

Un sistema per la gestione di riferimenti bibliografici con schede di catalogazione in XML

Riccardo Mazza, Roberto Gaspari, Davide Tam
SUPSI - Scuola universitaria professionale della Svizzera italiana
Dipartimento di informatica e elettronica
Tel. +41 91 6108531
Galleria 2
CH-6928 Manno (TI)
Switzerland
mazza@die.supsi.ch

Riassunto

Nelle comunità scientifiche il problema della gestione e classificazione dei riferimenti bibliografici è un problema critico. Questo documento descrive la realizzazione di un sistema di gestione documentale in grado di gestire e classificare riferimenti bibliografici utilizzando in input strutture dati in XML. Grazie a questo strumento il ricercatore può creare i propri riferimenti bibliografici tramite un editor XML, come una sorta di scheda di catalogazione come quelli presenti nelle biblioteche. Queste schede possono utilizzare una struttura che può essere ad esempio ripresa dallo schema bibtex di LaTeX. Le schede poi possono essere inserite in maniera automatica in un repository centralizzato. Il sistema dispone di una interfaccia Web-based con il quale l'utente (o chi altro è interessato) può sfogliare le proprie risorse bibliografiche, eseguire delle interrogazioni sui dati strutturati, fare delle ricerche di tipo fulltext. Il sistema ha il vantaggio di consentire agli utenti di effettuare l'editing delle schede XML senza essere necessariamente connessi ad Internet (quindi a casa, in vacanza, ecc.) e allo stesso tempo dare la possibilità ad altri di consultare ed interrogare il proprio database bibliografico tramite il Web.

Keywords

XML, bibliografie, Mysql, xml-dbms, Cocoon

Descrizione

Una delle attività principali del lavoro di un ricercatore scientifico consiste nel recuperare e gestire tutti i riferimenti bibliografici utili per le proprie ricerche: articoli di pubblicazioni, tesi, manuali tecnici, rapporti di ricerca Tutto ciò che può costituire un riferimento utile per il proprio lavoro di ricerca deve essere in linea di principio letto, classificato, catalogato ed infine, se utilizzato, deve essere citato come fonte di riferimento bibliografico nelle proprie pubblicazioni o, come nel caso degli studi universitari, nella propria tesi laurea.

Molte volte si ha la necessità di condividere le proprie risorse bibliografiche con altre persone. I docenti universitari ad esempio di solito consegnano un elenco di risorse bibliografiche ai propri tesisti dove possono trovare del materiale utile per il proprio studio. I ricercatori invece possono desiderare di comunicare ai propri colleghi quali riferimenti utilizzare per approfondire un determinato tema comune a cui essi stanno lavorando. In particolar modo, le comunità scientifiche di ricerca (dipartimenti universitari, centri di ricerca aziendale, gruppi di lavoro su progetti di ricerca ecc.) possono trarre un utile vantaggio disponendo di un database unico dove poter condividere ed accedere alle risorse bibliografiche.

Sistemi di questo tipo esistono già (ad esempio, il software eprints [1]), ma hanno lo svantaggio di richiedere agli utenti di utilizzare un browser ed inserire i dati tramite una connessione in Internet. Ciò è difficile da attuare se una persona desidera inserire dei dati mentre è a casa oppure se durante una vacanza o un viaggio di lavoro trova una risorsa interessante e vuole inserirla all'interno del sistema di gestione.

Per questo motivo abbiamo realizzato un sistema di gestione di riferimenti bibliografici, utile per archiviare, classificare e condividere all'interno di un gruppo di lavoro i riferimenti bibliografici delle proprie ricerche o studi. L'idea di base dalla quale siamo partiti è la necessità di avere uno strumento flessibile che permettesse ad una comunità di persone di :

- disporre di un archivio centralizzato di riferimenti bibliografici memorizzati sotto forma di documenti strutturati, chiamiamo questo archivio con il termine "repository";
- interrogare il repository tramite delle interfacce di query in modalità remota, possibilmente tramite interfacce Web;
- creare nuovi riferimenti bibliografici senza la necessità di essere connessi al sistema (in modalità off-line) ed avere in seguito la possibilità di inserire questi riferimenti nel repository tramite procedure semplici e veloci.

Un sistema di questo tipo può essere estremamente utile sia in un ambito didattico che in un ambito di una comunità di ricerca scientifica per la condivisione di riferimenti. Inoltre può essere utile per classificare i documenti ed effettuare

delle ricerche specifiche su documenti strutturati. Infine, la natura dei dati strutturati in XML rende agevole l'esportazione di questi in altri formati, come ad esempio bibtex, per essere utilizzati nella stesura di pubblicazioni o tesi.

Il sistema di gestione da noi sviluppato permette all'utente di creare un repository basato su una banca dati che li dovrà immagazzinare (il DBMS è MySQL[2]), la quale viene creata automaticamente a partire da un DTD (Document Type Definition) con un'operazione richiamabile da un'interfaccia Web. Questa procedura comprende sia il trasferimento del DTD sul server che la sua analisi (parsing) con un'applicazione scritta in Java basata sulle classi "XML-DBMS[3]". Quest'analisi produce in uscita le istruzioni SQL per la creazione delle tabelle della banca dati e da ultimo queste vengono eseguite.

Una volta creato il repository, l'utente può in seguito utilizzare un'altra procedura automatica che gli consente, sempre tramite interfaccia Web, di trasferire i propri documenti XML all'interno della banca dati. Anche in questo caso la procedura si preoccupa di trasferire sul server il documento XML, di scomporlo in singoli elementi e di trasferirlo nella banca dati, il tutto sempre tramite un'applicazione Java basata sulle classi "XML-DBMS".

Una volta che i documenti si trovano nella banca dati, questi sono automaticamente accessibili tramite Web grazie allo strumento Cocoon[4] affiancato al Web server Apache[5]. Il compito di questo strumento è quello di estrarre le informazioni dalla banca dati, di inserirle come risultato di una richiesta SQL in un documento XML, e, grazie ad un documento XSL (eXtensible Stylsheet Language) che specifica lo stile di rappresentazione, di trasformare quest'ultimo in una pagina HTML visibile da tutti i browser.

Per questa parte di visualizzazione abbiamo previsto una serie di viste che producono informazioni diverse relative ai documenti immagazzinati, e la possibilità di eseguire ricerche di parole chiave in alcuni elementi dei documenti. Il tutto è stato reso possibile grazie alle notevoli potenzialità di Cocoon che permette di creare pagine con contenuti altamente dinamici. Infatti offre la possibilità di inserire del codice java all'interno delle pagine XML, utile per l'analisi dei parametri immessi dall'utente e per la creazione della richiesta SQL correlata, anch'essa inserita direttamente nel documento XML.

Con questa applicazione l'utente deve solo compiere delle semplici operazioni: creare la banca dati caricando sul server il documento DTD e trasferire i propri documenti XML partendo da comode interfacce Web. Grazie a queste caratteristiche l'utente del sistema può preparare tutti i propri documenti XML off-line con un comune editor XML, ed in seguito trasferirli nella banca dati in una volta sola e in pochissimo tempo.

Attualmente la nostra applicazione supporta un'unica struttura dati, determinata dal DTD dato in input per la creazione del repository. In futuro prevediamo di sviluppare una nuova versione che supporti più DTD su un unico repository con interfacce uniche per l'interrogazione di questi.

Riferimenti

[1] <http://www.eprints.org>

[2] <http://www.mysql.com>

[3] <http://www.rpbouret.com/xmldbms/index.htm>

[4] <http://xml.apache.org/cocoon>

[5] <http://www.apache.org>