

# Progettazione e analisi delle funzionalità di un nuovo software nell'ambito dell'editoria digitale

**Paolo Arvati, Riccardo Ferrari**

CILEA, Segrate

## *Abstract*

L'articolo presenta una panoramica sull'attività di progettazione di un nuovo prodotto di interfaccia per l'editoria elettronica che permetta di integrare servizi distribuiti e di guidare l'utenza dalla fase di consultazione delle banche dati a quella di reperimento degli articoli originali. Segue un breve cenno alle nuove tecnologie e paradigmi di progettazione adottati presso il CILEA per far fronte all'elevato livello di complessità dei progetti moderni.

*Keywords:* Beni culturali, Progetto software, Editoria digitale, UML, PubMed, Web Service.

La crescente necessità di scambio e reperimento di informazione da parte dell'utenza e la sempre maggior disponibilità da parte degli editori a favorire la diffusione di materiale in formato elettronico, stanno delineando nuove esigenze e funzionalità che hanno spinto il CILEA ad impegnarsi nell'analisi e realizzazione di un prodotto che possa dare una risposta unitaria in termini di interfaccia tra l'utenza scientifica e i servizi di editoria elettronica disponibili sulla rete.

Le motivazioni che hanno indotto alla progettazione di un nuovo prodotto per l'editoria digitale presso il CILEA sono state: mettere a disposizione degli utenti la possibilità di interrogare e consultare contemporaneamente diverse banche dati, fornire la capacità di interfacciarsi con servizi remoti.

L'interesse primario da cui si è partiti è stata la possibilità di fornire un accesso personalizzato, tramite i nostri servizi, a tutta la banca dati Medline e PreMedline consultabile tramite il servizio PubMed (<http://www.ncbi.nlm.nih.gov>) messo a disposizione dalla National Library of Medicine (<http://www.nlm.nih.gov>). Si è poi aggiunta fin da principio l'esigenza di integrare PubMed con altri servizi e banche

dati eventualmente ospitati presso il nostro centro.

L'obiettivo del progetto è dunque quello di voler dare, come valore aggiunto rispetto ai servizi esistenti, la possibilità di utilizzare una interfaccia unica e adeguata alle esigenze professionali di ricerca dell'utenza e in secondo luogo di fornire agli utenti vari percorsi attraverso i quali giungere all'acquisizione del full-text degli articoli oggetto delle proprie ricerche o comunque ad una indicazione dei servizi che ne hanno disponibilità.

## **Caratteristiche e funzionalità**

Il centro del servizio attorno al quale si articola la serie di funzionalità, di moduli applicativi, di personalizzazioni, è rappresentato dall'inter-faccia attraverso la quale l'utente è messo nelle condizioni di poter costruire interrogazioni articolate verso servizi di editoria elettronica, a volte tra loro diversissimi nella struttura e nell'impostazione, senza distinzioni apparenti tra servizi esterni o interni al consorzio.

L'analisi dell'interfaccia è stata condotta su più fronti: da un lato lo studio di esperienze già presenti in rete e diffusamente utilizzate da biblioteche e privati con i quali abbiamo avuto diverse occasioni di discussione. A

questo proposito citiamo due tra i servizi a cui abbiamo fatto maggiormente riferimento: PubMed (<http://www.ncbi.nlm.nih.gov>) e OVID. Dall'altro lato il confronto con l'utenza, a cui intendiamo rivolgere il prodotto che si sta sviluppando, ci ha consentito (e ci consente dato che il progetto è tuttora in costruzione) di calibrare e quotare le funzionalità in risposta alle esigenze e all'esperienza maturata in questi anni dall'utenza stessa. Questo, in sintesi, significa implementare nell'interfaccia quelle funzionalità più utili e significative presenti nelle varie esperienze analizzate.

La caratteristica tra le più qualificanti, su cui si fonda la struttura della ricerca e si focalizza l'intero servizio, è la costruzione, la memorizzazione e il riuso delle interrogazioni costruite in maniera incrementale dagli utenti. Ad ogni utente, dopo ogni connessione al servizio, viene assegnata una sessione di lavoro durante la quale il sistema memorizza e gestisce tutte le sue strategie di interrogazione. La "storia" delle ricerche non viene tuttavia persa al termine della sessione di lavoro, ma rimane a disposizione dell'utente in modo permanente o fino a quando l'utente stesso decide di azzerarla e di riaprirne una nuova. L'utente può, a questo riguardo, riutilizzare una ricerca, semplicemente richiamandola con un click dalla tabella della "storia" (funzione di Display); oppure può affinare una ricerca imponendo altre condizioni di filtro con lo scopo di puntare ai documenti interessati in un modo più preciso e diretto (funzione di Limits); oppure può combinare tra loro due o più ricerche (presenti nella "storia") utilizzando i comuni connettori logici (AND, OR, NOT) e realizzare così una strategia di ricerca effettivamente complessa e articolata (funzione di Combine). Infine è sempre l'utente che dispone delle sue ricerche storicizzate, anche nella decisione di eliminare, e quindi azzerare, la sessione di ricerca, oppure di ripulire selettivamente quelle ricerche che sono ritenute non significative o errate nella loro formulazione.

Un'altra tra le caratteristiche fondanti della struttura dell'interfaccia di ricerca, è rappresentata dalla capacità e dalla possibilità di indirizzare un'unica interrogazione verso servizi e database diversi.

All'utente vengono forniti gli strumenti necessari per costruire e impostare un'interrogazione rispettando le regole di una sintassi generale (definita in accordo con l'utenza), che viene successivamente tradotta negli specifici linguaggi dei vari servizi.

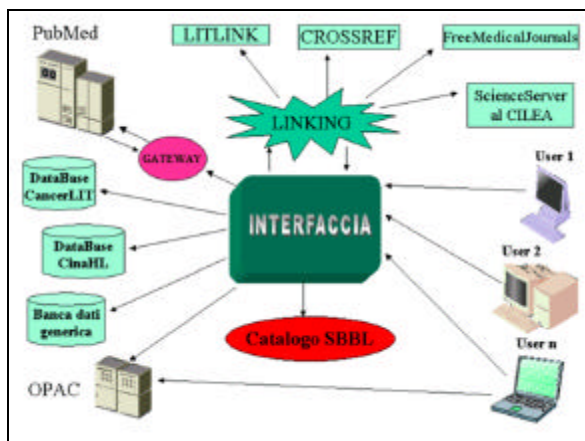
Nel caso specifico si sono concordate alcune regole sintattiche per la formulazione delle interrogazioni: individuazione dei campi di ricerca, regole per l'uso dei connettori logici e delle parentesi per la regolazione dei livelli di precedenza computazionale, definizione dei campi, delle regole e dei valori per le funzioni di Limits, definizione delle regole per la funzione di Combine, e altre opzioni ancora. Ogni ricerca viene memorizzata nella tabella della "storia" e il sistema traduce la sintassi utilizzata dall'utente in una meta-sintassi generalizzata ed astratta dai servizi particolari a cui viene rivolta.

A questo punto il sistema, utilizzando apposite funzioni di parsing, opportunamente predisposte per ogni servizio o database a cui la ricerca viene indirizzata, traduce la sintassi "astratta" negli specifici linguaggi dei motori cui viene rivolta l'interrogazione (ad esempio: la sintassi per PubMed o una sintassi per interrogazioni di database gestiti con linguaggio SQL).

### **Interrogazione e motori di ricerca**

La modalità con cui viene costruita ed eseguita una ricerca dipende dalla tipologia delle risorse cui ci si deve interfacciare, essenzialmente divisibili in due categorie principali:

1. Banca dati locale ospitata presso il CILEA come database installato sulle nostre macchine. Solitamente si tratta di banche dati relazionali interrogabili tramite il linguaggio SQL.
2. Servizio disponibile via rete presso un determinato indirizzo (web service) che può essere interrogato tramite regole prefissate e che fornisce i risultati in un formato definito.



Le interrogazioni alle varie banche dati e servizi vengono lanciate in contemporanea e in maniera asincrona. L'aspetto su cui si sta concentrando l'attenzione è la diversa modalità di visualizzazione dei risultati delle varie banche dati e il trattamento dei risultati pervenuti.

Oltre infatti ad essere raro l'utilizzo di identificativi comuni degli articoli fra servizi e banche dati differenti, resta altresì complesso definire delle regole per l'ordinamento o la relazione di uguaglianza tra articoli basati su aspetti quali per esempio il titolo o l'elenco autori che spesso vengono trattati con modalità differenti (dal set di caratteri al numero di autori memorizzato).

Alle pagine di visualizzazione dei risultati segue la pagina di dettaglio dei singoli articoli presso la quale è disponibile la citation completa dell'articolo selezionato e i parametri identificativi dell'articolo quali ISSN, numero di volume, numero di pagina ecc..

In questo progetto la citation dell'articolo assume un ruolo centrale in quanto contiene le informazioni sufficienti per connettersi al servizio di LINKOUT, sezione cui si è data molta importanza nello sviluppo del prodotto. Il LINKOUT è un vero e proprio servizio a sé stante che fornisce in una pagina, oltre ai campi descrittivi principali riferendosi all'articolo (titolo, autori ..), il maggior numero di percorsi possibili sul WEB attraverso i quali giungere al full-text dell'articolo.

Esso è anche un vero e proprio servizio di content delivery ospitato presso il CILEA e configurabile in base alle esigenze dell'utenza.

Alla sezione di LINKOUT per il singolo articolo è possibile accedere in due modi:

1. Seguendo i link disponibili presso il sito di cui stiamo trattando nel presente articolo.
2. Costruendo una richiesta diretta al servizio seguendo le dovute regole (regole di costruzione dell'URL).

Nel secondo caso il servizio si comporta come web service (regole per l'interrogazione e formato della risposta) e si prevede a breve di implementare un formato XML di output in modo da fornire i risultati in formato standard.

Seguono alcuni prodotti o servizi di cui sono già state valutate la fattibilità e l'utilità e che si prevede verranno inclusi nel servizio di LINKOUT:

- LitLink (<http://www.litlink.com>), sarà probabilmente il prodotto di punta che permetterà di interfacciarsi a decine di servizi distributori di full-text.
- CrossRef (<http://www.crossref.org>).
- ScienceServer ospitato presso il CILEA (<http://scienceserver.cilea.it>).
- I Free Medical Journals (<http://www.freemedicaljournals.com>).
- Collegamento ai servizi di Linkout e di 'related articles' di PubMed oltre al reindirizzamento a PubMedCentral (<http://www.pubmedcentral.nih.gov>).

### Analisi e sviluppo del progetto software

Affrontare un progetto di questo genere impone l'utilizzo di metodi e strumenti adatti a gestire le fasi che costituiscono l'attività di produzione del software dai preliminari fino alla sua messa a servizio.

L'elevato numero di funzionalità richieste e la complessità della logica applicativa del prodotto, unitamente alla progressiva evoluzione delle specifiche iniziali richiedono l'impiego di metodologie progettuali e produttive collaudate e certificate.

Dalla primavera di quest'anno il nostro team ha scelto il Rational Unified Process della Rational Software Corporation

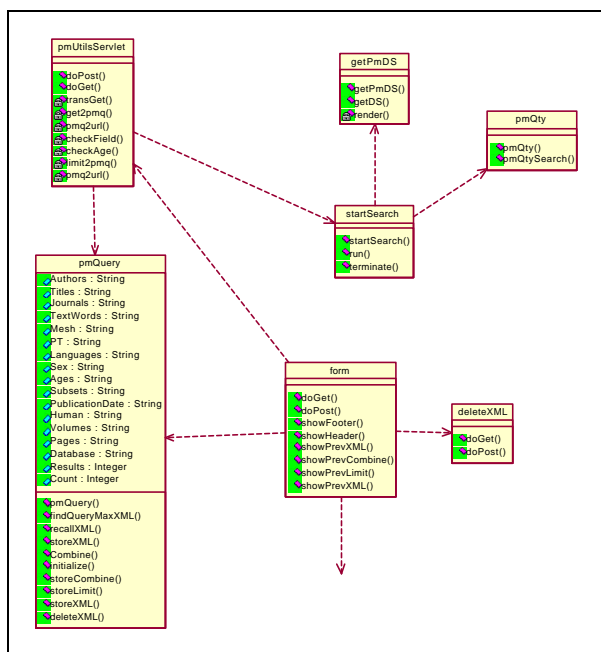
(<http://www.rational.com>)

come proprio processo di sviluppo. Allo scopo il CILEA ha acquisito il prodotto Rational Suite AnalystStudio in modo da avere gli strumenti necessari per affrontare tre problematiche centrali in questo contesto: modello iterativo, linguaggio UML,

coordinamento e comunicazione all'interno del team.

Nel modello iterativo tutte le fasi che portano alla realizzazione di una release del prodotto vengono ripetute per ogni release. Questo approccio si è rivelato particolarmente utile nel nostro caso a causa della carenza di specifiche iniziali, e della mutabilità delle funzionalità richieste, che in certi casi si evolvono con l'evolversi del prodotto stesso. Si è quindi puntato come prima fase allo sviluppo di un prototipo in modo da poter valutare la fattibilità di tutti gli aspetti critici e l'organizzazione delle principali funzionalità del prodotto. All'accettazione del prototipo ha seguito la fase per la produzione della prima release. Ovviamente, in coerenza con le metodologie scelte, il passaggio dal prototipo alla prima release e dalla prima release alla successiva ha comportato e comporterà il dover riaffrontare tutte o quasi le fasi del progetto.

L'uso del linguaggio UML (Unified Modelling Language) per la descrizione dei vari aspetti del progetto si è reso utile nella fase di raccolta delle specifiche e funzionalità del servizio (Use Case Diagram) e nella fase di modellazione e strutturazione del codice: Class Diagram, Collaboration Diagram e Sequence Diagram.



Tralasciando i dettagli sulle singole funzionalità offerte dai tools della Rational Suite va tuttavia sottolineato come l'utilizzo di un linguaggio universale e collaudato quale UML per la descrizione del progetto ha permesso di avere: maggiore chiarezza sulla architettura del sistema, possibilità di comunicazione all'interno del team facendo riferimento ad un linguaggio standard e infine capacità di produrre una tipologia di documentazione con caratteristiche riconosciute come lo stato dell'arte da parte della comunità informatica.

E' infatti ormai consolidato il forte legame tra i prodotti e i processi proposti da Rational e le piattaforme principali che IBM e Microsoft promuovono nel campo dello sviluppo software. Senza tralasciare che UML sta diventando uno standard in numerosi contesti al di là dell'utilizzo o meno delle Rational Suite.

L'utilizzo di JAVA come linguaggio di programmazione è stato dettato, in primo luogo, dall'esigenza di avere a disposizione un linguaggio e delle librerie capaci di soddisfare alcune specifiche di difficile implementazioni riguardanti i protocolli di rete e l'interscambio di informazioni tra servizi distribuiti sul WEB, in secondo luogo dalla volontà di prestare particolare attenzione ai parametri di qualità del software quali riusabilità, manutenibilità e portabilità.

Nel caso specifico di questo progetto sono stati considerati come obiettivi fondamentali la possibilità di riutilizzare parti del codice scritto per altri progetti simili (riusabilità) aventi problematiche comuni a livello di tecnologie di rete e connettività alle banche dati; la possibilità di intervenire in tempi brevi e bassi costi sul codice già scritto (manutenibilità) al fine di correggere eventuali errori rilevati in fase di testing e di poter partire dal codice di una release per implementare quella successiva senza passare per la riscrittura dei moduli.

E' stata inoltre data particolare attenzione alla fattibilità di installare il servizio su macchine equipaggiate con hardware e software differenti (portabilità) in modo tale da garantire:

- Possibilità di utilizzare il prodotto per un periodo temporaneo su macchine a basso

costo, per poi passare in un secondo momento a sistemi più performanti e tecnologicamente differenti.

- Possibilità di fare più installazioni contemporaneamente (magari per utenze differenti) e lievemente diversificate dello stesso prodotto su macchine differenti.

In linea più generale dal punto di vista informatico sono state scelte tecnologie aperte, in forte evoluzione e non legate all'hardware e al sistema informativo sottostante.

L'utilizzo di JAVA come linguaggio e della tecnologia SERVLET (ormai uno standard del settore) ci ha infatti permesso di poter impiegare come WEB server e gestore di servlet un prodotto quale Orion Application Server (<http://www.orionserver.com/>) che sta avendo ampia considerazione nel settore e che recentemente è stato scelto come piattaforma JAVA/SERVLET/JSP da parte di Oracle.

Punto cruciale che ha determinato la scelta di questa soluzione, oltre al fatto di poter spostare il servizio con tutta la piattaforma su macchine tecnologicamente diverse, è soprattutto la possibilità di poter portare il prodotto implementato e finito su altre piattaforme aderenti agli stessi standard (tipo BEA WebLogic, IBM WebSphere, Tomcat, ecc..) che via via si vanno delineando e di cui si volessero eventualmente sfruttare i vantaggi in fasi successive.