

Progetto indici e sommari in OPAC

A cura del **Gruppo di lavoro tecnico-informatico del CIB**: Teresa De Benedictis, Raffaella Gaddoni, Antonio Puglisi, Marialaura Vignocchi.

Premessa:

L'accesso al contenuto dei documenti è sempre più strategico al recupero dell'informazione pertinente. Avere, poi, a disposizione dati di spoglio per pubblicazioni periodiche e opere collettive rappresenta un ampliamento della base dati di ricerca bibliografica indispensabile per poter aspirare a risultati più esaustivi. L'indicizzazione semantica e la catalogazione analitica da parte di professionisti, sole procedure con garanzia di qualità dei risultati di ricerca, necessitano tuttavia di grandi investimenti e rappresentano una soluzione non sempre sostenibile.

Il progetto del C.I.B. rappresenta per gli utenti istituzionali un arricchimento del potenziale informativo del catalogo di Polo, da un punto di vista sia analitico che semantico, a basso investimento per le biblioteche di ateneo.

Rispetto ad esperienze analoghe avviate in passato (IUAV di Venezia) o tuttora in corso (BNCF - Progetto ARSBNI), il progetto del C.I.B. si presenta innovativo sia per l'organizzazione gestionale, distribuita e di tipo collaborativo, sia per la soluzione tecnologica adottata che offre alle biblioteche di ateneo una piattaforma comune centralizzata per il trattamento dei materiali e dei dati.

Finalità e oggetto del progetto:

Il progetto si propone di ampliare e potenziare l'accesso al contenuto delle collezioni cartacee disponibili nelle biblioteche di ateneo rendendo consultabili dall'OPAC di Polo gli indici e i sommari dei documenti catalogati mediante:

- la visualizzazione delle immagini digitalizzate degli indici attraverso il link ipertestuale presente sul singolo record;
- la ricerca per parola negli indici attraverso un apposito campo sulla home page dell'OPAC di Polo che consente il recupero delle informazioni dalle pagine di indice preventivamente digitalizzate e poi trattate all'OCR.

Un **gruppo di lavoro tecnico-informatico del CIB**, si incarica di realizzare:

- un database centralizzato per lo storage e la gestione e il trattamento dei file .jpeg .djvu e .txt (compresa eventuale produzione di metadati per le immagini archiviate);
- interfaccia web per la procedura di caricamento dei file sul database ad uso delle biblioteche (procedura di autenticazione operatore, campi per il caricamento dei files...);
- richiesta modifiche all'OPAC a Data Management;
- procedura di dialogo fra database ORTI e Opac di polo e ORTI e gestionale Sebina;
- interfaccia di visualizzazione e navigazione immagini per l'utente;
- assistenza tecnica per l'implementazione hw per digitalizzazioni e OCR;
- indicazioni di requisiti minimi di qualità grafica per le immagini digitali;
- test di verifica delle modifiche funzionali richieste a Data Management.
-

Un gruppo di lavoro **tecnico-bibliotecario incaricato dallo SBA di Ateneo** avrà il compito di:

- individuare il bacino di utenza e analizzare i bisogni degli utenti;
- elaborare le linee guida del servizio: utenti, scelta dei materiali, standard tecnici e biblioteconomici;
- definire gli indicatori per la misurazione della qualità del servizio;
- sperimentare le procedure

- testare la maschera di visualizzazione e navigazione per l'utente;
- formare gli operatori;
- promuovere il servizio;
- formare gli utenti;
- misurare e valutare la qualità del servizio.
-

Obiettivi

Il progetto si propone di ottenere i seguenti vantaggi in termini di servizio per gli utenti di ateneo:

- l' ampliamento dei risultati conseguibili attraverso una di ricerca per argomento nel catalogo;
- la possibilità di visualizzare "linkati" al record bibliografico dell'OPAC, pagine significative del documento come indici, sommari e altre pagine preliminari, fondamentali per l'accesso semantico, che normalmente sarebbero visibili solo aprendo fisicamente il documento e quindi recandosi personalmente presso una biblioteca;
- un'unica modalità di visualizzazione e navigazione delle immagini collegate al record bibliografico.

Il progetto si propone di ottenere i seguenti vantaggi per le biblioteche di ateneo:

- una piattaforma tecnologica centralizzata per lo stoccaggio, il trattamento, la visualizzazione e la navigazione delle immagini associate al record bibliografico che consente di realizzare un abbattimento dei costi di implementazione e manutenzione di hw e sw;
- una procedura automatizzata per la realizzazione del link multimediale fra record bibliografico e immagini digitali che non comporta l'intervento dell'operatore sul singolo record nel sw gestionale di Sebina, rendendo possibile l'impiego di personale non professionale;
- l' elaborazione di policy comuni e condivise per il trattamento dei dati e la soluzione di problemi.

Sviluppo del progetto ed integrazione con altri progetti del SBA

E' ipotizzabile un utilizzo delle procedure automatizzate sviluppate nell'ambito del presente progetto, ai fini di un possibile trattamento delle informazioni in formato elettronico (sintesi di contenuto, abstract, table of content, indici, sommari, prefazioni) che si possono ottenere direttamente dagli editori come conseguenza di accordi contrattuali all'interno di servizi di fornitura delle risorse bibliografiche cartacee sia di ateneo che di singola struttura.

Requisiti

Hw e sw

centralizzati:

- per il server e il database: in fase di avvio si possono utilizzare le dotazioni del CIB per la gestione dei progetti di digitalizzazione;
- uno scanner e un sw ocr (in dotazione con lo scanner) già acquisiti dal CIB;
- Si può prevedere l'acquisizione centralizzata di uno scanner per grandi formati (il costo dipende dal formato richiesto);

per le biblioteche:

- uno scanner di buona qualità di dimensioni variabili a seconda del formato richiesto;
- un sw ocr di solito in dotazione con lo scanner .

Personale:

il progetto prevede la costituzione di 2 team di sviluppo:

- un gruppo con competenze bibliotecomico-informatiche;
- un gruppo con competenze prettamente biblioteconomiche;

- il raccordo fra i due gruppi è garantito dal coordinamento e dalla presenza contemporanea nei due gruppi di un bibliotecario e di un informatico.

L'integrazione con altri progetti di ateneo avverrà a livello di coordinamento.

Il servizio a regime richiederà invece:

- un amministratore centrale di sistema in forza al CIB;
- un coordinatore centralizzato del servizio in forza al CIB;
- un operatore non professionale a livello di biblioteca.

Criticità

- Tempi e costi di implementazione modifiche all'OPAC e gestionale Sebina a carico di una ditta esterna;
- Legge sulla tutela del diritto d'autore che restringe il campo di applicazione del progetto;
- Percentuale di errore del sw di OCR che limita l'accuratezza delle ricerche testuale.

Punti di forza:

- possibilità di riutilizzo di competenze e procedure utilizzate per il progetto DigLib;
- possibilità di utilizzare hw già tarati su elevati livelli di servizio;
- requisiti minimi di hw e sw a carico delle biblioteche;
- possibilità di impiego di operatori non professionali da parte delle biblioteche;
- notevole miglioramento del potenziale informativo del catalogo;
- possibilità di evoluzione del progetto grazie a possibili accordi con editori e fornitori per l'acquisizione alla "fonte" delle informazioni che al momento le biblioteche devono digitalizzare dalle copie analogiche.

Costi

A livello centralizzato:

hw e sw

- modifiche all'opac e al gestionale sebina (7000+iva euro);
- acquisto degli scanner e sw ocr centralizzati (djvu+ocr) (2350 euro);
- eventuale potenziamento della dotazione hw centralizzata (il costo dipende dalla scelta e dal dimensionamento dell'hw fatta in base ai risultati della sperimentazione);

personale:

gruppo di lavoro informatico:

- 1 informatico a contratto di collaborazione esterna dedicato part-time al progetto;
- 1 informatico di Data Management nel ruolo di raccordo fra CIB e DM nel gruppo di lavoro;
- 1 bibliotecario del CIB nel ruolo di coordinamento;
- 1 bibliotecario del CIB nel gruppo di lavoro.

Gruppo di lavoro bibliotecario:

- il coordinatore del gruppo di lavoro informatico;
- il bibliotecario del gruppo di lavoro informatico;
- l'informatico part-time del gdl informatico;
- bibliotecari dello SBA.

I costi per il personale vengono calcolati in termini di tempo.

A livello di biblioteca:

hw e sw:

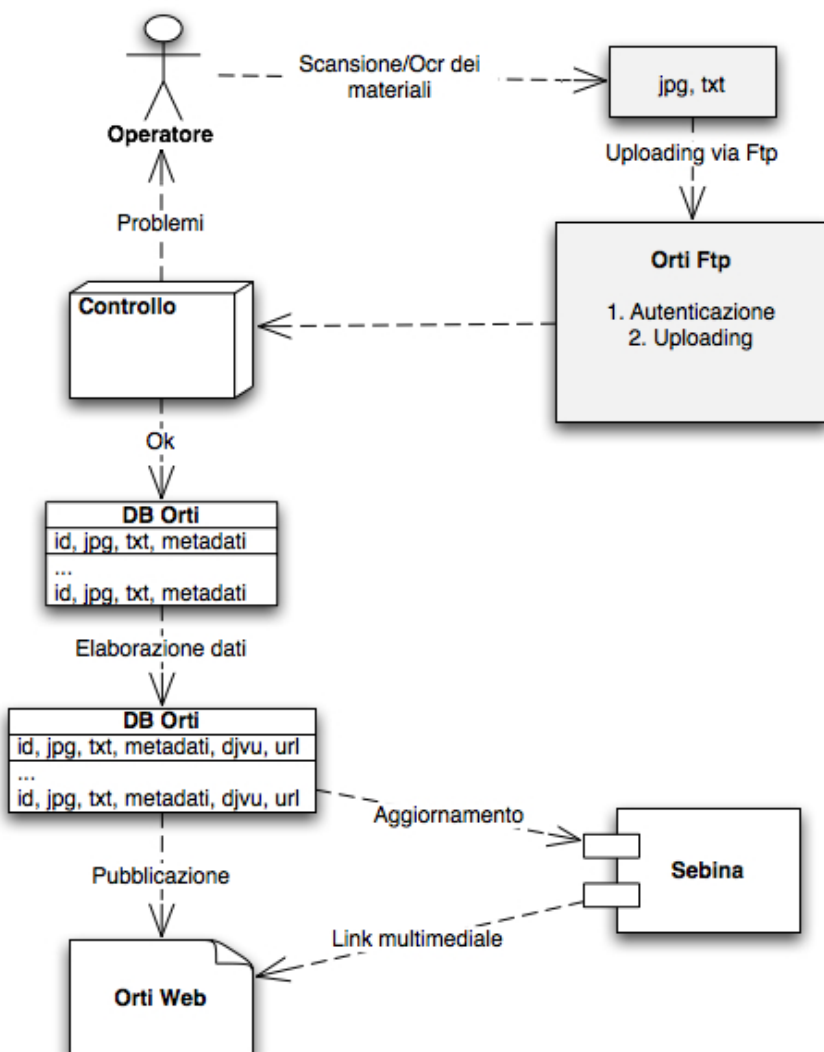
- scanner e ocr 200 euro;

personale:

operatore non professionale: il costo dipende dal carico di lavoro assegnato e dal tipo di contratto dell'operatore.

Indicativamente è possibile scansionare, trattare all'ocr e salvare come file .jpeg e .txt circa 20 pagine all'ora a questo si devono aggiungere i tempi di caricamento dei file nel database centralizzato.

Flusso di lavoro



- Produzione materiali.
- Raccolta/controllo materiali.
- Elaborazione.
- Aggiornamento Opac.
- Pubblicazione contenuti.