

Università degli Studi di Messina
Corso di Laurea in Informatica (Classe 26)

IMPLEMENTAZIONE DI UN WEB DINAMICO PER LA MISURAZIONE DELLE STATISTICHE D'USO DELLE R.E.R.: IL PORTALE CIBER-STATS

Tesi di Laurea di Mariella Smedile
Anno Accademico 2005/2006

PREFAZIONE

di Nunzio Femminò

L'avvento delle collezioni digitali in biblioteca ha prodotto un cambiamento di forte impatto nelle modalità di utilizzo delle risorse informative e nei servizi offerti all'utenza.

All'interno di questo contesto, la rilevazione dei dati statistici costituisce un'attività centrale. Attività utile per la valutazione delle scelte adottate nell'acquisto di collezioni elettroniche, in grado anche di offrire risposte attendibili sulle preferenze dei lettori, e tramite indispensabile per l'analisi dei comportamenti degli utenti delle biblioteche digitali.

I report statistici possono infatti fornire strumenti di rilevazioni importanti sull'uso effettivo delle risorse, di chi le utilizza, di cosa gli utenti utilizzano e di come lo utilizzano, di cosa preferiscono. Strumenti che, affiancati da altri metodi di indagine, come questionari e interviste, possono costituire una solida base per una più efficace organizzazione dell'offerta informativa e dei servizi ad essa connessi.

L'idea di implementare un portale dinamico di misurazione delle statistiche d'uso di risorse elettroniche remote è nata da un'esigenza concreta di lavoro e da

un'attività di collaborazione tra il C.A.B. (Centro di Ateneo per le Biblioteche) di Messina, il Consorzio CIBER di cui l'Università di Messina è membro, e CASPUR.

L'attività sperimentale qui pubblicata si colloca all'interno di questo contesto collaborativo e ha condotto alla realizzazione di uno strumento di elaborazione di dati statistici di cui da tempo si avvertiva grande esigenza all'interno della comunità consortile CIBER ed è stato ufficialmente presentato durante la "sessione statistiche" del Seminario residenziale CIBER, svoltosi a Messina, villa Pace, dal 12 al 14 giugno 2006.

Già molto sentita e più volte sottolineata anche dal gruppo di lavoro statistiche del CIBER, vi era non solo la necessità di adeguare i report statistici provenienti da CASPUR agli standards internazionali attuali, soprattutto a COUNTER, ma anche quella di facilitare il reperimento dei dati attraverso l'implementazione di una procedura interattiva.

In questa direzione il Settore automazione biblioteche del CASPUR si era già prodigato per un "parsing" dei log files dei servers di Emeroteca virtuale CIBER producendo un grande risultato in alternativa ai reports prodotti col sistema tradizionale (SAS) in uso al CASPUR e totalmente compatibili con lo standard COUNTER (R2).

La produzione semi-automatica dei report con questo nuovo sistema consiste nel lancio "batch" - con cadenza mensile - di una procedura che provvede al parsing ed al successivo deposito in area riservata (ciber-repository, uno spazio FTP riservato agli utenti CIBER) dei risultati, che possono essere successivamente recuperati dai singoli enti e, contemporaneamente produce una tabella riepilogativa di tutti i download effettuati.

Per quanto semplice nell'idea e nell'utilizzo, la fase di collegamento all'area FTP, il download del report, la successiva importazione del file CSV in excel sono sembrate, ai più, molto laboriose e non troppo chiare.

Inoltre, ci si era resi conto che i log files contenevano molte più informazioni di quante fossero necessarie per la realizzazione dei Journal Report 1 e 2 previsti da COUNTER (R2): infatti, oltre alle classiche data e ora e alla rivista da dove era stato effettuato il download, essi contenevano informazioni su chi aveva scaricato l'articolo e, soprattutto, dettagli su "cosa" aveva scaricato; c'erano i riferimenti per realizzare una "statistica del cosa" e non solo una "statistica del quanto".

Questa idea, emersa già nel corso del seminario residenziale CIBER del novembre 2004, è stata resa percorribile grazie alla realizzazione, da parte del CASPUR, di una ulteriore tabella riepilogativa di tutti gli articoli depositati in Emeroteca Virtuale contenente la suddivisione di tutte le riviste in fascicoli, gli articoli del fascicolo e, soprattutto, l'anno di pubblicazione dell'articolo stesso.

L'introduzione, poi, di elementi di valutazione "esterni", come le classi disciplinari MIUR e la soggettazione della Library of Congress per i periodici, che erano stati "raccolti" in un precedente lavoro dal C.A.B. di Messina, hanno reso il portale estremamente duttile ed in grado di rispondere alle esigenze più disparate anche grazie alla sezione "query libera" che permette di "interrogare" il data-base secondo parametri personalizzati.

Da un punto di vista più squisitamente tecnico, il portale è stato disegnato e costruito utilizzando tecnologie open source, in linea con le prerogative del C.A.B. di Messina, da sempre molto sensibile e vicino ai temi dell'open access.

L'attività sperimentale della tesi di Mariella ha dato un apporto notevole alla realizzazione del portale. Lavoro che ha preso avvio da una serie di precedenti attività che hanno prodotto una base di dati che ha costituito il nucleo centrale e il punto di partenza della tesi stessa.

Ad un'attenta e complessa analisi è seguita la costruzione dello schema del portale e la scrittura di script in grado di eseguire query, talvolta molto sofisticate, al database di MySQL dove sono state depositate le tabelle di lavoro ma anche di realizzare grafici ad istogramma. Sono state adottate diverse misure di sicurezza ed accorgimenti per rendere quanto più possibile agile la risposta del server.

L'esito del lavoro può considerarsi di alto profilo, soprattutto in termini di soddisfazione dell'utenza, ma anche in termini di qualità. Esso infatti è un tassello di una più estesa attività che ha raggiunto oramai una eco internazionale. Nel contesto CIBER, l'importanza del portale è strettamente collegata con l'esigenza di integrare reports provenienti da più fonti. Nel nostro caso, dalla piattaforma di accesso Emeroteca Virtuale di CASPUR, che archivia in modo permanente su propri server i files dei pacchetti editoriali acquisiti da CIBER, e dai siti degli editori, molti dei quali consentono il doppio accesso.

La duplicazione dei report a fronte di editori uguali e l'integrazione dei loro dati è una problematica molto delicata sulla quale già da tempo si sono avviate diverse iniziative anche di carattere internazionale. Sono già nati progetti (es. JURO – Journals Usage Report Online, sviluppato nell'ambito della HKUST Library di Hong Kong) e standards (es. SUSHI - Standardized Usage Statistics Harvesting Initiative, promosso dal NISO: National Information Standards Organization) per approfondire e sperimentare soluzioni alla problematica.

Anche il C.A.B. di Messina ha intrapreso alcune attività di sperimentazione volte a realizzare una integrazione dei dati statistici quanto più automatizzata a cominciare dalla adesione al progetto SUSHI, adeguando i propri reports allo standard XML, e con una installazione di JURO che si sta personalizzando allo scopo, grazie al suo carattere open source.

Nunzio Femminò