

MOP/GAS - un sistema multiplatforma per l'accesso integrato a più OPAC via Internet

Stefania Biagioni e Giuseppe Romano

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Area della Ricerca CNR
Via Moruzzi 1, 56124 Pisa
{stefania.biagioni, giuseppe.romano}@isti.cnr.it

Abstract. Il Meta Opac Pisano è una implementazione, del sistema Isis/GAS. Esso permette di accedere, in maniera integrata, ai cataloghi della maggiori biblioteche delle Istituzioni di ricerca e universitarie dell'area pisana. Il sistema permette di navigare all'interno di ogni singolo catalogo, di accedere ai relativi indici di autori, soggetti, titoli e, permette, inoltre di accedere ad un indice globale, somma degli indici singoli, delle organizzazioni che partecipano al progetto. Vengono descritte le motivazioni che hanno portato alla sua realizzazione, viene descritta l'organizzazione logico/fisica del sistema e la tecnologia utilizzata.

1 Introduzione

Il Meta Opac Pisano¹ (MOP) è la realizzazione informatizzata del Catalogo Collettivo del Sistema Bibliotecario Pisano, primo risultato del protocollo d'intesa stipulato nel gennaio del 2003 tra CNR, Scuola Normale di Pisa, Scuola Superiore di Studi Sant'Anna e Università di Pisa.

Alla base della sua realizzazione è la convinzione che esistano le condizioni e le potenzialità per sviluppare un servizio di elevata qualità in grado di utilizzare al meglio le risorse disponibili e le nuove opportunità che si aprono con lo sviluppo dell'informatizzazione e delle reti.

La strategia di riferimento è l'estensione della collaborazione tra le biblioteche ad un ampio ventaglio di attività e obiettivi comuni quali per esempio: i) la creazione di nuovi servizi informativi, indirizzati all'utente, in un ambito di condivisione di risorse e sistemi superando le logiche di appartenenza istituzionale e gestionale delle informazioni; ii) la realizzazione di un sistema che permetta di superare le barriere imposte dai singoli sistemi d'automazione bibliotecaria in uso (Aleph [4], Libero [5],

¹ Home page di MOP <http://leonardo.isti.cnr.it/metaopac/mich/mop/mop.html>

2 Stefania Biagioni e Giuseppe Romano

Isis [6]), svincolando l'utente dal dover conoscere interfacce diverse e dover gestire diversi risultati; iii) l'alleggerimento del carico di accessi ai singoli OPAC con conseguente miglioramento dell'usabilità dei sistemi; iv) l'abbattimento delle spese per le licenze web e i server Z39.50; v) la realizzazione di una copia alternativa di riserva dei singoli cataloghi e dei servizi ad essi connessi.

Il MOP è uno strumento che realizza i suddetti obiettivi: è in grado di interrogare, contemporaneamente o singolarmente, più cataloghi tramite un'interfaccia unica e semplificata che guida l'utente remoto alla ricerca delle informazioni in modo veloce, facile ed efficace e alla navigazione tra i risultati; è un punto di accesso comune che permette all'utente di scegliere dove effettuare la ricerca e sapere, in una volta sola, se un'opera è presente in una o più biblioteche cittadine (Fig. 1).

Il MOP è risultato essere anche un utile strumento per il lavoro di revisione dei singoli cataloghi, sia dal punto di vista dei contenuti, sia nell'ottica della razionalizzazione degli acquisti.

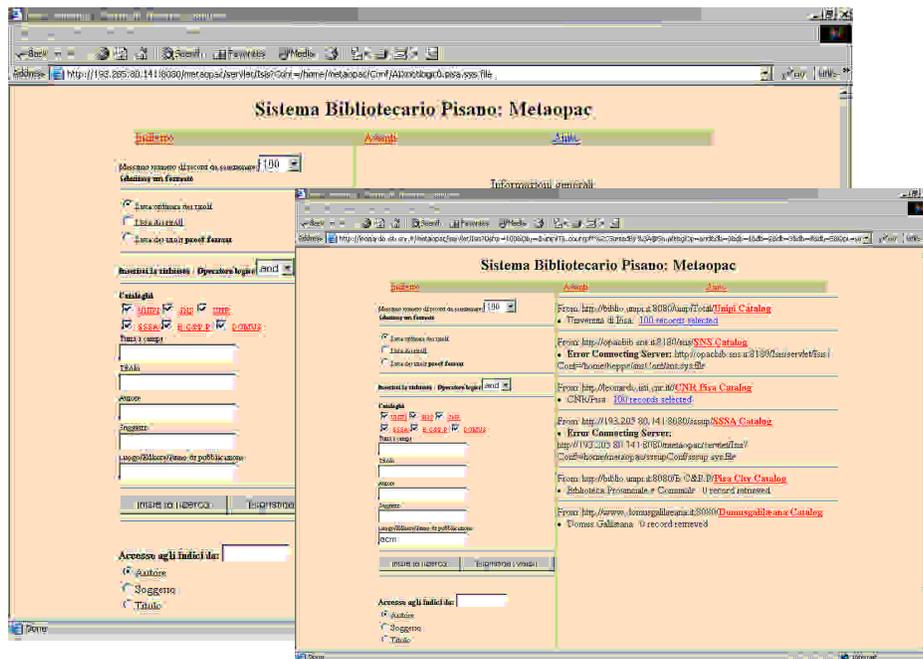


Fig. 1 Interfaccia comune del MOP

Attualmente il MOP rende disponibili i cataloghi delle biblioteche dei quattro Enti che hanno stipulato il protocollo d'intesa e delle Biblioteche Comunale e Provinciale e della Domus Galilaeana e gestisce, a oggi, circa 900.000 record bibliografici relativi a materiale diverso quale monografie, pubblicazioni seriali, e pubblicazioni del personale di singoli enti (ISTI e IFC CNR). Attraverso il MOP è, inoltre, possibile accedere al documento intero, in tutti i casi in cui è permesso.

2 Organizzazione logica e fisica del MOP

Il MOP è una istanza specializzata del software Isis/Gas [1] [7], la cui architettura verrà descritta in seguito. Isis/GAS è un sistema per l'accesso a basi di dati testuali collocate nella rete Internet che utilizza come contenitore di informazioni il sistema di accesso CDS/Isis².

Ogni Istituzione partecipante al MOP installa su un suo *Internet Server* il software Isis/Gas. A cadenza regolare, ogni Istituzione scarica i dati dal proprio sistema di automazione bibliotecaria, in formato ISO2709, e li ricarica automaticamente sul sistema Isis/Gas. Il caricamento dei dati nel data base server viene fatto localmente con procedure ad hoc. In tempo reale essi sono resi disponibili agli utenti.

Ogni catalogo (data base) è descritto al software Isis/Gas mediante un file di configurazione che mediante la definizione di un insieme di parametri permette di gestire l'accesso e la presentazione dei dati estratti. Per cui ogni database fisico è descritto da uno specifico file di configurazione.

Un insieme di database fisici, che risiedono sullo stesso *Internet Server* oppure su *Internet Server* diversi, possono essere definiti come appartenenti a un insieme logico, chiamato database logico, anch'esso descritto al software Isis/Gas. Il software Isis/Gas accede automaticamente alle risorse descritte come appartenenti ad un database logico, e restituisce all'utente finale la lista dei risultati selezionati.

Un database logico può contenere nella sua descrizione, oltre che database fisici, anche database logici. La ricorsività è assicurata dal software, mentre non viene controllato un eventuale loop ricorsivo generato da un'erronea definizione dei database logici descritti.

Uno stesso *Internet Server* può gestire più data base logici e più database fisici contemporaneamente.

Un *Internet server* può contenere anche un solo database logico, nel qual caso fa le funzioni di database gateway tra l'utente finale e i database componenti il database logico.

Il MOP è organizzato come un contenitore logico di informazioni ed è definito a Isis/Gas come un data base logico contenente i data base fisici componenti. Il contenitore è gestito dal software Isis/GAS che cura l'accesso ai cataloghi fisici in modo trasparente e sottopone all'utente finale la lista dei risultati che soddisfano la richiesta.

² CDS/Isis sviluppato presso l'Unesco

4 Stefania Biagioni e Giuseppe Romano

Ogni catalogo fisico è indipendente e può essere utilizzato, in un altro contesto, da un'altra applicazione.

Il software che gestisce il MOP fa riferimento a un file di configurazione che descrive, nel suo insieme, l'applicazione e i suoi componenti. Per motivi di sicurezza e recovery esistono più installazioni che agiscono come MOP:

- il server bibserv.isti.cnr.it:8080,
- il server 146.48.84.6:8080,
- il server 193.205.80.141:8080.

Nello schema del MOP, della Fig. 2, le sigle hanno il seguente significato:

- DS equivale a *Database Server* e rappresenta la funzione di accesso al catalogo locale;
- IS equivale a *Index Server* e rappresenta la funzione di accesso agli indici estesi ottenuti sia localmente che cumulati come somma di tutti i cataloghi e utilizzabili per accedere la base di dati locale o le basi di dati remote;
- AS equivale a *Access Server* e rappresenta la funzione di accesso a insiemi logici e/o fisici di database remoti e/o locali.

Il MOP, attualmente, è costituito da sei Internet server che installano il software Isis/GAS [2]:

1. bibserv.isti.cnr.it:8080 (solaris 9) che gestisce il database logico di recovery del MOP, l'accesso alla lista/indici globali per autore, titolo e soggetto e il data base fisico di recovery del CNR.
2. biblio.unipi.it:8080 (solaris 8) che gestisce il database fisico dell'Università di Pisa e della biblioteca Comunale e della biblioteca Provinciale.
3. opacbib.sns.it:8180 (debian Linux) che gestisce il database fisico della Scuola Normale Superiore.
4. 193.205.80.141:8080 (RedHat Linux) che gestisce il database fisico della Scuola Superiore Sant'Anna e il data base logico operativo di MOP.
5. 146.48.84.6:8080 (Windows2000 Professional) che gestisce il database fisico operativo dell'Area della Ricerca del CNR di Pisa e il data base logico di recovery di MOP e l'accesso alla lista/indici globali.
6. www.domusgalilaeana.it:8080 (RedHat Linux) che gestisce il database fisico della Domus Galilaeana di Pisa.

In conclusione, come si evidenzia nella Fig. 2, la funzioni di *Index Server* globale sono supportate dal server bibserv.isti.cnr.it:8080, mentre le funzioni di *Access Server* sono eseguite dal server 193.205.80.141:8080 scelto per le sue caratteristiche hardware molto "performanti". Il server 146.48.84.6:8080 è utilizzato, invece, come server di recovery nel caso di crash dei servizi a regime.

Il bilanciamento delle operazioni descritte è realizzato definendo particolari file di configurazione del MOP che risiedono sui relativi server.

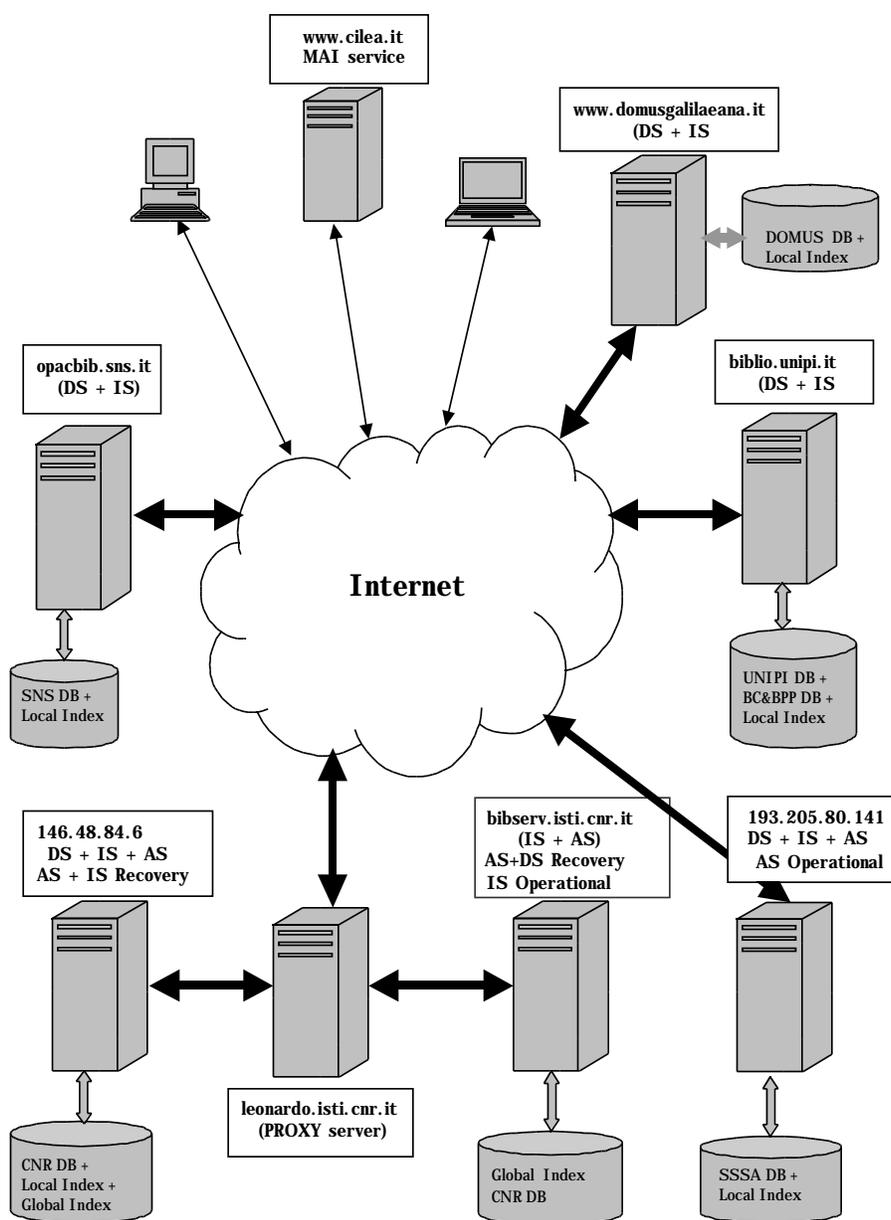


Fig. 2 Organizzazione logico-fisica del MOP

3 Architettura del sistema Isis/Gas

Il MOP utilizza il software Isis/GAS ed un insieme di programmi specifici per la sua gestione. Il software Isis/Gas è un sistema per l'accesso all'informazione memorizzata su database CDS/ISIS distribuiti su rete IP. Il software è costituito da un insieme di moduli java che costituiscono il servlet ISIS per l'accesso a database locali e per la gestione dell'interfaccia utente, altri moduli java per l'accesso ai database remoti via protocollo http (servlet GAS), e da un modulo eseguibile, di dominio pubblico, per l'accesso alle banche dati Isis importate sui sistemi Microsoft Windows e/o Unix.

Il sistema permette l'accesso a database locali, cioè installati sullo stesso host su cui è installato il servlet Isis e a insiemi di basi di dati logiche costituite da più database fisici collocati o sulla stesso host o indifferentemente su altri host remoti che implementano il servlet GAS.

Ogni database sia logico che fisico viene identificato da un file di configurazione che deve essere installato sulla stesso host su cui viene installato il servlet Isis. Il file di configurazione che definisce un data base logico, visto come collezione di database fisici e/o logici distribuiti, contiene la *Logic Database Table* che è un file, noto al servlet, che contiene le informazioni necessarie per accedere a tutti i data base che costituiscono l'insieme database logici [1].

E' compito del sistema Isis/GAS eseguire ricorsivamente gli accessi ai server garantendo l'integrità dei risultati. E', invece, responsabilità del gestore del sistema che la ricorsività sia finita.

Il servlet Isis viene utilizzato per accedere al database locale (cioè installato sullo stesso host che implementa il server) e per accedere a liste ordinate costruite su dati estratti dal database stesso mediante l'utilizzo di funzioni specializzate. L'accesso alle liste permette la selezione di elementi che possono essere utilizzati per una successiva selezione sul database stesso o su altri database locali e/o remoti.

Il sistema svolge quattro funzioni fondamentali (Fig. 3):

1. *User Interface Manager* (UIM): gestione dell'interfaccia utente;
2. *Database Server* (DS): per la selezione e l'accesso al database locale;
3. *Index Server* (IS): per la selezione e l'accesso agli indici (liste ordinate) relative al data base;
4. *Access Server* (AS): per l'accesso a database remoti, cioè collocati su host diversi dall'host su cui è operativo il servlet Isis.

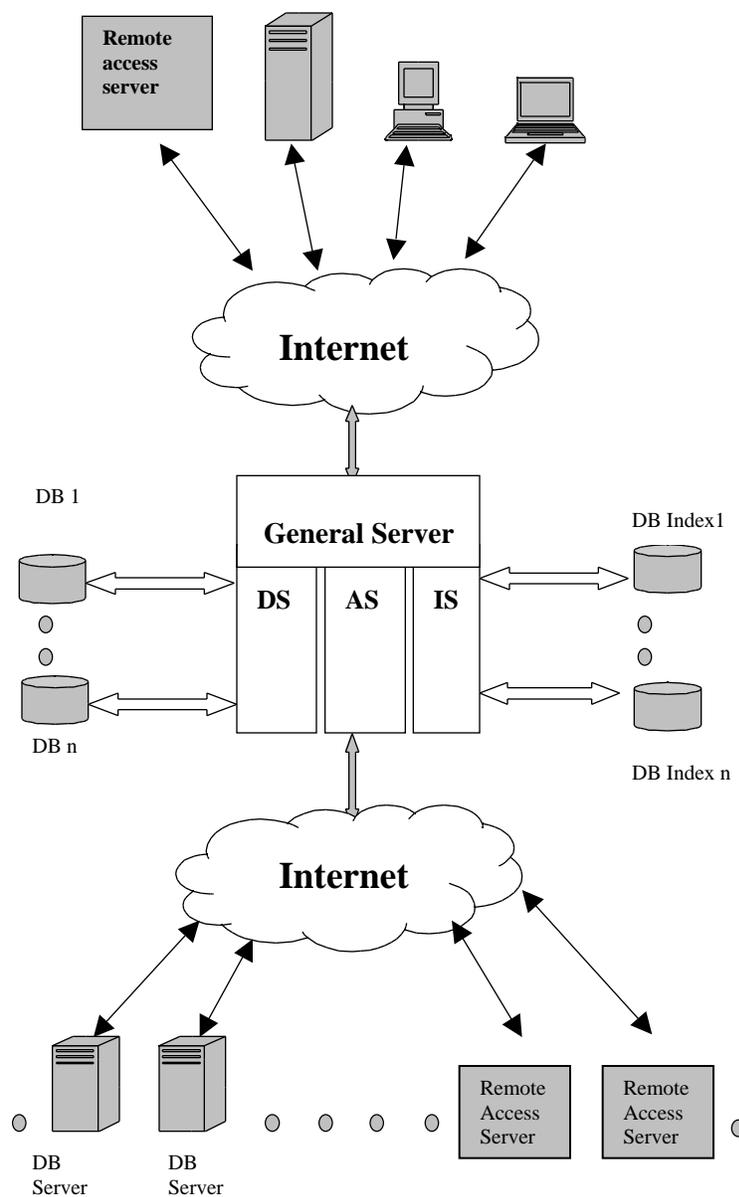


Fig. 3 Organizzazione generale del sistema

3.1 La User Interface Manager (UIM)

Il sistema viene attivato tramite il protocollo http, sulla macchina server, utilizzando il metodo GET che può essere attivato sia dalla interfaccia utente standard del sistema, sia da un qualunque applicativo, tra cui il servlet GAS stesso, sia, infine, da una URL posizionata in qualunque locazione di una pagina HTML che venga generata staticamente o dinamicamente. All'attivazione del servlet il sistema carica la definizione del database stesso, della sua locazione, delle modalità di interazione con le liste/indici precostituiti, e delle modalità di attivazione delle selezioni/ricerche.

La descrizione del database viene memorizzata su un file di testo e consta di un insieme di definizioni che possiedono per la maggior parte valori default. Poiché la descrizione viene caricata automaticamente ogni volta che viene attivato il servlet, le eventuali modifiche divengono operative in tempo reale e quindi permettono una rapida messa a punto del sistema.

Il software che implementa il sistema ha inglobate, al suo interno, le ulteriori tre funzionalità principali sopra menzionate.

3.2 Il Database Server (DS)

Il Database Server accede ai database locali eseguendo tre operazioni fondamentali:

1. *Ricerca*: selezione degli oggetti che soddisfano le richieste e sottoporre all'utente una lista ordinata/non-ordinata degli oggetti stessi affinché l'utente possa o perfezionare ulteriormente la sua selezione o accedere alla presentazione parziale o completa degli oggetti selezionati.
2. *Navigazione*: selezione degli oggetti a partire da uno specifico oggetto e successiva visualizzazione della lista ordinata/non-ordinata del primo oggetto e dei successivi in modo che l'utente possa accedere alla presentazione parziale o completa degli oggetti presenti nella lista.
3. *Visualizzazione*: presentazione del contenuto parziale o totale dell'oggetto. Il contenuto viene presentato in forma testo o immagine ed eventualmente ipertestualizzato selettivamente in modo da permettere una ulteriore navigazione - ricerca - visualizzazione nel database stesso.

La *ricerca* avviene per campi. Tra ogni campo l'operatore logico definito per default è "and"; sono possibili anche gli operatori "or" e "not". I risultati della ricerca vengono presentati ordinatamente/non-ordinatamente a seconda che si sia optato per un formato di restituzione ordinato e non, con una organizzazione in pagine successive fino a completare la lista dei risultati che soddisfano la richiesta. Ogni elemento della lista selezionata può contenere uno o più puntatori ipertestuali all'oggetto stesso o ad altri oggetti esterni al sistema o collocati su altri servlet Isis/GAS remoti.

La *navigazione* avviene previo posizionamento su un oggetto all'interno del database. E' possibile posizionarsi su due tipi di oggetti:

1. Su uno specifico record utilizzando il numero di sistema che identifica uno specifico documento.
2. Su una specifica istanza dell'albero binario che costituisce la lista invertita del database.

In entrambi i casi, come risultato, si producono liste ordinate per numero identificativo di sistema o istanze ordinate degli elementi della lista invertita. Naturalmente ogni elemento della lista prodotta può essere un semplice elemento passivo (cioè vuoto d'eventuale successiva informazione) o un elemento attivo che può, cioè, espandere il suo contenuto per selezionare ulteriormente classi di oggetti dal database o da altre sorgenti esterne e/o interne al sistema.

La *visualizzazione* avviene tramite l'acquisizione delle informazioni contenute dall'oggetto stesso. L'oggetto può essere acceduto singolarmente o unitamente a un insieme di altri oggetti. In entrambi i casi l'oggetto/i è identificato dal suo numero di sistema o da una lista degli stessi. Il contenuto dell'oggetto può essere scaricato, a seconda dell'applicazione e delle necessità, in un formato *XML well-formed* e quindi inviato al browser di utente con gli eventuali fogli di stile (xsl file) necessari per la presentazione, o scaricato in un formato XML e successivamente convertito dal server in formato HTML per la visualizzazione dal browser di utente. La traduzione XML/HTML viene fatta utilizzando un insieme di regole che sono memorizzate su uno specifico file noto al servlet Isis/GAS

3.3 L'Index Server (IS)

L'Index Server è quella parte del sistema che permette l'accesso agli indici (liste omogenee ordinate) del database, per permettere all'utente di visualizzarle e/o utilizzarle nel contesto stesso del database sia per accedere a ulteriore informazione che per raffinare e correggere il contenuto stesso degli oggetti memorizzati. Gli indici possono essere originati da due tipi di applicazioni diverse:

1. Dagli elementi generati per la costruzione delle liste invertite del database, e in questo caso, al di là della omogeneità complessiva estesa su tutte le liste invertite del database, è possibile costruire classi omogenee di elementi utilizzando le tecniche di indicizzazione rese disponibili dal sistema stesso: per esempio, prefissando i valori che appartengono alla stessa classe con un prefisso univoco. Questo tipo di indice ha un costo molto limitato, appartiene infatti alla gestione del database stesso, ma presenta due grandi limitazioni: la lunghezza della stringa non può superare i trenta caratteri e la stringa stessa viene presentata in caratteri maiuscoli e, per problemi di normalizzazione, è bene non contenga caratteri diacritici. Ha, di contro, un unico vantaggio, quello di avere disponibile il numero delle occorrenze della stringa all'interno del database.
2. Da stringhe di caratteri estratti, secondo necessità specifica, dal database stesso (o da più data base locali e/o remoti) e convertite, con un software ad hoc, in un database parallelo a quello principale da utilizzare come supporto. In questo

modo vengono costruiti gli indici globali di MOP. Questo tipo di indice è notevolmente più costoso sia per la complessità della messa a punto della sua costruzione che per lo spazio disco e il tempo di calcolo richiesto; inoltre, non rende disponibili immediatamente il numero di oggetti cui la stringa appartiene. Offre invece il vantaggio di non avere limitazione nella lunghezza della stringa e di potere utilizzare i caratteri diacritici e i caratteri di interpunzione nella stringa stessa. La stringa così ottenuta è completa da un punto di vista informativo e inoltre può essere facilmente utilizzata per la messa a punto dei dati e la eliminazione di omonimie.

Il posizionamento sull'indice di partenza viene fatto, per default, a partire dal valore più basso dell'indice per un numero limitato di istanze dell'indice stesso. Intervenedo esternamente, con specifiche ad hoc (javascript, o pagine statiche) è possibile partire a navigare nell'indice da posizioni diverse da quella iniziale. Il sistema genera automaticamente i comandi per la navigazione dell'indice (back, more) e, utilizzando il formato di presentazione specificato (parametro Obj) genera l'http-pointer per selezionare oggetti a partire da quella particolare stringa dell'indice.

Anche in questo caso, la lista degli elementi dell'indice che sono stati selezionati dalla lista invertita o dall'indice alternativo viene scaricata in un formato XML e successivamente convertita dal server in formato HTML per la visualizzazione dal browser di utente.

3.4 L'Access Server (AS)

L'*Access Server* è quella parte del sistema che accede ai database che sono collocati su host diversi da quello su cui è attivo. In realtà non esiste limitazione sulla collocazione del database remoto: esso potrebbe trovarsi anche sullo stesso host ma viene, in realtà, visto come un oggetto remoto.

L'*Access Server* è attivato quando, all'interno della descrizione del database, è definita una *Logic Database Table*. In questo caso il database descritto viene identificato come un database logico costituito da un insieme di database a loro volta fisici e/o logici.

Alla sua attivazione l'*Access Server* verifica la consistenza formale della query che viene sottoposta: nel caso di errore esso viene segnalato all'utente. Nel caso di query accettata è caricata la descrizione del database logico: *Logic Database Table*.

La *Logic Database Table* è costituita da tante righe quante sono i database che costituiscono il database logico. Ogni riga contiene (deve contenere) sei parametri:

1. Il primo parametro contiene la URL di identificazione del servlet Isis.
2. Il secondo parametro identifica il database (il file di configurazione che descrive il database remoto).
3. Il terzo parametro identifica il massimo numero di millisecondi che il server può attendere per ricevere la risposta dal server remoto (timeout).
4. Il quarto rappresenta il massimo numero di oggetti che devono essere restituiti dal server remoto.
5. Il quinto e il sesto parametro rappresentano due stringhe che sono utilizzate nella presentazione dei risultati provenienti dal server remoto.

Il servlet GAS attiva tanti processi concorrenti simili (thread) quanti sono i database presenti nell'ordine in cui sono descritti nella *Logic Database Table*, e attende le risposte dai server remoti verificandone lo stato dei server ogni 100 millisecondi fino a che non abbiano risposto tutti o sia scaduto il massimo timeout. A questo punto se ci sono ancora connessioni attive esse vengono chiuse e, infine, il sistema si prepara a valutare le risposte ottenute.

Le risposte ottenute sono memorizzate in file temporanei sul server chiamante sia che esse contengano risultati che messaggi di errore.

L'accesso al server remoto può produrre quattro risultati diversi:

1. La URL costruita localmente non è corretta: viene inviato un messaggio di URL malformata
2. La URL è corretta ma il server remoto è irraggiungibile: viene inviato un messaggio tipo: Error Connecting Server
3. La URL è corretta, il Server è raggiungibile ma il Server è sovraccarico e non riesce a eseguire la ricerca nel tempo dovuto: viene inviato un messaggio del tipo "Timeout Expired connecting server".
4. La URL è corretta, il Server è raggiungibile e invia una risposta plausibile.

4 Sviluppi futuri e conclusioni

Anche se nato dall'iniziativa degli Enti su citati il MOP, in armonia con la sua natura, è aperto a tutte le biblioteche cittadine che ne condividano la filosofia e le finalità. Hanno già chiesto di aderire all'iniziativa: la Biblioteca Universitaria, la Domus Mazziniana, la biblioteca Serrantini e Bibliolandia (insieme di biblioteche della provincia di Pisa)

Data la flessibilità del sistema Isis/GAS, l'esistenza di programmi di scaricamento e caricamento da sistemi di automazione bibliotecari più diffusi come Aleph, Libero e Isis stesso, e la possibilità di realizzazioni a costi contenuti, si potrebbe pensare di estendere il MOP a realtà simili in ambito nazionale.

Le funzionalità del sistema verranno ampliate dando la possibilità di creare delle viste particolari a partire dagli stessi dati per creare collezioni e punti di accesso specializzati. Un primo esempio è la realizzazione dell'interfaccia del catalogo delle pubblicazioni dell'ISTI del CNR³ (Fig. 4).

³ <http://leonardo.isti.cnr.it/Isis/servlet/Isis?Conf=.%2FCnf%2Fcnr-pub-list.sys.file>

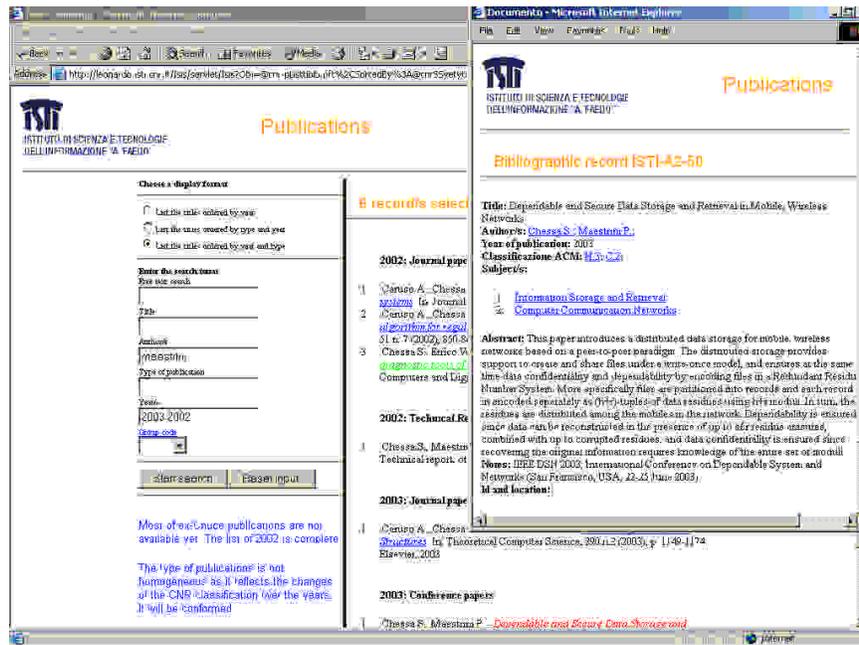


Fig. 4 CNR-ISTI Publications Home page

Riferimenti

1. Giuseppe A. Romano, Isis/GAS : manuale di installazione e d'uso
2. URL: <http://leonardo.isti.cnr.it/metaopac/IsisGASInstall.rtf>
3. Giuseppe A. Romano, Il Metaopac Pisano: descrizione tecnica e procedure di recovery
URL: <http://leonardo.isti.cnr.it/metaopac/MetaopacPisano.rtf>
4. Giuseppe A. Romano, Il Metaopac Pisano: l'organizzazione logica e il mapping fisico
URL: <http://leonardo.isti.cnr.it/metaopac/MetaOpacLogic.pdf>
5. <http://www.libris.it/index.php?lang=it&menu=a%3A2%3A%7Bi%3A0%3Bi%3A1%3Bi%3A1%3Bi%3A4%3B%7D>: Aleph.
6. <http://www.infologic.it/ita/libero.htm>: LIBERO.
7. http://portal.unesco.org/ci/ev.php?URL_ID=2071&URL_DO=DO_TOPIC&URL_SECTION=201&reload=1079025510&PHPSESSID=9c9bb4c0426b1e9c50390922cf4b1bdd: CDS/Isis.
8. Giuseppe A. Romano, Isis/GAS: un sistema multi piattaforma per l'accesso a basi di dati testuali via Internet - URL: <http://leonardo.isti.cnr.it/metaopac/IsisGAS.rtf>