

Politecnico di Milano

Facoltà di Ingegneria dell'Informazione

Corso di Laurea in Ingegneria Informatica

Dipartimento di Elettronica e Informazione



SPEAKIN' ABOUT – UNO STRUMENTO PER L'ANNOTAZIONE SEMANTICA DI TESTI

Relatore: Prof. Marco COLOMBETTI
Correlatore: Ing. Davide EYNARD

Elaborato di Laurea di:
Marco POLONIO
Matteo REBULLA

Sommario

- ★ Il Web
- ★ Cos'è un'annotazione?
- ★ Cos'è un'ontologia?
- ★ Definizione del problema
- ★ Analisi
- ★ Architettura
- ★ Sequence Diagram e Esempi di utilizzo
- ★ Conclusioni

Il Web

★ Presente

Rete che contiene un gran numero di documenti con molte informazioni non strutturate in maniera logica.

★ Futuro

Rete in cui le informazioni contenute nei documenti sono associate al loro significato e alla loro area del sapere (Web semantico).

Cos'è un'annotazione?

- ★ **Annotazione manuale**

Annotazione aggiunta da un utente.

- ★ **Annotazione automatica**

Annotazione aggiunta dal computer senza consultazione dell'utente.

- ★ **Annotazione semi-automatica**

Annotazione suggerita dal computer che ha bisogno della conferma dell'utente per essere aggiunta.

- ★ **Annotazione semantica**

Annotazione che descrive il significato di una parte di testo o che è essa stessa descritta tramite semantica.

Cos'è un'ontologia?

Organizzazione logica di una certa area del sapere.

- ★ Documento RDF o OWL
- ★ Computer-readable
- ★ Utilizzabile da software di ragionamento
- ★ Divisa in TBOX (conoscenze generali) e ABOX (conoscenze particolari)

Definizione del problema

★ Problemi del Web semantico

- L'utente medio del Web non conosce le nuove tecnologie;
- Le informazioni su Internet non sono associate al loro significato.

★ Risoluzione del problema

- Il nostro software, tramite interfaccia grafica, permette a un qualunque utente del Web di utilizzare le tecnologie indirettamente e di associare una porzione di testo al suo significato.

Analisi

Step 1

HARRY POTTER

Harry Potter è anche un fenomeno cinematografico: la major Warner Bros, acquistati i diritti sull'opera nel 1999, realizza le trasposizioni cinematografiche di ogni singolo libro.

Speakin' About

Copy

Select All

Search Google for "Harry Potter"

View Selection Source

Properties

Step 2

Stringa selezionata:

Harry Potter

URL:

http://it.wikipedia.org/wiki/Harry_Potter

Selezionare l'area di interesse...

Selezionare il concetto...

Submit

Cancel

Libri

Cinema

Videogiochi

...

Film

Attore

Regista

...

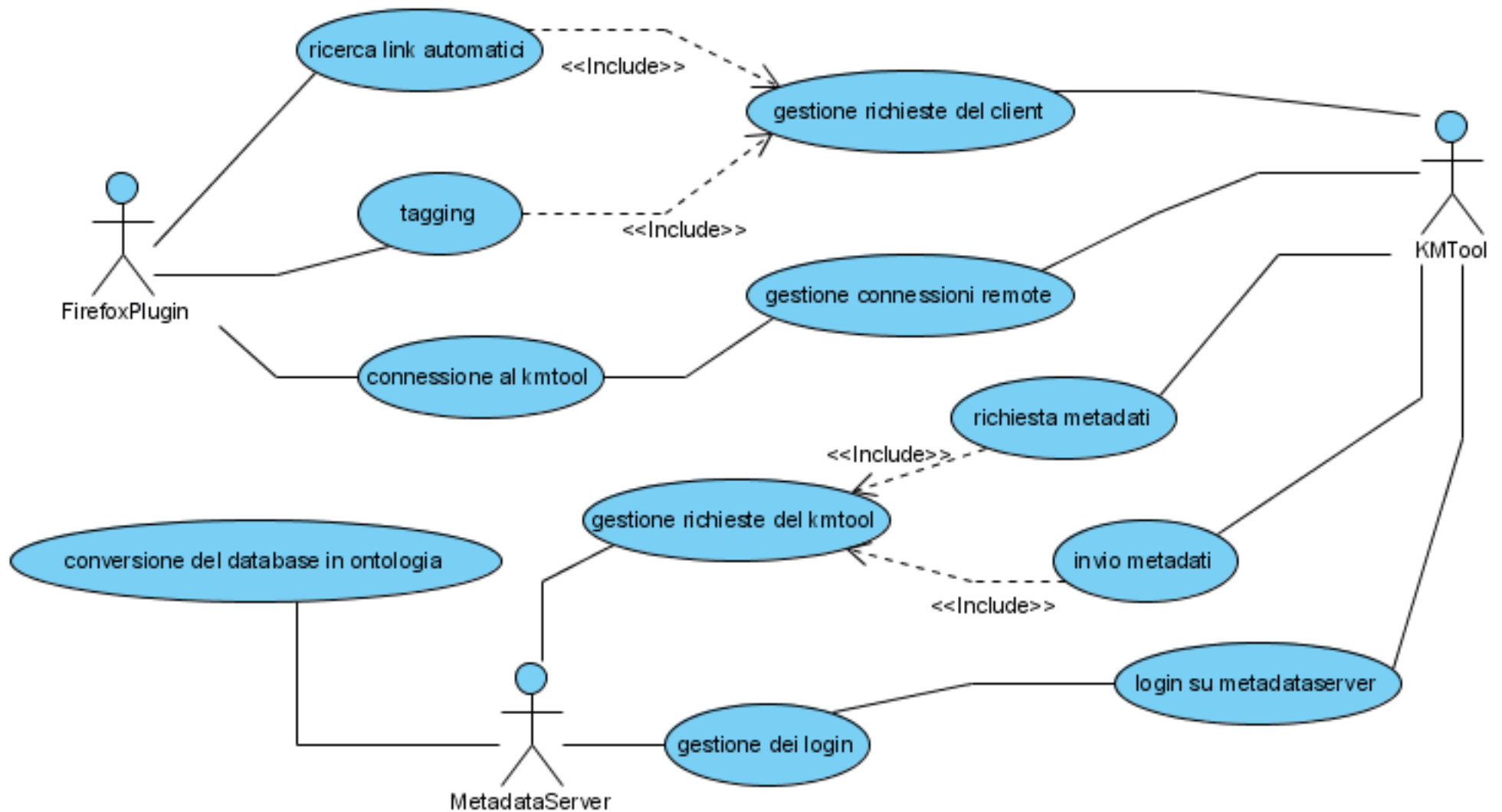
Step 3

HARRY POTTER

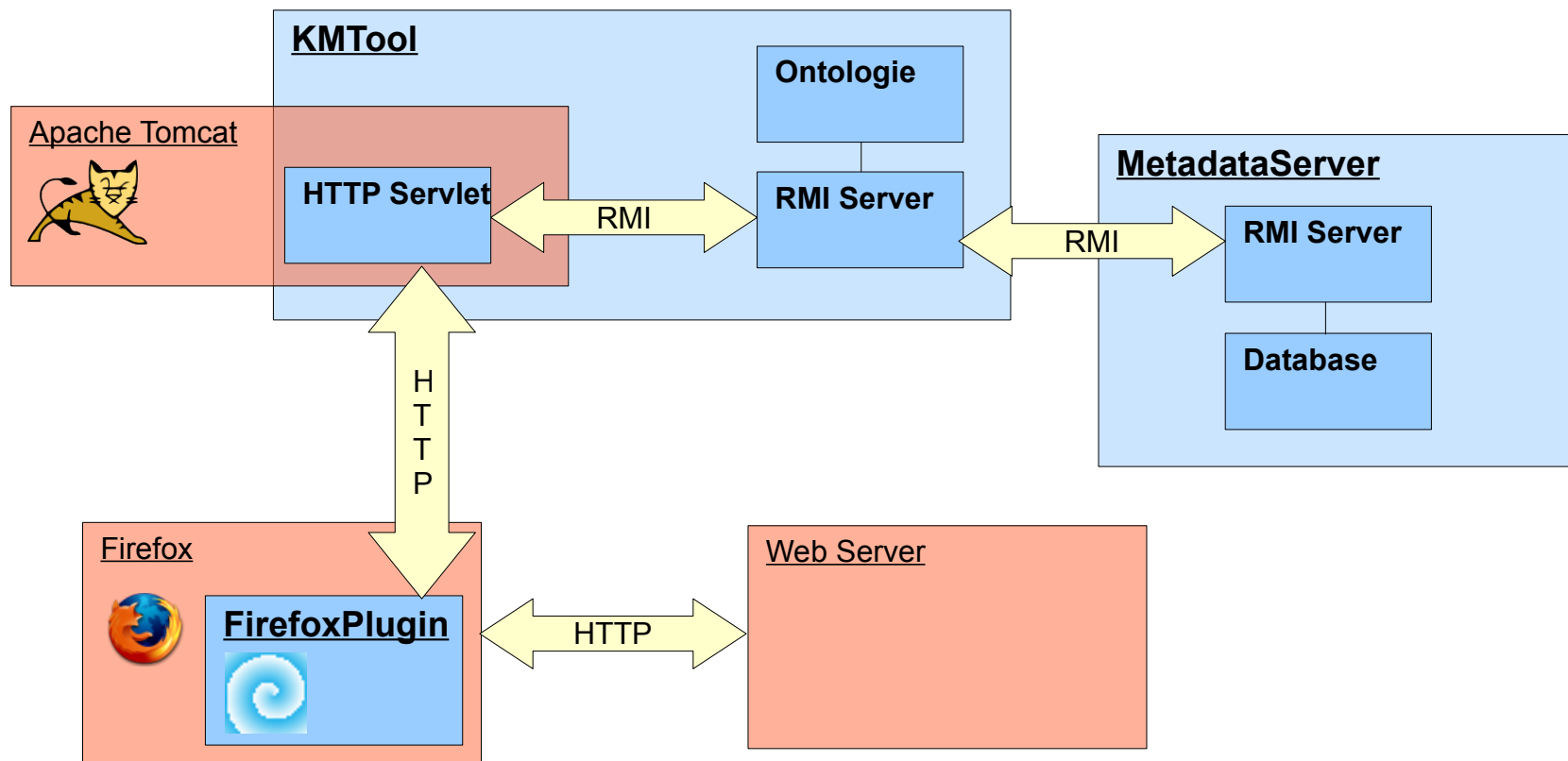
[Harry Potter](http://harrypotter.warnerbros.com) è anche un fenomeno cinematografico: la major Warner Bros, acquistati i diritti sull'opera nel 1999, realizza le trasposizioni cinematografiche di ogni singolo libro.

<http://harrypotter.warnerbros.com>

Analisi



Architettura



MetadataServer

★ RMI Server

- Gestione connessioni KMTool
- Gestione database dei metadati

★ Database

- Database SQL di una sola tabella:

METADATO(ID, URL, stringa, ontologia, individuo)

KMTool

★ RMI Server

- Gestione connessione al MetadataServer
- Gestione connessione in locale all'HTTP Servlet
- Lettura delle ontologie

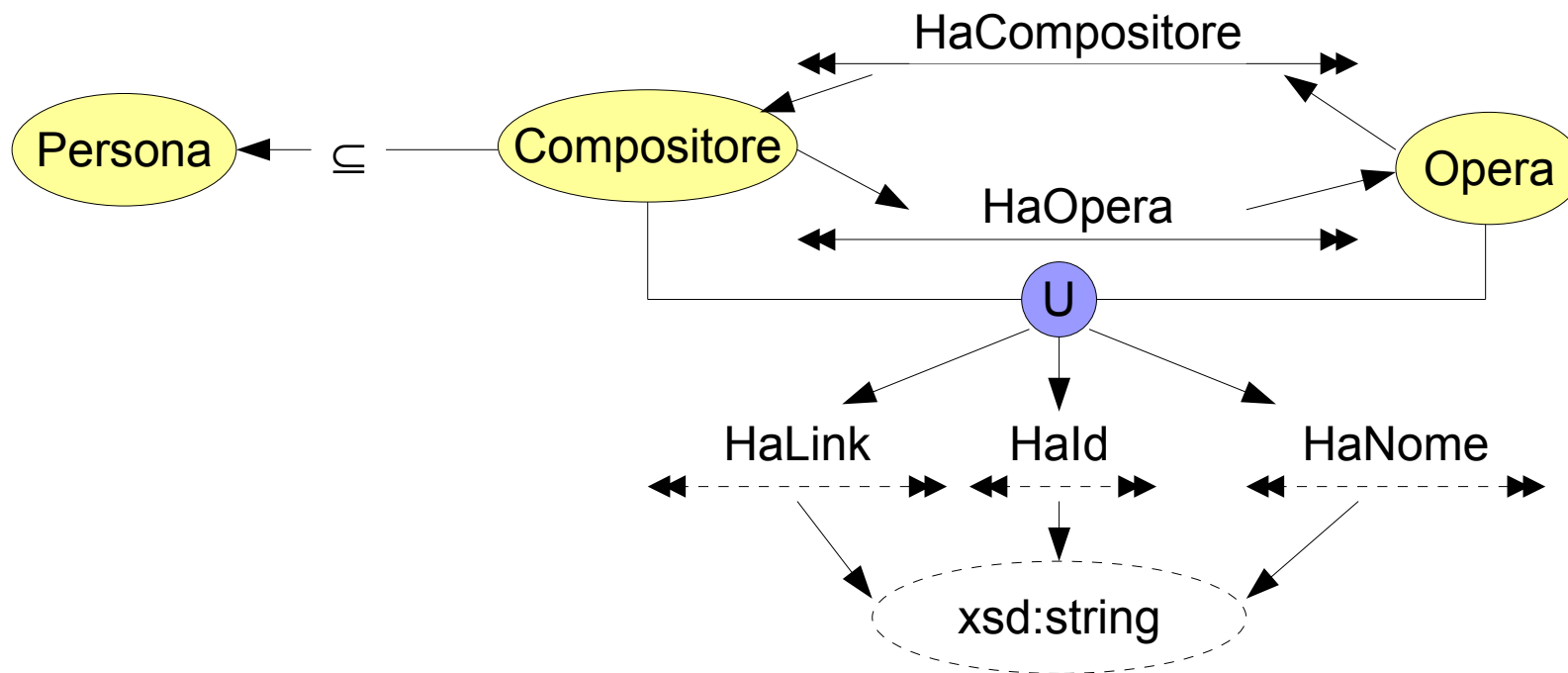
★ HTTP Servlet

- Gestione richieste FirefoxPlugin
 - KmServlet, per la fase di tagging (invio metadati)
 - AutoLinkServlet, per la fase di autolinking (recupero metadati)

★ Ontologie (slide seguente)

Ontologie del KMTool

★ Musica_Classica.owl

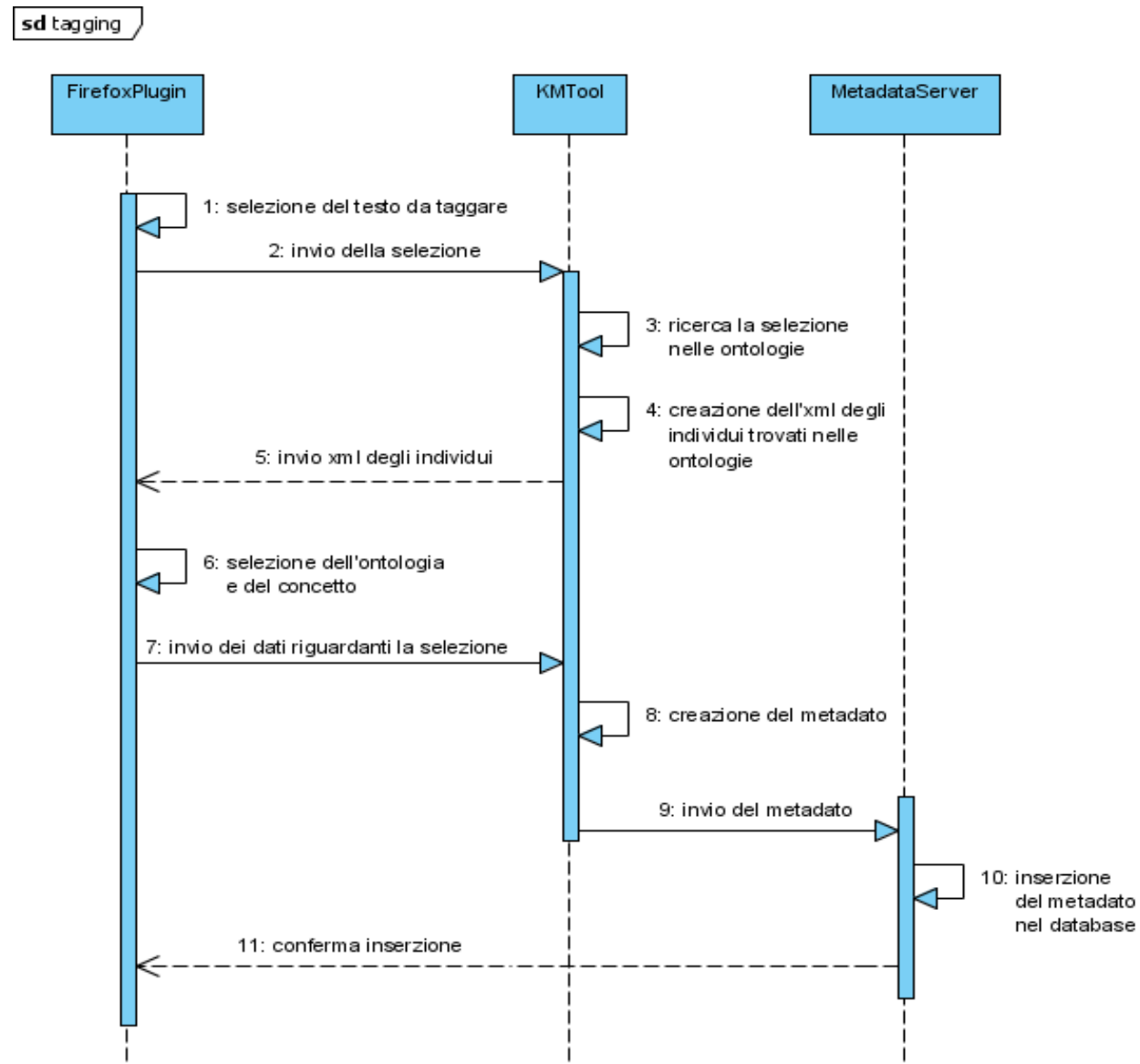


FirefoxPlugin

★ File `speakinAboutOverlay.js`

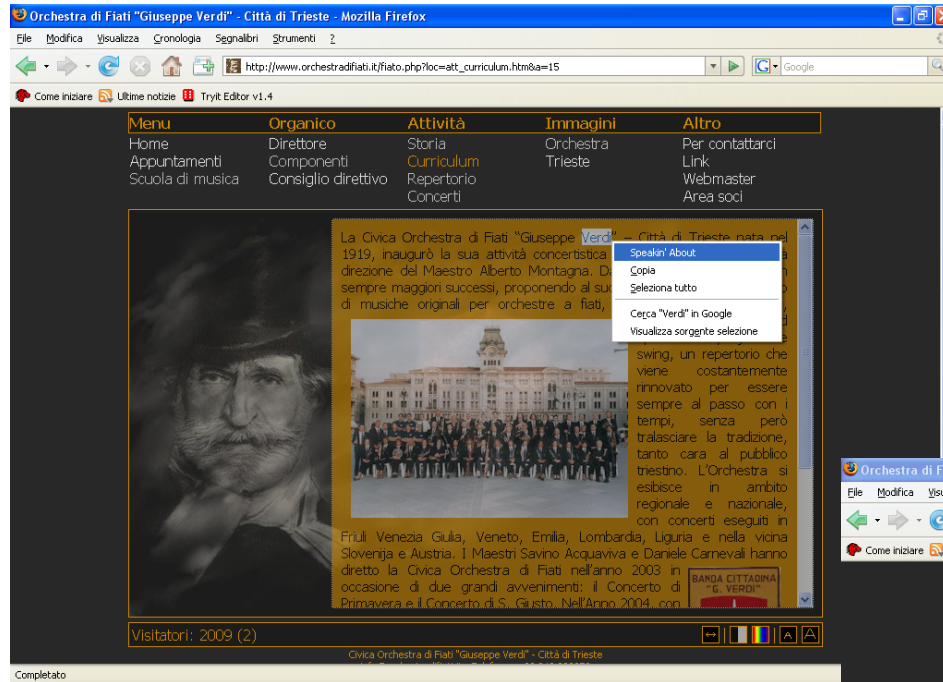
- Function `speakinabout`: Avvia la fase di tagging connettendosi al sito del `KmServlet` (`index.jsp`);
- Function `pageLoad`: Avvia la fase di autolinking intercettando l'evento di apertura di una nuova pagina;
- Function `connect`: Richiede i metadati della pagina connettendosi all'`AutoLinkServlet`;
- Function `genLink`: Genera i link automatici nella pagina.

Sequence Diagram

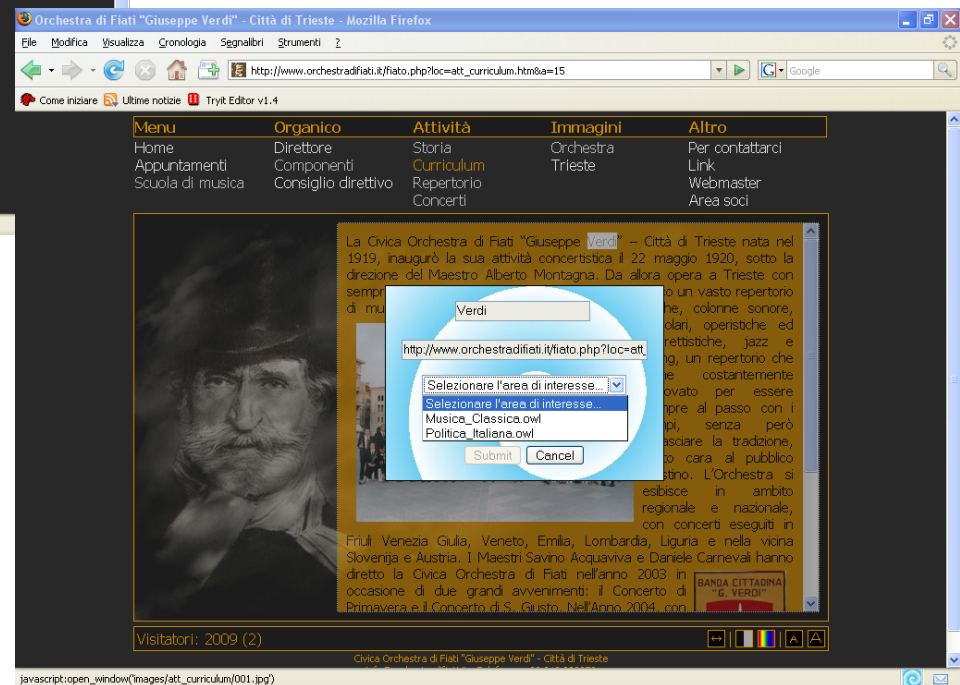


Esempio di utilizzo

1: selezione del testo da taggare

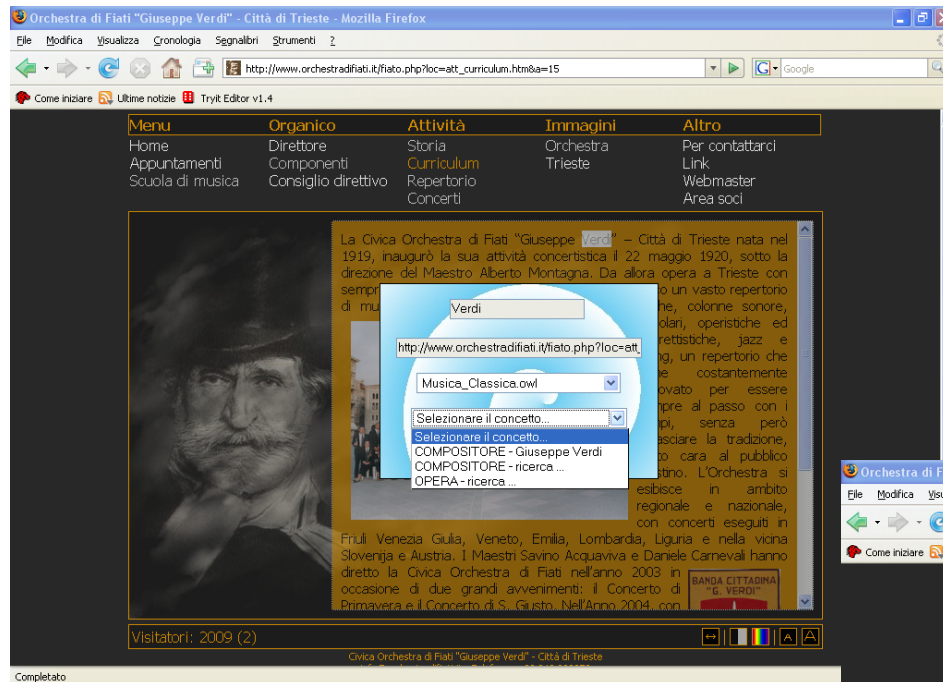


6a: selezione dell'ontologia

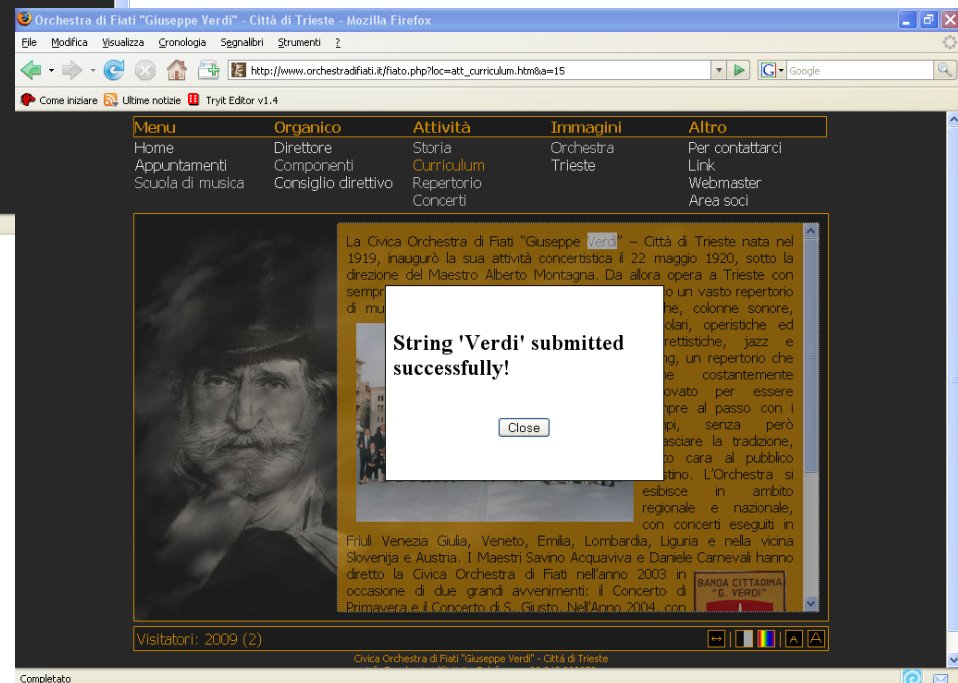


Esempio di utilizzo

6b: selezione del concetto

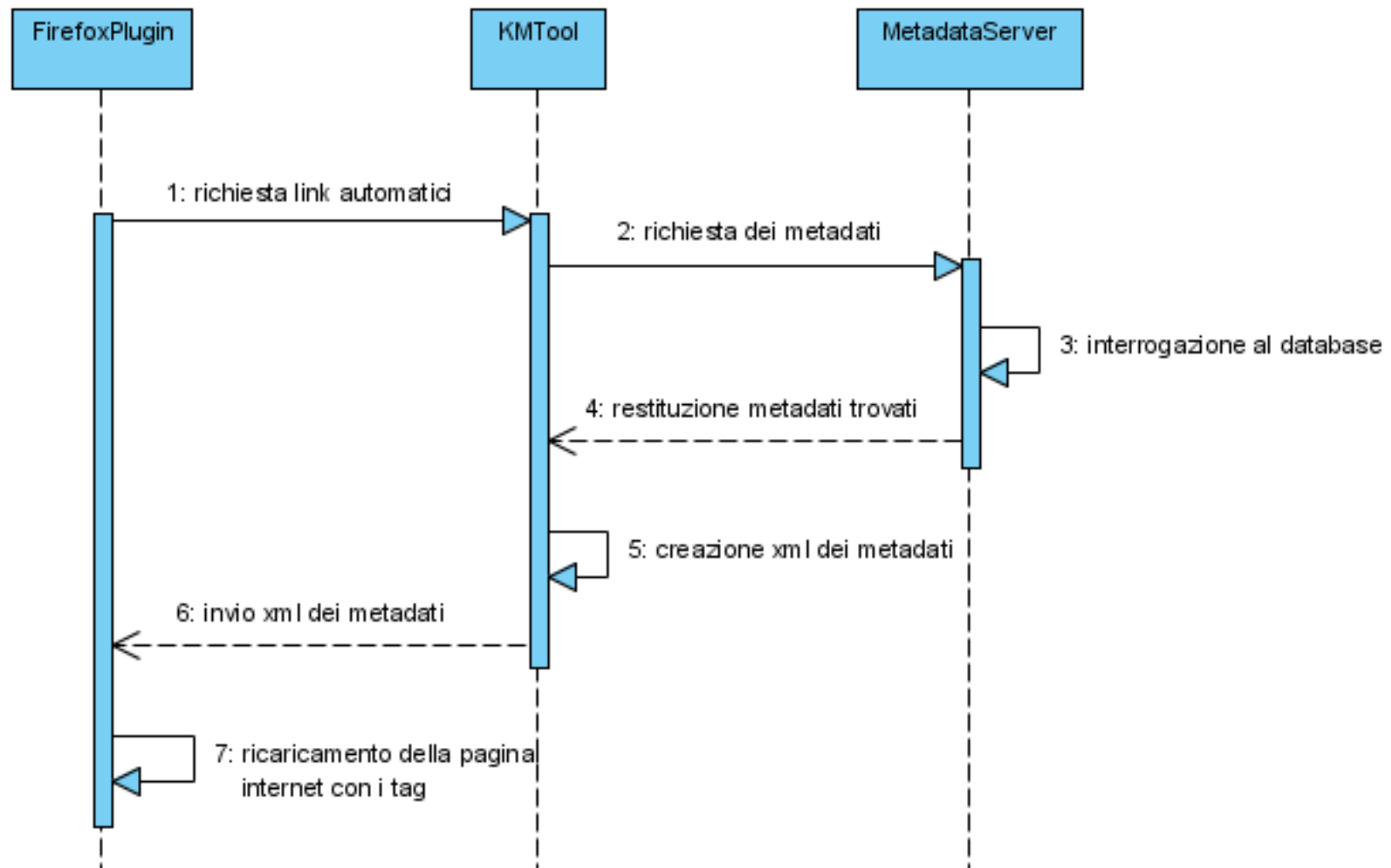


11: conferma inserzione

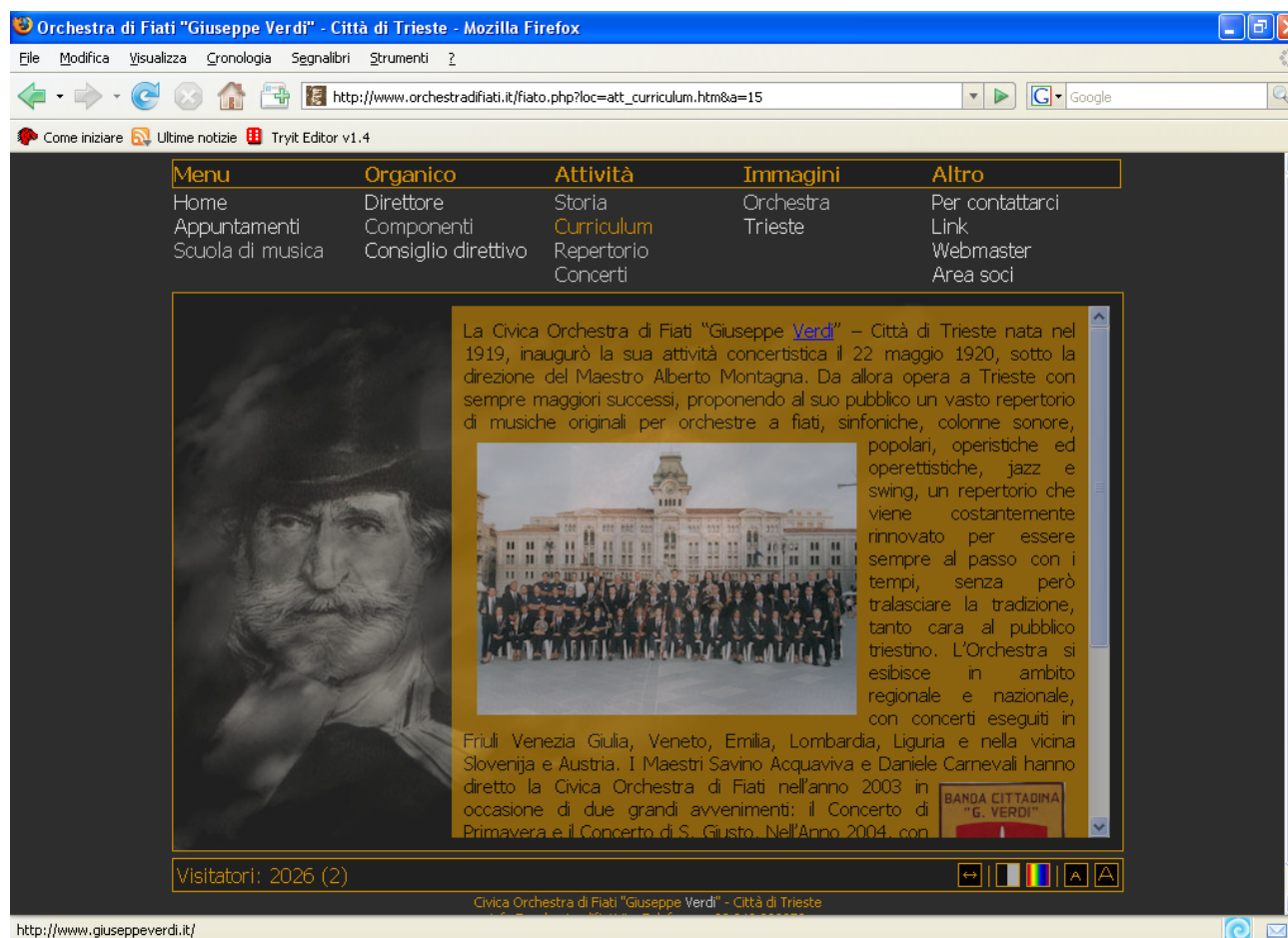


Sequence Diagram

sd autolinking



Esempio di utilizzo



7: ricaricamento della pagina internet con i tag

Conclusioni

★ Pregi del software

- Software modulare facilmente manutenibile ed estensibile
- Interfaccia grafica semplice e intuitiva

★ Limiti del software

- Nessun supporto multiutente, nessuna personalizzazione
- Possibilità di assegnare un significato errato alle parole

★ Sviluppi futuri

- Supporto multiutente
- Sviluppo template