

Formare e gestire collezioni con i discovery tools

I bibliotecari e la qualità dei dati nei sistemi di ricerca bibliografica

ANTONELLA TROMBONE

Università della Basilicata, Potenza
antonella.trombone@gmail.com

I *discovery tools* sono portali specializzati per la ricerca bibliografica molto diffusi nelle biblioteche che posseggono collezioni eterogenee di risorse elettroniche e digitali. Le istituzioni accademiche e di ricerca dedicano una percentuale significativa del proprio bilancio agli abbonamenti a risorse elettroniche in formato digitale, pubblicazioni che non sono, perciò, possedute dalle biblioteche, ma ricercabili in virtù del pagamento di canoni che danno diritto all'accesso ai loro contenuti. I *discovery tools* permettono di effettuare una ricerca simultanea sia tra le collezioni dalla biblioteca, su supporto fisico o elettronico, che tra le risorse digitali alle quali la stessa è abbonata. La caratteristica che distingue questi strumenti dai software per la ricerca federata tra le risorse online è il loro indice centrale, una definizione mutuata dall'inglese¹ che indica la banca dati alla quale le biblioteche hanno accesso indirettamente,

cioè grazie al contratto stipulato col produttore del *discovery tool*.²

Trascorsi più di vent'anni dalle prime presenze delle biblioteche sul web, si può provare a tracciare un'evoluzione dei loro strumenti di ricerca bibliografica che corre parallela, da una parte, alla dematerializzazione dei supporti bibliografici e, dall'altra, all'eterogeneità dei documenti ai quali la ricerca bibliografica dà accesso. A partire dagli anni Novanta le biblioteche adottano i cataloghi online (gli OPAC), compilano gli elenchi di siti web da proporre agli utenti per la ricerca bibliografica e pubblicano le liste *A to Z* dei titoli dei periodici elettronici ai quali le istituzioni sono abbonate. Quando già sono sul mercato i motori di ricerca, le biblioteche cominciano ad adottare i così detti cataloghi di nuova generazione che, in realtà, non sono sistemi di catalogazione bensì di ricerca bibliografica – e di visualizzazione dei risultati – che

L'articolo propone un'ulteriore analisi rispetto al contenuto del seminario tenuto dall'autrice a Milano, presso lo IULM, il 22 maggio 2018. Si ringrazia Giuseppe Marino e il Gruppo di lavoro Discovery tool dell'Associazione ITALE per l'invito e tutti i presenti per gli spunti di riflessione emersi in quella sede. I siti web e gli URL sono stati controllati l'ultima volta il 2-1-2019. In merito ai riferimenti, diretti o indiretti, a banche dati, servizi e software commerciali, l'autrice precisa che i marchi sono registrati dai rispettivi produttori.

segnano uno stadio intermedio tra gli OPAC e i *discovery tools*. Le nuove interfacce tentano di risolvere il problema della ricerca dei contenuti che non sono di proprietà ma concessi in licenza alle biblioteche, come l'editoria elettronica, e che non sono gestiti dai sistemi di catalogazione automatizzata né da un punto di vista catalografico né amministrativo; tali pubblicazioni, che verranno poi comunemente definite "risorse online" delle biblioteche, rappresentano una parte intangibile delle loro collezioni, alla quale si accede in base agli abbonamenti e alle licenze d'uso. I cataloghi di nuova generazione ricercano simultaneamente in tutte le collezioni, possedute o accessibili tramite una biblioteca, e nello stesso tempo cominciano a proporre agli utenti interfacce che avranno caratteristiche sempre più simili a quelle dei motori di ricerca e dei siti più popolari del commercio online.³

La letteratura biblioteconomica afferma che il primo catalogo di nuova generazione che adotta i principi del web 2.0⁴ è installato nel 2006 per le biblioteche della North Carolina State University e adotta il software Endeca. Il catalogo comincia ad assumere alcune delle caratteristiche dei siti commerciali quali la presenza di immagini, la navigazione tramite faccette, la velocità delle risposte e l'ordinamento per rilevanza. Nel sito web istituzionale, la stessa definizione di "nuovo catalogo online" è affiancata a quella di "nuovo motore di ricerca":

the NCSU Libraries' new catalog provides the speed and flexibility of popular online search engines while capitalizing on existing catalog records. As a result, students, faculty, and researchers can now search and browse the NCSU Libraries' collection as quickly and easily as searching and browsing the Web, while taking advantage of rich content and cutting-edge capabilities that no Web search engine can match.⁵

Il concetto di collezione

Ai fini della ricerca nei DT il lavoro dei catalogatori converge sia con quello dei bibliotecari che si occupano degli archivi istituzionali e dei dati della ricerca, sia con quello dei bibliotecari che lavorano per la consultazione e l'accesso alle risorse digitali. I *discovery tools* integrano i metadati delle collezioni fisiche, che

sono tradizionalmente catalogate negli OPAC, con i metadati degli articoli delle riviste in formato elettronico, quelli dei capitoli dei libri e delle altre tipologie di documenti contenuti negli archivi digitali. La ricerca acquisisce specificità e si arricchisce di titoli, autori, soggetti e parole chiave che descrivono documenti che nei cataloghi non erano invece ricercabili. La corretta indicizzazione di questi dati è favorita da una catalogazione granulare, cioè da una descrizione analitica che adotti dei metadati standard. Un risultato negativo ottenuto dall'integrazione dei metadati rende difficile sia localizzare le pubblicazioni di cui si ha esattamente bisogno, sia la possibilità di definire con precisione le ricerche attraverso le faccette. Eppure, al momento della sottoscrizione di un contratto per un DT la qualità e la tipologia dei suoi metadati sono valutate raramente, mentre invece viene posta particolare attenzione al funzionamento e all'implementazione dei software.

Sarebbe invece utile esaminare il modo in cui le registrazioni bibliografiche in formato MARC e i metadati che vi arrivano da altre fonti sono importati nel DT. Anche la qualità dei metadati ha un forte impatto sull'efficienza del sistema di *discovery*. Le registrazioni nei formati MARC spesso derivano da trascrizioni di cataloghi a schede che non contengono le informazioni dettagliate previste dai cataloghi online, come per esempio quelle sui formati bibliografici. Se privati di alcune informazioni che li descrivono, i documenti possono diventare invisibili una volta che sono ricercati attraverso un DT.

Terminologia e concetti chiave

Allo stato attuale possiamo definire tre fasi di sviluppo dei sistemi di *discovery*. La prima generazione è quella degli strumenti di ricerca federata che permettono che una singola ricerca sia condotta contemporaneamente in più database e che mostri i risultati in un'unica lista di riferimenti bibliografici. La seconda generazione è quella degli aggregatori di ricerca, cioè uno sviluppo della ricerca federata che si occupa prevalentemente dell'interfaccia, quindi del sistema di visualizzazione del DT. La terza generazione dei DT vede lo sviluppo di sistemi complessi di recupero dell'informazione che sono sempre più assimilabili a dei motori di ricerca specialistici costituiti da di-

versi elementi tra di loro interconnessi e interagenti. L'indice centrale contiene una raccolta di riferimenti bibliografici che è interrogata da un sistema di ricerca, definito *discovery layer*, che, basandosi su una traduzione letterale, potrebbe essere spiegato come il "livello di scoperta" del sistema e, perciò, associato al concetto di individuazione casuale e fortuita di informazioni che è spesso indicato tra le caratteristiche distintive dei DT. L'interfaccia di ricerca è ricca di funzionalità e permette di effettuare un'unica interrogazione che agisca contemporaneamente sulle collezioni locali, quelle in abbonamento e quelle in accesso aperto.⁶

Il sistema di *discovery* è quindi composto da un indice centrale di contenuti e da un'interfaccia di ricerca ricca di funzionalità.

L'indice centrale

L'indice dei *discovery tools* contiene citazioni e metadati che descrivono collezioni eterogenee, che appartengono ai seguenti macro-insiemi:

- i metadati, le citazioni bibliografiche e gli abstract che provengono dagli editori e dagli aggregatori delle pubblicazioni in formato elettronico alle quali la biblioteca è abbonata;
- i metadati, le citazioni bibliografiche e gli abstract della collezione di risorse online che appartiene al sistema di *discovery*;
- i metadati delle collezioni in accesso aperto;
- le registrazioni bibliografiche del catalogo della biblioteca, in formato MARC.

Se l'indicizzazione dei documenti non è completa, oppure se i metadati non sono disponibili per essere indicizzati, o ancora se gli algoritmi influenzano i risultati della ricerca, il risultato finale è ingannevole e manca di trasparenza. L'indice centrale del DT è il maggior elemento di competizione tra i produttori di questi software. Le sue dimensioni sono difficili da definire, mentre è più semplice stabilire quali siano i suoi contenuti, il tipo di metadati e di documenti che include. La quantità di metadati indicizzata dipende dagli accordi e dalle licenze esistenti tra le biblioteche e i loro fornitori di contenuti digitali. In genere, gli accordi con gli editori prevedono la visualizzazione di tutte le informazioni e permettono di indicizzare

l'intero insieme di metadati; può anche accadere che le licenze limitino la visualizzazione di alcune banche dati, e quindi che una biblioteca acquisti più risorse di quelle che visualizza. L'indice centrale del DT può contenere tutti o solo alcuni dei seguenti metadati:

- i metadati forniti dalla biblioteca (come le registrazioni catalografiche nei formati MARC), quelli prodotti per la descrizione di collezioni locali, fondi archivistici, progetti di digitalizzazione di fondi e collezioni;
- i metadati di pubblicazioni in accesso aperto o di dominio pubblico con il testo completo; i soli riferimenti bibliografici di documenti in *open access*;
- i metadati che provengono dagli editori e che danno accesso ai testi completi: sono forniti dagli autori e dagli editori e comprendono i riferimenti bibliografici per la citazione dei documenti, gli abstract e le parole chiave; se la biblioteca ha diritto di accedere al contenuto allora sono visualizzabili articoli e parti di libri, in caso contrario sono disponibili solo i metadati;
- i metadati relativi ai documenti a cui si accede attraverso la licenza del DT: questo tipo di contenuti – considerato il nucleo dell'indice – proviene da diverse fonti in base agli accordi stipulati dai produttori di *discovery tools* per indicizzare e avere accesso completo o parziale a contenuti editoriali di terze parti; l'accesso alle informazioni e ai contenuti è regolato dagli accordi raggiunti con gli editori e gli autori;
- i metadati dei contenuti con licenza reciproca, cioè visualizzabili sia attraverso l'indice centrale del DT sia tramite il sito web istituzionale della biblioteca; in genere si tratta di licenze e abbonamenti che la biblioteca sottoscrive ma che sono inclusi anche negli accordi che i produttori di DT cercano di stipulare di continuo con editori e aggregatori editoriali, con lo scopo di indicizzare i loro metadati e arricchire gli indici dei propri *discovery tools*.

La ricerca mediata da un sistema di *discovery* può condurre gli utenti solo ai metadati dei documenti oppure permettere il collegamento diretto anche ai loro testi completi. Occorre sempre ricordare che i diritti d'accesso ai metadati e ai documenti completi sono regolati da due tipologie di accordi, tra loro indipendenti: quelli stipulati dall'istituzione e quelli stipulati

dal produttore del *discovery tool* adottato dalla biblioteca.

Tutti questi fattori sono fondamentali per valutare quantitativamente il contenuto dell'indice centrale, anche se bisognerebbe analizzare con maggiore attenzione la qualità e la tipologia dei dati e dei contenuti che include. Non a tutti è noto che alcune banche dati impediscono ai DT di indicizzare i propri metadati, come per esempio è il caso di SciFinder Scholar, il portale d'accesso alle banche dati per la chimica dell'American Chemical Society – che permette anche la consultazione di MedLine – definito a sua volta da alcune biblioteche uno strumento di *discovery*.⁷

Può anche accadere che, sebbene una rivista non sia inclusa nell'indice centrale di un DT, il produttore affermi che essa vi si trovi in parte indicizzata. Ciò significa che, anche senza i diritti d'accesso al testo completo, nell'indice ci sono i metadati di base, di livello citazionale, di molti articoli. Questi contenuti di livello citazionale sono aggiunti all'indice centrale e forniscono informazioni che vengono indicizzate, ma che sono prive dei metadati analitici e del collegamento al *full text*. I soggetti, gli abstract e i metadati più analitici elaborati da chi indicizza, in questo caso, non si trovano nell'indice centrale. Tra i contenuti editoriali che sono dichiarati come inclusi nell'indice del DT occorre, perciò, fare anche questa distinzione. Per scegliere con consapevolezza un sistema di *discovery* i bibliotecari dovrebbero sempre distinguere tra le diverse tipologie di accordi tra le case produttrici di questi sistemi e gli editori: non tutti gli indici hanno gli stessi contenuti perché alcuni accordi prevedono l'accesso ai metadati completi delle pubblicazioni, altri invece permettono solo l'estrazione delle citazioni bibliografiche.

Metadati e inclusione dei documenti

Nei sistemi di ricerca che definiamo *discovery tools* i contenuti di tutti i database, locali e remoti, vengono re-indicizzati, compresi quelli sottoscritti dalla biblioteca con accordi commerciali e i dati del catalogo. I risultati delle ricerche sono una fonte di informazione secondaria, in quanto frutto di un'attività di re-indicizzazione dei metadati catalografici e di quelli delle risorse online interrogate dal *discovery*: per visualizzare le descrizioni originali occorre effettuare,

sempre attraverso l'interfaccia del sistema, il collegamento alla fonte del dato, quindi al catalogo online o alla singola banca dati, a cui generalmente si accede selezionando la descrizione dettagliata dalla lista dei risultati della ricerca. L'accesso ai documenti dipende, perciò, direttamente dai metadati che li descrivono che sono, a loro volta, mediati dal processo di re-indicizzazione del *discovery tool*. Premesso che la ricerca delle pubblicazioni in questi portali avviene attraverso il loro indice centrale e che dipende dai diversi gradi di qualità e completezza delle informazioni che i sistemi di *discovery* ricevono da fonti differenti, ci sono alcune domande che i bibliotecari dovrebbero porsi per definire i livelli di inclusione possibili di documenti e metadati nei DT.⁸

Per prima cosa occorre stabilire se l'indice centrale include i metadati originali di un documento, oppure i metadati che provengono da un altro indice che contiene quel documento. Questa analisi dovrebbe riguardare ogni tipo di collezione ricercabile tramite il DT.

Per tipologie di pubblicazioni come le monografie e gli atti di convegni, è importante sapere se i metadati descrivono solo i dati di base, oppure anche i soggetti e le parole chiave di ogni singolo contributo; in altri termini, se esiste un livello di inclusione relativo ai metadati del contenitore e uno relativo ai metadati del contenuto. Inoltre, deve essere definito il livello di inclusione delle risorse online per le quali la biblioteca paga, e di quelle non a pagamento alle quali la biblioteca dà accesso.

Quelle descritte di seguito sono alcune cause che possono limitare l'inclusione dei documenti. Ci sono editori e aggregatori di contenuti digitali che possono decidere di non condividere i metadati delle proprie risorse col produttore del *discovery*.⁹ Queste informazioni devono essere comunicate con evidenza agli utenti tramite il sito web della biblioteca perché, occorre ricordarlo, i portali di *discovery* sono usati da utenti autorizzati per consultare le risorse online di una biblioteca da remoto, tramite internet, in genere senza alcuna mediazione da parte dei bibliotecari. Solo attraverso le informazioni fornite dal sito istituzionale gli utenti sanno quali collezioni sono ricercabili attraverso il sistema di *discovery* e quali ne sono escluse.

È possibile che una biblioteca si abboni a una collezione digitale e che l'editore ceda i metadati al *discovery*

tool, ma alcune informazioni possono perdersi nella traduzione da un linguaggio all'altro per vari motivi:

- i metadati originali non sono rappresentati nello stesso modo nel DT, né possono essere attivati selettivamente in esso; questo è il caso degli ebook appartenenti a pacchetti a cui la biblioteca è abbonata che non possono essere attivati individualmente nel *discovery*, che indicizza solo i metadati del pacchetto;
- un abbonamento che appare come un'entità unica per la biblioteca è in realtà composto da più componenti aggregate; i metadati di queste parti componenti potrebbero non essere resi disponibili per l'indicizzazione nel DT;
- nel caso dei periodici, occorre sempre conoscere i diritti d'accesso e la copertura cronologica degli abbonamenti; inoltre ricordiamo che gli accordi con gli editori possono essere stipulati dalla biblioteca o dal produttore del DT o da entrambi i soggetti, perciò è necessario verificare se la copertura dei contenuti editoriali è quella prevista complessivamente da ogni tipologia di accordo in essere;
- può capitare che l'indice centrale del DT sia aggiornato meno frequentemente di quello della risorsa nativa e ciò comporta un ritardo nell'inclusione dei metadati; in questo caso si verifica una mancanza di sincronia tra il DT e le collezioni originali.

Per tante ragioni, quindi, la ricerca dello stesso contenuto nella risorsa nativa e nel sistema di *discovery* può evidenziare differenze nei risultati. Le ragioni di tali discrepanze dovrebbero essere sempre comunicate ai bibliotecari dai produttori dei sistemi di *discovery*.

I metadati indicizzati dai *discovery tools* sono eterogenei perché provengono da diverse comunità scientifiche e professionali. Il caso più comune a cui si può far riferimento è quello dei nomi degli autori. Negli archivi di periodici elettronici a volte i metadati contengono i nomi registrati con le sole iniziali, in ordine diretto e indiretto; nei cataloghi invece i nomi sono registrati in ordine diretto nel campo del titolo e dell'indicazione di responsabilità, in ordine indiretto e in forma completa nei campi degli autori.

La provenienza dei metadati ha caratteristiche di elevata eterogeneità. I sistemi di *discovery* importano i metadati provenienti da più modelli e quasi sempre li

normalizzano all'interno di un proprio schema di indicizzazione, se non li trasformano almeno li adattano attraverso delle mappature.¹⁰ Le informazioni col livello di inclusione più basso nel *discovery* sono quelle dei vocabolari controllati e quelle relative all'affiliazione degli autori.¹¹

Le collezioni locali nei *discovery tools*

L'indicizzazione dei contenuti del catalogo e delle collezioni che fanno parte del patrimonio della biblioteca nell'indice centrale dei DT avviene attraverso i metadati che le descrivono. Se i dati catalografici non sono corretti e nel caso in cui non si adottino dei modelli standard di metadati, l'indice centrale del DT non li indicizza e non li include. I problemi legati al recupero dell'informazione bibliografica attraverso i DT non sono dovuti solo ai criteri di ordinamento per rilevanza del sistema di ricerca, ma derivano anche dalla qualità delle registrazioni catalografiche e delle descrizioni delle collezioni locali.

I *discovery* raccolgono e indicizzano metadati provenienti da varie fonti, tra cui i servizi di *abstracting* e *indexing*. Anche se i due maggiori database generalisti di questo tipo (Web of Science e Scopus), sono oggi in genere disponibili per l'indicizzazione nei DT, il loro uso mediato da questi sistemi di ricerca può andare a discapito delle statistiche d'uso delle singole banche dati che, essendo a pagamento, sono soggette a continui riscontri di questo tipo. Inoltre può capitare che la terminologia semantica molto precisa dei servizi di A&I non venga rappresentata in modo altrettanto puntuale nei *discovery tools*.

Le collezioni in accesso aperto sono spesso un punto debole dell'indicizzazione nei sistemi di *discovery* a causa della qualità dei metadati che le descrivono che può variare molto in base agli editori o agli autori.

Facendo riferimento alle collezioni di una biblioteca, l'indice centrale gestisce e indicizza: i contenuti in abbonamento da editori e aggregatori; i servizi di *abstracting* e *indexing* in abbonamento; i contenuti locali (catalogo, archivi istituzionali e digitali); i contenuti in accesso aperto; termini, locuzioni, parole chiave che derivano dai contenuti forniti dall'editore. Tuttavia, nessun sistema di *discovery* indicizza e permette l'accesso complessivo a tutte le collezioni di una biblioteca. L'approccio di ricerca ibrido, cioè la ricerca

federata unita alle potenzialità dell'indice centrale, risolve in parte il problema. Sono in genere esclusi dall'indice del *discovery tool* i contenuti dei siti web, i contenuti didattici dei corsi e i contenuti degli editori contrari a farsi indicizzare.¹²

Trasparenza e standardizzazione

Trattandosi di società che operano in concorrenza, la riservatezza delle case produttrici dei sistemi di *discovery* è naturale oltre che perfettamente comprensibile, tuttavia non gioca a favore degli utenti finali delle biblioteche. Proprio per questo motivo occorre stigmatizzare il ruolo dei bibliotecari in favore della trasparenza dei contenuti degli indici centrali e per la standardizzazione delle procedure di implementazione di questi sistemi. Una prima proposta di uniformità è quella pubblicata dalla National Information Standards Organization (NISO) nel 2014,¹³ dopo aver raccolto, a partire dal 2011, un insieme di raccomandazioni rivolte ai bibliotecari sui sistemi di *discovery* basati su indici. Durante la selezione di un DT, in relazione ai contenuti del suo indice centrale, occorre valutare la presenza di contenuti con licenza doppia, dell'istituzione e della casa produttrice del DT. Inoltre, è importante chiarire se il DT indicizza i database specialistici che forniscono citazioni e abstract di vari settori della letteratura scientifica, quali include e con quale livello di specificità dei metadati. Un altro elemento da analizzare è il livello d'indicizzazione dei testi completi.

Nel valutare un DT è necessario osservare se vengono esportati tutti i metadati completi che si trovano nel suo indice centrale, così come quelli degli abbonamenti sottoscritti dalla biblioteca. Il sistema dovrebbe essere in grado di indicizzare tutti i contenuti e ordinarli senza che l'indice centrale diventi dominante nei risultati della ricerca. Sempre in fase di scelta bisogna analizzare il tipo e la quantità di riferimenti bibliografici e di abstract che sono inclusi nel sistema di *discovery* e anche la presenza di citazioni e documenti dei quali la biblioteca non ha bisogno. Documenti e informazioni non necessari possono influenzare negativamente l'efficacia della ricerca.

Le domande sui contenuti dell'indice centrale dei DT dovrebbero sempre ricevere delle risposte esplicite specie in relazione alle banche dati specialistiche,

alla possibilità di effettuare la ricerca in base ai titoli delle riviste e al tipo di ricerca che si può effettuare in relazione al soggetto, al contenuto e al tipo di documento.

Per quanto concerne i periodici elettronici, occorre sempre aver presente il problema dei numeri arretrati delle riviste. Gli abbonamenti delle biblioteche possono aver avuto un andamento discontinuo, quindi deve essere chiaro anche per le annate arretrate a quali pubblicazioni la biblioteca ha diritto d'accesso completo.

L'interfaccia di ricerca

L'interfaccia di ricerca (*discovery layer*) dei sistemi di *discovery* deriva le sue caratteristiche da cataloghi di nuova generazione come Encore, Innovative Interface e AquaBrowser: un unico punto d'accesso per la ricerca che interroga tutte le collezioni della biblioteca; la possibilità di definire i termini del quesito attraverso faccette e limitatori; l'ordinamento dei risultati per rilevanza e la possibilità di collegarsi al testo completo dall'elenco dei risultati. Attraverso uno stesso portale è anche possibile interrogare gli indici centrali di diversi produttori commerciali di *discovery tools*.¹⁴ Analisi condotte mediante Google analytics event tracking, abbinate con l'analisi dei termini di ricerca associati al contenuto delle collezioni, dimostrano che i *discovery tools* sono molto usati per la ricerca di documenti noti. Le statistiche attestano anche che la ricerca più frequente è quella per argomento, seguita dalla ricerca di documenti noti, e che Google scholar è il maggiore concorrente dei *discovery tools*. Chi fa ricerca si rivolge a Scholar spesso prima e più che ai sistemi di *discovery*.¹⁵

Le impostazioni di visualizzazione dell'interfaccia di ricerca hanno un'ampia possibilità di parametrizzazione. Per esempio, non tutte le biblioteche decidono di mostrare attraverso il *discovery tool* anche i documenti non disponibili in testo completo. Alcune biblioteche richiedono di separare i risultati della ricerca per classi di contenuto e di non mostrarli in una lista unica, ma divisi in settori nella stessa pagina.¹⁶ I metadati sono gli elementi di base per la definizione delle faccette, gli strumenti che limitano e definiscono i risultati della ricerca. Il nucleo della ricerca e dell'accesso intellettuale al sistema di *discovery* è, perciò, collegato al grado di richiamo e di precisio-

ne dei metadati. I metadati specialistici dei soggetti e dei sistemi di classificazione producono ulteriori raggruppamenti, faccette e limitatori. Dopo l'azione di recupero delle informazioni, mediata dai metadati, l'ordinamento dei risultati della ricerca avviene in genere in base ad algoritmi di rilevanza, anche se spesso la ricerca di un documento noto non lo restituisce nella prima pagina dei risultati; altro punto dolente è la ricerca per soggetto, che a volte non ottiene risposte precise. Questi ultimi due tipi di ricerca, quella di un documento noto e quella per soggetto, possono creare problemi agli utenti dei sistemi di *discovery*.¹⁷ Il criterio di rilevanza dipende anche dai metadati disponibili per i contenuti a testo completo. I *discovery tools* ricevono i metadati e i testi completi dai fornitori dei contenuti. Tali metadati a volte sono deboli, limitati a titolo, autore e soggetto (come può avvenire nel caso degli ebook). In altri casi sono completi, con soggetti, indici e recensioni tratte da servizi di *abstracting e indexing*. Un altro caso che si verifica spesso è che i testi completi non siano disponibili per l'indicizzazione; a volte il *discovery* può includere materiali non testuali come i video.¹⁸

I contenuti ai quali si applica l'algoritmo di rilevanza possono essere suddivisi in quattro macro-aree.

- I contenuti che hanno metadati completi ma non sono consultabili a testo completo, per esempio i servizi di *abstracting e indexing* come Scopus e Web of Science.
- I contenuti che hanno metadati completi, che sono consultabili a testo completo e che sono perfezionati dai metadati dei servizi di *abstracting e indexing*.
- I documenti che hanno metadati deboli e non sono consultabili a testo completo, per esempio le collezioni di video, quelle senza full text e alcuni archivi istituzionali. Questi contenuti sono i più difficili da scoprire attraverso i *discovery tools*.
- I contenuti che hanno metadati deboli ma sono consultabili a testo completo.

La ricerca di un documento noto si basa sulla precisione e, se il documento esiste, il risultato deve combaciare al 100% con la domanda. Il richiamo non ha valore in una ricerca di questo tipo. I cataloghi con metadati deboli o le risorse elettroniche e i libri il cui testo completo non è disponibile determinano problemi in fase di ricerca se hanno titoli brevi e comuni.

Infatti, può essere difficile che la ricerca di un documento noto nel caso, per esempio, delle registrazioni catalografiche che hanno metadati deboli, produca risultati rilevanti.¹⁹

Alcune conclusioni

Delle due parti di cui si compone un *discovery tool* – l'indice centrale e l'interfaccia di ricerca – è alla seconda che è stata dedicata grande attenzione da parte delle biblioteche in fase di scelta e di parametrizzazione di questi sistemi di ricerca. Le faccette, i filtri e i limitatori, i sistemi di raccomandazione e di filtraggio collaborativo, hanno potenziato la ricerca federata facendo sì che per gli utenti aumentasse la sensazione di usare un motore di ricerca, piuttosto che un portale per la ricerca bibliografica che interroga alcune precise banche dati. Il fatto che questo tipo di sistemi di ricerca stia diventando l'unico portale d'accesso alle collezioni delle biblioteche, specie in ambito accademico, dovrebbe spostare l'attenzione dei bibliotecari sul nucleo dei *discovery tools*: il loro indice centrale, cioè una vasta banca dati che non contiene documenti ma metadati che re-indirizzano l'utente ai database originari contenenti le informazioni e le pubblicazioni. I dati contenuti negli indici centrali non sono fonti primarie, perché sono il risultato di una nuova indicizzazione dei metadati delle collezioni locali e remote di una biblioteca, delle collezioni proprie del sistema di *discovery* e delle collezioni che sono frutto di accordi reciproci.

Ci troviamo di fronte, dunque, a un nuovo concetto di collezione, difficile da delimitare, composto da metadati e non da documenti, caratterizzato da una forte eterogeneità e fluidità di contenuti (intesi nel senso di obiettivi di interrogazione). Ciò nonostante, uno sforzo deve essere fatto in questo senso considerato che, in molti casi, anche gli OPAC delle biblioteche sono stati sostituiti dai *discovery tools*. Nelle fasi di scelta e in quelle successive di controllo di questi sistemi i bibliotecari dovrebbero verificare la qualità e la tipologia dei metadati che risiedono nell'indice centrale insieme al grado di inclusione dei contenuti almeno rispetto alle collezioni locali e remote di una biblioteca. Se si verificano perdite di informazione, se ci sono differenze di risultato tra le notizie ricercabili nelle collezioni primarie e quelle ricercabili nel

sistema di *discovery*, bisognerebbe sempre comunicarlo agli utenti e fornire un accesso diretto alle banche dati originarie, senza mediazione. Di tutto questo i bibliotecari possono accorgersi solo se attuano verifiche costanti sulla qualità del sistema di ricerca.

Un altro tema è quello relativo alle risorse in accesso aperto, che occupano grande spazio nei risultati delle ricerche mediati dai *discovery tools*, fornendo spesso risposte duplicate e confusionarie. Le cause possono essere diverse: i metadati originali delle risorse a libero accesso spesso non seguono alcuno standard; i *discovery tools* non de-duplicano e non correggono i dati che provengono da fonti diverse ma che indicano i medesimi documenti; le informazioni tratte da archivi *open access* che, a loro volta, indicizzano risorse primarie, sono considerate nuove informazioni rispetto a quelle originali, tratte direttamente dalle testate o dai siti web. Comunque si delimiti il problema, un ruolo maggiormente propositivo dei bibliotecari nei confronti dei produttori di *discovery tools* può determinarne la soluzione. Il discrimine si colloca tra un atteggiamento che accetta passivamente quello che viene proposto da case produttrici che sono sul mercato (e si sa che in questo senso ogni parte gioca il suo ruolo), tra una altrettanto passiva corsa ai contenuti – che avviene da entrambe le parti – che consiste nell’aggiunta all’indice centrale di qualsiasi collezione, sperando che sia l’algoritmo di rilevanza a risolvere ogni problema; oppure, in alternativa, nell’assunzione di un ruolo attivo e propositivo da parte dei bibliotecari, finalizzato a individuare i problemi e richiedere le modifiche necessarie per risolverli. Quest’ultimo compito non può che essere a carico delle biblioteche perché qui si discute della formazione e della gestione di nuove tipologie di collezioni, che contengono i metadati che mediano l’accesso alle risorse fisiche e online senza i quali tali risorse rimarrebbero invisibili per gli utenti.

NOTE

¹ La terminologia italiana legata a questi sistemi di ricerca bibliografica, o motori di ricerca specialistica (che in questo articolo sono a volte abbreviati in DT, o definiti sistemi, piattaforme, portali di *discovery*) merita una riflessione. Trattandosi di un linguaggio e di un ambito specialistico, si

è scelto di lasciare in inglese alcuni termini, anche flessi al plurale con -s, proprio per evidenziare che non sono entrati nella lingua italiana e che, a parere di chi scrive, devono essere ricercate delle definizioni per riferirsi a strumenti ormai così diffusi nelle biblioteche. Per le proposte “strumenti per la scoperta dei contenuti” e “piattaforme specializzate per la ricerca di dati bibliografici”, rinvio ad ANDREA MARCHITELLI, *Orientarsi tra le informazioni in biblioteca*, Milano, Editrice Bibliografica, 2015, p. 50, 81-82; *Id.*, <https://andreamarchitelli.it/actual/discovery-tool-guida-per-gli-editori-riviste-scientifiche>.

² Con quest’espressione si intende una ricerca simultanea in più basi dati, si veda in proposito YONGMING WANG, JIA MI, *Searchability and discoverability of library resources: federated search and beyond*, “College & undergraduate libraries”, 19 (2012), 2-4, p. 229-245.

³ Per un’analisi dei modelli di catalogo mi permetto di rinviare al mio, *L’analisi della presenza delle monografie negli strumenti di discovery*, in MARIA TERESA BIAGETTI, *Valutare la ricerca nelle scienze umane e sociali: potenzialità e limiti della library catalog analysis*, con scritti di Antonella Iacono e Antonella Trombone, Milano, Editrice Bibliografica, 2017, p. 133-147: 133-136; per una bibliografia generale di riferimento cfr. *Ivi*, p. 239-243.

⁴ Sui concetti e sulle interpretazioni fuorvianti di web semantico e web 2.0 e anche sulle differenze di portato informativo e di interazione con gli utenti delle susseguenti “fasi numeriche” del web si rinvia a LUCIANO FLORIDI, *Web 2.0 vs. the semantic Web: a philosophical assessment*, “Episteme”, vol. 6 (2009), n. 1, p. 25-37; si veda, inoltre, FABIO METTIERI, *Il grande inganno del Web 2.0*, Roma - Bari, Laterza, 2009.

⁵ Cfr. <https://www.lib.ncsu.edu/endeca>. Sebbene Endeca sia stato sostituito da un *discovery tool* nel 2016, sul sito web della biblioteca si legge la storia e la motivazione delle scelte adottate per i software della biblioteca, dalle quali chi scrive ha tratto queste notizie e una bibliografia di riferimento. Tali informazioni sono sicuramente una fonte per la storia delle biblioteche. Sull’implementazione di Endeca alla NCSU si veda anche KAREN G. SCHNEIDER, *Toward the next gen catalog*, “ALA TechSource”, 03-10-2006, <http://www.ala.org/tools/article/ala-techsource/toward-next-gen-catalog>; CHRISTINE DEZELAR-TIEDMAN, *Redefining library resources in discovery systems*, in *Managing metadata in web-scale discovery systems*, ed. by Louise F. Spiteri, London, Facet Publishing/ALA Editions, 2016, p. 91-112: 93.

⁶ Questa definizione è liberamente tratta da un documento tra i primi a proporre uno schema chiarificatore e definitorio di questi sistemi: ATHENA HOEPPNER, *The ins and outs of evaluating web-scale discovery services*, “Computers in Libra-

ries”, 32 (2012), n. 3, p. 6-10.

⁷ Si veda per esempio come lo definisce il MIT: <https://libguides.mit.edu/scifinder>; è molto chiaro in proposito il sito dell’Università di Catania, che annovera CAS (il portale d’accesso a SciFinder Scholar) tra le risorse online non interrogate dal DT adottato dalla biblioteca: <http://www.sida.unict.it/content/summon>.

⁸ Sul processo di indicizzazione si veda MARSHALL BREEDING, *The future of library resource discovery*, Baltimore, NISO, 2015, p. 18; MARY M. SOMERVILLE, *Digital age discoverability: a collaborative organizational approach*, “Serials review”, 39 (2013), 4, p. 234-239.

⁹ Il problema è affrontato in NISO, *Open discovery initiative: promoting transparency in discovery. Recommended Practice of the National information standards organization*, prepared by the Open discovery initiative working group, Baltimore, National information standards organization (NISO), 2014.

¹⁰ Sul concetto di inclusione nei DT si rinvia all’analisi di MONICA MOORE, *But is my resource included? How to manage, develop, and think about the content in your discovery tool*, “The serials librarian”, 70 (2016), n. 1-4, p. 149-157; alcuni paragoni tra i metadati originali e quelli indicizzati si trovano in A. Hoepfner, cit.

¹¹ Cfr. NISO, *Open discovery initiative: promoting transparency in discovery*, cit.

¹² Un’analisi approfondita si trova in ANGELA KROEGER, MARSHALL BREEDING, HEATHER MOULAISSON SANDY, *Sharing metadata across discovery systems*, in *Managing metadata in integrated discovery systems*, ed. by Louise F. Spiteri, London, Facet Publishing/ALA Editions, 2016, p. 17-55.

¹³ NISO, *Open discovery initiative: promoting transparency in discovery*, cit.

¹⁴ Il caso è riferito da DeZelar-Tiedman, *Redefining library resources in discovery systems*, cit., p. 147.

¹⁵ Sulle abitudini di ricerca e le statistiche si vedano, nell’ordine: MONICA MOORE, *But is my resource included? How to manage, develop, and think about the content in your discovery tool*, “The serials librarian”, 70 (2016), n. 1-4, p. 149-157; NADINE P. ELLERO, *Integration or disintegration: where is discovery headed?*, “Journal of library metadata”, 13 (2013), n. 4, p. 311-329; AARON TAY, *Managing volume in discovery systems*, in *Managing metadata in web-scale discovery systems*, ed. by Louise F. Spiteri, London, Facet Publishing/ALA Editions, 2016, p. 113-135.

¹⁶ Si vedano i portali d’accesso ai *discovery tools* di Duke University: <https://library.duke.edu/find>; North Carolina University: <https://search.lib.ncsu.edu/?q=obama+care>; Stanford University: <http://library.stanford.edu/all/?q=obama+care&op=Search>.

¹⁷ M. BREEDING, *The future of library resource discovery*, cit.; ROGER C. SCHONFELD, *Does discovery still happen in the library?*, 24 September 2014, <https://sr.ithaka.org/publications/does-discovery-still-happen-in-the-library-roles-and-strategies-for-a-shifting-reality>.

¹⁸ Nel rapporto NISO, *Open discovery initiative*, cit., p. 16-21, si vedano le tabelle riassuntive delle tipologie di metadati forniti dagli editori ai produttori di sistemi di *discovery*, suddivise per formati bibliografici.

¹⁹ Una tabella che suddivide i contenuti in base ai concetti di richiamo e precisione si trova in A. TAY, *Managing volume in discovery systems*, cit., p. 124-125.

ABSTRACT

Discovery tools are specialized portals for bibliographic research widely used in libraries with heterogeneous collections of electronic and digital resources. The feature distinguishing them from software for federated search among online resources is their central index, that is a database to which libraries have access thanks to the contract stipulated with the discovery tool software companies. Between the two parts composing a discovery tool – the central index and the search interface – librarians have paid great attention to the second during the selection and parametrization of these research systems. The fact that discovery tools are becoming the only access portal to library collections, especially in the academic field, should shift the attention to their core tool: the central index, that is, a large collection not containing documents but metadata redirecting the user to the original databases containing information and publications.

DOI: 10.3302/0392-8586-201901-011-1