

Infrastrutture software per un portale customizzabile

Enrico Cavalli

CILEA, Segrate

Abstract

In questo articolo evidenzieremo gli aspetti tecnici e infrastrutturali del portale informativo www.supercomputing.it. Cercheremo di illustrare le motivazioni e gli obiettivi strategici che hanno portato a determinate scelte nel progetto del database e del motore software del portale, ponendoci sia nell'ottica dell'utente finale, sia nell'ottica del curatore di area che redige brevi abstract su notizie attinenti vari settori del mondo del Supercalcolo.

Keywords: Supercalcolo, portale, database, MySQL, PHP.

Premessa

www.supercomputing.it vuole proporsi come una guida qualificata per chi è interessato al mondo del supercalcolo. È un contenitore di informazioni variamente catalogate, che l'utente può customizzare in base ai propri interessi. Obiettivo primario dello sviluppo del portale è la predisposizione di un'infrastruttura in grado di fornire notizie il più possibile confacenti agli utenti che si registrano (vedi [1] per un approfondimento sulle motivazioni delle scelte).

Particolare attenzione è stata inoltre posta nel rendere le varie parti del portale coerenti dal punto di vista dell'aspetto grafico e dei contenuti.

Il portale è interamente costruito con strumenti OpenSource, dal database MySQL [2], al linguaggio di web-scripting PHP [3], fino ai programmi accessori di manutenzione scritti in PERL [4]. Il risultato è un sistema di Content Management (seppur non estremamente sofisticato), che consente ad un gruppo di esperti in vari rami del supercalcolo, di presentare notizie specifiche ad un pubblico specialistico. Le pagine web sono costituite da uno schema le cui varie parti vengono composte e riempite a partire dai dati inseriti nel database. La gestione dei contenuti è

semplificata da una facile interfaccia web



verso il database.

Figura 1- L'home page non customizzata del portale

Quest'ultimo contiene tre dati fondamentali, ovvero:

1. gli utenti registrati;
2. le aree tematiche di interesse;
3. le notizie, costituite da un titolo, una breve descrizione – una sorta di abstract redatto da un esperto del settore – e un puntatore alla notizia completa.

Un visitatore può registrarsi sul sito, e dichiarare i propri interessi in determinati ambiti

del mondo del supercomputing. In questo modo può usufruire di una home-page personalizzata in base alle proprie richieste. Gli esperti dei vari settori hanno a disposizione una comoda interfaccia web per inserire le notizie. L'interfaccia permette la catalogazione nelle varie aree tematiche e consente di inserire una breve recensione sulla notizia che si sta inserendo. La stessa interfaccia consente di inserire le notizie di primo piano (gli Highlights), e di aggiornare l'archivio di corsi e conferenze (anch'essi ovviamente classificati nelle rispettive aree tematiche).

Il punto di vista dell'utente

Sussistono alcune ovvie relazioni tra i dati di cui abbiamo parlato. Innanzitutto un utente può essere interessato a più aree tematiche e d'altra parte un'area tematica può raccogliere l'interesse di più utenti. Nel linguaggio dei database relazionali abbiamo una relazione *n:m* tra utenti ed aree di interesse. Lo stesso tipo di relazione intercorre tra notizie e ambiti di interesse, in quanto una notizia può appartenere a più aree tematiche, e un'area di interesse raccoglierà ovviamente più notizie.



Figura 2 - L'home-page customizzata

Fin qui niente di strano. L'utente registrato può impostare in maniera banale un proprio profilo scegliendo, con ordine, le aree tematiche che predilige. In questo modo il portale presenterà le notizie più rilevanti nell'ordine scelto dall'utente, ponendo in secondo piano quelle non rientranti in un'area selezionata.

L'utente che lo desidera, potrà inoltre scegliere di ricevere una newsletter, contenente le ultime notizie inserite, relative alle aree tematiche scelte. In Fig. 1 vediamo la home page che si presenta ad un utente "anonimo", ossia qualsiasi visitatore che non si sia ancora registrato nel portale.

La prima volta che un utente visita il sito, si apre una finestra che lo invita a registrarsi, oppure a fornire le proprie credenziali (username e password) qualora si fosse precedentemente registrato.

La home page cambia quando l'utente definisce i propri interessi, come si nota nell'esempio di Fig.2. Ogni area è identificata da un proprio colore, omogeneo in tutte le zone del portale: dalle pagine *Courses and Conferences*, ai banner tematici nella parte destra della home page.

Il comportamento dell'utente

Il portale è predisposto per costruire un "profilo" dell'utente. Questo è in parte costruito dichiarativamente, cioè in base alla scelta delle aree tematiche da parte dell'utente, come abbiamo già spiegato. Il concetto è sinteticamente espresso in Fig. 3, dove è rappresentato con un diagramma entità-relazioni: la parte rilevante del database.

E' stata inoltre predisposta un'infrastruttura che permetterà di catturare alcune informazioni dal comportamento dell'utente e di integrarne il profilo.

Ogni notizia inserita nel database compare infatti nel portale con un link che l'utente può decidere di visitare qualora desideri approfondire l'informazione presentata. Quando un utente clicca un link, l'evento viene registrato dal portale. In questo modo si può fare tutta una serie di statistiche, e apprendere utili informazioni sul comportamento del visitatore. Supponiamo ad esempio che, dopo essersi registrato, questi dichiara di essere interessato alla categoria "Economy and Finance". Nel caso il portale notasse un numero statisticamente significativo di clic su notizie relative ad aree tematiche estranee al mondo "Economy and finance", potrebbe ad esempio decidere di inviarle all'utente nella newsletter, benché non dichiarate tra le sue preferenze.

Questa funzionalità è tuttora oggetto di studio e non ancora implementata in forma definitiva.

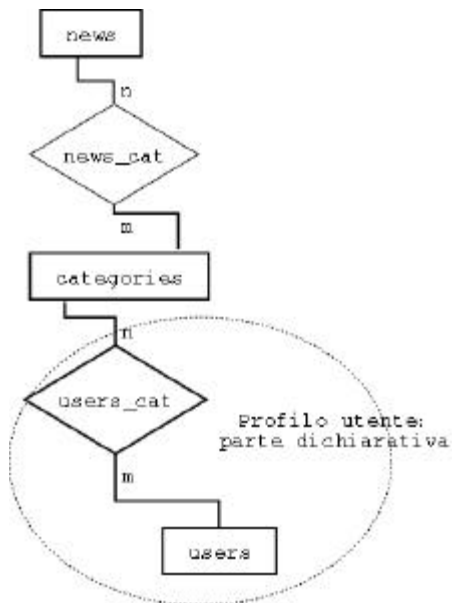


Figura 3 - Il profilo dell'utente all'interno della struttura del database

Aree tematiche gerarchiche e CDL

Esaminiamo ora un dettaglio tecnico, nell'intento di illustrare una caratteristica che potrebbe rivelarsi molto utile in edizioni successive del portale.

La tabella delle categorie è predisposta per creare un albero gerarchico di aree tematiche, in quanto dotata dei seguenti campi:

```
cat_id tinyint(4),
parent_id tinyint(4),
cat_name varchar(40)
```

Ogni categoria quindi, oltre ad un nome e ad un identificativo univoco, ha un campo che specifica quale sia la categoria padre. Se ad esempio la categoria *Computational Fluid Dynamics* avesse *cat_id=4*, le sue sottocategorie avrebbero *parent_id=4*. Il portale può quindi prestarsi ad una categorizzazione molto fine delle notizie inserite, per poter rispondere ad una futura esigenza di fornire agli utenti una categorizzazione dei contenuti estremamente granulare. La domanda a questo punto è: in cosa si potrebbe concretizzare una maggiore granularità? Quali i vantaggi per l'utente? Per come è strutturato ora il portale, l'utilità pare non sussistere. Al contrario si indisporrebbe l'utente costringendolo a interfacciarsi con una

selva di categorie e sottocategorie che comparirebbero nella pagina di compilazione del proprio profilo. Tutto questo senza evidenti vantaggi.

Ma proviamo a ragionare in termini più generali. Il CILEA offre un servizio estremamente interessante denominato CDL (CILEA Digital Library, [5]): un database dei journal scientifici più prestigiosi. Collegando la fine capillarità delle categorie del portale con la catalogazione bibliografica dei giornali della Digital Library, si potrebbe a questo punto realizzare un servizio estremamente utile: l'alerting bibliografico. L'utente potrebbe infatti essere avvisato via email dell'uscita su CDL dell'ultimo numero di una rivista specialistica, collegata al proprio specifico interesse scientifico.

In questo contesto ci sembra sicuramente più significativa l'adozione di una maggiore specificità nella definizione delle aree tematiche.

Nella prima edizione del portale si è scelto tuttavia di adottare una struttura orizzontale di macro-categorie, corrispondenti a grandi aree del mondo del supercalcolo. Non è detto che in futuro non si sfrutti questa possibilità, cui l'infrastruttura del portale è già abilitata. Tutti i programmi Perl o le pagine PHP infatti, tengono già conto di questa opzione.

Il punto di vista del redattore

Come già accennato in precedenza, gli esperti dei vari settori incaricati della redazione delle notizie da inserire nel portale, dispongono di una comoda interfaccia web per svolgere il proprio compito.

L'interfaccia consente di avere varie viste sui dati inseriti, permettendo ad esempio una semplice vista in ordine cronologico, oppure una vista suddivisa in base alle aree gerarchiche impostate come si vede Fig. 4.

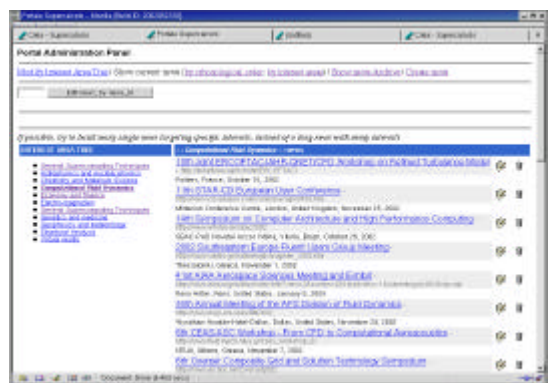


Figura 4 - Selezione delle notizie per area tematiche

Cliccando sulla voce *Create News* si apre la pagina di inserimento delle notizie e dei corsi. Il redattore potrà inserire il titolo, un puntatore al sito web dove reperire maggiori informazioni e un breve abstract. Ultimata la prima fase di compilazione, si passa a classificare la notizia. La pagina che ne risulta è visibile in Fig. 5.



Figura 5 - L'inserimento di una notizia nel portale

Qui l'esperto può stabilire se si tratta di una notizia che andrà in Highlights, oppure se si tratta della segnalazione di un corso o una conferenza. Ovviamente potrà anche classificare l'oggetto che sta inserendo nel portale, scegliendo in quali categorie inserirlo. Non è stato trascurato il fatto che il portale può essere un'utile strumento di promozione delle iniziative del CILEA. Per questo motivo le notizie inserite dagli esperti possono essere etichettate impostando il campo `max_priority=Y`. Una notizia di questo tipo verrà comunque visualizzata con particolare enfasi, e spedita via posta elettronica a tutti gli utenti, indipendentemente da come la notizia stessa sia stata catalogata nelle aree tematiche.

Lo scopo di tutto questo è molto semplice: il CILEA vuole per esempio promuovere il CAPI, oppure annunciare l'installazione di un nuovo supercalcolatore? Basterà inserire la notizia etichettandola in maniera speciale, per raggiungere l'intera comunità degli utenti registrati nel portale.

Tra le scelte che l'esperto è chiamato a compiere nel momento in cui inserisce un elemento nel portale, c'è anche l'indicazione di una *deadline*. Questo ci introduce direttamente al prossimo argomento.

Meccanismi di automanutenzione

Abbiamo individuato almeno due aspetti chiave per il buon successo di un portale. In primo luogo, ovviamente, è necessaria un'elevata qualità delle informazioni proposte ai visitatori. Questo è chiaramente demandato alle capacità degli esperti nelle varie aree di interesse: nessun programma potrà supplire alle capacità delle persone coinvolte nel lavoro. D'altra parte è essenziale che le news siano il più possibile aggiornate.

Anche se è comunque compito dei redattori aggiornare le notizie inserite nel portale, è pur vero che è possibile automatizzare alcune operazioni di supporto a questo processo.

In quest'ottica interviene la *deadline*, una data che segna il termine di validità delle notizie inserite nel motore del portale. Pensiamo ad esempio alla segnalazione di un convegno: non ha più senso che tale notizia appaia nelle highlights dopo il termine del convegno stesso. Facciamo notare che ogni oggetto inserito nel portale non viene cancellato una volta superata la *deadline*, bensì viene inserito in un archivio che resterà sempre disponibile all'utente.

Un altro semplice meccanismo di automanutenzione è costituito dai reminder agli esperti. Osservando la data dell'ultimo inserimento in una determinata area tematica, viene automaticamente avvisato il curatore di area qualora venga superata una soglia prestabilita.

Questi due semplici accorgimenti, uniti alle capacità dei redattori, dovrebbero garantire un costante aggiornamento del portale. Un portale ricco di notizie qualificate e il più possibile aggiornate può diventare uno strumento molto utile nel lavoro quotidiano dei ricercatori e appassionati che vorranno registrarsi su www.supercomputing.it.

Bibliografia

- [1] S. Mussi, "Portal features for heterogeneous user communities", Bollettino del CILEA, n. 84, Settembre 2002
- [2] MySQL, URL: <http://www.mysql.com>
- [3] PHP, URL: <http://www.php.net>
- [4] PERL, URL: <http://www.perl.com>
- [5] Cilea Digital Library, URL: <http://cdl.cilea.it>