



Pochi vantaggi, alcuni limiti ed errori nel Catalogo Storico per Autori digitalizzato della BIASA

Francesco Franco

L'OPAC SBN (Catalogo del Sistema Bibliotecario Nazionale) ha avuto 80.643 visitatori nel febbraio 2008 (link); una quantità scarsa considerato il numero di ricercatori, docenti e studenti (della scuola e dell'università). Gran parte del mondo della scuola, purtroppo, ancora ignora l'esistenza dell'ICCU (Istituto Centrale per il Catalogo Unico) e, soprattutto, le enormi potenzialità offerte da questo OPAC, capace di interrogare simultaneamente un grandissimo numero di biblioteche italiane.

Da qualche anno, però, un altro progetto cerca di rendere disponibili anche i vecchi cataloghi cartacei, senza riversarli nell'OPAC SBN.

Il progetto BDI (Biblioteca Digitale Italiana) di digitalizzazione del Catalogo storico (cartaceo) della Biblioteca di Archeologia e Storia dell'Arte di Palazzo Venezia di Roma (BIASA), si inserisce nell'ambito del «Programma di digitalizzazione dei cataloghi storici» che ha coinvolto ben 35 importanti biblioteche italiane, a opera dell'ICCU.

La possibilità di consultare l'intero patrimonio librario delle biblioteche, da qualsiasi postazione internet, ha un'influenza enorme sulla rapidità e sull'abbassamento dei costi della ricerca italiana e internazionale, soprattutto di quella umanistica, ancora troppo lontana dalla rete. Non doversi recare fisicamente nelle singole biblioteche delle varie città italiane, per consultare manualmente le schede, è forse l'unico enorme vantaggio del progetto. Come utente (e non sono il solo) avrei preferito il riversamento diretto dei cataloghi storici nell'OPAC SBN (sopportando il piccolissimo peso, inevitabile e lentamente sanabile, della catalogazione multipla di migliaia di libri identici), invece di dover consultare in internet i cataloghi cartacei scheda dopo scheda.

Potendo definirmi uno studioso, ma non avendo attualmente competenze né di biblioteconomia né di software, in questo articolo mi limiterò a segnalare alcuni limiti e due impedimenti (attivi da alcuni mesi) nel Catalogo Storico Digitalizzato della BIASA per Autori e Titoli, che ho consultato poco (verificandone l'iniziale approssimativa efficacia, insieme ad alcuni funzionari, a partire dai primi di dicembre 2007).

L'operazione di foto-riprodurre o di scannerizzare, in formato digitale, le schede o i microfilm dei cataloghi delle biblioteche, per poi limitarsi a indicizzare gli autori (cognome e nome) e, in caso di opere anonime, anche i titoli, comporta una mole di lavoro tale da rendere maggiormente produttivo, a mio avviso, il riversamento diretto manuale (o semi-manuale) delle schede nell'OPAC SBN. Tale riversamento, nei prossimi anni, credo che verrà comunque effettuato. L'ICCU, infatti, pochi anni fa, comunicava:

«La scelta di convertire in formato digitale i cataloghi storici delle biblioteche italiane si fonda sulla considerazione, emersa nello Studio di Fattibilità, che in alcuni paesi, con particolare riferimento a quelli dell'area dell'Europa orientale e di lingua tedesca, tale attività ha rappresentato quasi sempre il punto di partenza per la realizzazione della Biblioteca Digitale».

<http://www.iccu.sbn.it/genera.jsp?id=261> (pagina aggiornata al 15/12/2006).

Non mi è chiaro se per «Biblioteca Digitale» si intendesse una «Biblioteca on-line full text» (testi leggibili direttamente in rete) o semplicemente un OPAC storico (un catalogo cartaceo reso accessibile on-line), ma non è questa la sede più idonea per affrontare questo tema.

Sarà utile, comunque, per realizzare una «Biblioteca Digitale» aver eseguito alcune centinaia di migliaia di acquisizioni digitali non indicizzate? (per limitarmi al solo catalogo della BIASA) Sarà vantaggioso copiare (via dattilografica) i dati da immagini in formato «tiff», invece che da schede cartacee?

Non sarebbe stato più opportuno usare direttamente i migliori software di OCR (Riconoscimento Ottico dei Caratteri), in grado di ridurre al minimo il lavoro delle correzioni da effettuare poi in via dattilografica, per catalogare i risultati, ripuliti, nell'OPAC SBN? (evitando migliaia di record multipli tra i vari cataloghi storici e avendo un solo rapido strumento nazionale principale).

Oppure non si poteva creare almeno un unico Catalogo Storico Italiano, ben indicizzato, invece di realizzare 35 cataloghi di biblioteche che, sostanzialmente, devono essere consultati uno ad uno, adattandosi alla lentezza e alla scarsa versatilità del software che li gestisce?

Il Comitato Guida della BDI ha costituito un Gruppo di lavoro che ha formulato un unico protocollo tecnico:

«Le immagini saranno prodotte nel formato grafico TIFF CCITT G4, di norma in bianco e nero. La loro dimensione su disco potrà giungere indicativamente a 15-20kb per le schede e a 90-100kb per le pagine di volume. La risoluzione sarà di solito intorno ai 200-300 ppi (pixel per inch). Va sempre previsto il ricorso a programmi di miglioramento e fotoritocco [...]. Le immagini sono destinate alla trasmissione via Internet e perciò dovranno essere il più possibile compatte, ma la buona qualità e la buona leggibilità a video - anche sotto ingrandimento - sono requisito indispensabile del processo di acquisizione. Durante la realizzazione del prototipo [...] il Responsabile dell'attività per conto della Biblioteca concorderà con il Fornitore i parametri medi più opportuni, sia per lo spazio-disco che per la risoluzione, prevedendo se necessario dimensioni maggiori e - nei casi più difficili - anche il ricorso alla scansione a colori». http://www.iccu.sbn.it/upload/documenti/capitati_digicat.pdf

Rispetto al protocollo tecnico non sembra che siano stati sempre rispettati tali requisiti, almeno nel caso del Catalogo della BIASA. Molte schede cartacee difficilmente leggibili sono risultate totalmente illeggibili nel formato digitalizzato (acquisito, credo, quasi sempre dai microfilm) e quelle illeggibili non sono state integrate, neppure manualmente prima della scansione.

La risoluzione di acquisizione, non alta in origine, evidentemente non è stata aumentata nei casi di scarsa leggibilità (come poter pretendere ciò da operatori costretti, quasi sempre in questi appalti, a produrre un numero alto di file in breve tempo?).

Non ho mai visto, per esempio, nessuna immagine a colori nel catalogo della BIASA, come previsto dal Gruppo di Lavoro della BDI (ma credo che tale operazione non ne avrebbe migliorato comunque la leggibilità). Erano previste «verifiche periodiche» e la possibilità «di chiedere la ripetizione delle scansioni difettose»: http://www.iccu.sbn.it/upload/documenti/capitati_digicat.pdf.

Stabilite le obiezioni di principio e alcuni limiti ora passiamo al primo errore, anzi a un impedimento contenuto nel catalogo preso in esame: «Autori e Titoli».

Partiamo da un esempio di ricerca concreta di un autore che ho personalmente effettuato ai primi di dicembre 2007.

Il catalogo è complesso: un utente medio da casa, senza l'aiuto di un bibliotecario, è propenso a utilizzare la «ricerca base». Così ho fatto. Giunto a questa pagina

http://cataloghiarcheologica.librari.beniculturali.it/code/cent_seq_elenco_cataloghi.asp?TipoPercorso=SKD&ResetFilter=Y (vi prego di verificare insieme la ricerca) ho premuto il tasto «ricerca base» e poi ho selezionato la casella «ricerca libera». Volevo visualizzare le schede relative a «Valsecchi Marco», un autore molto produttivo.

Nel catalogo, fino al 8/3/2008, appariva la seguente dicitura:

«Risultati per: valsecchi marco

Tipo di ricerca: Ricerca base libera | Cataloghi esaminati: visualizza l'elenco

Pagina 1 di 12 | Primi 300 risultati di 369»

A questo punto ho sfogliato le pagine che portavano alle relative schede, ma arrivato alla pagina 12, mi sono arrestato al 300° risultato, senza poter andare oltre. La freccia (→) di passaggio alla pagina successiva spariva, insieme a quella per andare all'ultima pagina (→ |). Del resto ero a «pagina 12 di 12» e i risultati erano finiti. Allora dovevo credere che «Valsecchi Marco» avesse 300 schede ? Perché in alto nella pagina leggevo: «primi 300 risultati di 369» ?

Come visualizzare nel catalogo on-line le altre 69 schede ?

Ho tentato questa procedura: ho inserito il numero 13 nella piccola finestrella in basso a destra e premuto (→). Mi sono chiesto: «perché devo digitare la pagina 13 ? Le pagine non sono 12, come è scritto in alto ?»

No, le pagine erano solo "nascoste" e per visualizzare i restanti risultati del nostro «Valsecchi Marco» dovevo digitare il numero di una pagina inesistente !

Ero riuscito ad andare, leggendo in basso, a «pagina 13 di 12»; e poi a «pagina 14 di 12», per scorrere i record fino alla fine a «pagina 15 di 12». In questa pagina (che secondo la matematica più consueta non doveva esistere), in alto, leggevo: «pagina 12 di 12», ma io ero a «pagina 15 di 12» (come leggevo in basso).

Ho segnalato questa situazione, ai primi di dicembre 2007, senza preoccuparmi di sapere in modo certo chi fosse materialmente in grado di eliminare l'inconveniente e avesse l'autorità per farlo. Confesso di non essermene occupato in seguito, per alcune settimane, pensando a un problema "momentaneo" e usufruendo del consueto catalogo cartaceo. Come studioso, Dottore di Ricerca, lavorando quotidianamente con i cataloghi, ho cercato di essere d'aiuto a colleghi, docenti e studenti, segnalando il problema singolarmente ad alcuni di loro.

Spero che l'impedimento venga eliminato prima della pubblicazione di questo articolo (non sono in grado di dire esattamente da quanto tempo persistesse, prima del dicembre 2007). Chi ha fatto ricerche con il metodo sopra descritto, nel Catalogo Storico Digitalizzato, su autori o parole di titolo che superano i 300 record sarebbe opportuno che le rieseguisse. Argan, Crispolti (per esempio) e tutti gli altri termini indicizzati nel titolo ricorrenti in oltre 300 schede, attraverso la modalità «ricerca libera», non possono essere visualizzati nella loro interezza.

Comunque è sempre stato possibile visualizzare tutte le schede relative al nostro «Valsecchi Marco» (analogamente a tutti gli altri autori o termini nel titolo con più di 300 schede in catalogo). Si doveva e si deve tuttora procedere nel modo seguente. Non selezionare «ricerca libera» (tasto che forse sarebbe opportuno eliminare), ma selezionare «ricerca di un gruppo». Poi, individuato il gruppo di schede in cui il cognome del nostro autore è compreso alfabeticamente (nel nostro caso i gruppi VALLEN-VALSEC e VALSEC-VANUXEN), vanno visualizzare fino alla fine del primo "cassetto virtuale", come se fosse un "cassetto reale" (anche a salti). Poi si passa al secondo gruppo, come se re-infilassimo il cassetto nello schedario e prendessimo il secondo cassetto. Qual è il vantaggio di questa catalogazione ?

Si elimina la fatica di sfilare materialmente i cassettei dallo schedario (ho sperimentato, personalmente, che il tempo per sfogliare il cartaceo è più breve di quello necessario per il catalogo on-line).

Attenzione, però, perché anche in questo caso si rischia di intraprendere una strada fuorviante nella ricerca. Dobbiamo digitare il nome «Valsecchi Marco» nella casella in basso a sinistra «cerca il gruppo». Non dobbiamo assolutamente scrivere il nome «Valsecchi Marco» nella casella in alto a sinistra (dal nome quasi identico) «ricerca di un gruppo»; altrimenti andiamo a consultare solo le schede relative a un solo cassetto (senza che sia arguibile che «Valsecchi Marco» è presente con molte altre schede anche in un altro cassetto). Anche in questo caso vi invito a provare (magari aiutandovi nell'operazione con i cassettei a carrello della BIASA).

Suppongo che il primo impedimento segnalato sia derivato dall'aver previsto un numero massimo di 12 pagine per la visualizzazione dei record (partendo dal percorso «ricerca libera» del Catalogo per Autori e Titoli).

La società che ha realizzato il Software è la Amanuense di Firenze (il cui dominio linkabile dalla BIASA attualmente risulta scaduto), che ha realizzato importanti progetti nell'ambito della Biblioteca Digitale Italiana. Sembra che ora sia denominata Softeam Ware S.r.l. (<http://www.archiviodistatovenezia.it/index.php?id=118>) e, oltre a possedere quote di altre società (<http://www.assointernet.it/assointernet/chisiamo/soci/bassnet>), ha vinto molte gare d'appalto, direttamente o come società mandataria, presso importanti biblioteche, università e archivi italiani (fra le

varie:<http://www1.bncf.firenze.sbn.it/notizie/testi/aggiudicazionefondimusicali.htm>;<http://www.bncf.firenze.sbn.it/notizie/testi/avvisoositogara.htm>;<http://www.bncf.firenze.sbn.it/notizie/testi/aggiudicazionebnbn.htm>). La Softeam Ware, che ha un portfolio clienti assolutamente eccellente (http://www.softeamware.it/beni_realizzazioni_recenti_1.html), vanta modelli operativi di ottimo profilo (http://www.softeamware.it/beni_catalogazione.html). Anche se personalmente non mi è chiaro se gli errori siano di sua responsabilità, pur non pregiudicando in modo assoluto la riuscita dell'interrogazione, sarebbe estremamente utile eliminarli.

Questo articolo, terminato in data 8/3/2008, vorrebbe avviare un dibattito generale sulla creazione dei software per la ricerca, che dovrebbero essere testati, in primo luogo, da studenti, ricercatori e docenti, prima che da bibliotecari esperti e da i loro creatori (che ovviamente ne conoscono le modalità di utilizzo).

Spero di aver dato un piccolo contributo al miglioramento delle potenzialità di ricerca del sistema, a favore di colleghi, ricercatori, docenti, bibliotecari e laureandi.