Web-scale discovery services

Principi, applicazioni e ipotesi di sviluppo

Roberto Raieli

Prefazione di Riccardo Ridi

Roma
Associazione italiana biblioteche
2020

Per la valutazione *ex ante* delle pubblicazioni monografiche l'Associazione italiana biblioteche ricorre a due esperti del settore, di cui almeno uno individuato all'esterno del Comitato scientifico.
Il testo viene riesaminato da almeno uno dei due esperti dopo la revisione richiesta agli autori.

Il Comitato scientifico è composto da Giovanni Di Domenico, Anna Galluzzi, Alberto Petrucciani.

Editing Palmira M. Barbini

Una versione epub e cartacea è disponibile in vendita all'indirizzo <http://www.aib.it/negozio-aib/>

© 2020 Associazione italiana biblioteche
Produzione e diffusione: Associazione italiana biblioteche
Viale Castro Pretorio 105 - 00185 Roma
Tel. 064463532, fax 064441139
e-mail aib@aib.it, http://www.aib.it
ISBN 978-88-7812-298-7 (ebook)

Indice

Prefazione (Riccardo Ridi) 5

Introduzione. Ambito, strumenti, attori e valori della scoperta di conoscenza 13

1. Evoluzione dei sistemi di ricerca
1.1 Il rinnovamento degli OPAC 33
1.2 Panorami di ricerca, interazione e scoperta 52
1.3 Problematiche dei sistemi di discovery 62

2. Gli strumenti di ricerca e scoperta
2.1 Definizione degli strumenti e delle risorse 81
2.2 I principali sistemi di discovery 107
2.3 Problematiche di implementazione dei sistemi e di visualizzazione dei dati 131

3. Principi e teorie
3.1 La metodologia dei linked data e il progetto del web semantico 149
3.2 Possibilità e criticità dei nuovi metodi 178
3.3 Verso una nuova definizione di risorsa 190

4. Information discovery e information literacy
4.1 Information retrieval e discovery 203
4.2 Discovery tool e information literacy 222
4.3 L'information literacy e le persone 237

Conclusioni. Da quando abbiamo Google e Sci-Hub, che bisogno c'è delle biblioteche? 245

Bibliografia e letture consigliate 251

Prefazione

Sempre più spesso chi, navigando nei siti web delle università sia italiane che straniere, riesce a raggiungere la sezione dedicata alle biblioteche (purtroppo spesso assente dai menu principali) non trova più, come prima cosa, un link ben evidenziato al catalogo online, ma una maschera di ricerca 'alla Google' che consente di inserire uno o più termini senza né specificare se si riferiscono ad autori, titoli, soggetti o che altro, né collegarli fra loro con gli operatori booleani, ottenendo una lunghissima lista di descrizioni di documenti dei quali solo talvolta è possibile visualizzare online (a seconda dei propri diritti di accesso) il testo completo oppure localizzarlo presso le collezioni fisiche delle biblioteche di quella università.

A tali software, che fanno parte della più ampia famiglia dei cosiddetti *web-scale discovery services* e che vengono prevalentemente denominati – dopo un iniziale periodo di incertezza – discovery tool (ma c'è ancora chi li chiama, più genericamente, 'portali', specificando solo talvolta 'per la ricerca bibliografica'), e che in certi paesi da qualche anno stanno cominciando a diffondersi anche nelle biblioteche di altre tipologie, Roberto Raieli ha dedicato questo libro – di cui si cominciava a sentire il bisogno – molto chiaro, completo, aggiornato e documentato, nel quale trovano spazio anche considerazioni piuttosto ampie e pertinenti su tematiche collegate come l'open access, il web semantico, i linked data e l'information literacy.

Il gradimento dei discovery tool da parte degli utenti è notevole e crescente, tanto da indurre la grande maggioranza delle biblioteche che li adottano a evidenziarli rispetto al tradizionale catalogo online, che talvolta viene addirittura abbandonato, limitandosi magari a offrire un link a un

OPAC collettivo cui la biblioteca aderisce. I principali motivi di tale successo sono, a mio avviso, tre:

1) L'eccezionale semplicità d'uso, modellata su quella tipica dei motori di ricerca del web di tipo generalista e di cui gli utenti – soprattutto quelli più giovani – non sanno più fare a meno.
2) La grande vastità dei risultati mediamente ottenuti, che va al di là delle dimensioni della collezione locale (sia fisica che digitale) rispecchiata dall'OPAC, individuando anche documenti meno tradizionali e riducendo il rischio di frustranti ricerche dai risultati esigui o nulli, senza però rischiare di imbattersi nei documenti ancora più numerosi e spesso di scarsa qualità offerti dai motori di ricerca generalisti.
3) La percezione, illusoria ma diffusa, che l'uso esclusivo di uno di tali strumenti consenta una ricerca bibliografica esaustiva, o comunque sufficiente per gran parte delle esigenze di – mettiamo – uno studente universitario medio, dispensandolo dall'individuare, imparare a usare e infine interrogare anche altri repertori.

A fronte di tali vantaggi c'è chi, soprattutto fra bibliotecari e biblioteconomi, nota però anche numerosi difetti, o comunque problematicità, alcuni dei quali paradossalmente causati o comunque assecondati dai bibliotecari stessi che hanno il compito di adattare i discovery tool alle esigenze di ciascuna specifica biblioteca o università, mentre su altri aspetti le possibilità di intervento dello staff locale sui software, spesso acquistati da grandi aziende internazionali, sono oggettivamente limitate. Fra tali criticità vorrei ricordare almeno le seguenti:

A) È estremamente difficile, non solo per gli utenti ma persino per i bibliotecari, conoscere con certezza e completezza quali sono esattamente le banche dati e le altre fonti coperte dall'indice del discovery tool che è stato acquisito, le quali oltretutto mutano spesso a seconda degli accordi

commerciali fra il produttore del discovery tool e i fornitori delle fonti bibliografiche.

B) Ancora più imperscrutabili sono i criteri con cui il discovery tool ordina i risultati delle ricerche effettuate, che in linea generale si ispirano agli algoritmi di *relevance ranking* tipici dei motori di ricerca generalisti, originariamente concepiti per essere applicati a testi integrali non strutturati di varia natura e qualità prodotti da chiunque (come le pagine web) e non a insiemi di metadati omogenei e strutturati creati da professionisti (come quelli spesso presenti nelle banche dati bibliografiche). Solo talvolta i bibliotecari sono in grado di modificare tali criteri *a priori*, ad esempio valorizzando le risorse locali catalogate nell'OPAC, e agli utenti è consentito di cambiarli *a posteriori*, riordinando l'elenco dei risultati in base alla data o all'autore.

C) Poiché talvolta l'azienda che produce il discovery tool distribuisce anche fonti bibliografiche, potrebbe nascere il sospetto che esse vengano privilegiate nell'ordinamento o comunque valorizzate in qualche modo rispetto a quelle distribuite da aziende concorrenti.

D) Poiché, sebbene le pubblicità dei produttori talvolta sembrino suggerire il contrario, nessun discovery tool include nel proprio indice la totalità delle banche dati e delle altre fonti bibliografiche esistenti sul mercato e in open access; capita spesso che una biblioteca paghi l'abbonamento per fonti informative che non sono coperte dall'indice del discovery tool adottato. In tal caso quelle fonti dovranno essere interrogate dagli utenti attraverso la loro interfaccia nativa, oppure la biblioteca dovrà procurarsi uno strumento di ricerca da gestire parallelamente al discovery tool, in entrambi i casi col rischio che le fonti in questione risultino sottoutilizzate.

E) Inversamente, l'indice di ciascun discovery tool copre quasi sicuramente anche fonti informative a testo integrale alle quali la singola biblioteca non è abbonata. Deselezio-

narle dall'indice stesso, rendendole invisibili ai propri utenti, sarebbe per i bibliotecari un lavoro impegnativo e non sempre possibile, che comunque priverebbe gli utenti anche dei puri metadati bibliografici, in certi casi comunque utili. Non deselezionarle, d'altra parte, aumenterebbe la percentuale delle descrizioni bibliografiche recuperate durante la ricerca che non conducono al testo integrale del documento descritto e, quindi, la frustrazione degli utenti.

F) Per cercare di ridurre i rischi legati ai punti D ed E alcune biblioteche potrebbero essere tentate di acquisire soprattutto fonti informative commercializzate dalla stessa azienda che produce il loro discovery tool o, almeno, incluse nell'indice del discovery tool stesso. Ciò potrebbe condurre, sul lungo periodo, a uno snaturamento delle collezioni o, quanto meno, alla penalizzazione e progressiva marginalizzazione di preziose fonti informative pubblicate da piccoli editori.

G) La natura doppiamente ibrida dei discovery tool (in parte bibliografie dell'esistente e in parte cataloghi del posseduto, ma anche repertori sia di documenti cartacei, di cui forniscono sempre solo i metadati, che di documenti digitali, dei quali offrono talvolta anche il testo integrale), unita al secondo corno del dilemma descritto nel punto E, riduce la percentuale dei documenti rintracciati al cui testo digitale integrale gli utenti possono immediatamente accedere. Ciò può, paradossalmente, risultare particolarmente frustrante proprio per quegli utenti più giovani a cui i discovery tool si rivolgono preferenzialmente e per i quali ciò che non è immediatamente fruibile sullo schermo risulta quasi inesistente.

H) Gli algoritmi di ricerca dei discovery tool privilegiano l'operatore booleano OR rispetto a quello AND e, più in generale, il richiamo rispetto alla precisione, producendo elenchi di risultati estremamente lunghi che, dopo un attimo di sollievo per 'aver trovato qualcosa', generano negli utenti (soprattutto quelli meno giovani) prima

un'ansia da 'sovraccarico informativo' e poi irritazione per la significativa percentuale di documenti rintracciati che si rivelano completamente irrilevanti.

I) La maschera di ricerca 'avanzata' dei discovery tool, che permetterebbe di incentivare la precisione e di ridurre il richiamo, è spesso estremamente meno ricca e sofisticata di quella dei migliori OPAC ed è sempre molto meno visibile della *single search box* in stile Google, che scatena invece in automatico gli algoritmi 'moltiplicatori di risultati'.

J) I metadati presenti nell'indice del discovery tool sono eterogenei e spesso sia qualitativamente peggiori che quantitativamente minori rispetto a quelli delle banche dati originarie da cui provengono. Inoltre il loro riversamento nell'indice non sempre è contemporaneo e tempestivo. Ciò produce disfunzioni sia nei meccanismi di ricerca che nelle descrizioni dei documenti e, in particolare, rende meno efficaci i filtri applicabili *a posteriori* alla lista dei risultati per cercare di segmentarla in insiemi coerenti di documenti.

K) Oltre alla ricerca sul loro intero indice, spesso i discovery tool permettono anche di indagare solo su alcuni suoi sottoinsiemi che dovrebbero mutualmente escludersi, ma segmentazioni irrazionali e terminologie ambigue (spesso attribuibili più ai bibliotecari locali che alle aziende produttrici) rendono tale opzione di incerta applicazione. Ad esempio: se la scelta è fra 'risorse cartacee' e 'risorse online', dove stanno CD e DVD? E se fra le alternative ci sono 'riviste elettroniche' e 'articoli', dove devo cercare un articolo pubblicato in una rivista elettronica?

L) I nomi fantasiosi con cui ciascuna biblioteca o istituzione denomina il proprio discovery tool rendono più difficile agli utenti capire che si tratta comunque sempre della stessa tipologia di strumento e sottolineano talvolta (come nel caso di Tutto o di Uno per tutto) proprio quella pretesa esaustività della ricerca da cui invece i bibliotecari dovrebbero mettere in guardia.

M) Il processo di progressivo incremento dell'ipertestualizzazione (e, quindi, della contestualizzazione dei documenti e della libertà di scelta dei percorsi di navigazione bibliografica da parte degli utenti), che ha accompagnato l'evoluzione delle varie generazioni di OPAC dalla loro nascita fino a oggi, si arresta e regredisce nei discovery tool, nei quali sia la lista dei risultati che le descrizioni dei singoli documenti recuperati presentano in genere un numero minore di link rispetto ai più recenti OPAC.

N) I discovery tool sono strumenti concepiti per la ricerca bibliografica e non per una più complessiva gestione dei documenti bibliografici. Essi sono quindi spesso carenti di servizi tipicamente offerti agli utenti dagli OPAC, come la gestione dei prestiti, la segnalazione personalizzata delle nuove acquisizioni, il mantenimento di uno 'scaffale virtuale personale' ecc.

Per cercare di contrastare almeno una parte di questi problemi, Raieli propone e argomenta varie strategie, fra le quali riassumo qui sommariamente le principali:

I) I discovery tool andrebbero utilizzati prevalentemente nella fase iniziale della ricerca bibliografica, per passare poi agli OPAC (ammesso che non siano stati, nel frattempo, dismessi) e alle banche dati specializzate per i successivi approfondimenti.

II) Il successo di strumenti 'amichevoli' come i discovery tool non riduce l'importanza dell'information literacy, su cui occorre continuare a investire affinché tutti conoscano costi, limiti e illusioni di tale amicizia, nonché l'intero ventaglio disponibile dei mezzi e delle tecniche di ricerca e, soprattutto, di selezione e valutazione delle informazioni.

III) Per quanto gli algoritmi di ricerca e le interfacce di interazione con l'utente siano importanti in qualsiasi strumento bibliografico digitale, neppure quelli dei discovery tool possono fare miracoli quando la materia prima a cui vengono applicati è di scarsa qualità e troppo eterogenea. Per migliorare efficacia e affidabilità dei discovery tool

sarebbe quindi indispensabile, prima di tutto, aumentare il livello qualitativo e l'omogeneità dei metadati che vengono riversati nei loro indici.

IV) I bibliotecari non dovrebbero illudersi che il costoso (in termini finanziari) acquisto di un discovery tool li esima dall'ulteriore costo (in termini di risorse umane) di un impegnativo e costante lavoro di implementazione e di personalizzazione rispetto alle specificità dell'utenza e della collezione locale.

V) Non solo i bibliotecari dovrebbero essere maggiormente coinvolti e impegnati nell'information literacy, nella produzione e nel controllo dei metadati e nell'implementazione dei software, ma, più in generale, andrebbe riconosciuta, valorizzata e potenziata – come scrive Raieli nelle ultime righe della prima parte del libro – «la necessità della presenza delle conoscenze bibliotecarie in ogni fase della creazione, strutturazione, archiviazione, ricerca e visualizzazione dei dati» gestiti dai discovery tool. Tali conoscenze risulteranno, ad esempio, particolarmente preziose nel garantire appropriatezza e coerenza dei termini utilizzati per indicare agli utenti funzioni, segmentazioni, operatori, filtri e maschere di ricerca disponibili.

Sono tutti consigli corretti e condivisibili, anche se personalmente sono forse un po' più pessimista di Raieli sulla loro efficacia, perché sarà sempre più difficile per una categoria professionale in crescente crisi (numerica, economica, culturale, identitaria, ecc.) incidere significativamente sugli interessi, gli orientamenti e i comportamenti degli utenti, delle aziende fornitrici e persino delle proprie istituzioni di afferenza. In particolare, nel mio ruolo di docente che tiene agli studenti universitari anche corsi sulle basi della ricerca bibliografica, incontro notevoli difficoltà nell'insegnare loro a comprendere e utilizzare al meglio strumenti talmente imperscrutabili e scarsamente 'orientabili' come i discovery tool.

In ogni caso, insegnanti di bibliografia e bibliotecari devono per forza aggiornarsi sui discovery tool e sforzarsi di

trasmettere a studenti e utenti ciò che apprendono da studi come quello che state per leggere e dalle loro esperienze sia davanti che dietro le interfacce dei sistemi di discovery, tenendo realisticamente conto del clima creato sia dal contesto socio-culturale (ovvero da ciò che chi effettua ricerche bibliografiche, a torto o a ragione, preferisce) che da quello economico-tecnologico (ossia dalle potenti spinte impresse delle aziende sul mercato informatico in generale e sui software bibliotecari in particolare).

Concludo questa prefazione con un ultimo, breve, elenco di rapide considerazioni:

- Vista la natura preliminare e introduttiva che dovrebbero rivestire le ricerche bibliografiche effettuate con i discovery tool, credo che sarebbe più corretto (ma, lo capisco, anche meno accattivante) chiamarli 'strumenti di esplorazione' anziché 'di scoperta'.
- È paradossale (anche se comprensibile dal punto di vista tecnologico ed economico) che il successo dei discovery tool sia nato e tuttora spicchi proprio nell'ambiente universitario, dove ricercatori e studenti non dovrebbero spaventarsi troppo delle complessità della ricerca bibliografica, visto che per i primi si tratta di una parte rilevante del proprio lavoro e delle proprie competenze e per i secondi costituisce una delle cose principali che stanno cercando di imparare (ma che difficilmente impareranno davvero, se viene eccessivamente semplificata e automatizzata).
- Forse i discovery tool potrebbero essere un buon esempio, nell'ambito della ricerca bibliografica, di quanto espresso in linea generale dall'aforisma tradizionalmente attribuito ad Albert Einstein secondo cui «tutto dovrebbe essere reso il più semplice possibile, ma non più semplice ancora».

Riccardo Ridi
Università Ca' Foscari, Venezia
ottobre 2019