

Catalogare Internet

Un'esperienza di indicizzazione delle risorse web di viaggi, arte, musei e natura maturata in ambiente bibliotecario

di Davide Bassi e Gianfranco Bettoni

Il paradosso di un "catalogo universale" ha tormentato il sonno dei bibliotecari per vari secoli. L'ipotesi della totalità della conoscenza rispecchiata da un ordine bibliografico universale non smette di esercitare il suo fascino neppure oggi, appare però involontariamente ingenua. La struttura della conoscenza è mutata e con essa i supporti in grado di registrarla e scambiarla. Un bibliotecario alchimista del XVI secolo o un erudito del Settecento concepivano il sapere come un'acquisizione immutabile. Il Novecento ce ne ha svelato drammaticamente la relatività. I nuovi media comunicativi ci inducono a pensare la conoscenza non solo come dato relativo, ma anche provvisorio: le informazioni in Internet impiegano qualche decina di secondi a manifestarsi, ma sono tanto effimere da evolvere o diventare obsolete nel giro di qualche giorno o settimana. L'aspirazione al catalogo universale ha progressivamente perso di valore di fronte all'impatto delle tecnologie dell'informazione, che allo stato dell'arte permettono a chiunque di produrre, riprodurre, diffondere, simulare "conoscenze". Le tecnolo-

gie dell'informazione hanno prodotto Internet, in luogo del catalogo universale. E i bibliotecari moderni di fronte a questo strumento non soffrono più tanto il problema di disciplinare l'interrezza dello scibile, quanto quello di ricavare informazioni usabili da una mole di dati "universale".

Nella ricerca sul web il rumore è generalizzato e imponente, tanto da rendere realmente inquantificabile il tempo che si impiegherà ad ottenere qualcosa di utile. E questa dispersione è tanto più frustrante, quanto si percepisce la ricchezza potenziale del mezzo che si usa.

Internet è, non solo metaforicamente, assimilabile alla biblioteca: entrambe sono raccolte di documenti (i siti altro non sono che questo). Il web però non dispone di un catalogo che organizzi le sue risorse informative, rendendole sistematicamente fruibili. I motori di ricerca rimangono strumenti indispensabili, ma operano su dati non preventivamente indicizzati e per questo non possono offrire grandi garanzie riguardo alla selettività dei risultati. I bibliotecari sanno che la complessità e l'accumulo multidirezionale delle informazioni della re-

te mettono in crisi la natura indiscriminata della ricerca automatica, come già rilevava Clifford Lynch nel lontano 1997.¹

Dalla constatazione di come sia oggi più importante di ieri valutare per selezionare l'informazione, scartare ciò che è superfluo per promuovere i contenuti meritevoli di recupero, discernere ciò che risulta stabile nel tempo e rigoroso nel metodo da ciò che è effimero e portatore di "rumore" è nata un'esperienza pratica di catalogazione del web sfociata nella produzione di un database su cd-rom e nel suo dimostrativo come sito-guida Internet, utilizzabile altresì da aggiornamento del database.²

Un gruppo di catalogatori di Mantova ha voluto sperimentare una possibile metodologia indicizzatoria, applicandola ad una tipologia particolarmente rappresentativa della vitalità del web, quella dei siti di viaggi, arte, musei e cultura ambientale. Il progetto ha tratto origine dall'esperienza di Itineroteca Ulisse, base dati di spoglio dei periodici distribuita a lettori e biblioteche dal 1992 e patrocinata sino all'anno 2000 dalla Biblioteca comunale di Viadana.³

Vogliamo qui rilevare le differenze morfologiche e contenutistiche dell'approccio catalografico al web rispetto allo spoglio tradizionale, ancorché automatizzato, dei periodici di viaggi.

Come necessaria premessa deve essere sottolineata la fonte delle informazioni. L'input è derivato dalle citazioni di indirizzi Internet che da qualche anno corredano i servizi giornalistici delle riviste indicizzate da Itineroteca Ulisse. La selezione effettuata da giornalisti italiani, europei e statunitensi è stata considerata garanzia della affidabilità dei siti. Questi vengono valutati con scrupolo dal team di catalogazione prima di trovare ospitalità nella base dati: viene esaminata la significatività dell'informa-



maticamente si viene trasportati al nuovo indirizzo. Solo il decesso del sito, in genere, rende inefficace il dominio. Una questione decisiva riguardava poi il modo di strutturare il contenuto dell'unità bibliografica sito. È parso subito evidente che i criteri della soggettazione e del-

mente verso ciò che gli è realmente utile.

Per converso, l'indicizzazione deve essere tanto spinta da segnalare e rendere recuperabile ogni aspetto significativo del sito. Ciò può comportare l'adozione di dieci, quindici, venti soggetti per siti particolarmente complessi. In concreto, si è approntato uno schema di categorie, o classi, organizzate gerarchicamente. Si è fatto in modo che queste categorie rispecchiassero l'organizzazione ipertestuale, "a menù", dei siti migliori e più navigabili.

In fase di implementazione, ad ogni classe selezionata vengono abbinate parole chiave, necessarie ad esaurire l'argomento che la classe tratta. Per fare un esempio concreto: se il sito <www.barcelona.com> ospita buone pagine sulla Sagrada Familia, verrà scelta, tra le altre, la classe Architettura religiosa, alla quale andrà legata la parola chiave "cattedrale", seguita dal nome proprio del monumento, ecc. Sarà possibile quindi effettuare la ricerca in base alla semplice località, oppure associare alla località una categoria o più parole chiave, ottenendo un risultato selettivo e puntuale. Il meteo, le webcam, le mappe, l'indicazione di bibliografie specifiche, un buon corredo fotografico e la presenza di link ci sono parsi attributi "formali" salienti e ne abbiamo permesso una ricerca separata e facilitata.

Il controllo terminologico dei soggetti attiva poi relazioni sinonimiche strutturate, trasparenti all'utente. La ricerca inoltrata a partire ➤

zione e la ricchezza documentaria, nonché la rilevanza come crocevia verso altre risorse, attraverso l'articolazione dei link proposti. Altre fonti di indirizzi sono il portale Yahoo (si è optato per uno spoglio ragionato delle directory viaggi e turismo delle diverse lingue), i molteplici portali nazionali di turismo, arte, musei, parchi, nonché le guide cartacee di risorse on line.⁴

Il primo problema squisitamente catalografico che si è dovuto affrontare concerneva l'affidabilità delle URL. Gli indirizzi dei siti, chiunque navighi lo ha sperimentato, non sono stabili. Si corre costantemente il rischio di costruire un record catalografico che si basa su un link evanescente o non più valido.

L'esperienza dell'ultimo biennio ci ha insegnato che le URL delle pagine interne a un sito variano a distanza di qualche mese, mentre il dominio nudo è, nella maggior parte dei casi, permanente. Anche nel caso in cui questo muti, vengono attivati di norma meccanismi di ridirezzionamento per cui auto-

la classificazione libraria erano poco adeguati. L'utente di Internet dispone di tempi limitati rispetto all'utente di biblioteca e, in funzione di questi tempi limitati, cerca un tipo di informazione diversa: pratica, circostanziata, immediatamente spendibile. Può aver bisogno di sapere il numero di telefono di un ristorante o di un affittacamere a Firenze, la mappa di un percorso naturalistico, lo stato di innevamento di una pista di sci, gli orari e le collezioni di un museo, la foto di un bene architettonico o il meteo di Dublino. Si deve dunque offrire all'utente una griglia di opzioni ampia e capillare in modo da focalizzare le sue esigenze e guidarlo immediata-

Tab. 1 - e-viaggiatori base dati siti web

Luoghi	Siti catalogati	Accessi per classe
Italia	2.500	12.400
Europa (esclusa Italia)	4.000	18.500
Asia	350	3.500
Americhe	780	5.200
Oceania	210	3.100
Africa	150	1.800

da “alberghi” mette in azione automaticamente la ricerca sui termini “albergo” e “hotel” in quanto sinonimi stretti e restituisce una tabella di opzione dalla quale è possibile vagliare termini affini di rimando (i *related terms* delle forme tesaurali) adottati nella base dati, come “resort”, “bed & breakfast”, “paradores”, ecc. L’accuratezza dell’intelaiatura sinonimica costituisce un vero e proprio dizionario in linea, che, contemplando traduzioni per le voci di più diffuso utilizzo, è di fondamentale ausilio per l’utenza non di lingua italiana.

Ma di tutte le differenze tra catalogazione semantica delle fonti tradizionali a stampa delle risorse virtuali la più rilevante è quella legata ai toponimi. In una base dati orientata ai viaggi e al turismo, le liste di autorità dei nomi geografici hanno una rilevanza essenziale. Per questo stati, regioni, province (dipartimenti, contee, ecc.), regioni

storiche e geofisiche, sono state ordinate gerarchicamente in base all’appartenenza amministrativa e territoriale. Questa struttura tesaurale ha permesso di mettere in atto una prestazione vantaggiosa in fase di interrogazione. Richiedendo Aix-en-Provence potremo anche non trovare il sito Internet specifico della località, in questo caso verremo rimandati al livello superiore della gerarchia geografica (il dipartimento Bouches du Rhone) e, nel caso malcapitato di ulteriore risultato nullo, direttamente alla regione Provence, dove l’utente potrà comunque essere soddisfatto dei suoi primi bisogni informativi. Inutile aggiungere che nell’ulteriore risultato negativo su Provence (che renderebbe inaccettabile la povertà della base dati) il rimando porterebbe alla Francia e, infine, all’Europa. L’authority file attuale è il prodotto dei 10 anni di indicizzazione di Itineroteca Ulisse.⁵

Ma come si presentano i siti ben costruiti, ad esempio quelli degli enti turistici nazionali o regionali, o degli enti locali? I migliori siti hanno una vocazione enciclopedica. Si riferiscono ad un territorio generale, ma, nelle pagine interne a partire da quella di snodo, vanno a “spogliare” tutte le località del territorio in questione. Come si può trattare catalogograficamente questa peculiarità con un’indicizzazione pertinente senza trascurare alcun elemento di sostanza e al contempo senza creare “fantasmi virtuali”? (passateci il doppio paradosso).

Tutte le località trattate significativamente nelle pagine interne del sito vengono citate in un campo apposito, recuperandone il nome nella forma accettata dall’authority file, e sono rese disponibili in una fase di ricerca avanzata. Le informazioni sui siti dei musei di S. Gimignano, infatti, possono essere albergate tanto sul portale regionale, quanto sul sito del comune. Nel primo caso, il record catalogografico riporterà “Toscana” come identificatore geografico, e “S. Gimignano” risulterà ugualmente recuperabile dalla ricerca avanzata, una volta inserito nel campo degli “Altri toponimi”. Nel secondo caso, l’identificatore geografico del sito sarà direttamente “S. Gimignano”, attingibile sia dalla ricerca semplice che da quella avanzata.

Il metodo strutturale di ricerca è l’AND. Combinando “Parchi” e “Toscana” troveremo tutti i siti delle aree protette di questa regione ma anche le pagine relative ai parchi toscani presenti su altri siti di natura “enciclopedica”.

Attivando la scelta sulle icone delle utilità standard più frequenti nella tipologia dei siti turistici (mappe, meteo, foto, webcam, bibliografia) potremo selezionare le mappe dei quartieri di una città metropolitana, consultare le pagine di bibliografia su un determinato territorio, cono-

Tab. 2 - e-viaggiatori base dati articoli riviste

Principali riviste indicizzate	Copertura cronologica	Spogli bibliografici
Airone	1981-2000	5.010
Archeo	1990-2000	3.010
Archeologia viva	1988-2000	1.140
Bell'Europa	1993-2000	2.750
Bell'Italia	1986-2000	6.180
Dove	1991-2000	2.960
Gente Viaggi	1982-2000	5.120
Geo (ed. francese)	1995-1998	620
Geo (ed. tedesca)	1995-2000	960
Globo	1995-1999	790
Grands Reportages	1995-2000	1.350
Gulliver	1993-2000	2.010
In viaggio	1997-2000	930
Itinerari e luoghi	1992-2000	1.510
Merian	1995-2000	860
Meridiani	1988-2000	1.940
National Geographic	1983-2000	1.150
National Geographic (ed. italiana)	1988-2000	560
National Geographic Traveler	1995-1999	490
Oasis	1995-2000	1.590
Panorama Travel	1998-2000	1.100
Plein air	1991-2000	4.750
Qui Touring	1984-2000	3.200
Rivista del trekking	1984-2000	1.780
Tuttoturismo	1987-2000	4.600
Weekend Viaggi	1983-2000	5.100

scere la situazione meteorologica aggiornata di ciascuna area. Una performance consentita da un'indicizzazione esaustiva con assoluta garanzia di pertinenza. L'indicazione delle lingue dei siti e la loro possibile selezione informa preventivamente l'utente per meglio orientarlo sui siti più comprensibili e a lui più congeniali.

All'interno dello stesso ambito geografico (dallo stato all'unità amministrativa più piccola qual è il quartiere metropolitano) i risultati vengono presentati in base al voto attribuito dal bibliotecario documentalista. Ciò permette all'utente di disporre di una lista di siti ordinati in base alla loro importanza.

La valutazione dell'indicizzatore (con scala da 1 a 5) è data dalla sintesi tra valore contenutistico, sua grafica e usabilità del sito. A questo primo giudizio si aggiunge

quello sulla quantità dei link, previa verifica della loro attendibilità alla data di indicizzazione, riportata automaticamente sulla scheda e riferita alla prima indicizzazione e ai successivi editing.

Un sito poco elaborato dal punto di vista del contenuto proprio può tuttavia essere importante come crocevia verso le risorse Internet grazie ad un buon reticolato di link. I portali organizzati per *directories* e la nuova guida Internet <www.e-viaggiatori.com>, della quale abbiamo trattato, fanno parte di questa categoria di siti.

Non possiamo concludere senza auspicare un largo coinvolgimento dei bibliotecari nel trattamento delle informazioni e della documentazione presente in rete con la consapevolezza che (ciascuno nel contesto in cui esercita la propria *mission*) occorre prendersi cura della biblioteca senza catalogo. ■

Note

¹ CLIFFORD LYNCH, *La ricerca su Internet*, "Le Scienze", ed. italiana di "Scientific American", ottobre 1997, 350, p. 60.

² Il sito Internet di cui si parla è <www.e-viaggiatori.com> oppure <www.e-viaggiatori.org>.

³ ANTONIO ALIANI - GIANFRANCO BETTONI, *L'itineroteca Ulisse esplora nuovi territori*, "Biblioteche oggi", 12 (1994), 9, p. 32-35; GIANFRANCO BETTONI, *Viaggiare... tra le righe*, "Biblioteche oggi", 15 (1997), 10, p. 32-33.

⁴ Citiamo per tutti MARCO CERIANI, *Viaggi e turismo on line*, Milano, Alpha test/Hoepli, 2000.

⁵ Queste le autorità bibliografiche e telematiche utilizzate: *Compact:Enciclopedia geografica De Agostini*, Novara, 1989 e successive edizioni aggiornate; TOURING CLUB ITALIANO, *Annuario generale dei comuni e delle frazioni d'Italia*, Milano, 1993; NATIONAL GEOGRAPHIC SOCIETY, *Atlas of the World*, 1994; *Getty Thesaurus of Geographic names*, in <http://shiva.pub.getty.edu>.