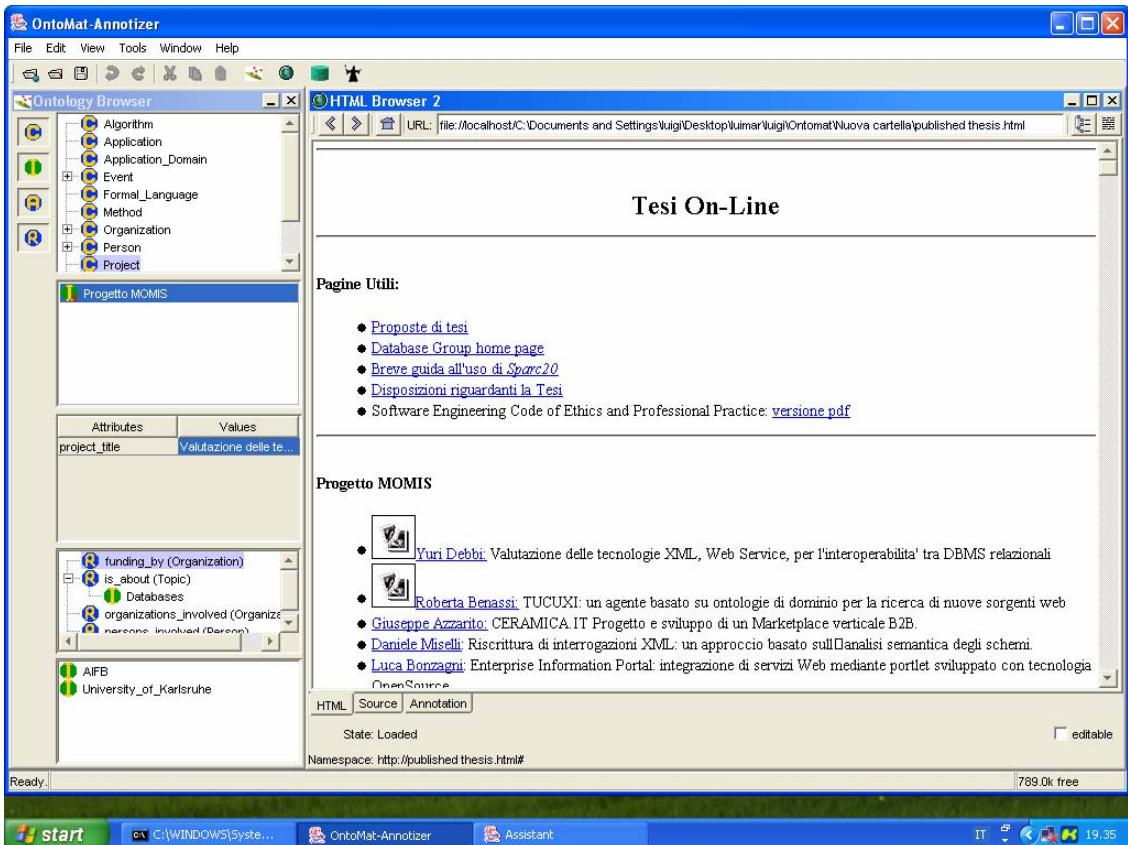


Figura 20.

3.4 Archiviazione delle annotazioni

Con OntoMat le annotazioni vengono direttamente salvate nel codice HTML della pagina Web sotto forma di codice RDF.

Quindi nel file verranno specificate tutte le informazioni dell'annotazione, dai riferimenti alle ontologie usate, il corpo vero e proprio dell'annotazione, e il riferimento all'URL della pagina web.

Nella figura 21 si può notare il codice RDF della pagina web annotata, in particolare:

```
<Association rdf:about="http://DataBase Group members.htm#Full professor Sonia Bergamaschi ">
```

```
  <name>
    Sonia Bergamaschi
  </name>
</Association>
```

che specifica la creazione dell'ontologia selezionata (Association) e il valore di "input dialog" cioè il commento (Full professor Sonia Bergamaschi).

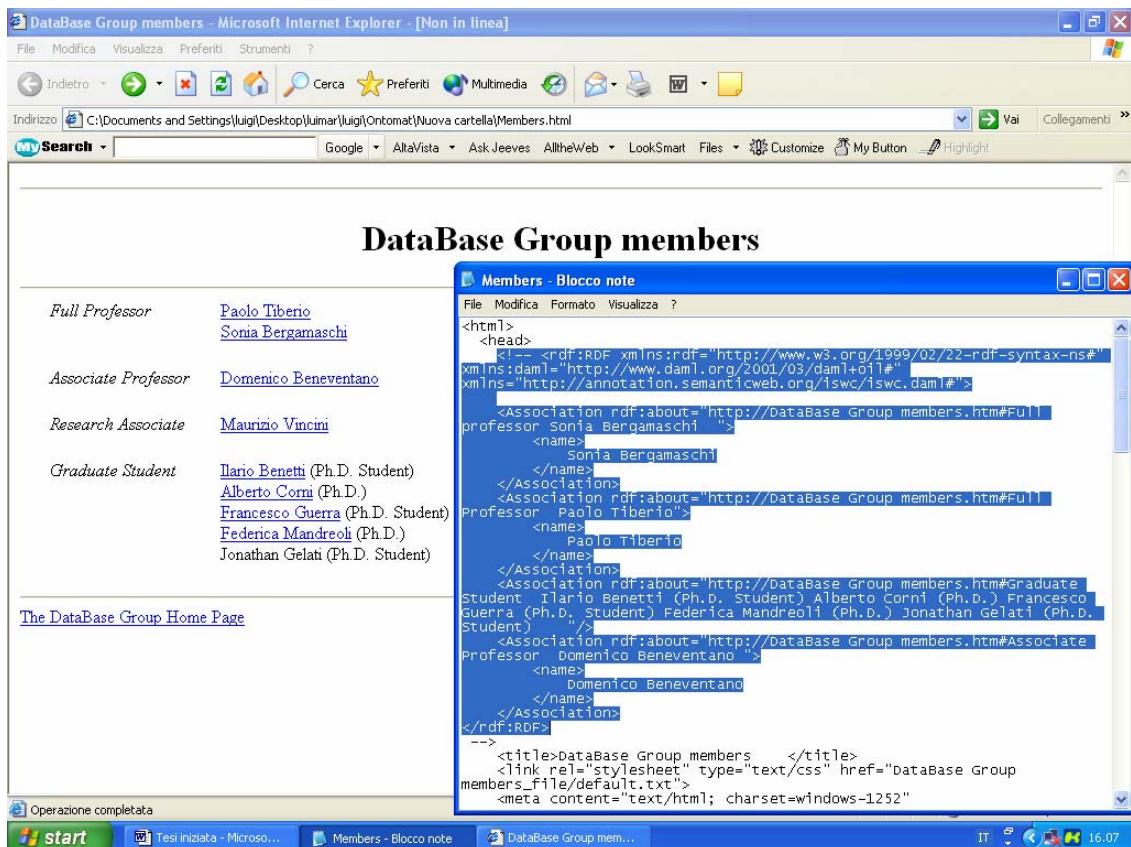


Figura 21:visualizzazione del codice RDF relativo all'annotazione

CAPITOLO 4

Confronto fra i due tools

	Facilità d'uso	Portabilità	Aggiornamento/ sviluppo	Ontologie	Archiviazione delle annotazioni
Annotea	Molto semplice e di facile comprensione	Finora è possibile utilizzarlo sia con Windows che con Linux	Attualmente è disponibile una versione con capacità un po' limitate, comunque il codice è disponibile per eventuali estensioni. Si prevede però un motore di ricerca che sfrutti l'annotazione delle pagine	Non vengono menzionate, le annotazioni non sono fatte rispetto a nessuna ontologia	In locale o anche in remoto, espresse in RDF
OntoMat	E' presente una guida che rende di facile uso le operazioni principali	Funziona sia con Windows sia con Linux. Recentemente sono uscite delle versioni anche per altre piattaforme	E' sempre possibile aggiornare le Ontologie scaricandole dal sito, o crearne delle proprie. In continuo sviluppo.	Sono divise in classi, e coprono vari aspetti di interesse.	Anche qui è possibile salvarle in locale o in remoto, sempre in RDF

CAPITOLO 5

Gli altri tools

5.1 Smore

SMORE: Semantic Markup, Ontology and RDF Editor [18]

SMORE è uno strumento che permette all’utente di arricchire i propri documenti in RDF usando ontologie Web in relazione a specifici termini ed elementi dell’utente stesso.

Lo scopo di questo software è il seguente:

- Fornire all’utente un ambiente flessibile nel quale può creare la sua pagina web senza troppi ostacoli.
- Permettere all’utente di arricchire i propri documenti con minime conoscenze dei termini e della sintassi RDF.
Comunque, l’utente deve essere in grado di classificare semanticamente i suoi dati per annotazioni;
- Fornire un riferimento a ontologie esistenti su Internet, per permettere di usare più precisi riferimenti sulla propria pagina web. L’utente può anche creare la sua ontologia da appunti e prendere in prestito termini da ontologie già esistenti.
- Garantire un accurato e completo RDF, con lo scopo di rendere semplici le eventuali modifiche.

Questo tool è sviluppato all’interno del progetto MINDSWAP [19].

Consente di annotare anche immagini (o parti di esse) con PhotoSMORE. Vuole ridurre la differenza tra la creazione dei contenuti (creazione di pagine web – authoring) e la fase di annotazione. SMORE ha all’interno un html editor. Il processo di annotazione è più o meno simile agli altri: si seleziona una parte del testo e si creano delle triple (tipo RDF). Gli elementi delle triple possono essere associati agli elementi di ontologie anche in un momento successivo alla creazione della tripla stessa.

C’è un componente – MailSMORE – che consente di arricchire le mail scritte rispettando certe strutture, di triple collegate ad ontologie relative alla posta elettronica. Questo framework consente di modificare, navigare, cercare ontologie, aiutando in questo modo la ricerca dell’ontologia più adatta da associare all’annotazione che si sta effettuando.

E’ inoltre possibile cercare dinamicamente l’ontologia che più si addice al contesto specifico. C’è un sistema che inferisce alcuni tipi di triple e le propone all’utente che può eventualmente cancellarle.

Scaper è un modulo che consente di analizzare il contenuto di pagine web alla ricerca di strutture predefinite quali tabelle, campi, liste e creare automaticamente triple RDF.

C'è un sistema per la gestione delle ontologie – Parka-DB – che si affida ad un DBMS e gestisce le triple RDF per fare inferenza.

Il sito internet da cui prendere ulteriori informazioni è:

<http://www.mindswap.org/> oppure <http://www.mindswap.org/~aditkal/editor.shtml>

5.2 MnM

MnM: Ontology Driver Semi-Automatic and Automatic Support for Semantic Markup [20]

MnM è uno strumento per l'annotazione che è in grado di supportare l'annotazione sia automatica che semi-automatica di pagine web con contenuto semantico.

MnM integra un web browser con un editor per l'ontologie e fornisce APIs da collegare ai server per le ontologie e per l'integrazione di informazioni tramite strumenti per l'estrazione delle informazioni.

E' un progetto molto recente, infatti solo in questi periodi hanno creato la home page del progetto ed è in rapida evoluzione.

Le cose più interessanti sono:

- possibilità di annotare rispetto ad ontologie ben precise;
- possibilità di navigare le varie ontologie rispetto a cui annotare i documenti;
- possibilità di annotare automaticamente dei documenti dopo una fase di "learning";
- controllo di validità dei risultati rispetto alle ontologie usate.

Gli autori del progetto dicono che il loro sistema è simile ad **OntoMat**.

Il tool richiede diversi pacchetti software per funzionare (per adesso funziona solo su Windows 2000).

Questo tool fa uso di Annie e Amilcare, per trovare i token dei documenti (parole, frasi, etc.) e per indurre le regole di estrazione.

In futuro prevedono di utilizzare l'annotazione così creata per migliorare le ricerche facendo anche uso di agenti.

Il sito ufficiale per avere ulteriori informazioni è <http://kmi.open.ac.uk/projects/akt/MnM/index.html>.

CAPITOLO 6

Conclusioni

In questa tesi ho presentato il lavoro fatto su due tools, **Annotea** ed **OntoMat**, che sono stati utilizzati per realizzare delle annotazioni su alcune pagine del portale DBGROUP.

Sono stati fatti vari tipi di confronti fra i due tools e ho rilevato che entrambi, sono due strumenti molto utili per effettuare questo tipo di operazioni, e rappresentano due standard che sicuramente acquisteranno grande importanza nello sviluppo futuro del Semantic Web.

I risultati da me ottenuti sono stati sicuramente soddisfacenti, perché ho verificato le potenzialità di questi strumenti e capito l'importanza che hanno nel mondo del Semantica Web, visto che i due tools rappresentano bene la loro visione di Semantic Web e adottano degli approcci diversi l'uno dall'altro.

Basta dare un'occhiata alla griglia (vedi capitolo 4), per verificare che Annotea non permette di gestire le ontologie tramite il suo editor Amaya, al contrario di OntoMat che basa l'operazione di annotazione proprio sulle ontologie; per questo motivo direi che OntoMat è sicuramente più completo e rispecchia di più la visione del Semantic Web.

Per quanto riguarda l' esportazione dei dati, Annotea esporta i dati in RDF schema così come OntoMat, quindi adottano entrambi il sistema di esportazione più avanzato.

Nelle prospettive di lavoro future diciamo che esistono anche altri tipi di tools che permettono di migliorare e di ampliare la visione del Semantic Web, proprio perché ogni gruppo di lavoro segue una sua linea guida, un proprio tipo di approccio che permette sicuramente di espandere il Semantic Web.

Sarebbe interessante realizzare delle Ontologie, e per quanto possibile per svariati argomenti, e cercare di pubblicarle o di renderle disponibili presso una banca dati.

Ancora più interessante sarebbe poi realizzare un motore di ricerca che sfrutti tutte queste disponibilità, cioè che metta insieme le Ontologie e generi delle risposte alle ricerche sul web che non siano basate sulla sintassi (Keyword ecc..) ma bensì sulla semantica.

I vantaggi di questa soluzione sarebbero enormi, da tanti punti di vista (tempo, denaro, risorse, ...) e snellendo così una buona parte del Web che oggi è pieno di informazioni e documenti ammassati e intrecciati fra di loro.

BIBLIOGRAFIA

[1]Annotea. Home page: <http://www.w3.org/2001/Annotea>

[2]OntoMat. Home page: <http://annotation.semanticweb.org/tools/ontomat>

[3]W3C. Home page: <http://www.w3.org>

[4]SemanticWeb. Home page: <http://www.w3.org/2001/sw>

[5] T. R. Gruber. A translation approach to portable ontology specifications. In Knowledge Acquisition, volume 5, pagina 199.

[6] N. Guarino. Formal ontologies and information systems. In Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS'98), Trento, 1998.

[7]Riferimento al sito <http://dev.w3.org/cvsweb/checkout/Amaya/doc/amaya/Attic/>

[8] RDF. Riferimento al sito <http://www.w3.org/TR/rdf-schema/>

[9] XML. Riferimento al sito <http://www.w3.org/XML/>

[10]Annotea protocol. Riferimento al sito
<http://www.w3.org/2002/12/AnnoteaProtocol-20021219>

[11]Annotea Bookmark Schema. Riferimento al sito
<http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707>

[12]Amaya. Home page: www.w3.org/amaya

[13]Amaya public release. Riferimento al sito: www.w3.org/amaya/user/bindist.html

[14]Amaya, Annotation-Server. Riferimento al sito:
<http://www.w3.org/1999/02/26-modules/user/annotation-HOWTO.html>

[15]CREAM. Riferimento al file pdf del sito:
http://www.aifb.uni-karlsruhe.de/WBS/sha/papers/aa_cream_www11.pdf

[16]S-CREAM. Riferimento al file pdf del sito :
<http://www.aifb.uni-karlsruhe.de/WBS/sst/Research/Publications/ekaw2002scream-sub.pdf>

[17]Installazione di OntoMat :

<http://annotation.semanticweb.org/ontomat/installV004.html#windows>

[18]SMORE. Home page: <http://www.mindswap.org/>

[19]Mindswap. Home page al sito: <http://www.mindswap.org/>.

[20]MnM. Home page: <http://kmi.open.ac.uk/projects/akt/MnM/index.html>