

UNIVERSITÀ CA' FOSCARI - VENEZIA

FACOLTÀ DI LETTERE E FILOSOFIA

CORSO DI LAUREA SPECIALISTICA IN ARCHIVISTICA E BIBLIOTECONOMIA

Opac di nuova generazione: il confronto con il Web

RELATORE
PROFESSOR RICCARDO RIDI

LAUREANDA
LUCIA TRONCHIN
MATRICOLA 820493

ANNO ACCADEMICO 2009-2010

INDICE

INTRODUZIONE	5
1. IL CONTESTO	7
1.1 <i>Le prime generazioni di opac</i>	7
1.2 <i>L'ambiente informativo</i>	14
1.3 <i>La Google generation e i comportamenti di ricerca</i>	16
1.4 <i>Next generation opacs</i>	20
2. GLI OPAC SOCIALI.....	24
2.1 <i>Tagging e recupero dell'informazione semantica</i>	25
2.2 <i>Massa critica e volontà di partecipazione</i>	36
2.3 <i>Tag e altre caratteristiche sociali negli opac in uso</i>	41
3. RELEVANCE RANKING E RAGGRUPPAMENTI DINAMICI.....	48
3.1 <i>Il relevance ranking</i>	50
3.2 <i>Il concetto di rilevanza</i>	51
3.3 <i>Il relevance ranking negli opac</i>	59
3.4 <i>Classificazione a faccette e faceted browsing</i>	67
3.5 <i>Le faccette negli opac</i>	70
3.6 <i>Nuvole di parole e grafi</i>	81
3.7 <i>Ranking e faccettazione nel giudizio degli utenti</i>	85
4. OPEN DATA, LINKED DATA.....	90
4.1 <i>Il Web arricchisce il catalogo</i>	90
4.2 <i>Il catalogo arricchisce il Web</i>	93
4.3 <i>Il Web semantico</i>	99
4.4 <i>Componenti del Web semantico</i>	104
4.5 <i>Linking Open Data initiative e Linked data</i>	110
4.6 <i>Web semantico, Linked Data e biblioteche</i>	115
4.7 <i>LIBRIS il catalogo delle biblioteche Svedesi</i>	122
CONCLUSIONE	125
BIBLIOGRAFIA	127

INTRODUZIONE

OPAC è l'acronimo di Online Public Access Catalog creato per designare i cataloghi elettronici delle biblioteche. Questi strumenti hanno avuto nel corso della loro storia ormai trentennale il destino di essere, nell'opinione di quelli che maggiormente li usano, bibliotecari e utenti delle biblioteche, alternativamente “nella polvere” e “sull'altare”. Nati in America negli anni 70 del Novecento sono stati all'inizio un sottoprodotto dei processi di automazione delle procedure gestionali all'interno di grandi e ben dotate istituzioni, poco apprezzati nella loro funzione di recupero dei documenti, usati principalmente dai bibliotecari e solo occasionalmente messi a disposizione del pubblico che normalmente continuava a servirsi dei cataloghi cartacei per le difficoltà rappresentate da interfacce poco amichevoli e metodi di interrogazione complessi. Solo nel corso degli anni '80 lo sviluppo delle tecnologie dei minicomputer ha consentito la loro diffusione nelle biblioteche di più piccole dimensioni e li ha resi un comune strumento di accesso ai cataloghi anche per gli utenti. L'introduzione nei software delle tecniche di information retrieval, che hanno consentito più articolate interrogazioni del catalogo, associata ai miglioramenti nel disegno delle interfacce, ha fatto sì che gli opac diventassero nei primi anni '90 estremamente graditi agli utenti. Di diverso avviso i bibliotecari che, pur consapevoli delle migliori funzionalità di recupero e della facilità d'uso delle nuove interfacce a menù e in seguito di quelle grafiche, evidenziavano i limiti di uno strumento che non aveva saputo restituire pienamente la rete di relazioni rappresentate nei cataloghi cartacei e, specie nella ricerca per soggetti, non aveva trovato il modo di sfruttare pienamente la tecnologia per sviluppare modalità innovative di esplorazione dei dati catalografici.

La nascita della rete internet ha consentito agli opac di acquistare visibilità al di fuori delle mura delle biblioteche e raggiungere una platea più vasta facilitando la vita degli utenti nel consentire loro di fare ricerche sulle risorse documentarie delle biblioteche da casa o dal loro posto di lavoro e nel limitare la necessità di recarsi in biblioteca al solo recupero fisico del documento. Nello stesso tempo, specie con l'avvento del World Wide Web, la rete è diventata il metro di misura con cui gli utenti valutano gli opac. Le tecniche di ricerca dei search engine, gli stili di presentazione delle informazioni dei siti commerciali, i modi di creazione e condivisione di contenuti tipiche dei social network hanno modificato le abitudini degli utenti e retrocesso gli opac a strumenti che poco corrispondono alle loro aspettative, alle loro abitudini di ricerca e utilizzo dei documenti e delle informazioni. La grande disponibilità di risorse provenienti da innumerevoli fonti ha creato inoltre un ambiente informativo diverso dal passato, nel

quale le biblioteche rappresentano solo una piccola parte di tutto ciò che è disponibile, e una parte poco appetibile se consideriamo che la maggior parte delle biblioteche è in grado di mostrare sul Web solo metadati mentre non riesce ancora, se non in modo limitato, a fornire l'accesso diretto ai documenti siano testi, immagini, registrazioni sonore o visive. Gli opac di nuova generazione devono rispondere a questi cambiamenti se vogliono continuare a mantenere una certa rilevanza per gli utenti. Per farlo hanno adottato molte delle caratteristiche dei sistemi oggi diffusi sul Web: presentano interfacce di ricerca semplici, costituite da una sola casella in cui si immettono parole chiave; mostrano i risultati ordinati secondo criteri di relevance ranking; organizzano il set di risposta in insiemi che evidenziano particolari caratteristiche dei documenti favorendo il processo di selezione delle risorse; adottano visualizzazioni grafiche di parole significative che consentono la navigazione del catalogo. Cercano inoltre di arricchire i dati bibliografici presentando copertine, indici, abstract, recensioni dei documenti e consentono agli utenti di interagire col catalogo attribuendo i propri tag o lasciando commenti e suggerimenti, favorendo il crearsi di comunità di utenti intorno alle biblioteche. Uno sforzo ancora maggiore per mantenere la centralità del catalogo elettronico quale punto di accesso a molteplici risorse è fatto da quelle biblioteche che consentono attraverso l'opac l'utilizzo di banche dati e altre risorse digitali selezionate dalla biblioteca a favore dei propri utenti.

Tutto questo non sembra ancora essere sufficiente a garantire un posto stabile per i cataloghi elettronici delle biblioteche nell'attuale panorama informativo, complesso e in rapida evoluzione. Gli opac di nuova generazione cominciano ad essere adottati in diverse biblioteche ma non costituiscono ancora lo standard per tutte; alcune caratteristiche, come quelle legate al concetto di socialità sono state adottate in modo superficiale; nuovi dispositivi come gli iPhone e gli E-Book, e tutti quelli che verranno, pongono nuove sfide tecnologiche ma soprattutto rilanciano le problematiche legate all'identità e al ruolo degli opac, a come le informazioni bibliografiche, e tutte le altre, vengono cercate e distribuite nel contesto informativo attuale. Il cambiamento sostanziale che gli opac sembrano dover affrontare è trasformarsi da database di record bibliografici "stand-alone" a insiemi di dati interconnessi che possano interagire con altre risorse nel World Wide Web, integrati nel normale flusso di lavoro delle persone. Questa visione è ancora vaga e non è possibile dire dove questi cambiamenti porteranno ma è un fatto che le grandi istituzioni bibliotecarie si stanno preparando per questo futuro: l'idea di essere parte integrante e non separata di questo universo interconnesso è stata recepita e il processo di riorganizzazione dei dati è avviato.

1. IL CONTESTO

1.1 Le prime generazioni di opac

Gli opac¹ delle biblioteche hanno oggi una storia più che trentennale. Nati dagli strumenti di automazione del lavoro di back office delle biblioteche, principalmente delle procedure di acquisizione e di gestione della circolazione dei documenti, e dal trasferimento in ambiente elettronico delle procedure manuali di catalogazione dei documenti e di produzione delle schede catalografiche, si sono evoluti fino a diventare prodotti estremamente complessi dal punto di vista logico e tecnologico in funzione di cambiamenti che hanno interessato tutte le componenti del sistema. La moltiplicazione dei supporti documentari, la natura ibrida delle collezioni, l'aumento esponenziale dei documenti disponibili, l'evoluzione dei codici catalografici, la creazione di formati standard per la registrazione e il trasferimento dei dati bibliografici, la molteplicità delle attività degli utenti da gestire attraverso l'opac, l'evoluzione dei componenti hardware, software e di rete (dalla creazione dei personal computer all'applicazione delle tecniche di information retrieval, dall'adozione delle interfacce grafiche allo sviluppo delle reti informatiche fino al nuovo ambiente informativo costituito dal Web 2.0) hanno determinato innovazioni nel contenuto degli opac, richiesto un'estensione delle loro funzionalità, imposto un continuo sforzo di aggiornamento e adattamento ai nuovi contesti biblioteconomici e tecnologici.

Nel ripercorrere velocemente le tappe di questa storia non si possono trascurare i pionieri, coloro che hanno aperto la strada in un contesto tecnologico costituito da computer main frame, terminali a tubo catodico, interfacce a caratteri e in un contesto professionale talvolta ostile. I primi tre Opac americani furono quello dell'Online Computer Library Center

¹ OPAC è l'acronimo di On Line Public Access Catalog. Sembra che il termine sia stato coniato nel 1981 da Charles Hildreth durante un incontro del Council on Library Resources (CLR) a Washington. In quell'incontro le principali istituzioni bibliotecarie americane, tra cui la Library of Congress (LC), OCLC, l'Association of Research Libraries (ARL), la New York University (NYU) Libraries, programmavano di condurre uno studio sui loro cataloghi online. Prendendo appunti Charles Hildreth scrisse per la prima volta OPAC sottolineando i termini public e access. Ricordando l'episodio nel 1994 in un suo messaggio all'Open Lib/Info science Education Forum egli racconta che propose di utilizzare il termine per distinguere questi nuovi sistemi di ricerca ad accesso pubblico dagli allora popolari (tra i bibliotecari) sistemi di catalogazione online interni come quelli di OCLC e di RLIN che occasionalmente venivano girati agli utenti. Ma molti non avevano ancora compreso la sostanziale diversità dei due strumenti e questo suggerimento non fu accolto mentre fu adottato il termine Computer catalog. Non poteva immaginare quanto "OPAC" avrebbe preso e sarebbe diventato il termine universalmente scelto e riconosciuto in tutto il mondo (Hildreth 1994).

(OCLC) nel gennaio 1974, del Research Libraries Group/ Research Libraries Network (RLG/RLIN) nel dicembre 1974, e dell'Ohio State University Library nel gennaio 1975 (Hildreth 1982 p. 3). Gli anni '70 furono un periodo di sperimentazione di questa tecnologia: nessuna grande istituzione aveva ancora pensato di chiudere i propri cataloghi cartacei, tra i bibliotecari molti erano scettici sulla possibilità che questi strumenti potessero sostituire i cataloghi tradizionali e giudicavano negativamente gli informatici, molti cataloghi automatizzati furono abbandonati e persino alcune biblioteche che erano state pioniere nello studio dei sistemi di automazione li abbandonarono considerandoli troppo costosi (Su 1994 p. 136). Quello che diede una svolta decisiva alla diffusione degli opac fu il cambiamento che avvenne nella tecnologia dei computer e nel relativo mercato nel corso di quegli stessi anni '70. I computer divennero più veloci, con maggiori capacità di memoria, meno costosi. La diffusione sul mercato dei minicomputer determinò un cambiamento radicale nelle prospettive delle aziende di software, che cominciarono ad offrire sistemi di automazione "chiavi in mano" anche a biblioteche più piccole². All'inizio degli anni '80 si erano create le condizioni perché la questione opac diventasse di interesse per tutte le biblioteche negli Stati Uniti: le piccole biblioteche poterono pensare di sviluppare o adattare in proprio sistemi adeguati alle loro specifiche esigenze mentre le grandi istituzioni ebbero la possibilità di offrire direttamente ai loro utenti strumenti di accesso al catalogo (Yee - Layne 1996 p. 154-155). Parallelamente alla diffusione degli opac si moltiplicarono gli studi che ne analizzavano i diversi aspetti tra cui l'analisi del comportamento degli utenti³. Questi studi rivelarono che l'ostilità degli utenti verso i cataloghi on line, temuta da alcuni bibliotecari, non esisteva. Al contrario gli utenti apprezzavano gli opac e li preferivano al catalogo cartaceo anche quando ne ottenevano risultati di qualità inferiore (Su 1994 p. 137).

Ripercorrendo la storia degli opac dal punto di vista delle capacità funzionali, Hildreth parla di due generazioni di prodotti sostanzialmente diversi e di una terza che a metà degli anni '90 si stava facendo strada ma non era ancora diffusa nel mercato dei sistemi per

² Per un'analisi dello sviluppo negli opac in America negli anni '60, '70 e '80 in relazione allo sviluppo delle tecnologie e alla cooperazione tra biblioteche si veda De Gennaro (De Gennaro 1982).

³ Uno specifico programma del Council on Library Resources (CLR) promosse studi su diversi aspetti degli opac tra cui l'analisi dei sistemi esistenti, lo sviluppo di metodi di analisi dei costi e per la gestione, lo sviluppo di tecniche di comunicazione tra diversi sistemi oltre all'impatto sugli utenti (Li 1985 p. 326). Questi studi culminarono in un rapporto finale (Matthews-Lawrence-Ferguson 1983) nel quale il 1982 fu dichiarato l'anno del catalogo on line.

biblioteche (Hildreth 1989, 1991, 1995). La prima generazione di opac sarebbe stata un prodotto secondario derivato dall'automazione delle procedure di circolazione dei documenti, poco di più che una meccanizzazione del catalogo cartaceo con possibilità di accesso minime. Normalmente in questi primi opac era possibile la ricerca per autore e per titolo, mentre la ricerca per soggetto non era prevista; il catalogo era organizzato secondo i principi della ricerca preordinata e l'accesso ai dati avveniva immettendo chiavi derivate o attraverso la corrispondenza esatta di una parola o di una frase; non era possibile usare filtri di ricerca, quali la data o la lingua del documento; non era possibile raffinare progressivamente la ricerca sulla base dei primi risultati ottenuti; la ricerca era un processo lineare in due o tre passaggi obbligati; i dati catalografici che venivano proposti come risposta erano una rappresentazione in formato ridotto del contenuto di una scheda cartacea.

La seconda generazione di opac, che apparve intorno alla metà degli anni '80, rappresentò un salto qualitativo notevole: era caratterizzata da prestazioni migliori dovute all'aumento dei punti di accesso, alla possibilità di scorrimento di liste alfabetiche, all'introduzione della ricerca per soggetti che era stata individuata dagli utenti come una necessità e quindi realizzata nei sistemi. In questi opac si era inoltre cercato di applicare e sfruttare al massimo i metodi di accesso ai dati tipici dell'information retrieval. La sfida che si poneva agli sviluppatori era costituita dalla necessità di combinare la semplicità d'uso, per rendere autonomi anche gli utilizzatori meno esperti, con la possibilità di sfruttare le nuove potenzialità di ricerca (Hildreth 1995). I risultati conseguiti con questi sviluppi consentivano a Hildreth di affermare

Today's online catalogs are very popular with users. This young library access technology may even be, in some sense, addictive - users return to it again and again. In study after study it is reported that users much prefer online library catalogs over card or microfilm catalogs. These same studies also report that users are usually satisfied with the search performance of OPACs and enjoy using them. This development alone is reason enough to pause and celebrate the tremendous OPAC-related accomplishments of the past fifteen years. We all - librarians, researchers, innovative system designers, and vendors - have come a long way in a very short period of time. Reflecting on the early, cautious visions of the card catalog online [...], we

can now point with pride to the achievements gained and the expanded service benefits provided by the best of today's installed OPACs (Hildreth 1991 p. 17)⁴

Secondo Weston, che esplicitamente fa riferimento agli studi di Hildreth (1989, 1991), la terza generazione di opac è rappresentata dai cataloghi che, a partire dall'inizio degli anni '90, hanno adottato l'interfaccia grafica e l'architettura client-server. L'interfaccia grafica segna un radicale cambiamento nelle funzionalità del catalogo semplificando il dialogo tra utente e computer grazie alla presenza di icone che sostituiscono i comandi e grazie all'uso del mouse. L'architettura client-server consente la suddivisione delle elaborazioni tra il personal computer locale e la macchina centrale server e, nel ridurre al minimo la trasmissione dei dati nelle due direzioni, ottimizza l'uso dei canali di trasmissione. Nel contesto degli opac il client prepara l'interrogazione e una volta ricevuta la risposta dal server la presenta all'utente, mentre il server si occupa della ricerca e della trasmissione del risultato. Questa architettura consente un ampio margine di flessibilità nella presentazione dei dati all'utente, sia un loro impiego successivo. Consente inoltre di svolgere contemporaneamente più procedure, come ad esempio interrogare più basi dati esterne o spedire messaggi di posta elettronica. Lo sviluppo di queste potenzialità rende possibile l'accesso ipertestuale all'informazione e consente l'interattività tra utente e sistema creando un ambiente in cui l'utente può manipolare i dati e non esserne solo passivo fruitore. Questa nuova generazione di opac ha dunque le potenzialità per trarre pieno vantaggio dalla logica dei computer, per diventare un dispositivo di vera e propria mediazione bibliografica (Weston 2002 p. 54).

L'adozione delle interfacce grafiche e dell'architettura client-server non sembra tuttavia esaurire i requisiti che Hildreth (1995) aveva indicato per gli opac di terza generazione. Il catalogo E³ (enhanced, expanded, extended) migliorato nella funzionalità e nell'usabilità

⁴ I cataloghi on line di oggi sono molto popolari tra gli utenti. Questa giovane tecnologia di accesso alla biblioteca può anche, in un certo senso, creare dipendenza: gli utenti vi ritornano continuamente. Studio dopo studio si conferma che gli utenti preferiscono di gran lunga i cataloghi online rispetto a quelli cartacei o in microfilm. Gli stessi studi rilevano anche che gli utenti sono normalmente soddisfatti delle performance di ricerca degli OPACs e hanno piacere di usarli. Già questi sviluppi sono ragione sufficiente per fermarsi e celebrare i formidabili risultati legati agli OPAC degli scorsi quindici anni. Noi tutti, bibliotecari, ricercatori, innovativi designer di sistemi e venditori, abbiamo fatto molta strada in poco tempo. Riflettendo sulle prime caute visioni del catalogo cartaceo online, possiamo guardare con orgoglio ai risultati conseguiti e ai benefici nell'estensione dei servizi consentiti dai migliori tra gli OPACs installati oggi.

attraverso le GUI (Graphical User Interface), espanso fino a diventare lo strumento di accesso a tutta la collezione indipendentemente dalla tipologia dei materiali e con l'inclusione di bibliografie, sommari, indici provenienti da diverse fonti, esteso attraverso link e reti fino a diventare gateway verso collezioni di altre biblioteche e di altri sistemi informativi e fino al full-text dei documenti, non è ancora uno strumento che possa pienamente soddisfare le esigenze di ricerca degli utenti. Hildreth, infatti, nonostante il compiacimento per le realizzazioni più recenti e avanzate, non manca di rilevare le insoddisfacenti capacità di recupero dei dati degli opac in commercio a metà degli anni '90 dovute, a suo avviso, ad un inadeguato modello concettuale della ricerca, specie della ricerca per soggetto. Il punto centrale che egli mette in discussione è l'esperienza di ricerca che l'utente può fare nel sistema. Gli opac, riflettendo il modello dei sistemi di information retrieval, assumono che al momento in cui si inizia una ricerca esista un bisogno informativo definibile, che sia rappresentato attraverso una query, che i materiali rilevanti rispetto a quel bisogno informativo siano indicizzati attraverso parole chiave o descrittori. Il set di risposta è dato dalla corrispondenza (best-matched) tra l'espressione della query e l'indicizzazione dei documenti. Assumono inoltre che per ogni query ci sia un singolo set di risposte che meglio incontra il bisogno informativo. Secondo Hildreth questo paradigma, anche se ha migliorato i suoi risultati introducendo tecniche più probabilistiche che riconoscono e incorporano gli elementi di incertezza intrinseci al processo di recupero delle informazioni, rappresenta solo uno dei possibili percorsi di ricerca ed è concettualmente inadeguato a spiegare la varietà delle situazioni di ricerca e dei comportamenti messi in atto dagli utenti⁵. Una via per creare una nuova generazione di opac si troverà concentrando l'attenzione sulle modalità di ricerca che complessivamente possono essere definite come "browsing" e su come possano essere estese e migliorate.

La lista delle funzionalità che dovrebbero essere il segno di questa terza generazione di opac comprende:

- espressione delle interrogazioni in linguaggio naturale

⁵ Una disamina di queste problematiche è ripresa anche da Yee (Yee 2006).

- conversione automatica dei termini/strumenti di matching dei termini (correzione dell'ortografia, stemming, soundex, tavole di sinonimia⁶)
- tecniche di retrieval di tipo closest o best-match (diversamente dalle query di tipo booleano non si richiede che le chiavi siano esattamente corrispondenti affinché i documenti siano recuperati. Il grado di corrispondenza viene pesato come criterio di ranking)
- ranking dei risultati (secondo diversi criteri: il più rilevante, il più recente, il più citato, il più prestato)
- metodi di relevance feedback (mostrami altri simili a questo, mostrami altri sullo stesso argomento)
- ipertestualità, searching e browsing di registrazioni correlate
- integrazione tra approcci di ricerca di tipo keyword, vocabolari controllati e basati su classificazioni
- massima copertura ed estensione del catalogo (il catalogo come strumento di accesso all'intera collezione).

I cataloghi di terza generazione dovrebbero essere un tipo sostanzialmente nuovo di sistemi di retrieval basati su un modello più rappresentativo degli effettivi comportamenti di ricerca degli utilizzatori (Hildreth 1995 parte 6 p. 2).

A quindici anni di distanza possiamo oggi dire che esistano cataloghi di terza generazione? Le caratteristiche che abbiamo appena descritto sono presenti negli opac? Solo in parte. Markey individua una serie di motivi per cui queste funzionalità non sono state applicate:

- l'ossessione della professione per la catalogazione descrittiva
- l'aumento dei costi di catalogazione
- l'aver scelto altre priorità di intervento tra cui la conversione del retrospettivo e il controllo di autorità
- la mancanza di un consenso generale sulle necessità più immediate del catalogo e l'incapacità di proporre soluzioni economicamente sostenibili

⁶ Gli algoritmi di stemming consistono nel prendere la parola immessa come chiave di ricerca e, aggiungendo prefissi o suffissi e facendo variazioni ortografiche, generare altre parole significative che sono trattate, in modo trasparente all'utente, come ulteriori chiavi di ricerca associate a quella di partenza. Quelle di soundex applicano un principio analogo a parole che hanno un suono simile.

- uno spostamento di interesse dagli aspetti tecnici ai servizi agli utenti
- la riduzione dei fondi disponibili per le biblioteche
- il vertiginoso costo per l'incremento delle collezioni specie per l'acquisto di licenze per l'accesso a fonti online
- gli alti costi di mantenimento degli ILS (Integrated Library System)
- l'incapacità dei venditori di ILS di monitorare i cambiamenti nelle tecnologie di information retrieval e di aggiornare conseguentemente i sistemi (Markey 2007 p. 3).

Quali che siano i motivi, il fallimento nel raggiungere questi obiettivi ha portato gli opac a cedere il loro primato quale punto di partenza nelle ricerche degli utenti. Sembra infatti assodato che gli opac abbiano perso l'appello di cui godevano negli anni '90. Nonostante tutti i progressi, in alcuni casi veramente notevoli, gli opac attuali non godono più di buona reputazione, né fra gli utenti, né fra i bibliotecari⁷. Nel 2005 una ricerca di OCLC ha rivelato che solo il 2% degli studenti si rivolge ai cataloghi delle biblioteche per iniziare una ricerca mentre l'89% si rivolge ad un motore di ricerca (OCLC 2005)⁸. Pur giudicando rilevanti e affidabili le risorse elettroniche messe a disposizione dalle biblioteche, gli utenti ne hanno in realtà una scarsa conoscenza e continuano ad associare la biblioteca solo ai libri; conoscono invece bene i motori di ricerca e li considerano altrettanto degni di fiducia e superiori per qualità, quantità e velocità di accesso. Le biblioteche rientrano in gioco solo quando gli utenti si trovano nella necessità di individuare la copia fisica di un documento.

Il rapporto descrive la situazione in questo modo:

Another way of stating this is that trends indicate a dissonance between the environment and content that libraries provide and the environment and content that information consumers want and use (OCLC 2005 p. 7)⁹.

⁷ Tra gli studi più importanti sugli opac si possono consultare quello della Library of Congress (Calhoun 2006), dell'Università della California (University of California Bibliographic service task force 2005), delle biblioteche universitarie di Wisconsin Madison (University of Wisconsin-Madison Libraries 2008) oltre a interventi di diversi autori tra i quali Fast-Campbell (2004), Breeding (2007).

⁸ La ricerca mirava a raccogliere informazioni sui comportamenti di ricerca degli utenti, sulla loro familiarità con gli strumenti e le risorse messe a disposizione dalle biblioteche e su quanto le biblioteche rispondessero alle loro esigenze in confronto ad altre risorse informative elettroniche. Ha coinvolto, attraverso la compilazione di un questionario online, 3348 utenti fra i 14 e 65 anni e oltre, di lingua inglese, provenienti da USA, Canada, Australia, India, Singapore, Gran Bretagna.

⁹ Un altro modo per affermare questo è che le tendenze indicano una dissonanza tra l'ambiente e il contenuto che le biblioteche forniscono e l'ambiente e il contenuto che i consumatori di informazioni vogliono.

Un precedente rapporto, in relazione al diverso contesto informativo in cui si muovono gli utenti, affermava:

It has become increasingly difficult to characterize and describe the purpose of using libraries [...] The relationships among the information professional, the user and the content have changed and continue to change (OCLC 2004 p. 9)¹⁰.

Prima delle critiche a singole componenti degli opac e dei suoi meccanismi di funzionamento sembra dunque essere oggi dominante una critica generale alla capacità dei cataloghi delle biblioteche, anche nelle loro forme più evolute, a competere, nella scelta degli utenti, con le altre fonti informative disponibili. Dopo essere stati per anni i soli punti di accesso alla conoscenza soffrono oggi la concorrenza di altri strumenti.

1.2 L'ambiente informativo

Dalla fine degli anni '90 il World Wide Web è cresciuto in modo esponenziale e già dall'inizio degli anni 2000 i motori di ricerca, che paradossalmente hanno adottato proprio le tecniche di ricerca che si volevano realizzate negli opac, stanno conseguendo il loro obiettivo di organizzare il mondo della conoscenza (Markey 2007) scalzando gli opac¹¹. Secondo Dempsey (2006) l'architettura generale dei cataloghi è stata tracciata in un mondo pre-network. In quel mondo i materiali erano distribuiti in diverse localizzazioni fisiche ed era molto probabile che vi accedessero principalmente coloro che vi si trovavano più vicini fisicamente. La collezione disponibile corrispondeva in linea di massima a quella rappresentata nel catalogo cioè era quella presente localmente. Si può affermare che in quel contesto le risorse disponibili fossero scarse e disponessero di un alto grado di attenzione: le persone erano disposte a spendere del tempo in biblioteca o nel consultare cataloghi, bibliografie, sommari, indici e altri strumenti di ricerca che normalmente si affiancavano al catalogo. Questo era il modo per trovare le cose. Oggi l'ambiente è cambiato: le risorse informative sono relativamente abbondanti e l'attenzione degli utilizzatori scarsa. In rete sono presenti numerosi e diversi strumenti di ricerca e in particolare i motori di ricerca

¹⁰ E' diventato sempre più difficile caratterizzare e descrivere il fine nell'uso delle biblioteche. [...] Le relazioni tra i professionisti dell'informazione, gli utenti e i contenuti sono cambiate e continuano a cambiare.

¹¹ Markey comparando due ricerche (Cooper 2001) (Sherman 2003) stimava che nel 2000 gli accessi giornalieri a Google superassero di 700 volte gli accessi mensili al catalogo Web dell'Università della California (un catalogo collettivo che comprende le risorse di tutte le biblioteche universitarie e di altre istituzioni dell'intero stato californiano) (Markey 2007 p. 3).

costituiscono un'aggregazione di risorse imponente. L'utente è dunque sommerso di possibilità ed è la rete nel suo insieme, piuttosto che la singola risorsa, ad essere al centro del suo interesse. Il concetto di "collezione disponibile" non coincide più con la collezione localmente disponibile. Questo era vero anche in passato, e vi si sopperiva con la creazione di bibliografie e cataloghi di ogni genere, ma la tecnologia ha alzato le aspettative degli utenti e sembra essere cambiata la loro disponibilità a fare gli sforzi necessari per creare i legami tra risorse disponibili in diversi luoghi: la responsabilità di fare questo attraverso la rete è una sfida per le biblioteche ma non passa necessariamente attraverso il catalogo.

So, the catalogue emerged when patterns of distribution and use of resources, and corresponding behaviours, were very different than they are now. The catalogue was a response to a particular configuration of resources and circumstances. The question now is not how we improve the catalogue as such; it is how we provide effective discovery and delivery of library materials in a network environment where attention is scarce and information resources are abundant, and where discovery opportunities are being centralised into major search engines and distributed to other environments (Dempsey 2006 p. 3)¹².

Grazie ai motori di ricerca il Web si presenta, mentitamente, come un unico insieme di risorse accessibile con strumenti semplici, veloci, efficienti. I risultati di una ricerca, impostata attraverso poche parole chiave, sono presentati in modo standardizzato adottando semplici convenzioni che si ripetono in diversi contesti, consolidando i comportamenti degli utenti, mentre la presentazione dei risultati secondo i criteri di relevance ranking sembra sollevare anche dal lavoro di valutazione delle risorse posizionando in evidenza le risorse ritenute più significative¹³. Chiunque abbia un minimo di competenza informativa è in grado di distinguere tra le apparenze e la realtà: il Web non esaurisce l'universo delle conoscenze e i motori di ricerca tendono a conferire omogeneità a informazioni che sono invece provenienti da fonti estremamente eterogenee per struttura, scopi, aggiornamento (Weston 2007 p. 247) appiattendone il valore, non distinguendone il grado di accuratezza e autenticità. La molteplicità, la varietà e la complessità delle informazioni e delle conoscen-

¹² Quindi, il catalogo è nato quando i modelli di distribuzione e uso delle risorse, e i corrispondenti comportamenti, erano molto diversi da come sono ora. Il catalogo era la risposta a una particolare configurazione di risorse e di circostanze. La questione ora non è come migliorare il catalogo in quanto tale; è come noi forniamo un modo efficace per rinvenire e distribuire i materiali librari in un ambiente di rete dove l'attenzione è scarsa e le risorse informative sono abbondanti e dove le opportunità di scoperta sono state centralizzate in grandi motori di ricerca e distribuite ad altri ambienti.

¹³ Secondo Nielsen il modello mentale consolidato è costituito da una casella in cui si immettono delle parole, un bottone con scritto Search da cliccare per avviare la ricerca, una lista di risultati ordinati che compare in una diversa pagina (Nielsen 2005).

ze oggi disponibili non sono dominabili solo attraverso questi strumenti: esistono grandi quantità di risorse analogiche ed enormi quantità di risorse digitali, il cosiddetto deep Web, non accessibili ai motori di ricerca e il giudizio di valore sui singoli documenti non può banalmente essere risolto da pur sofisticati criteri di ranking¹⁴. Tuttavia va riconosciuto che diversi tipi di utenti, o lo stesso utente in situazioni differenti, hanno diseguali esigenze informative che possono essere soddisfatte da strumenti diversi. Va anche riconosciuto che, piuttosto che cercare inutilmente di convincere gli utenti a cambiare i propri comportamenti di ricerca e a fare riferimento alla biblioteca per trovare informazioni apprezzabili, è opportuno far sì che queste informazioni siano presenti online nei luoghi in cui gli utenti vivono, giocano, imparano (Coyle 2007 b p. 414).

1.3 La Google generation e i comportamenti di ricerca

Questo nuovo ambiente informativo è definito anche dalla presenza di una particolare componente umana che sembra presentare caratteristiche e attitudini specifiche: i cosiddetti millennial, digital native, net gen o Google generation, persone cioè che sono nate e cresciute con tecnologie quali i computer, la rete internet, i telefoni cellulari, gli MP3 e di conseguenza hanno un approccio peculiare all'istruzione, all'informazione, ai documenti, alle biblioteche diverso da quello delle generazioni precedenti¹⁵. Abram e Luther (2004) ne tracciano le caratteristiche attraverso nove parole chiave, alcune delle quali sembrano particolarmente significative in relazione alle biblioteche e i loro cataloghi.

Agnostici rispetto al formato: non vedono particolare differenza nel formato, stampa o multimediali hanno uguale peso sia nella ricerca di informazioni che di intrattenimento. Le loro opinioni possono essere modificate indifferentemente da libri, giornali, tv, blog, siti Web. In questa prospettiva offrire la possibilità di limitare la ricerca a particolari formati è privo di senso, così come selezionare la fonte prima di iniziare la ricerca. Viceversa vanno incontro a questa caratteristica le tecniche di ricerca federata e strumenti di ricerca che integrino testi, suoni, immagini e video.

Nomadi: questa generazione si aspetta di avere le informazioni o l'intrattenimento che cerca esattamente quando e dove le cerca attraverso wi-fi, telefoni cellulari e palmari. Le

¹⁴ Si vedano sul tema della valutazione le considerazioni presenti in Metitieri-Ridi 2005 p.74-77.

biblioteche devono essere in grado di raggiungere questi utenti nei dispositivi di loro scelta, nei formati e con gli standard che essi adottano e offrire servizi remoti, interattivi e di qualità. Se i servizi delle biblioteche, portali, opac, database, non saranno accessibili da questi dispositivi rischiano di essere del tutto irrilevanti per questa generazione.

Multitasking: una delle abilità di questa generazione è di integrare perfettamente diverse applicazioni e di utilizzarle in modo simultaneo come un unico ambiente. Le biblioteche dovrebbero inserirsi in questo flusso di lavoro e consentire che l'integrazione di applicazioni sia possibile al loro interno senza creare blocchi all'utilizzo di questa o quella tecnologia¹⁶.

Collaborativi: questa generazione è cresciuta dialogando con gli Instant Message, giocando on line, interagendo in classi virtuali e nelle chat room. Ha desiderio e capacità di collaborare.

Integrati: tecnologie e contenuti sono inseparabili, le tecnologie di comunicazione, IM, Web cam, blog, cellulari con macchina fotografica, hanno rotto le barriere tra pubblico e privato e tra ambiente informativo e ricreativo. I bibliotecari devono immergersi in questo ambiente e assumervi un ruolo per esempio con servizi di reference digitale sincroni.

Secondo i due autori questa generazione pone alle biblioteche una sfida maggiore di quella rappresentata da Internet.

Se l'approccio dei nativi digitali sia effettivamente caratteristico è stato esaminato, in relazione alle biblioteche e alle istituzioni educative, in uno studio del 2008 (University college London 2008)¹⁷. Alcune delle affermazioni comunemente accettate su questa generazione si sono dimostrate veritiere, altre frutto di pregiudizi e miti. E' vero che sono più competenti tecnologicamente, anche se usano applicazioni complessivamente semplici, hanno molte aspettative rispetto l'ICT, preferiscono strumenti interattivi, preferiscono informazioni visive piuttosto che testuali, sono una generazione copia-incolla incline al

¹⁵ Per una disamina delle caratteristiche di questa generazione specie nel contesto educativo si veda Prenski (2001 a, 2001 b).

¹⁶ Hanlon e Robertson (2009) rilevano, per esempio, che nelle biblioteche scozzesi la diffusione delle tecnologie 2.0 è ostacolata da firewalls istituzionali e sistemi di filtering (Hanlon e Robertson 2009 p. 2).

¹⁷ Lo studio, promosso dalla British Library e da Jisc, aveva lo scopo di verificare se in ambiente Web i comportamenti di ricerca dei nativi digitali fossero peculiari, se questi comportamenti potranno influenzare il loro stile di ricerca nel momento in cui diventeranno ricercatori maturi nei prossimi anni, se esista oggi un modo specifico di cercare i contenuti diverso da quello normalmente usato dagli studiosi, qual è il futuro delle biblioteche nell'era di internet.

plagio, non rispettano la proprietà intellettuale o meglio capiscono il concetto ma non lo approvano. Non ci sono evidenze che siano multitasking, che essendo abituati all'intrattenimento si aspettino questo tipo di comunicazione nell'ambito dell'istruzione, che valutino i loro pari come fonti più attendibili, che ritengano che tutto sia nel Web, anche se c'è una certa evidenza che non riconoscono le informazioni organizzate dalle biblioteche e non hanno interesse ad usarle, che siano agnostici rispetto al formato, concetto che ha poco senso nel cyberspazio. Non è vero che si aspettino risposte immediate e non sappiano attendere il tempo necessario, che vogliano essere sempre connessi al Web, che acquisiscano le competenze necessarie all'uso dei computer velocemente solo provando ad usarle, che preferiscano informazioni nella forma di sunti già pronti piuttosto che il full text, che siano esperti ricercatori del Web. I risultati complessivi indicano che i nativi digitali, una volta diventati studenti universitari o giovani ricercatori, non hanno un'adeguata mappa mentale di internet, pensano che tutte le risorse online siano gratuite e fornite dai motori di ricerca, utilizzano strategie di ricerca elementari e poco efficaci e sono tuttavia molto soddisfatti dei risultati ottenuti (Metitieri 2009 p. 8). Proprio questa incapacità di cogliere la propria inefficienza nel condurre una ricerca, il gap tra quello che credono di saper fare e le loro competenze reali è il dato che più preoccupa le istituzioni educative e coinvolge la responsabilità delle biblioteche. Quello che emerge più chiaramente è che stiamo tutti diventando una Google generation. Ci sono forti evidenze che l'uso di internet e delle tecnologie 2.0 per una grande varietà di scopi sta rapidamente coinvolgendo persone di ogni età, riducendo le differenze generazionali e assimilando tutti nello stesso ambiente informativo e in modalità di approccio alla ricerca analoghe (University college London 2008 p. 21). La convergenza di comportamenti tra digitali nativi e immigrati era stata rilevata anche da uno studio di OCLC (2007 p. 1.1,8.1).

Non è solo la Google generation a mettere in discussione alcune idee consolidate nel mondo delle biblioteche e degli opac e a spingere nella direzione di approcci innovativi. Studi recenti sui comportamenti di ricerca smentiscono assunzioni che hanno da sempre costituito la base delle idee in quel campo, idee che sono state assimilate ed usate anche nel mondo delle biblioteche e delle istituzioni educative. In un testo comparativo dei diversi modelli di studio e interpretazione dei comportamenti di ricerca Case, reinterpretando e aggiornando al contesto attuale tesi già esposte negli anni '70 da Dervin (1976), sintetizza

in pochi punti alcune idee sui comportamenti di ricerca che sono particolarmente significative anche in relazione alle biblioteche e ai cataloghi (Case 2008 p. 8-9, 326-328):

- Le fonti formali e le ricerche basate su comportamenti razionali, in cui gli utenti prendono informazioni dall'ambiente, valutano le alternative e prendono decisioni ottimali, riflettono solo un aspetto dei comportamenti di ricerca. Molte persone usano fonti informali come gli amici, la famiglia, i colleghi insieme a quanto apprendono dai mass media e dal proprio ambiente. Non è opportuno giudicare irrazionali i comportamenti di ricerca guidati dal principio del minimo sforzo perché hanno una loro inerente efficienza, per esempio nel fatto di non utilizzare tempo in più rispetto a quello necessario per produrre un risultato soddisfacente ancorché non il migliore.

- Non sempre è meglio avere più informazioni. La maggiore preoccupazione degli esseri umani è filtrare, interpretare e comprendere la grande quantità di informazioni con cui si confrontano. E' quindi giustificato che informazioni, anche rilevanti, siano coscientemente filtrate o ignorate quando non c'è tempo o energia per considerarle tutte: in alcuni casi è una strategia razionale nella vita e nel lavoro.

- Il contesto è essenziale nel trasferimento delle informazioni. Le persone si sforzano di avere una visione olistica del proprio mondo e ignorano informazioni isolate che non riescono a inserire nel contesto. Talvolta non mettono in connessione informazioni esterne con la loro realtà interna perché non ne colgono la rilevanza; il loro giudizio sulla situazione influenza i loro bisogni tanto quanto la situazione reale. Quello che dà senso all'informazione non è solo dentro l'informazione ma anche fuori, nelle esperienze dirette e indirette di chi la riceve, nella sua comprensione delle parole e del linguaggio.

- Qualche volta le informazioni non aiutano, specie se si tratta di pacchetti di informazioni generalizzate. L'informazione in sé non soddisfa molti dei bisogni umani: spesso l'informazione non è volta alla comprensione ma a soddisfare bisogni basilari e le persone cercano soluzioni personalizzate piuttosto che risposte standard come quelle che sono in grado di dare un'agenzia o un sistema informativo. Anche quando hanno bisogno di vere informazioni la gran parte di quelle rappresentate da libri, articoli, programmi televisivi non fornisce loro le istruzioni o le risposte di cui hanno bisogno.

- Talvolta non è possibile rendere l'informazione disponibile o accessibile. Le istituzioni e i sistemi formali di informazione spesso non sono in grado di stare al passo dell'eccezionale e imprevedibile richiesta di informazioni che ricevono, tuttavia noi agiamo come se ci aspettassimo che fossero capaci di farlo.

- La ricerca di informazioni è un processo dinamico. I bisogni informativi sorgono rapidamente e possono essere soddisfatti o svanire. La natura delle domande poste può cambiare in vari momenti della ricerca. Soddisfare un bisogno informativo può dare origine ad una nuova domanda. La difficoltà di studiare questi cambiamenti di scenario induce a considerare ogni episodio di ricerca come qualcosa di semplice, lineare e completo, ma la ricerca di informazioni non è né semplice, né lineare, né completa. Per questo le più recenti teorie enfatizzano gli aspetti dinamici e la natura iterativa dei bisogni e della ricerca.

- La ricerca di informazioni non è sempre in relazione a un problema o ad una situazione problematica. Alcuni comportamenti connessi all'informazione non sono "in risposta a" ma hanno una origine genuinamente creativa. La frequenza con cui, nell'ambito degli studi sul comportamento di ricerca, si parla dell'incapacità di spiegare perché la gente agisca come agisce sembra indicare che la ricerca di informazioni sia molto più che risolvere un problema, trovare fatti, prendere una decisione: entra in gioco anche la creatività.

1.4 Next generation opacs

Di fronte alla situazione che abbiamo complessivamente descritto non manca la consapevolezza dei bibliotecari sulla necessità di cambiare gli opac perché possano inserirsi nel nuovo contesto e venire incontro al diverso approccio degli utenti. Nella letteratura e nella blogosfera professionale il dibattito sul futuro del catalogo è molto vivace¹⁸ e si svolge parallelamente ad una riflessione più ampia sul futuro delle biblioteche. Secondo Coyle gli argomenti catalogo, catalogazione e futuro delle biblioteche, non possono andare separatamente perché se le biblioteche cambiano il catalogo deve cambiare e la catalogazione deve assecondare questi cambiamenti (Coyle 2007 a p. 289). Oggi, come nelle riflessioni di Hildreth, si parla di nuova generazione di opac, più precisamente di Next generation o di

¹⁸ Si vedano nel wiki <<http://futurelib.pbworks.com/>> oltre a descrizioni di possibili scenari ed esempi di nuovi opac, i riferimenti ad alcuni blog e liste di discussione su questi argomenti.

cataloghi 2.0 con riferimento alle caratteristiche del Web 2.0 che dovrebbero essere adottate negli opac al fine di renderli adeguati al contesto attuale.

Il concetto di next generation opac non è chiaramente definito: sembra qualificarsi più in opposizione ai vecchi strumenti che in positivo attraverso una precisa definizione delle caratteristiche da sviluppare (Tosca Consultant 2008 p. 7). La direzione verso cui muoversi e i punti di vista su quali siano gli approcci più consoni a realizzare velocemente e fattivamente opac di nuova generazione sono molto diversi a seconda che l'attenzione sia focalizzata sugli utenti, sui dati, sulle funzioni del catalogo, sulle interfacce, se ne discuta in relazione ai cataloghi, più o meno collettivi, di biblioteche di pubblica lettura o di istituti universitari e di ricerca o in un contesto di biblioteca ibrida o di biblioteca digitale. Sul mercato esistono diversi prodotti commerciali e open source in costante evoluzione che hanno apportato sostanziali cambiamenti in uno o più aspetti degli opac tradizionali¹⁹. L'opac ideale sarebbe probabilmente la somma delle funzionalità migliori di ciascuno dei cataloghi esistenti poiché nessuno soddisfa tutte le caratteristiche desiderabili (Wisniewski 2009 p. 56). Alcune idee di base e funzionalità comuni sono tuttavia rintracciabili nei nuovi opac:

- sono software concepiti ad uso degli utenti, sin dal primo approccio non necessitano di alcuna conoscenza biblioteconomica per essere usati
- hanno scelto come principio la ricerca “alla Google” sia per quanto riguarda l'ergonomia della ricerca, con l'adozione di un'unica finestra di ricerca, sia per quanto riguarda la filosofia di base che preferisce il rumore al silenzio del sistema
- selezionano e mostrano i risultati secondo criteri di relevance ranking
- hanno adottato i raggruppamenti a faccette come strumenti per l'analisi dei risultati della ricerca, permettendo agli utenti di escludere risultati non pertinenti e offrendo numerose possibilità di raffinamento; le faccette, che danno contemporaneamente una visione d'insieme dei risultati e la possibilità di vederli secondo diversi punti di

¹⁹ Per una panoramica sui prodotti in commercio si vedano la pagina Web curata da Marshall Breeding <<http://www.librarytechnology.org/discovery.pl?SID=20100426520141601>>, il suo rapporto annuale (Breeding 2009, 2010) o il lavoro di Tosca Consultant orientato sui prodotti adottati in Francia (Tosca Consultant 2008). Per un rapporto completo sui sistemi di gestione delle biblioteche e sugli opac in UK si veda il rapporto JISC – SCONUL (2008).

vista e permettono di evidenziare sotto insiemi di risultati, sono considerate tra i principali miglioramenti conseguiti dai nuovi opac

- utilizzano strategie di rilancio nel caso in cui la ricerca si sia rivelata infruttuosa, proponendo nuovi termini di ricerca generalmente nella forma di nuvola o lista di parole affini per grafia o significato
- propongono risultati arricchiti da diversi tipi di contenuti da diverse fonti
- offrono agli utenti la possibilità di commentare un documento, fare delle note, redigere una critica , attribuire delle parole chiave
- sono indipendenti dal sistema di gestione della biblioteca
- non si limitano alla presentazione dei dati del catalogo, anche arricchiti di contenuti esterni, ma consentono l'accesso ad altri dati che possono essere ricopiati nel proprio DB interno o consultati in linea attraverso la ricerca federata
- sono il punto di accesso a numerosi servizi per l'utente e i dati statistici raccolti possono essere utilizzati per sistemi di raccomandazione (Tosca Consultant 2008 p. 7-11).

Coyle sintetizza le caratteristiche dei nuovi opac nei termini di facile, arricchito, capace di suggerire, interattivo, partecipativo, eterogeneo (Coyle 2007 b p. 415-416). Non manca inoltre di presentare una visione diversa del futuro catalogo:

Starting with the most extreme view of the future of the library catalog, some have a vision of a totally linked information universe where all resources interact. In this vision, libraries would continue to provide access services to quality resources, many of which are not available for open access on the public Web. However, rather than concentrating that access in a separate library system that users must consciously access, the information about the library's materials would be diffused across the information landscape. Users searching in the global pool of information resources would encounter library materials along with materials from other sources including the public sphere of the Web. The information Web is then a universal Memex where all resources are interconnected and all citations hot-link to their do-

documents. The report of the Denmark's Electronic Research Library (DEFF) calls this the "shift from collections to connections" (Coyle 2007 b p. 414)²⁰.

Nei prossimi capitoli vedremo alcuni aspetti dei nuovi opac: esamineremo le possibilità di partecipazione offerta agli utenti, considerando in particolar modo le attività di tagging in relazione alle funzionalità di recupero dei dati; esamineremo il relevance ranking, i raggruppamenti a faccette dei risultati, le nuvole di parole come strumenti per migliorare la presentazione dei risultati e per incrementare le funzionalità di recupero e di esplorazione e prenderemo in considerazione il concetto di Linked Data come possibile futuro del catalogo.

²⁰ Per cominciare con la visione più estrema del futuro del catalogo della biblioteca, alcuni hanno l'immagine di un universo informativo totalmente collegato in cui tutte le risorse interagiscono. In questa visione le biblioteche continueranno a fornire servizi di accesso a risorse di qualità, molte delle quali non sono disponibili ad accesso aperto nel Web. Tuttavia, piuttosto che concentrare quell'accesso in sistemi di biblioteca separati ai quali gli utenti accedono consapevolmente, le informazioni sui materiali della biblioteca saranno diffuse in tutto il panorama informativo. Gli utenti che cerchino nel mare globale delle risorse informative incontreranno i materiali delle biblioteche insieme a materiali di altra fonte compresa la sfera pubblica del Web. Il Web dell'informazione è allora un Memex universale in cui tutte le risorse sono interconnesse e tutte le citazioni sono hot-link ai documenti. Il rapporto delle Denmark's Electronic Research Library (DEFF) chiama questo il "passaggio dalle collezioni alle connessioni".

2. GLI OPAC SOCIALI

Possiamo definire opac sociali o sopac²¹ gli strumenti che consentono quell'interazione tra utenti e catalogo in cui gli utenti sono parte attiva, strumenti in cui gli utenti possono non solo cercare, organizzare e prelevare dati ma anche aggiungere propri contenuti, nella forma di tag, commenti, recensioni, valutazioni delle risorse contribuendo così ad aumentare il valore del catalogo a vantaggio di tutti gli utilizzatori. Così facendo gli utenti diventano anche parte di una rete di relazioni costruita intorno alla biblioteca e ai suoi ambiti di interesse: possono confrontarsi con persone che hanno un profilo simile al proprio e condividere opinioni e informazioni (Marchitelli - Piazzini 2008). La possibilità di svolgere sul Web attività di creazione e di condivisione di contenuti è uno degli aspetti tipici offerti dalle applicazioni tecnologiche più recenti. Gli utilizzatori possono creare una propria identità su MySpace o Facebook, possono votare il gradimento di un libro o scriverne una recensione su Amazon, possono commentare nei blog o scrivere le voci di Wikipedia. Sanno di trovare nella loro navigazione sul Web delle comunità e dei luoghi in cui possono condividere ciò che creano, si aspettano di interagire con le risorse informative e non esserne solo fruitori passivi. Sono contenti di condividere le loro fonti di informazione e combinarle con quelle di altri utenti, come nei siti di socialbookmarking, e vogliono che i siti che usano agevolino l'interazione non solo tra le persone ma anche tra le risorse consentendone l'esportazione e il riuso. I cataloghi delle biblioteche sono stati fino ad oggi privi di queste possibilità: normalmente consentono all'utente identificato di prenotare i libri, suggerire acquisti, controllare i propri prestiti, salvare le ricerche in uno spazio personale, esportare i risultati delle ricerche ma non consentono di aggiungere nulla al catalogo, forti di una tradizione basata sulla precisa distinzione tra specialisti e dilettanti, tra autorevolezza e popolarità. I dati bibliografici sono altamente strutturati e formalizzati, frutto di un lavoro specialistico e non è previsto che persone non esperte possano aggiungere, annotare, commentare, recensire, classificare introducendo nel catalogo elementi di soggettività che possano intaccarne la coerenza complessiva. Benché nessuno metta in discussione la

²¹ Sopac (Social Online Public Access Catalog) è il nome di un prodotto basato sul CMS Drupal che consente una integrazione tra dati del catalogo e dati gestiti da Drupal. Consente agli utenti di aggiungere tag, recensioni, rating delle risorse che vengono incorporati negli indici di ricerca. Si veda <www.thesocialopac.net>.

validità dei cataloghi e dei principi che ne regolano la creazione, principi che si stanno evolvendo nelle sedi istituzionali appropriate con i tempi e la ponderatezza necessaria quando sono coinvolti milioni di dati e migliaia di istituzioni bibliotecarie in ogni parte del mondo²², le riflessioni che abbiamo introdotto sul contesto informativo attuale spingono a cercare rapidamente modi di porsi più vicini allo stile degli utenti, alle esperienze che essi fanno comunemente nel Web. Il Catalog manifesto di Harris (2007) nella sua schematicità può darci una visione di sintesi di questo approccio ma molti altri autori, e tutto il movimento legato al concetto di Library 2.0, sottolineano diversi modi in cui l'opac dovrebbe aprirsi al mondo degli utenti piuttosto che illudersi che gli utenti si adattino agli standard delle biblioteche. Il primo punto del manifesto sostiene che il catalogo non è dei bibliotecari, benché i bibliotecari siano coloro che lo creano e lo mantengono: destinatari del catalogo sono gli utenti²³. Il quarto e ultimo punto del manifesto sostiene che il catalogo è dei lettori i quali, attraverso il catalogo, devono avere la possibilità di mettersi in relazione con altri utenti. Il catalogo deve rendere evidenti queste connessioni, deve imparare dai tag e dalle recensioni degli utenti, condividere con gli utenti un'esperienza di biblioteca sociale favorita dal catalogo. L'adesione a questa "visione sociale" ha portato ad analizzare le potenzialità degli strumenti sociali disponibili nel Web nel contesto dei cataloghi. L'attività di tagging è stata particolarmente studiata, non solo come fattore di coinvolgimento e socializzazione ma in relazione alle sue possibilità di impiego nel recupero delle informazioni.

2.1 Tagging e recupero dell'informazione semantica

L'attività di tagging consiste nell'associare ad una risorsa informativa sul Web (una pagina Web, un video, una foto, un record bibliografico ecc. qualsiasi cosa con un URL) dei termini che la descrivano, consentano la sua classificazione per parole chiave e ne favorisca-

²² Dopo l'elaborazione di FRBR, dal 2003 l'IFLA ha promosso la creazione di un codice internazionale di catalogazione, che sostituisce i Principi di Parigi del 1961, attraverso una serie di meeting di esperti internazionali. I nuovi principi sono stati pubblicati nel febbraio 2009 <<http://www.ifla.org/en/publications/statement-of-international-cataloguing-principles>>. Anche singoli stati stanno revisionando i proprio codici. Si veda per l'Italia le Reicat pubblicate nel 2009, per gli Stati Uniti d'America RDA che dovrebbero essere definitivamente rilasciate nel 2010 <<http://rdaonline.org/>>.

²³ Mi sembra tuttavia più corretta l'impostazione di OCLC che nella sua ricerca del 2009 "*Online catalogs: What users and librarians want*" considera anche i bibliotecari un gruppo specifico di utenti e ne esamina le aspettative (OCLC 2009).

no il recupero. Quando questa attività avviene in un ambiente sociale condiviso in cui gli individui e i gruppi creano, organizzano e condividono i tag, si parla di social tagging, tagging collaborativo o di classificazioni sociali. L'insieme dei concetti rappresentato dai nomi e dai termini usati come tag che sono risultato del tagging collaborativo viene definito *folksosonomy*²⁴ (Mendes – Quinonez - Skaggs 2009 p. 32). Il tagging si è sviluppato nei software di link manager, per esempio Delicious, come mezzo per organizzare il numero crescente di bookmark personali²⁵. I browser hanno sempre consentito la registrazione degli indirizzi delle risorse in cartelle organizzate gerarchicamente ma la crescita del Web ha mostrato i limiti di questo sistema e favorito la nascita di siti che consentono la memorizzazione degli URL preferiti associati a tag che ne facilitino il recupero. Secondo Sinha, questo sistema funziona bene perché non richiede un grande dispendio cognitivo: non impone di collocare la risorsa all'interno di un sistema di classificazione ed è sufficiente che le persone ricordino i termini che hanno assegnato (Sinha 2005). Le persone che usano i tag non devono essere esperti, imparare complesse regole e termini appropriati come accade quando si classifica una risorsa con un vocabolario controllato (Mathes 2004). Secondo Vander Wal le persone non sono interessate a classificare quanto a dare un significato nel proprio contesto, a creare un "gancio" che consentirà loro di recuperare le risorse nel tempo (Vander Wal 2007). Questa facilità e libertà d'uso spiegano perché i tag sono così popolari. Lo scopo dei tag è principalmente individuale: chi li usa lo fa in primo luogo per se stesso. I software sociali, tuttavia, poiché rendono visibili a tutti i tag usati, consentono il reciproco confronto delle scelte: le persone sono influenzate dai tag altrui e sono rinforzate nelle loro scelte se altri hanno usato i loro stessi tag. In breve si crea un vocabolario comune che non è preordinato ma nasce spontaneamente dall'uso all'interno della comunità. Man mano che cresce il numero degli utilizzatori ciascuna risorsa sviluppa una tag cloud, cioè un raggruppamento di tag popolari. La tag cloud è una rappresentazione visiva dei tag attribuiti ad una risorsa, nella quale i termini sono normalmente rappresentati con dimensioni dei caratteri proporzionali al loro peso mettendo così in risalto quelli più usati. Basta un colpo d'occhio alla tag cloud per rendersi conto di quali sono i tag più popolari in un particolare contesto.

²⁴ Altri termini usati sono *folk classification*, *ethnoclassification*, *distributed classification* o *social classification* (Hammond-Hannay-Lund-Scott 2005).

Diversi studi sulle caratteristiche dei tag (Goldberg - Huberman 2006, Kipp - Kampbell 2006, Mathes 2004) dimostrano che i tag più popolari attribuiti ad una risorsa costituiscono un'efficace rappresentazione della risorsa stessa: i tag descrittivi sono scelti in proporzione più grande rispetto a tag di altro genere e a quelli di tipo personale. Inoltre ogni risorsa ha alcuni tag scelti dalla maggioranza seguiti da una "lunga coda" di tag usati da pochi o singoli individui secondo una power law distributions. L'insieme dei tag assegnati da tutti gli utenti segue un modello stabile nel quale la frequenza di ciascun tipo di tag è in proporzione quasi fissa rispetto alla frequenza totale di tutti i tag usati. Questo si rileva per tutte le risorse, non solo per quelle più popolari, e si stabilizza a meno di 100 bookmark assegnati per ogni risorsa. La stabilità del modello fa sì che questi metadati²⁶ siano considerati utili per il recupero delle informazioni a livello generale: l'attività di tagging svolta per scopi individuali, collettivamente crea un insieme di descrittori per soggetto della risorsa che costituiscono uno strumento di accesso alla risorsa non solo a livello personale ma a livello generale (Thomas – Caudle – Schmitz 2009). Data questa caratteristica è naturale che le biblioteche siano interessate a capire come tag e folksonomie possano essere utilizzati nell'ambito del catalogo. Come abbiamo già accennato, l'area della ricerca semantica è sempre stata tra le più problematiche negli opac: gli utenti sono in difficoltà nel recuperare informazioni attraverso i termini di soggetto e le classificazioni perché, non avendo alcun riferimento conosciuto da cui iniziare la ricerca, come il titolo o l'autore, devono essere in grado di individuare i termini che il catalogatore ha utilizzato per descrivere il contenuto del documento. Questi termini appartengono ad un vocabolario controllato attinto da espressioni precise e specifiche delle diverse discipline cui il soggetto fa riferimento piuttosto che dal linguaggio comune. E' quindi abbastanza frequente che l'utente immetta un termine che non è stato usato dal catalogatore e ottenga una risposta nulla, che viene interpretata come "non esistono documenti su questo argomento" mentre potrebbe semplicemente significare "non hai usato il termine usato dal catalogatore". Se i problemi legati al linguaggio ostacolano le funzioni di searching, le interfacce fin qui realizzate negli opac non sono state in grado di favorire neanche le funzioni di browsing dei soggetti, di mostrare le relazioni tra i concetti o le strutture dei tesauri, che pure sono presenti nei data base,

²⁵ Per dettagli sulla nascita dei link manager e dei bookmark si veda Hammond-Hannay-Lund-Scott [2005].

²⁶ Per una definizione di metadati si vedano Ridi (2010 p. 15-22), Coyle (2010 a p. 6-10).

per favorirne l'esplorazione²⁷. Le folksonomie potrebbero aiutare a ridurre questa distanza tra il linguaggio dei catalogatori e degli utenti e la rappresentazione visuale dei tag potrebbe favorire le attività di browsing del catalogo. Non mancano tuttavia difficoltà all'adozione di questo approccio. Le critiche reciproche tra tag e soggetti strutturati, tra classificazioni e folksonomie sono numerose e note. Il nome stesso "folksonomia" suscita discussioni. E' un neologismo creato da Thomas Vander Wal nel 2004 (Vander Wal 2007) componendo il termine folk (gente) e tassonomia per indicare una classificazione creata dalla gente, dagli utilizzatori che attribuiscono i tag alle risorse. Le risorse non sono quindi classificate a priori ma la classificazione emerge spontaneamente dalle aggregazioni che si creano dall'attività dagli utenti. Secondo Santoro il concetto di tassonomia implica non solo la capacità di attribuire descrittori del contenuto ma quella "... di dar vita a distribuzioni categoriali di concetti". Implica sempre un'idea di distinzione, di ripartizione, di ordinamento in categorie e non una mera enumerazione di oggetti e fenomeni: per questo le folksonomie non dovrebbero essere considerate delle tassonomie e il nome è fuorviante.

Tuttavia il processo di classificazione non si limita a raggruppare gli elementi simili e a separarli da quelli dissimili ma, all'interno dei singoli gruppi, permette di effettuare ulteriori ripartizioni, fino ad arrivare a elementi non più divisibili; si creano in tal modo delle "sotto-classi", all'interno delle quali si ritrovano caratteristiche che identificano in maniera via via più precisa i propri membri. L'atto del classificare allora non produce solo raggruppamenti più o meno omogenei di oggetti o di idee, ma una serie di suddivisioni (solitamente di tipo gerarchico) in base al grado di somiglianza o di affinità dei suoi componenti. Nulla di tutto questo avviene nel mondo delle folksonomies, dal momento che i tags vengono assegnati in maniera piatta e orizzontale, senza tener conto delle relazioni - gerarchiche o di altra natura - che intercorrono fra i concetti; ciò non consente di far emergere delle classi, e quindi non giustifica l'idea che, per quanto fatte dalla gente, si tratta comunque di tassonomie (Santoro 2007 p. 2).

Un primo tema di confronto tra le folksonomie e le classificazioni è relativo alla capacità di rappresentare l'evoluzione delle conoscenze. Secondo Quintarelli gli schemi gerarchici sono più efficaci quando sono applicati ad ambiti disciplinari ristretti, poco interdisciplinari e non sottoposti a rapidi cambiamenti. Hanno infatti una certa rigidità che impedisce un veloce adeguamento ai nuovi contenuti. I professionisti che le usano hanno un approccio conservativo: prima di accogliere nuovi termini attendono che questi si siano consolidati

²⁷ Vanno oggi in questa direzione alcuni software come AquaBrowser (www.aquabrowser.com), il navigatore Dewey del catalogo BNCF (<http://opac.bncf.firenze.sbn.it/opac/help.jsp#navigatoredewey>) o la nuova versione beta del navigatore Dewey di OCLC (<http://deweybrowser.oclc.org/ddcbrowser2/>).

nella disciplina o abbiano sufficiente garanzia bibliografica²⁸. Inoltre le classificazioni riflettono i condizionamenti culturali dell'ambiente e del tempo in cui sono create (Quintarelli 2005). Le folksonomie invece rappresentano le prospettive correnti e sono in grado di accogliere subito i nuovi termini che le rappresentano perché non devono relazionarli a strutture preesistenti. Inoltre la loro flessibilità è in grado di soddisfare persone con background culturali e sociali diversi come è necessario nel Web. Tutti questi aspetti sono particolarmente significativi per gli opac che sono sistemi destinati ad una grande varietà di utenti non identificabili e che sono creati per incremento successivo della stessa base dati nel corso di un lungo periodo di tempo. L'adozione di folksonomie accanto alle classificazioni strutturate potrebbe mantenere la dinamicità del linguaggio e dei concetti. Campbell e Fast sostengono che una forza destabilizzante, come l'uso dei tag, incrementando la diversità e la flessibilità, può promuovere la sopravvivenza a lungo termine dell'ecosistema informativo (nel nostro caso la biblioteca, il catalogo) che viceversa potrebbe essere distrutto se si irrigidisse nelle sue strutture tradizionali (Campbell - Fast 2006).

Un secondo aspetto di confronto tra folksonomie e classificazioni è relativo ai costi. L'attività di creazione, mantenimento e sviluppo di vocabolari controllati è molto costosa perché implica un addestramento di alto livello professionale per i classificatori e un discreto investimento di tempo da parte degli utenti prima di essere in grado di padroneggiarne l'uso. In un ambiente informativo come l'attuale in cui le risorse digitali disponibili crescono esponenzialmente sembra essere impossibile applicare le classificazioni tradizionali a tutto. Le folksonomie potrebbe supplire a costi contenuti sfruttando l'attività degli utenti. Le folksonomie sarebbero dunque "meglio di niente" in situazioni come l'attuale (Sirky 2005 b). Inoltre risponderebbero meglio alle aspettative degli utenti in quanto semplicità e velocità nell'impiego costituiscono un risparmio di risorse dal punto di vista dell'utilizzatore. Il tema dei costi della catalogazione è molto sentito in ambito bibliotecario. Nel rapporto preparato per la Library of Congress, *The Changing nature of the catalog and its integration with other discovery tools* (Calhoun 2006), Karen Calhoun prospetta l'abbandono dell'indicizzazione semantica per il suo alto costo e la scarsa rilevanza che

²⁸ Il termine AIDS è stato registrato nell'ICD International classification of Diseases nel 1986, tre anni dopo la pubblicazione del primo articolo sull'argomento; il termine Hapartheid fu introdotto nel 1973 da un catalogatore della Hannepin Country Library nel suo catalogo locale ma fu adottato dal Library of congress Subject headings solo nel 1986, ben tredici anni dopo (Smith 2008 p. 64, 87).

avrebbe nell'attuale contesto informativo. Questa posizione è stata oggetto di forti critiche (Mann 2006), non è condivisa da tutti ma è significativa di una tendenza in atto anche nelle grandi istituzioni e spinge alla ricerca di alternative vuoi nel campo delle folksonomie vuoi in quello dell'indicizzazione automatica²⁹.

Un terzo aspetto, forse il principale tema di confronto, è il vocabolario usato in relazione alla capacità di recupero delle risorse. Nelle folksonomie, al contrario che nelle classificazioni, i termini usati sono liberamente scelti dagli utilizzatori senza alcun riferimento a schemi o vocabolari. Purtroppo i vocabolari non controllati hanno problemi di ambiguità e polisemia, controllo dei sinonimi, diverso livello di specificità, variazioni e anomalie lessicali nelle forme dei tag.

Ambiguità e polisemia: gli utenti usano lo stesso tag per identificare cose differenti perché lo usano in contesti diversi, applicato a risorse diverse. Gli acronimi sono un esempio di questo problema (la sigla AIB ha diversi significati nel campo bancario, del business, delle biblioteche). La polisemia è un tipo particolare di ambiguità (il termine porto può indicare un tipo di vino o un'infrastruttura)

Sinonimia: per definire lo stesso oggetto vengono usati nomi diversi (mac, apple, macintosh)

Diverso livello di specificità: non ci sono regole per indicare a quale livello di specificità attribuire i tag. Si può attribuire perl e javascript oppure programmi. Chi cercasse programmi non troverebbe mai perl e javascript poiché non c'è possibilità di indicare relazioni tra i termini e neanche l'articolazione gerarchica dei concetti.

Variazioni lessicali: singolare verso plurale, variazioni ortografiche, forme verbali, diverso trattamento delle parole composte, diverse convenzioni di punteggiatura, ordine delle parole.

Tutte queste problematiche sono invece gestite tanto nelle classificazioni basate su tesauri, che esplicitamente indicano le relazioni gerarchiche, di affinità e di equivalenza, quanto nei sistemi di indicizzazione basati su soggetti, attraverso le relazioni sindetiche tra i termini. L'indicizzazione attraverso questi strumenti consente di disambiguare i concetti secondo il contesto di riferimento, di riconoscere i collegamenti tra concetti, di eliminare i problemi di sinonimia, omografia, omonimia. Questo favorirebbe un miglior recupero delle risorse.

²⁹ Per qualche esempio di indicizzazione automatica nell'ambito delle biblioteche si veda Coyle (2008 a).

Secondo Mathes questo è vero nel caso di interrogazioni di tipo search, cioè nel contesto di una ricerca mirata, mentre non è vero nel caso delle attività di browsing che sono invece favorite dai tag (Mathes 2004). Secondo Shirky i vocabolari controllati sono meno efficienti nel recupero nel caso di sistemi ampi e caotici come il Web nel cui contesto è impossibile per gli esperti predire come verrà effettuata una ricerca ed è impossibile mantenere stabile nel tempo una classificazione (Shirky 2005 a, 2005 b). Probabilmente queste osservazioni sono altrettanto vere nel caso di grandi cataloghi collettivi online. Diversi autori propongono un'integrazione tra i due sistemi: le folksonomie potrebbero aggiungere metadati per aspetti non coperti dalle tradizionali classificazioni, come avviene per le immagini, senza che questo significhi abbandonare l'indicizzazione tradizionale (Matusiak 2006 p. 295 citato da Thomas – Caudle – Schmitz 2009). Questo viene già attuato con risultati soddisfacenti in diverse istituzioni tra cui il Metropolitan Museum of Modern Art, dove si è verificato che il 92 % dei tag inseriti dagli utenti nelle raccolte digitali aggiunge nuove informazioni che non erano presenti nelle fonti descrittive tradizionali (Datema 2007) in particolare consentendo l'espressione dei concetti in lingue diverse e dettagliando molto la descrizione³⁰.

Nella prospettiva di integrare o affiancare i due sistemi all'interno degli opac sono stati effettuati diversi studi proprio sul confronto del vocabolario di tag e classificazioni. Spiteri (Spiteri 2007) ha analizzato una serie di tag usati in Delicious, Furl, Technorati per vedere in che misura corrispondessero alle linee guida del NISO per la costruzione di tesauri che sono le regole oggi più accreditate per la creazione di vocabolari controllati (NISO 2005). E' evidente che non si possono applicare ai tag le regole per stabilire relazioni tra i termini ma le indicazioni sulle caratteristiche dei singoli termini da usare in un vocabolario controllato possono essere rilevanti anche nella creazione dei tag. Il risultato della ricerca conferma che nella struttura delle folksonomie ambiguità e polisemia costituiscono un problema anche se omografi e ambigui sono meno di un quarto dei tag usati: la maggioranza dei tag usati non sono ambigui e sono perciò conformi alle raccomandazioni NISO. Per quanto riguarda i concetti espressi dai termini (principalmente nomi, pochi avverbi e agget-

³⁰ La possibilità di inserire tag avviene grazie all'interfaccia del progetto Steve: The Museum Social Tagging Project <<http://steve.museum/>>. Questa iniziativa, nata nel 2005, ha lo scopo di studiare come il social tagging possa essere usato in ambito museale per incrementare l'accesso alle collezioni digitali superando le barriere semantiche e linguistiche tra indicizzatori professionisti e visitatori.

tivi), corrispondono alle raccomandazioni NISO e anche rispetto all'ortografia la maggior parte dei termini risulta essere corretta. Ci sono inoltre pochi termini gergali, così come previsto dalle raccomandazioni. Spiteri suggerisce che volendo adottare le folksonomie all'interno dei cataloghi dovrebbero essere date chiare istruzioni sulla scelta e la forma dei tag nel caso di nomi numerabili e non numerabili (specificando come l'uso di singolare e plurale influenzi il recupero dei dati) e sul modo di costruire tag multitermini poiché questi sono gli aspetti che si sono rivelati più critici. Suggerisce anche che il sistema potrebbe prevedere un link a un dizionario o ad una enciclopedia online per consentire agli utenti di determinare il significato dei termini, disambiguare gli omografi, scegliere la forma corretta da adottare nel caso di sigle (Spiteri 2007).

Uno studio più recente ha analizzato in che misura le folksonomie corrispondano a LCSH (Library of Congress Subject Heading) per verificare se potrebbero effettivamente affiancarsi al catalogo e aumentarne le capacità di recupero come suggerito da Spiteri (Thomas – Caudle – Schmitz 2009). Lo studio confronta tag e LCSH assegnati a 10 libri in 5 biblioteche e in Libray Thing. I libri scelti sono molto popolari e rappresentano una varietà di aree nelle quali LCSH mostra una certa debolezza. Le biblioteche sono state scelte perché utilizzano tutte LCSH e ciascuna usa un diverso applicativo che consente l'utilizzo di tag³¹. Library thing, un famoso sito di social cataloguing, è stato scelto per la sua ricchezza di tag. Di tutti i tag assegnati ai libri sono stati considerati e confrontati con LCSH solo quelli che rappresentano il soggetto del libro. Secondo Golder e Huberman (2006), ripresi anche da Smith (2008) i tag rappresentano sette diverse funzioni:

- descrivere: identificare di chi o di che cosa si tratta
- identificare il tipo di risorsa: libro, video, foto ecc.
- identificare chi possiede la risorsa: nome della persona ma anche di una biblioteca o di una collezione
- raffinare: tag che non vengono usati da soli ma per modificare altri tag
- identificare qualità o caratteristiche: tag che esprimono opinioni (come “divertente”)

³¹ Le biblioteche e gli applicativi scelti sono stati: University of Pennsylvania Libraries - Penn Tags; Queens Public Library or Oklahoma State Library - AquaBrowserW; Ann Arbor District Library - SOPAC; University of Michigan - MTagger; Birmingham Public Library - Encore.

- self reference: tag che hanno un significato per la persona che li usa (come “mio”, “personale”, “desiderata”)
- organizzare attività: esprimono attività da fare in relazione alla risorsa (come “da leggere”, “da comprare”).

Solo i tag delle prime due categorie sono stati considerati pertinenti all’area semantica e utilizzati nello studio confrontandone il grado di sovrapposizione con i soggetti di LCSH su una scala graduata (1 corrispondenza esatta, 2 sinonimo in LCSH, 3 sinonimo nel linguaggio naturale, 4 termine più ampio del soggetto LCSH assegnato al libro, 5 termine più ristretto, 6 termine correlato, 7 termine corretto in LCSH ma non assegnato, 8 termine correlato ma non previsto in LCSH, 9 termine senza alcuna relazione o troppo vago). I tag assegnati nella categoria 1-2 rappresentano la completa sovrapposizione tra linguaggio dei tag e di LCSH. I tag nelle categorie 3-8 rappresentano il linguaggio degli utenti. Lo studio assume che una grande quantità di tag nelle categorie 3-8 costituisca un ampliamento dei soggetti in aree e tipi di materiali non ben coperti da LCSH.

Dei 10 libri esaminati 4 non raggiungono i 100 tag che sono considerati il limite minimo per considerare stabile la classificazione attribuita ma gli altri 6 hanno un numero di tag così elevato che la classificazione può essere giudicata stabile (Golder – Huberman 2006 p. 205). L’analisi dei dati dimostra che il 46% dei tag ricade nelle prime due categorie di Golder e Huberman mentre nelle categorie 5-7, che rappresentano tag significativi solo per chi li ha usati, ricade il 40% dei tag. Si conferma quindi l’assunzione generale che gli utenti usano i tag prima per se stessi e solo in secondo luogo per gli altri ma si rileva tuttavia un’equa divisione tra tag di soggetto e tag personali. Rispetto al grado di sovrapposizione con LCSH emerge che solo il 14% dei tag ha una corrispondenza esatta con i soggetti LCSH o con sinonimi predefiniti, rientra cioè nei gruppi 1 e 2, mentre il 62% rientra nelle categorie da 3 a 8 e costituisce quindi, secondo l’ipotesi della ricerca, un arricchimento di nuovi metadati. In particolare il 35% dato dalla somma delle categorie 3 e 8 rappresenta tag che ampliano il vocabolario presente in LCSH.

Vediamo qualche dettaglio:

i gruppi 1-2, identico e sinonimo, costituiscono rispettivamente 6 e 8% dei tag;

il gruppo 3, sinonimi non previsti da LCSH, rappresenta il 13% dei tag;

i gruppi 4, 5 e 6 che esprimono relazioni BT, NT, RT con un termine LCSH, rappresentano solo l'1% com'è naturale in un sistema in cui spesso non c'è il termine base a cui fare riferimento per indicare concetti più larghi o più ristretti;

il gruppo 7, termini validi in LCSH ma non assegnati a quel libro, rappresenta un significativo 26 % che include termini assegnati a un diverso livello di specificità ma senza relazioni reciproche. Una percentuale così alta di tag che sono anche soggetti LCSH indica che chi usa i tag ha una certa familiarità con i vocabolari controllati; indica anche che il libro è coperto in tutti i suoi argomenti, anche quelli secondari che sono normalmente trascurati in una catalogazione standard che considera solo l'argomento principale;

il gruppo 8, termine correlato ma non previsto in LCSH, rappresenta il 22% dei tag usati.

Un terzo studio ha comparato i soggetti LCSH assegnati a 155 libri con i tag assegnati a questi stessi libri in Library Thing e in Amazon per vedere le intersezioni dei termini e quanto l'integrazione dei due sistemi potrebbe essere utile per le biblioteche e gli utenti (Lawson 2009). I record bibliografici sono stati tratti dal catalogo di OCLC scegliendo i 5 titoli più posseduti delle biblioteche di ciascuna delle 31 principali suddivisioni per argomento filtrate per data (2007), lingua (inglese), formato (libro), pubblico (adulti), genere (fiction e non fiction). I tag sono stati distinti in Oggettivi, che descrivono il contenuto del libro, e Soggettivi, che esprimono punti di vista dell'utente, personalizzazioni o altro non legato al contenuto del libro. Lo studio rileva che ogni libro ha in media 3 soggetti mentre ha 63 tag. In media il 51% dei tag sono oggettivi e il 49% sono soggettivi. Registra tuttavia una notevole differenza di queste percentuali a seconda dei campi disciplinari cui il libro appartiene: i documenti governativi, i libri di informatica o di biologia hanno tag in quantità inferiore ma in grande prevalenza oggettivi (rispettivamente 90%, 80%, 68%) mentre i libri di letteratura hanno molti tag e la percentuale di tag oggettivi si abbassa al 41%. I libri delle discipline come l'educazione fisica hanno sia pochi tag sia pochissimi tag di tipo oggettivo (19%) mentre quelli di storia hanno moltissimi tag e una percentuale del 60% di tag oggettivi. Lo studio non trae da questi elementi conclusioni pratiche al di là di una generica utilità nell'utilizzare i tag come complemento ai soggetti specie avendo lo studio verificato che molti dei tag oggettivi corrispondono a termini usati in LCSH. Sarebbe tuttavia necessario tener conto di questi aspetti in relazione alle problematiche della copertura del catalogo cioè alla necessità che tutte le discipline e tutte le risorse siano adeguatamente indicizzate non solo quelle più alla moda o quelle più comuni. La consapevolezza

delle asimmetrie che si creano nel catalogo a causa dell'attività di tagging dovrebbe indurre a studiare forme di bilanciamento per garantire l'imparzialità del catalogo.

Un diverso studio (Eckert – Hanger - Niemann 2009) ha testato attraverso un apposito software i risultati di una indicizzazione fatta da esperti, da utenti non esperti attraverso l'attribuzione di tag e da procedure automatiche nel contesto di una biblioteca accademica e su documenti digitali. Sono stati analizzati 372 articoli di argomento economico ai quali sono stati attribuiti solo termini derivati dal German Standard Thesaurus for Economics (STW). I bibliotecari hanno assegnato 1547 termini validi secondo il tesoro. Le procedure automatiche hanno utilizzato il motore di ricerca Collexis che ha generato 4135 parole chiave estraendole dagli abstract degli articoli. Una quantità così grande di parole chiave ha fatto subito pensare ad una migliore copertura di aree non coperte dall'indicizzazione umana. Si sono tuttavia rilevate annotazioni errate per termini non disambiguati e difficoltà dovute alla differenza tra lingua del tesoro, il tedesco, e lingua delle risorse, l'inglese. Da notare anche, rispetto all'indicizzazione fatta da esperti, la mancanza di designazioni geografiche dovuta al fatto che i termini corrispondenti non erano esplicitamente presenti negli abstract ma avrebbero dovuto essere dedotti dal contesto. Manca anche, nell'indicizzazione automatica, l'inclusione di termini che nascono dall'associazione di concetti o dal riferimento a concetti o teorie più generali. I tag generati dagli utenti (in questo esperimento rappresentati da un unico studente di Library and Information Science senza preparazione specifica nell'indicizzazione) sono stati 579. Si può osservare che mentre i bibliotecari hanno assegnato termini più specialistici, se disponibili, l'utente ha assegnato termini più generali. Non ci sono in questo caso errori dovuti a incomprensione di termini ambigui nel significato, ma è interessante notare che, anche in questo caso, sono stati assegnati solo termini esplicitamente presenti nei testi. La mancanza di addestramento fa sì che l'utente non sia in grado, o non sia motivato, a leggere tra righe, ad inferire concetti al di là dei termini presenti. Va precisato che i sistemi di tagging basano il loro successo su quella che viene definita "wisdom of the crowds", l'intelligenza collettiva di un insieme di utenti e quindi si presume che le carenze rilevate nello studio possano nella realtà essere mitigate. La conclusione generale dello studio è che la combinazione di indicizzazione professionale e automatica e di tag generati dagli utenti, può garantire un miglior recupero delle risorse senza compromettere la qualità della catalogazione. Nonostante la loro inferiore qualità, automatismi e tag possono migliorare l'esperienza di ricerca quando non sia possibile

un'adeguata soggettazione, per esempio perché l'articolo è appena stato pubblicato, o per il diverso livello di granularità (se la biblioteca non cataloga a livello del singolo articolo) (Eckert – Hanger - Niemann 2009 p. 568).

2.2 Massa critica e volontà di partecipazione

Nel valutare l'utilità di introdurre sistemi di tagging negli opac vanno prese in considerazione anche altre due problematiche tra loro correlate: la questione della massa critica e la volontà degli utenti di aggiungere tag in quel contesto. Come abbiamo visto, affinché da un insieme di tag emerga una descrizione appropriata della risorsa che possa essere valida per tutti e non solo per chi ha attribuito specifici tag, è necessario che sia attribuita una elevata quantità di tag poiché pochi tag, utili a livello personale, costituiscono solo rumore in un contesto di utilizzazione collettiva. Raggiungere la massa critica è dunque fondamentale. Secondo Spalding poiché gli utenti usano i tag per se stessi, per organizzare le proprie risorse, non c'è alcun interesse da parte degli utenti ad attribuire tag a risorse altrui come avviene nei siti commerciali, a meno che non sia loro offerto un incentivo convincente. Per questo l'attività di tagging nei siti commerciali sarebbe fallimentare (Spalding 2007). Una situazione analoga a quella dei siti commerciali si presenta negli opac: infatti, nonostante alcuni opac abbiano introdotto la possibilità di gestire contenuti generati dagli utenti, sembra che questa possibilità sia al momento sotto utilizzata e abbia portato pochi benefici ai cataloghi (Wenzler 2007 p. 1). Forse gli utenti non hanno trovato un motivo di interesse personale ad attribuire tag nel catalogo di una biblioteca e non sono disposti a lavorare per un risultato collettivo³². Secondo Smith dovrebbe sempre essere esplicitato chiaramente all'utente quali sono i vantaggi che potrebbe avere nello svolgere attività di tagging all'interno di un certo sistema ma non va dimenticato che il sistema è organizzato per propri scopi specifici, stabiliti da chi lo gestisce: dovrebbe dunque esserci un bilanciamento tra interesse degli utenti e scopi del sistema (Smith 2008 p. 22). Nel contesto dei cataloghi delle biblioteche un sistema di tagging ha lo scopo principale di ottenere metadati a basso costo grazie ai quali migliorare le capacità di recupero delle risorse. Anche altri scopi possono essere ricercati: aumentare la partecipazione con un'attività che per l'utente

³² Per un'analisi complessiva delle motivazioni alla collaborazione nei social network si veda Shirky (Shirky 2008).

è a basso rischio e comporta poca esposizione; identificare modelli di comportamento cioè inferire dall'attività di tagging il gradimento delle risorse, come vengono utilizzate, chi le utilizza. Quali potrebbero invece essere i vantaggi personali per gli utenti? Smith individua tra gli scopi degli utenti la possibilità di gestire informazioni personali, desiderio di collaborare e condividere senza tuttavia interagire direttamente con altri utenti, divertirsi, esprimere opinioni (Smith 2008 p. 23-28). Il sistema dovrebbe proporsi agli utenti in modo da soddisfare almeno uno di questi scopi che potrebbero essere diversificati a seconda del contesto, biblioteca pubblica, biblioteca universitaria o altro. In una biblioteca pubblica potrebbe essere sufficiente favorire gli aspetti ludici o di espressione della personalità mentre in una biblioteca universitaria la possibilità di creare uno spazio personale nel quale salvare propri bookmark delle risorse del sistema integrati con bookmark di risorse esterne potrebbe essere indispensabile per convincere gli utenti a svolgere attività di tagging. Nel contesto di biblioteche di ricerca il desiderio di creare una comunità di utenti potrebbe essere una motivazione sufficiente. Non è facile tuttavia indicare quali attese devono essere soddisfatte per quali comunità di utenti³³. Ciascuna di queste opzioni deve essere sostenuta da precise scelte nella gestione dei dati e delle interfacce: deve essere chiaro come si diventa membri del sistema, come si può lasciare il sistema senza perdere le attività che vi si sono svolte, quali sono le regole di privacy, cioè se i tag sono pubblici o privati, se i tag sono liberi o vengono proposte delle opzioni predefinite e così via. In ogni caso la regola base è che l'attività di tagging deve essere percepita come semplice, non richiedere grande investimento di tempo e di energia altrimenti è inevitabile che il sistema resti poco popolare e non riesca a raggiungere una quantità di dati significativa per i suoi scopi.

Alcune ricerche hanno cercato di indagare quale sia l'atteggiamento degli utenti nei confronti dei software sociali applicati al mondo delle biblioteche. Pur valutando corretta la posizione di Metitieri (2009 p. 8) secondo il quale la rapida crescita dei siti di social network rende difficile trarre conclusioni da qualsiasi ricerca, non si può eludere la conclusione cui perviene la ricerca di OCLC "Sharing, Privacy and Trust in Our Networked

³³ Il concetto di comunità di utenti è stato indagato in uno studio anche in relazione al problema di come la comunità influenzi il comportamento di tagging e di come questo abbia effetto sulle potenzialità di recupero all'interno del sistema (Collaborative 2008).

World” (OCLC 2007)³⁴ che inizia il capitolo dedicato ai rapporti tra biblioteche e social networking con un’affermazione lapidaria:

The general public respondents surveyed do not currently see a role for libraries in their new social networked world. The general public respondents do not see a role for libraries in constructing social sites, and most would not be very likely to contribute content, self-publish or join discussion groups if a library were to offer these services. Interest in participating in activities on a library-hosted social networking site was low among respondents. Most activities evaluated garnered the interest of less than 10% of the total general public who indicated they would be either extremely or very likely to participate. This general level of disinterest held very constant across countries surveyed (p. 5.1)³⁵.

Sembra dunque che gli sforzi che le biblioteche stanno facendo per inseguire l’approccio social non trovino corrispondenza negli utenti. Non sembra che gli utenti abbiano desiderio di contribuire con propri contenuti sui siti delle biblioteche o di creare delle comunità intorno alle stesse. Nello studio le due domande proposte agli utenti che più si avvicinano al nostro tema riguardano l’interesse degli utenti a descrivere le proprie raccolte personali e a vedere quelle altrui su strumenti predisposti dalle biblioteche: il primo caso raccoglie il 6% di consensi, il secondo il 7%.

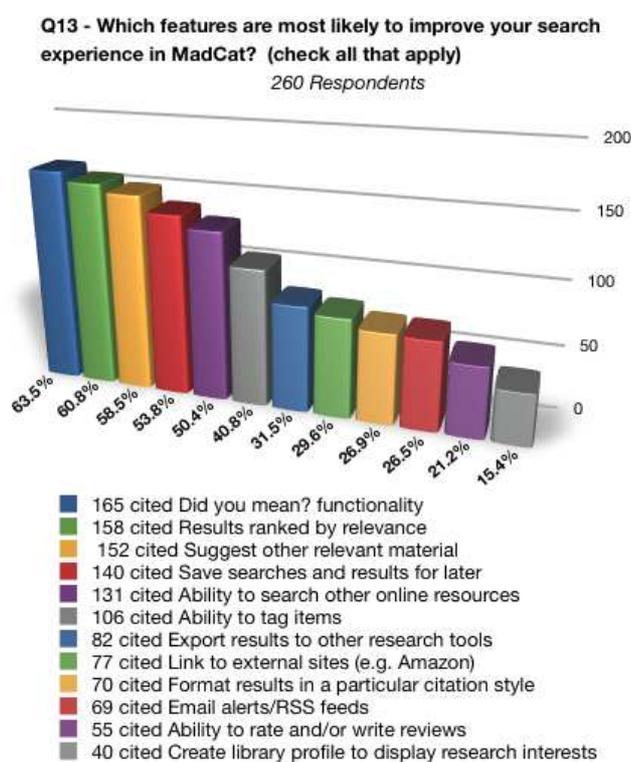
Studi più specifici, benché più limitati nel campione indagato, sono stati svolti da diverse istituzioni per capire le preferenze degli utenti in vista dell’adozione di nuovi opac o per testare prodotti in corso di sviluppo. Queste ricerche danno risultati tra loro non comparabili per la diversità dei contesti, degli scopi, della terminologia usata ma possono comunque dare un’idea delle valutazioni degli utenti riguardo all’uso dei tag nel contesto della biblioteche. In alcuni casi sembrano confermare lo studio di OCLC, in altri sembrano andare in direzione opposta.

Le biblioteche della Wisconsin – Madison University hanno svolto nel 2008 un’indagine attraverso un questionario on line proposto agli studenti e allo staff che ha ricevuto 301

³⁴ La ricerca ha coinvolto 6163 utenti di diverse età e di diversi paesi (America, Canada, Francia, Germania, Inghilterra, Giappone) oltre a 382 direttori di biblioteca negli stati Uniti.

³⁵ Gli utenti che hanno risposto all’indagine non vedono attualmente un ruolo per le biblioteche nel loro nuovo mondo di network sociali. Non vedono un ruolo per le biblioteche nel costruire siti sociali e la maggioranza non sarebbe molto incline a contribuire con contenuti, proprie pubblicazioni o ad aggregarsi a gruppi di discussione se fossero le biblioteche a offrire questi servizi. L’interesse a partecipare in attività in siti di social networking ospitati da biblioteche è basso fra coloro che hanno risposto. La maggior parte delle attività valutate hanno raggiunto l’interesse di meno del 10% di utenti che hanno indicato di essere estremamente disposti o molto disposti a partecipare. Questo generale livello di disinteresse si mantiene costante per tutti i paesi indagati.

risposte, seguito da diversi focus group nelle singole biblioteche. Lo scopo generale dell'indagine era sviluppare una nuova visione dell'ambiente di ricerca delle risorse per renderlo più adatto a supportare l'insegnamento e la ricerca scientifica nell'attuale contesto dominato dai motori di ricerca, individuandone caratteristiche e priorità. Le domande hanno riguardato i comportamenti degli utenti, i loro bisogni nella ricerca e le proposte per migliorare il processo individuazione e selezione dei materiali. La domanda più interessante per il nostro lavoro è quella che chiedeva di indicare quali caratteristiche migliorerebbero l'esperienza di ricerca. La possibilità di attribuire tag ai documenti è stata indicata 106 volte, ottenendo il 40,8% di citazioni. Le altre voci indicate sono visibili nella tabella che segue (University Of Wisconsin-Madison Libraries 2008 p. 20).



Il prototipo di un catalogo di nuova generazione ad Amburgo è stato accompagnato da una serie di test di gradimento delle innovazioni introdotte e di usabilità condotti attraverso focus group e indagini sul comportamento di ricerca degli utenti (Christensen 2009). Mentre le nuove caratteristiche legate ai metodi di ricerca, alla presentazione dei risultati, alla creazione di spazi personali, hanno trovato un'accoglienza favorevole, la possibilità offerta agli utenti di contribuire ai metadati generati dai bibliotecari con propri contenuti e di relazionarsi ad altri utenti ha trovato tiepida accoglienza. La motivazione degli utenti a

contribuire al catalogo con tag, recensioni o voti alle risorse è stata molto bassa. I partecipanti hanno espresso l'opinione che, ammesso e non concesso di aver voglia di contribuire al catalogo, non avrebbero piacere di condividere i propri contenuti con estranei. Interessante è anche rilevare che gli utenti hanno rigettato la possibilità di arricchire il catalogo con recensioni e votazioni provenienti da Amazon sostenendo la necessità di mantenere la neutralità del catalogo mostrando così di condividere il valore della terzietà nei processi di indicizzazione espresso in ambito bibliotecario per esempio da Ridi (2007 p. 167, 263; 2010 p. 55-57).

Un terzo studio si è svolto all'università di Sheffield (Tam – Cox – Bussey 2009). Ha coinvolto attraverso un'intervista 16 studenti stranieri con lo scopo di capire quali nuove caratteristiche sarebbe preferibile aggiungere all'opac attualmente in uso distinguendole tra quelle destinate a migliorare la ricerca (faccette, relevance ranking, spell checker), quelle relative alle tecnologie Web 2.0 (recensioni e voti, suggerimenti, RSS) quelle rivolte a migliorare la visualizzazione dei risultati (copertine, toc, tag cloud, informazioni di disponibilità e localizzazione). Nella valutazione complessiva di tutti gli elementi preferiti la tag cloud si posiziona al secondo posto dopo le faccette e prima del relevance ranking dei risultati. Va notato tuttavia che gli utenti non apprezzano la possibilità di aggiungere i propri tag quanto la possibilità di usare i tag presenti per restringere velocemente i risultati della ricerca o di utilizzare le parole che emergono dalla tag cloud, a cui non avevano pensato, come estensione della ricerca: in sostanza sembrano apprezzare gli aspetti funzionali piuttosto che quelli sociali.

Un quarto studio si è svolto all'università Copernico a Torun in Polonia e ha coinvolto 100 studenti di 20 diverse facoltà ai quali è stato somministrato un questionario (Wontorowska 2009). La ricerca era volta a valutare l'utilità di diversi elementi Web 2.0 nei siti delle biblioteche attraverso l'attribuzione di un valore scelto tra indispensabile, utile, inutile. Poiché le domande poste erano riferite in generale ai siti delle biblioteche senza ulteriori specificazioni, non è possibile trarre conclusioni specifiche per gli opac. Tuttavia rileva che i tag attribuiti dagli utenti a descrizione di risorse in Opac sono stati giudicati in molti casi indispensabili e in molti casi utili mentre solo in pochissimi casi sono stati considerati

inutili ottenendo complessivamente il punteggio più alto tra tutti gli strumenti sottoposti a giudizio precedendo persino le e-mail³⁶.

2.3 Tag e altre caratteristiche sociali negli opac in uso

Negli opac la possibilità offerta agli utenti di attribuire tag ai documenti è in genere associata ad altre funzionalità che vanno nella stessa direzione di consentire agli utenti di essere attivi nel sistema e contribuire così a creare un ambiente che favorisce una relazione e uno scambio tra biblioteca e utenti e tra utenti e utenti. Attraverso queste funzionalità viene data la possibilità di votare, recensire e commentare le risorse e di interagire con gli altri utenti instaurando un dialogo sui libri o su argomenti di proprio interesse. Anche i suggerimenti di lettura basati sui dati di circolazione, e quindi sulle attività degli utenti nel sistema, possono essere considerati parte di questo approccio in quanto contribuiscono a creare un senso di comunità e partecipazione intorno all'opac. I voti, le opinioni, i commenti e i suggerimenti, provenienti da una fonte meno professionale ma più vicina per linguaggio e approccio ai propri, possono costituire per l'utente una via più amichevole per esplorare le risorse presenti nell'opac e operare le proprie scelte: il giudizio dei propri pari, espresso attraverso un voto o un commento, può aiutare nel momento della selezione di un documento mentre i suggerimenti per ulteriori letture possono costituire un modo diverso per individuare risorse utili a proseguire una ricerca. Alcuni opac favoriscono l'esportazione dei dati bibliografici di un documento selezionato verso siti di socialbookmarking o siti di socialnetworking³⁷ superando la più tradizionale creazione di bibliografie personali da inviare al proprio indirizzo di posta elettronica che è stata la prima forma di esportazione dei dati consentita agli utenti degli opac. In questo modo consentono all'utente di usare quel dato all'interno dei propri sistemi di registrazione delle informazioni oppure di farne un uso condiviso nel proprio ambiente sociale online. In questo secondo caso la biblioteca

³⁶ Va notato che nel questionario la domanda si riferiva genericamente all'utilità dei tag (parole chiave o termini associati a parti di informazioni) e solo nel commento ai risultati si parla specificamente di tag attribuiti dagli utenti a descrizione di risorse in Opac rendendo a mio avviso dubbio il senso attribuito dagli utenti alla domanda e quindi la sua interpretazione. I punteggi di tutti gli elementi sottoposti a giudizio sono stati: tags 123, e-mail 120, link collection 90, wiki 90, e-learning 85, dedicated sites 80, adding news 77, forums 69, suggesting new items 63, streaming media 62, RSS 41, instant messengers 32, librarians' Weblog -1, social networking -6, users' Weblogs -24, entertainment -25.

³⁷ Vedi per esempio il catalogo di OCLC < <http://www.worldcat.org> > .

acquista indirettamente anche una sua visibilità in ambienti lontani dal proprio sito istituzionale o dall'opac.

Negli opac di nuova generazione le funzionalità che consentono i contributi degli utenti e altre forme di "socialità" sono adottate in misura diversa: in alcuni casi sono tutte presenti in altri ce ne sono solo alcune in una varietà di combinazioni nei diversi prodotti in commercio e nelle varie configurazioni effettivamente installate. In alcuni software queste funzionalità non erano state concepite nella prima versione e sono state realizzate con moduli aggiuntivi piuttosto che essere nativamente integrate nel sistema (Tosca Consultant 2008 p. 44-45)³⁸ rivelando forse una rincorsa della moda del momento piuttosto che una convinzione sull'utilità e possibilità di integrazione di questi aspetti nelle funzioni centrali dell'opac. Anche il modo in cui questi dati degli utenti vengono utilizzati dalle biblioteche nelle singole installazioni varia molto e produce effetti differenti. E' evidente che fornire suggerimenti di lettura basati sui dati di circolazione interni alla biblioteca è cosa diversa dall'usare dati creati dagli utenti di Amazon poiché che nel secondo caso si tratta di dati di vendita di un prodotto di cui non si conosce neanche esattamente il modo in cui sono ricavati né se ci siano liberi da interessi commerciali. L'uso di dati provenienti da fonti esterne al sistema viene fatto per sopperire alla scarsità dei dati interni dovuti alla finora breve storia di questi sistemi e alla scarsa predisposizione che gli utenti hanno dimostrato al loro uso.

In Italia le biblioteche che hanno adottato opac con le caratteristiche sociali che abbiamo descritto appartengono alla tipologia delle biblioteche pubbliche, istituzioni che sentono maggiormente la necessità di coinvolgere gli utenti e hanno anche una tradizione in questo senso legata alla promozione della lettura (Marchitelli – Piazzini 2008 p. 88). Si tratta più precisamente di sistemi di biblioteche e di cataloghi collettivi. Il primo in ordine di tempo è stato il sistema delle biblioteche della provincia di Verona³⁹ seguito da una serie di opac basati sul prodotto Sebina You⁴⁰ tra i quali l'opac delle biblioteche di Roma⁴¹ e delle biblio-

³⁸ E' il caso di AquaBrowser che usa il modulo supplementare MyDiscovery.

³⁹ Questo opac è basato sul software di gestione Clavis NG <<http://www.comperio.it/clavisng/>> e sul modulo Discovery NG <<http://www.comperio.it/discoveryng/>> della società Comperio <<http://www.comperio.it/>>.

⁴⁰ Sebina You <<http://www.sebina.it/SebinaYOU.htm>> è un prodotto della divisione Beni Culturali della società Data Management <<http://www.datamanagement.it/>>.

⁴¹ Vedi <<http://www.bibliotu.it/SebinaOpac/.do>>.

teche della Rete di Romagna⁴². L'opac delle biblioteche veronesi presenta nella home page le ultime recensioni scritte dagli utenti accompagnate da simboli che rappresentano il gradimento della risorsa; sempre nella home page è presente una nuvola dei tag più usati e l'accesso ad un blog (che subito ammonisce gli utenti su cosa si può o non si può scrivere!). Anche se spesso non è indicato espressamente, tutte le funzioni che consentono agli utenti di inserire propri contributi in qualsiasi forma, risultano riservate agli utenti iscritti alle biblioteche. Questa limitazione, presente in tutti i prodotti italiani e in molti stranieri⁴³, è dettata da esigenze di auto tutela delle biblioteche tese ad evitare l'utilizzo dell'opac per la pubblicazione di contenuti impropri⁴⁴. Nell'opac delle biblioteche romane, nella home page, è presente una rubrica "Tu cosa ne pensi?" per lasciare commenti e opinioni e una più inconsueta "BiblioScriviTu" che consente di leggere racconti scritti da altri utenti e di inviare un proprio testo. Come risposta ad una interrogazione per parole chiave, insieme al set di notizie recuperate viene presentata una nuvola di tag collegati preceduta dall'etichetta "Raffina per tag": cliccando su un tag la selezione precedente viene ristretta ai soli documenti che condividono quel tag. Scegliendo un documento è possibile leggere i commenti di altri utenti e avere suggerimenti per ulteriori letture basati sui dati di circolazione introdotti dalla formula "Chi ha letto questo ha letto anche...". Vengono mostrati anche i tag associati alla risorsa che in questo caso, se cliccati, hanno la funzione di lanciare una ricerca completamente nuova nel catalogo. Abbiamo qui un'esemplificazione di come i tag possano essere usati nello stesso sistema in entrambe le funzioni di raffinamento ed estensione della ricerca in momenti diversi. Va sottolineato che in questo opac, i tag non sono etichette attribuite dagli utenti desiderosi di partecipare e collaborare ma parole o stringhe estratte dall'equivalente verbale del numero di classificazione Dewey⁴⁵: sono

⁴² Vedi <<http://scoprirete.bibliotecheromagna.it/>>.

⁴³ Fanno eccezione per esempio l'opac della Hannepin library che consente anche all'utente non iscritto alla biblioteca di lasciare commenti o creare una propria lista di preferiti identificandosi solo tramite un indirizzo e-mail <http://www.hclib.org/pub/bookspace/> e gli opac basati su Sopac che vedremo in seguito.

⁴⁴ Nel caso della Pubblica Amministrazione italiana la fornitura di servizi in rete è consentita solo dopo l'autenticazione informatica dell'utente (art. 64 del Dlgs. 82/2005). Che in questi casi si tratti effettivamente di un servizio reso dalla P.A. e si rientri nei casi normati da quell'articolo è cosa dubbia e l'obbligo di identificare l'utente sembra piuttosto collegato a necessità di tutela del sistema e di rintracciabilità dell'utente in caso di abusi.

⁴⁵ Il software consente ai gestori del sistema la massima personalizzazione nel configurare la fonte dei tag: possono essere tag degli utenti, tag provenienti da fonti interne come keyword attribuite in fase di catalogazione, termini provenienti da soggetti, tesauri, classificazioni, termini provenienti da fonti esterne o dal web.

quindi un modo per facilitare il recupero del contenuto semantico del documento ma hanno poco a che vedere con gli aspetti caratteristici dei tag attribuiti dagli utenti. L'opac delle biblioteche di Romagna è una personalizzazione diversa dello stesso prodotto Sebina You e presenta funzionalità analoghe a quelle che abbiamo appena descritte. Si caratterizza per la presenza della faccettazione dei risultati, di cui parleremo in seguito, introdotta dall'etichetta "Naviga tra i risultati" posizionata colonna a sinistra dello schermo. Una volta selezionato un documento la stessa posizione viene occupata da un "Navigatore" che non è altro che la riproposizione in formato ridotto dell'intero set di risposta con la possibilità di scorrere i risultati: in questo modo l'utente ha contestualmente presente sia l'insieme dei risultati sia una particolare selezione. Una funzione presente in entrambi questi opac è il browsing dello scaffale: scegliendo un documento si possono vedere copertine e titoli dei documenti che lo seguono secondo la classificazione Dewey, un tentativo di riprodurre online l'esplorazione fisica degli scaffali della biblioteca e di facilitare l'incontro con documenti di cui non si conoscono dati precisi o di cui non si era ancora percepito il bisogno. Emerge da questa breve esplorazione che in questi opac i contributi offerti dagli utenti non sono integrati nei dati catalografici o ad essi affiancati: possono servire gli utenti come spunto alla lettura ma collaborano ben poco alla costruzione del catalogo. Non c'è dubbio che le interfacce si presentino in modo più accattivante, dinamico e coinvolgente per l'utente ma parlare di questi opac come strumenti collaborativi (Metitieri 2009 p. 10-11) corrisponde poco alla realtà e sembra corretta l'osservazione di Spalding:

Some of the biggest ideas—like social features—have been completely misunderstood and misapplied by the big companies. But all the big companies now has a supposedly "next gen" catalog. Some are even good (Spalding 2010)⁴⁶.

Una maggiore integrazione dei contributi degli utenti nei dati utilizzati in fase di costruzione degli indici di ricerca è scelta dalle biblioteche che usano il software Sopac (Social Opac)⁴⁷ creato da John Blyberg. Questo opac si definisce "a truly community-driven catalog system" e ha la caratteristica di essere un software open source (rilasciato con licenza

⁴⁶ Alcune delle grandi idee, come le caratteristiche sociali, sono state completamente fraintese e male applicate dalle grandi società. Ma tutte le grandi società ora hanno quello che si suppone essere un catalogo "next gen". Alcuni sono persino buoni.

⁴⁷ Vedi <<http://thesocialopac.net/>>. Le biblioteche che attualmente adottano questo software sono Darien Library <<http://www.darienlibrary.org/catalog/>>, Palos Verdes Library District <<http://www.pvld.org/catalog/>>, Ann Arbor District Library <<http://www.aadl.org/catalog/>> .

GNU Public License v3). In questo caso l'aspetto sociale può essere rilevato da molti punti di vista: nel fatto che il software si sviluppi grazie alla comunità degli utilizzatori e si basi su prodotti open source, che i contributi degli utenti siano effettivamente integrati nel catalogo, che le biblioteche che utilizzano questo sistema possano scambiarsi reciprocamente i dati immessi dagli utenti contribuendo collettivamente a creare la massa critica di dati che abbiamo visto essere necessaria al buon funzionamento dei sistemi collaborativi (Hadro 2008). Lo scopo dichiarato nella documentazione del prodotto Sopac è quello di sfruttare le interazioni tipiche del web 2.0 per costruire un ambiente di scoperta ed esplorazione più ricco e attraente (Blyberg 2008). Una prima idea applicata è quella di fondere in un'unica interfaccia unificata tutte le componenti dell'esperienza degli utenti nel sistema, dalla home page della biblioteca, al display del singolo record (Hadro 2008). Questa impostazione è visibile in tutte le installazioni che abbiamo citato e contribuisce a superare la sensazione che l'opac sia una componente esterna alla biblioteca che si ha normalmente quando la grafica della home page della biblioteca e quella dell'opac sono diverse, cosa che accade nella maggior parte dei casi. L'interfaccia di ricerca presenta una lista di "top tag" che costituiscono una prima forma di accesso ad una parte selezionata delle risorse. Sempre in prima pagina è presente una lista dei post e delle recensioni degli utenti scritti più di recente. Immettendo delle parole chiave nella maschera di ricerca i risultati vengono presentati in ordine di rilevanza e accompagnati da una faccettazione dei risultati che consente un progressivo affinamento del set di risposta. In corrispondenza di ogni documento è presente una lista di contenuti aggiuntivi, recensioni, TOC ecc., mentre i contributi degli utenti, definiti Community reviews, sono presenti nel dettaglio del singolo record dove è anche presente la possibilità di lasciare un proprio commento, voto o recensione, dopo essersi registrati (ma senza necessità di essere iscritti alla biblioteca). Sempre a livello di dettaglio del singolo record sono presenti i tag associati a quel documento, che in questo caso sono un modo per rilanciare una nuova ricerca sul termine selezionato, ed è presente la possibilità di contribuire con propri tag.

The screenshot shows a library catalog entry for a book. On the left, there is a sidebar with metadata: **Format:** Book; **Subjects:** [Potter, Harry \(Fictitious character\)](#), [Hogwarts School of Witchcraft and Wizardry \(Imaginary place\)](#), [Wizards -- Fiction](#), [Magic -- Fiction](#), [Schools -- Fiction](#), [England -- Fiction](#), [Fantasy fiction](#); **Tags:** [Fiction \(1\)](#), [Wizardry \(1\)](#), [Books made into movies \(1\)](#), [Harry Potter \(1\)](#), [Hogwarts \(1\)](#), [Best Book \(1\)](#). At the bottom of the sidebar is a [Login to add tags](#) link.

The main content area is divided into two sections:

- Reviews and Summaries:** A list of links including [School Library Journal Review](#), [Booklist Review](#), [Publishers Weekly Review](#), [Summary / Annotation](#), [Fiction Profile](#), and [Author Notes](#).
- Community Reviews:** A section showing a **Rating:** of five red stars, with the text "(2 votes - [Login](#) to add yours)". Below this is a **Best Book of the Series** badge, followed by the text "submitted by [kelpaige](#) on June 19, 2007, 9:15 pm". The review text begins with "In the already spellbinding Harry Potter series, this one outshines the rest" and is followed by a [Login to write a review](#) link.

E' difficile dire nella realtà quanto l'integrazione dei contenuti degli utenti negli indici di ricerca degli opac giovi alle capacità di recupero dell'opac. Abbiamo visto che gli utenti non sono così propensi a collaborare al catalogo ma potrebbe essere una fase momentanea considerando che le attività di socialnetworking, per quanto costantemente citate nella letteratura sui servizi delle biblioteche, hanno una nascita molto recente e sono nuove per gli utenti (Tam – Cox – Bussey 2009 p. 371): non si può escludere che in futuro possano essere maggiormente utilizzate. In questo caso la relazione tra dati di autorità e contributi generati dagli utenti dovrebbe essere considerata con attenzione. In linea teorica bisognerebbe essere molto attenti a mantenere un equilibrio tra dati di autorità e contributi generati dagli utenti, tra i record bibliografici e una somma indistinta di contributi che continuano ad accumularsi nel tempo (Phillipson 2006) aumentando il loro peso complessivo. Mentre potrebbe essere abbastanza semplice tenere distinte e presentare come forme complementari di accesso folksonomie e classificazioni strutturate, sembra più difficile poter presentare separatamente i risultati basati su una ricerca sui dati catalografici da una ricerca realizzata sull'insieme di tutti i termini presenti nel sistema, mantenendo un'interfaccia sufficientemente semplice e comprensibile. Anche il contesto in cui l'opac si colloca, il bacino di utenti cui fa riferimento incidono molto sulle scelte da operare. Più precisamente si dovrebbe parlare del contesto della specifica ricerca del singolo utente che può variare nel

corso del tempo e delle situazioni ma è evidente che in generale le aspettative degli utenti nelle biblioteche di pubblica lettura e nelle biblioteche delle istituzioni di studio e di ricerca sono significativamente diverse e vanno considerate. Se nel primo caso i giudizi soggettivi e arbitrari su un romanzo o i suggerimenti di lettura basati sui gusti personali hanno poche conseguenze per chi li segua, nel caso di ricerche svolte per motivi di studio l'esigenza di basarsi su informazioni corrette e di qualità è fondamentale ed è riconosciuta anche dagli utenti (Tam – Cox – Bussey 2009 p. 362-363). Non è quindi di secondaria importanza dotare gli opac di interfacce, forse meno semplici, ma che consentano agli utenti di scegliere se e quando servirsi del giudizio dei propri pari, della saggezza della folla e quando invece affidarsi ad informazioni create da catalogatori di professione.

3. RELEVANCE RANKING E RAGGRUPPAMENTI DINAMICI

Le problematiche relative alla presentazione dei risultati delle interrogazioni sono sempre state tenute presenti nella progettazione degli opac più recenti. Una delle tematiche più dibattute è stata la necessità di creare un ambiente informativo che non solo fosse in grado di rispondere a interrogazioni specifiche ma che favorisse le funzioni di browsing cioè l'esplorazione dei contenuti del catalogo in una situazione di indeterminatezza del bisogno informativo (Hildreth 1995, Bates 1989)⁴⁸. L'introduzione della navigazione tra termini correlati attraverso la tecnica dei link ha costituito un notevole progresso rispetto alla staticità dello scorrimento di liste ordinate che costituivano l'unica forma di browsing negli opac fino alla seconda generazione ma oggi questa tecnica risulta largamente insufficiente sia perché si confronta con soluzioni più innovative adottate nei siti commerciali sia perché il catalogo stesso si è evoluto. La vastità del catalogo e il suo essere in molte situazioni non solo un data base ma un indice di risorse eterogenee per formato, contenuto, granularità⁴⁹, livello di catalogazione e provenienza rendono necessario ordinare e presentare in modo coerente e significativo per l'utente un'enorme massa di risultati in modo che la quantità e la varietà non diventino disorientamento e impossibilità di selezione. L'esigenza di una presentazione significativa dei dati è stata accentuata anche dall'aver modificato le maschere di ricerca. Negli opac più tradizionali ancora oggi vengono presentate all'utente per l'immissione dei termini di ricerca diverse caselle che rappresentano differenti canali di

⁴⁸ Il browsing è un'insieme di molti tipi di attività. Per descriverla Hildreth individua tre categorie (Hildreth 1995 parte 5 pag. 1, 4): indiretto (indirected browsing, general browsing, serendipity browsing); semi-diretto (semi-directed browsing, general purposive browsing, general purpose browsing); diretto (directed browsing, specific browsing, search browsing). Alla prima categoria può essere ricondotta la più semplice attività di browsing che viene svolta in biblioteca: quella di curiosare tra gli scaffali cercando qualcosa, non si sa bene che cosa, nella speranza di trovare una lettura di proprio interesse. I documenti vengono esaminati in modo casuale per essere presi in prestito. E' un'attività occasionale, non strutturata. Non tutte le attività di browsing sono così indeterminate: alcuni tipi partono da una necessità informativa che, ancorché non perfettamente formulata, viene riconosciuta una volta che la si trovi. E' quello che avviene quando si esplorano gli scaffali della biblioteca in una zona precisa cercando documenti relativi ad una particolare area di interesse. Questo tipo di ricerca è di solito aiutata da forme di orientamento predisposte dai bibliotecari. La terza categoria di browsing è quella svolta da persone più padrone degli strumenti di ricerca quando cercano in un repertorio, in una bibliografia, in un abstract, nel catalogo, senza preparare una strategia di ricerca ma partendo da un punto e facendosi poi guidare dai riferimenti incrociati, strutturati e organizzati, presenti nei vari strumenti di ricerca. E' un'attività strutturata ma aperta a cogliere indizi e suggestioni che emergano nel corso della ricerca.

⁴⁹ Per una trattazione del concetto di granularità di veda Zani (2006).

ricerca, normalmente per autore, titolo, soggetto (e solo come ulteriore opzione per parole chiave), oppure una casella unica seguita dalla possibilità di selezionare i campi, o anche la tipologia di documenti, su cui la ricerca deve essere lanciata. Questo consente al sistema sin dall'inizio di contestualizzare la ricerca, di restringerla a particolari dati specificamente indicizzati e di presentare i risultati, che normalmente sono costituiti da un set relativamente ristretto, in liste di notizie, ordinate alfabeticamente o per data, che sono omogenee e coerenti rispetto al campo di immissione della richiesta scelto dall'utente. Queste modalità sono funzionali ad un contesto di ricerca in cui si assume che l'utente sia in grado di qualificare le proprie richieste e sono efficaci quando l'utente voglia recuperare materiali che gli sono noti mentre funzionano molto meno in un contesto di esplorazione delle risorse, quando la richiesta è imprecisa oppure generica, come nel caso delle ricerche per soggetto o per parole. Nel caso della ricerca libera per parole, non essendo in grado il sistema di inferire dalla richiesta immessa alcun criterio di ordinamento dei risultati, il criterio seguito da molti opac è stato di presentare per primi i documenti catalogati per ultimi nel sistema (Antelman - Lynema - Pace 2006 p. 129).

Negli opac di nuova generazione la presentazione di un'unica casella di ricerca, collocata al centro di una pagina dedicata dell'opac oppure posizionata in contesti diversi in cui costituisce un link implicito al catalogo (la home page di un'istituzione o la pagina dei servizi offerti, o quella degli strumenti didattici), sta diventando l'opzione principale⁵⁰. Per venire incontro alle esigenze degli utenti, i quali si aspettano di applicare negli opac le stesse modalità che usano nelle altre ricerche in rete, si è ritenuto necessario adattare la maschera di ricerca, ma anche il modello della ricerca, a quella presente nei motori del Web: search box unica e parole chiave messe implicitamente in relazione dall'operatore "and". Lo scopo di questa tipologia di maschera è di accogliere interrogazioni generiche, di non richiedere all'utente di definire i termini immessi, di non pretendere alcuna conoscenza dei meccanismi di funzionamento del sistema, di non far fallire la ricerca. Negli opac la ricerca dei termini immessi viene lanciata, di norma, su tutte le risorse del sistema, sia nei campi indicizzati, sia nelle descrizioni, sia nei full text, ma può anche essere caratterizzata

⁵⁰ Si vedano per esempio il polo SBN del Friuli (<http://sol.units.it/>), i cataloghi delle biblioteche della provincia di Verona e della Romagna (<http://sbp.provincia.verona.it/> ; <http://scoprirete.bibliotecheromagna.it/>), questa pagina dell'università di Pennsylvania (<http://www.library.upenn.edu/courseware/>), il catalogo della biblioteca di Nashville <<http://www.library.nashville.org/startpages/encorecatalog.asp>>.

in ciascun sistema di recupero grazie alle funzioni di configurazione messe a disposizione dei gestori. La risposta alle interrogazioni è costituita da un set molto ampio che, in assenza di indicazioni provenienti dall'utente, deve essere organizzato in qualche modo per poter essere esaminato e risultare significativo. I metodi adottati a questo scopo negli opac di nuova generazione sono principalmente il relevance ranking e i raggruppamenti a faccette ma esistono anche altre modalità come diverse forme di clusterizzazione e rappresentazioni visive dei risultati.

3.1 Il relevance ranking

Il concetto di "relevance ranking" (ordinamento in base alla rilevanza) si riferisce ad un insieme di metodi statistici per ordinare i risultati di una ricerca base al grado di corrispondenza con i termini di ricerca immessi. I criteri che stabiliscono la pertinenza e determinano il posizionamento sono molteplici, molti noti e in uso in tutti i sistemi, altri tenuti riservati, come riservati restano il mix di criteri e il peso loro assegnato che definiscono l'algoritmo complessivo di ordinamento, data la loro capacità di determinare la qualità delle performance del sistema e di distinguerlo dagli altri⁵¹.

Tra i principali elementi che concorrono a determinare la rilevanza di un documento negli algoritmi di relevance ranking Ridi (2008) individua i seguenti:

frequenza (numero di occorrenze dei termini utilizzati per la ricerca all'interno della singola pagina recuperata), **densità** (rapporto fra il numero di occorrenze dei termini all'interno della pagina e il totale delle parole contenute nella pagina), **rarietà** (rapporto inverso fra il numero di occorrenze dei termini utilizzati per la ricerca all'interno del database del motore e il totale delle parole contenute nel database stesso), **compresenza** (presenza, nella pagina, di più d'uno dei termini richiesti o addirittura di tutti), **prossimità** (presenza, nella pagina, di più d'uno dei termini richiesti nella medesima frase o comunque vicini fra loro), **posizione** (presenza dei termini in particolari zone della pagina), **priorità** (ordine in cui sono stati immessi i termini nel form di ricerca), **provenienza** (pagine provenienti dal

⁵¹ In un suo articolo Lewandowski (Lewandowski 2009 p.587) cita una pagina tecnica di Google secondo la quale nel ranking vengono presi in considerazione più di 200 "signals" (www.google.com/corporate/tech.html). La stessa pagina, ma nella versione italiana, indica oggi che PageRank esegue una misurazione oggettiva dell'importanza delle pagine Web risolvendo un'equazione con oltre 500 milioni di variabili e 2 miliardi di termini.

medesimo sito vengono spesso mostrate l'una vicina all'altra, talvolta con un rientro a partire dalla seconda e nascondendo le ultime se troppo numerose. Anche la provenienza dell'interrogazione, ovvero l'IP number del computer dal quale si effettua la ricerca, può influenzare il ranking), **età e aggiornamento** (a parità di requisiti, alcuni motori privilegiano le pagine aggiornate più recentemente e che esistono da più tempo al medesimo URL), **popolarità** (numero dei link "in entrata" provenienti da altre pagine contenute nel database del motore e numero delle selezioni della pagina effettuate dagli utenti a partire dalle schermate di risposta del motore), **tariffazione** (posizionamento garantito dal pagamento di una tariffa, un tempo occulto è oggi palesato agli utenti)⁵².

L'introduzione del concetto di relevance ranking negli opac non è privo di problematicità. Molti autori che hanno rilevato le insufficienze degli opac tradizionali sostengono l'adozione del relevance ranking senza giustificare e valutare questa scelta se non in rapporto alla necessità di adeguarsi all'ambiente proposto dai siti commerciali ma ad alcuni i criteri di popolarità e novità adottati in quei siti non sembrano adeguati alle biblioteche dopo che fiumi di inchiostro sono stati spesi sul concetto di rilevanza (Šauperl - Saye 2009 p. 508). Altri giudicano l'introduzione del relevance ranking una scelta poco sensata se non è sostenuta da tecniche che leghino il concetto di rilevanza a specifici utenti (Bade 2007); su questa scia anche Buizza (2008) che vede un passaggio dal concetto di rilevanza soggettiva ad una rilevanza oggettivata da criteri standard funzionali all'omologazione; altri propongono criteri specifici di rilevanza per il materiale delle biblioteche (Lewandowski 2009, Dellit - Boston 2007).

3.2 Il concetto di rilevanza

Esaminare il concetto di rilevanza potrebbe aiutare a capire quali aspetti devono essere considerati nella creazione di algoritmi di ranking significativi negli opac. Per fare questo ci serviamo di due articoli di Saracevic che, rivedendo la letteratura sul tema degli ultimi 30 anni, fanno il punto sulla nozione di rilevanza nel campo della scienza dell'informazione (Saracevic 2007 a, 2007 b).

Il concetto di rilevanza è una nozione di base nelle scienze dell'informazione e in particolare nell'information retrieval. Tutti i sistemi di IR sono disegnati per fornire informazioni

⁵² Ulteriori criteri sono indicati in Gambari-Guerrini (2002) p. 278-279.

o oggetti informativi potenzialmente rilevanti per le persone: attraverso un complesso set di operazioni basate su algoritmi sofisticati e costantemente migliorati recuperano e mostrano la loro versione di cosa può essere rilevante. Le persone dal canto loro si comportano secondo le proprie caratteristiche e hanno una propria visione di rilevanza. Ci sono dunque due categorie di base di rilevanza: umana e di sistema. Sistemi di IR e persone trattano entrambi la rilevanza come una relazione ma ci sono premesse di base differenti nel valutare cosa sia effettivamente coinvolto nella relazione.

Nel contesto dell'IR il concetto di rilevanza nasce intorno al 1950 in relazione ai cosiddetti "false drops" informazioni non volute ma recuperate dai sistemi: si trattava di identificare la non-rilevanza, cioè le informazioni errate rispetto alla query, per pulire e limitare i set di risposta. Di rilevanza in senso proprio si parla invece dal 1959 quando la nozione fu proposta per essere usata nella coppia richiamo-rilevanza come misura delle capacità di recupero dei sistemi: in seguito si preferirono i termini richiamo e precisione ma è evidente che la rilevanza è il criterio che sottende a questa misura⁵³.

Secondo Saracevic il concetto di rilevanza è intuitivo per gli esseri umani (altra cosa è stabilire che cosa sia rilevante), così come è intuitivo capire che essa comprende sempre una relazione: la rilevanza comprende esplicitamente o implicitamente un "per", un contesto. Superando l'intuizione si può dire che la rilevanza è una nozione che consiste di un numero di parti che interagiscono e hanno una relazione basata su alcune proprietà o criteri. Le parti possono essere oggetti tangibili (documenti, macchine, processi), intangibili (idee, concetti, informazioni) o una combinazione di entrambi (compiti, situazioni, responsabilità). Le proprietà (aderenza al soggetto, utilità ecc..) forniscono la base e il contesto per stabilire una relazione. La rilevanza è anche una misura della forza di queste relazioni. Nelle scienze dell'informazione la rilevanza è dunque una relazione tra informazione o oggetto informativo da una parte e contesto dall'altra nonché la misura di questa relazione. Il contesto è qualcosa di complesso che comprende stati e situazioni cognitive e affettive (bisogni informativi, intenti, argomenti, problemi, compiti, oggetti informativi): è l'interazione dinamica tra numerosi aspetti interni ed esterni del sistema e di chi lo utilizza.

⁵³ Richiamo e precisione misurano la probabilità di accordo tra quello che il sistema recupera come rilevante (system relevance) e quello che l'utente valuta come rilevante (user relevance) (Saracevic 2007 a p. 1918).

Come si stabilisce una relazione di rilevanza e come si misura? Nella scienza dell'informazione si dice che i sistemi creano la rilevanza mentre gli utenti la derivano: in ogni caso è considerata un'inferenza. I sistemi di IR prendono una query, la confrontano con gli oggetti informativi presenti nel sistema secondo certi algoritmi e forniscono quello che considerano rilevante: in questo senso creano la rilevanza, perché ne sono la causa, la producono, ne sono l'origine. La gente deriva la rilevanza dalle informazioni o dagli oggetti informativi che ha ottenuto: interpreta e collega le informazioni al problema che sta esaminando, al proprio stato cognitivo e ad altri fattori, cioè prende i risultati e ne deduce che cosa può essere rilevante per sé. L'utente legge nei risultati molto più che una semplice corrispondenza tra parole o frasi (quello che fa il sistema). Questo fa capire che esistono diversi tipi di rilevanza che vedremo in seguito. Fa capire anche che per creare o derivare rilevanza è necessario un processo di selezione: le informazioni o gli oggetti informativi sono selezionati come rilevanti tra tanti altri disponibili e la selezione viene fatta con l'intento di massimizzare il risultato e/o minimizzare lo sforzo nell'uso dei risultati. Il processo di selezione implica una serie di interazioni nel corso delle quali l'interpretazione degli attributi e il contesto possono cambiare. Saracevic sintetizza in questo modo quanto detto fino ad ora:

We consider relevance as having a number of dimensions or attributes: Relevance is a relation. Relevance is a property. Relevance is a measure. Relevance has a context, external and internal. Relevance may change. Relevance has a number of manifestations or kinds. Relevance is not given. Relevance is inferred. Relevance is created or derived. Relevance involves selection. Relevance involves interaction. Relevance follows some intentionality. (Saracevic 2007 a p. 1920)⁵⁴.

Mentre manca ancora oggi una teoria della rilevanza applicabile all'IR che sia fondata su basi concettuali solide e sia testabile, esistono invece diversi modelli descrittivi che indicano gli elementi o le variabili connesse al concetto di rilevanza. Negli anni fra il 1959 e il 1976 gli studi si sono orientati verso l'individuazione della rilevanza inerente di documenti e query e solo in seguito si è compreso e formalizzato il concetto di rilevanza in senso più soggettivo, dinamico e multidimensionale. Tradizionalmente vengono proposti due modelli

⁵⁴ Consideriamo che la rilevanza abbia un certo numero di dimensioni o attributi: la rilevanza è una relazione. La rilevanza è una proprietà. La rilevanza è una misura. La rilevanza ha un contesto, interno ed esterno. La rilevanza può cambiare. La rilevanza ha un numero di manifestazioni o tipi. La rilevanza non è data. La rilevanza è dedotta. La rilevanza è creata o derivata. La rilevanza implica la selezione. La rilevanza implica una interazione. La rilevanza segue una qualche intenzionalità.

rilevanza che ricalcano la disputa tra due diversi modelli di IR: user model e system model. Il modello di IR cosiddetto di sistema considera l'IR dal punto di vista del sistema ignorando l'utente: l'enfasi è posta sui metodi di elaborazione degli oggetti dell'informazione e sui procedimenti che determinano la corrispondenza con le query. L'utente è rappresentato solo dalle query e non si prende in alcun modo in considerazione l'interazione che avviene tra utente e sistema. Viceversa il modello di IR di tipo user considera il sistema come dato e si concentra sull'utente cercando di incorporare le dimensioni cognitive e sociale e l'interazione. Il principale attacco al modello di sistema è venuto dagli studi sui comportamenti di ricerca di Dervin e Nilan (1986) che hanno sottolineato che l'orientamento al sistema è scorretto e hanno invitato ad un cambio di paradigma: secondo questi autori osservando i tipi di criteri che gli utenti usano nel valutare le informazioni possiamo avere una comprensione della rilevanza e dobbiamo usare questa comprensione per modificare il disegno dei sistemi di IR. I sostenitori di questo approccio non sono stati però in grado di esplicitare in che modo questa comprensione si dovesse concretizzare, come richiesto dai sostenitori del modello di sistema, dunque essa non ha prodotto cambiamenti negli applicativi. Saracevic sostiene che la visione dualistica, umana e di sistema, non sia appropriata e blocchi entrambi i contendenti in un vicolo cieco. Egli propone un modello di rilevanza, che cerca di integrare le due visioni: entrambe le componenti hanno una moltitudine di variabili che si relazionano attraverso un'interfaccia. La domanda centrale dovrebbe essere: Come possiamo far sì che utente e sistema lavorino insieme a beneficio di entrambi? In questo modello stratificato, rappresentato nell'immagine che segue, il concetto di rilevanza è posto in un contesto di IR. Le interazioni di IR coinvolgono un certo numero di livelli o strati: le deduzioni di rilevanza sono create o derivate nell'interazione tra i diversi strati. Il modello assume che l'utente interagisca col sistema per utilizzare le informazioni e che l'uso delle informazioni sia connesso con aspetti cognitivi, in situazioni applicative e in un contesto, nel senso più ampio e plurale del termine, cioè sia connesso con la rilevanza. Le interazioni tra utente e sistema sono rappresentate a livello dell'interfaccia ma l'interfaccia non è il focus dell'interazione, nonostante il disegno dell'interfaccia possa supportare o frustrare l'interazione. L'interazione deve essere pensata come una serie di processi che avvengono in diversi livelli fra loro interconnessi: è un dialogo tra gli elementi associati all'utente e gli elementi associati al sistema che ha lo scopo di modificare lo stato cognitivo dell'utente affinché possa utilizzare informazioni rilevanti in connessione

con compiti immediati. Il dialogo può essere iterativo e includere vari tipi di feedback. Va precisato che in un sistema di IR la rilevanza si manifesta in ogni strato in forme diverse: possiamo quindi parlare di diversi sistemi di rilevanza.

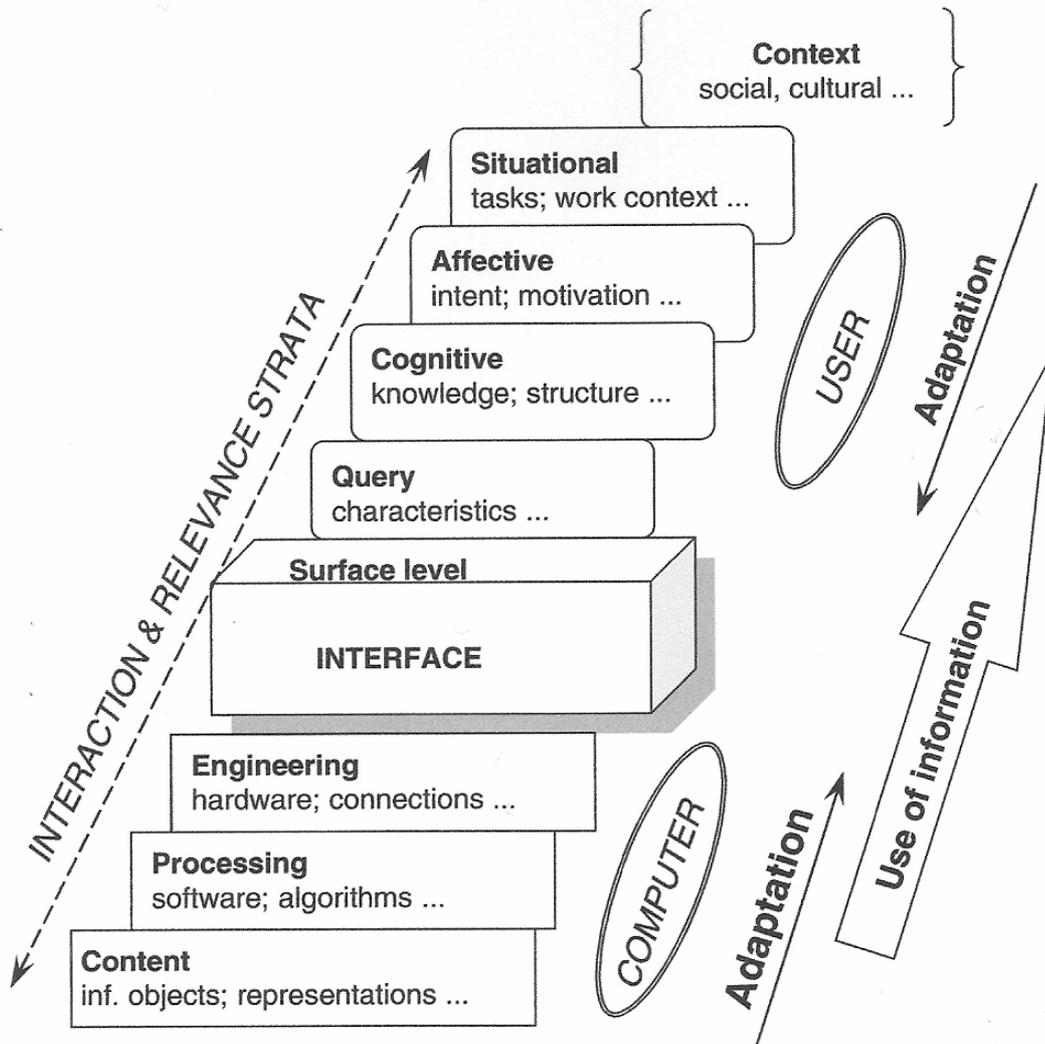


FIG. 1. Stratified model of relevance interactions.

Lo sforzo per specificare le diverse forme di rilevanza o, in altre parole, le manifestazioni della rilevanza, si è concentrato nell'identificare quali particolari oggetti siano collegati da un particolare tipo di rilevanza. Una prima distinzione è stata fatta tra la rilevanza **sogettiva** (user relevance) e la rilevanza **oggettiva** (topic relevance). La prima è stata considerata come un aspetto non trattabile: ancora oggi non è stato possibile includere aspetti di rilevanza soggettiva negli oggetti informativi e negli algoritmi di retrieval. Si parla di user

relevance **interna** (**psicologica o cognitiva** come la relazione tra gli oggetti informativi e lo stato cognitivo dell'utente oppure **motivazionale o affettiva** come relazione con intenti, mete, motivazioni, frustrazioni dell'utente) o **esterna** (**situazionale** come relazione con la situazione, i compiti o i problemi che l'utente sta affrontando oppure **socioculturale** in relazione ai compiti percepiti nel loro contesto socioculturale).

La rilevanza oggettiva è basata sulle relazioni logiche tra una richiesta formulata e i documenti recuperati: è la relazione tra informazione o oggetti informativi e un argomento o un soggetto considerato. Può essere dedotta dall'output del sistema di retrieval. Non è mai perfetta: diversi algoritmi producono diversi output per la stessa query e dallo stesso set di documenti presenti nel sistema.

Alcuni studi parlano anche di **rilevanza documentaria**, intendendo una rilevanza per argomento ma ristretta ai documenti in quanto testi, e di **rilevanza bibliografica** intesa come relazione tra i metadati (per esempio di un catalogo) e l'argomento.

Saracevic propone una lista riassuntiva dei tipi di rilevanza (Saracevic 2007 a p. 1931)

Rilevanza di sistema o algoritmica

É la relazione tra una query e l'informazione in un sistema che viene recuperata, o non recuperata, da una procedura o algoritmo. Ciascun sistema ha metodi e strumenti attraverso i quali i diversi oggetti sono rappresentati, organizzati e confrontati con una query. Include un'assunzione di rilevanza in quanto lo scopo è recuperare un set di oggetti che il sistema deduce come rilevanti rispetto alla query. Nella rilevanza algoritmica il criterio è l'efficacia relativa nel dedurre la rilevanza.

Rilevanza topica o di soggetto

É la relazione tra il soggetto o l'argomento espresso nella query e il soggetto o l'argomento coperto dall'informazione o dagli oggetti informativi recuperati o presenti nel sistema. Si assume che query e oggetti possano essere identificati come relativi ad un argomento. In questo caso il criterio di rilevanza è il concetto di circolarità.

Rilevanza cognitiva o pertinenza

É la relazione tra lo stato cognitivo dei fini conoscitivi dell'utente e l'informazione o gli oggetti informativi recuperati o presenti nel sistema. I criteri attraverso i quali si deduce in questo caso la rilevanza sono la corrispondenza cognitiva, l'attualità, la qualità dell'informazione e altri criteri simili.

Rilevanza situazionale o utilità

É la relazione tra la situazione, i compiti, i problemi che si stanno affrontando e gli oggetti informativi recuperati o presenti nel sistema. I criteri attraverso i quali si deduce in questo caso la rilevanza sono l'utilità per prendere decisioni, l'adeguatezza dell'informazione per risolvere un particolare problema, la riduzione dell'incertezza e simili. Si possono estendere fino ad includere fattori sociali e culturali.

Rilevanza affettiva

É la relazione tra gli intenti, gli scopi, le emozioni, le motivazioni dell'utente e le informazioni recuperate o presenti nel sistema. La rilevanza è dedotta attraverso criteri come soddisfazione, successo, realizzazione e simili. E' evidente che è in relazione con altre manifestazioni di rilevanza, in particolare quella situazionale.

Un altro aspetto da considerare è che cosa renda l'informazione o gli oggetti informativi rilevanti, in particolare quali criteri gli utenti usino per capire che cosa è rilevante e che cosa no. Diversi studi, incomparabili per metodologia e differenti per contesto di osservazione, hanno formulato molteplici schemi e categorie variamente etichettate, tuttavia il set di definizioni e criteri di rilevanza risulta sostanzialmente simile. Da questi studi si evince che i criteri per dedurre la rilevanza sono un numero non elevato e che sono molto simili anche in utenti diversi ma diverso è il valore che le persone assegnano a ciascun criterio a seconda dei compiti che stanno svolgendo e del modo in cui il compito progredisce nel tempo. Questo valore può essere misurato per gradi, anche se tende a polarizzarsi ai due estremi, rilevante o non rilevante (e nei sistemi di IR viene sempre ricondotto a questa forma binaria). I criteri secondo cui la rilevanza viene dedotta possono essere così raggruppati:

- **contenuto:** contenuto, qualità, profondità, estensione, aggiornamento, trattazione, chiarezza
- **oggetto:** caratteristiche degli oggetti informativi cioè tipo, organizzazione, rappresentazione, formato, disponibilità, accessibilità, costo
- **validità:** accuratezza delle informazioni, autorità, attendibilità delle fonti, verificabilità
- **corrispondenza alla situazione o all'utente:** appropriatezza al caso o agli scopi, usabilità, necessità
- **corrispondenza cognitiva:** comprensibilità, aggiornamento, sforzo mentale richiesto

- **corrispondenza affettiva**: risposta emotiva alle informazioni, divertimento, frustrazione, incertezza

- **corrispondenza alle convinzioni personali**: credito dato alle informazioni, fiducia (Saracevic 2007 b p. 2130).

Questi gruppi di criteri non sono indipendenti ma vengono applicati secondo le caratteristiche dell'informazione e della persona (o dei gruppi omogenei di persone) poiché, come in ogni altro processo cognitivo che coinvolga l'elaborazione di informazioni, i fattori individuali sono prominenti. I criteri orientati al contenuto sono di maggiore importanza per l'utente, ma non sono i soli e comunque interagiscono sempre con gli altri criteri. Gli studi hanno rilevato che l'importanza assegnata ai diversi criteri varia se all'utente viene presentato il full text del documento oppure una sua rappresentazione, ma non sono stati in grado di stabilire un modello certo di questa variazione. Un singolo studio ha evidenziato inoltre che le informazioni visuali consentono di inferire la rilevanza più velocemente rispetto alle informazioni testuali. Così come cambiano i criteri per inferire la rilevanza in funzione delle diverse persone o gruppi di persone, anche i giudizi di rilevanza rispetto ad un set di risposte variano: lo stesso set di risposte è giudicato in modo variabile da diversi utenti o gruppi di utenti. Sembra che il giudizio di rilevanza sia tanto più stabile quanto più coloro che giudicano sono esperti della materia.

Un argomento di studio e discussione molto pertinente rispetto al tema degli opac che stiamo trattando è se la valutazione di rilevanza del singolo oggetto informativo sia indipendente o sia influenzata dalla presentazione, che equivale a chiedersi se l'ordine di presentazione e le dimensioni del set di risposta influenzino il giudizio di rilevanza. Gli studi in proposito non sono molti ma se ne possono trarre alcune conclusioni:

- l'ordine di presentazione dei risultati influenza il giudizio: gli oggetti presentati per primi hanno la maggiore probabilità di essere giudicati rilevanti, tuttavia quando il set di risposte è molto ristretto l'ordine di presentazione non è significativo
- i diversi formati dell'informazione (titolo, abstract, termini indicizzati, full text) hanno influenza sulla deduzione di rilevanza; il giudizio di rilevanza cambia se si aggiungono informazioni, per esempio se si passa dal titolo all'abstract o ad altre rappresentazioni supplementari; i titoli non sono importanti tanto quanto gli abstract o i full text nella deduzione di rilevanza .

Un'ultima serie di annotazioni fatte da Saracevic (2007 b p. 2139-2140) può essere interessante: gli studi sul concetto generale di rilevanza sono in declino poiché la maggior parte delle ricerche si sono orientate all'informazione nel contesto dei computer, intraprese da esperti nel campo delle tecnologie informatiche e quindi ben poco sta migliorando nella comprensione del rapporto tra rilevanza umana e rilevanza per un sistema.

L'IR che, prima era legato principalmente al recupero di informazioni associate all'ambito scientifico, tecnico, professionale, è oggi applicato in contesti quotidiani su applicazioni destinate a milioni di persone che hanno bisogno di trovare informazioni rilevanti. Questo ha portato a chiedersi se i criteri di rilevanza da applicare siano gli stessi, se la vastità di questo pubblico possa essere in qualche modo segmentata rispetto al concetto di rilevanza; parallelamente ha evidenziato la necessità di un approccio diverso alla rilevanza nel contesto accademico.

La rilevanza sta diventando sempre più una questione proprietaria. Le tecniche di IR usate nei motori di ricerca sono note nei principi ma proprietarie e quindi sconosciute nei dettagli. Si sa che le aziende conducono propri studi sulla rilevanza ma i risultati non vengono diffusi. E' evidente che stanno cercando di incorporare l'utente e il contesto nelle tecniche di recupero e per farlo devono aver condotto degli studi sul comportamento degli utenti ma dei risultati non abbiamo conoscenza, se non cercando di rintracciare la loro applicazione nei sistemi. Saracevic nota l'ironia della situazione: il Web è salutato come libero universale e democratico ma mantenere proprietarie le ricerche sulla rilevanza è tutt'altro.

3.3 Il relevance ranking negli opac

Gli algoritmi di relevance ranking sono stati sviluppati principalmente per trattare documenti full text. Come abbiamo visto, utilizzano vari criteri associati alla frequenza di un termine e alla sua rarità all'interno di una pagina e nell'insieme di riferimento. Questi criteri possono essere fuorvianti se applicati pedissequamente ai record bibliografici. Nei record bibliografici i campi sono corti, contengono poco testo, alcune informazioni possono essere ripetute senza che questo implichi che siano rilevanti come avverrebbe nel caso di testi più lunghi. Si pensi ad esempio ai nomi degli enti: possono contenere molti termini, essere ripetuti nelle parti descrittive e nelle intestazioni autore magari in più forme varianti, in sostanza presentare ripetute occorrenze dello stesso termine senza che questo renda quel record più significativo di altri in cui lo stesso termine è meno presente. La struttura dei

record bibliografici fornisce molti dati che possono essere usati per il relevance ranking ma questi devono essere letti in modo adeguato (Dellit - Boston 2007 p. 5). L'applicazione intelligente dei criteri di ranking ai record bibliografici può essere una risposta alle posizioni di Bade sull'uso della ricerca per parole chiave. Secondo Bade nei cataloghi ogni elemento di un record rappresenta l'interpretazione e il giudizio di un essere umano; l'intestazione di soggetto non è altro che un'asserzione di rilevanza fatta dal catalogatore secondo una tassonomia predefinita; il formato stesso è pensato per massimizzare i risultati di una ricerca secondo il tipo di informazione desiderata: autore, titolo, soggetto. Usare la ricerca per parole chiave negli opac non è quindi sensato:

Using a general keyword search ignores the information supplied by acts of intelligence isomorphic with the searchers intelligence (i.e. not by statistical weighing of data but by judgements based on human knowledge and experience in information seeking); using algorithms to weigh the results is simply an attempt to produce a substitute for that judgement using methods different in kind (Bade 2007 p. 836)⁵⁵.

Una critica alla ricerca per parole chiave come sostituto della ricerca attraverso metadati è presente anche in Beall. Secondo questo autore l'applicazione nei sistemi di retrieval della ricerca per parole chiave, che è avvenuta negli ultimi 15 anni, ha avuto come conseguenza una diminuzione della qualità dell'IR deleteria specie nel contesto accademico perché questo tipo di ricerca produce risultati inaffidabili, incompleti, insidiosamente imprecisi non essendo in grado di padroneggiare molti aspetti linguistici (sinonimia, omonimia, varianti ortografiche, abbreviazioni, diversità di lingua, obsolescenza dei termini) e di altro genere (impossibilità di generare raggruppamenti, incapacità di selezionare per faccette) non ultimo l'individuazione del contenuto semantico dei documenti (Beall 2008).

Saper usare correttamente i metadati bibliografici per creare specifici algoritmi di relevance ranking consentirebbe di lasciare all'utente la possibilità di usare la ricerca libera, anziché imporgli la ricerca per campi, ma nello stesso tempo sfruttare al massimo le attività intelligenti svolte durante la creazione del record: alcune applicazioni vanno in questa direzione. Nel 2006 la National Library of Australia ha svolto dei test per introdurre modifiche nell'opac che gestisce Libraries Australia, un catalogo di circa 16 milioni record

⁵⁵Usando una ricerca generale per parole chiave si ignorano le informazioni fornite da atti di intelligenza isomorfa all'intelligenza del ricercatore (cioè non dal peso statistico dei dati ma dal giudizio basato sulla conoscenza ed esperienza umana nella ricerca di informazioni); usare algoritmi per pesare i risultati è semplicemente un tentativo di produrre un sostituto di quel giudizio usando metodi di diverso tipo.

bibliografici di documenti di diversi generi e supporti tra i quali anche materiali unici. Lo scopo principale era creare uno strumento che rendesse facile la ricerca e l'ottenimento dei documenti: l'introduzione del relevance ranking come forma di presentazione dei risultati è stata la prima tappa del percorso. Per ottenere buoni risultati di ordinamento sono stati individuati criteri di rilevanza basati sui record MARC e su alcune informazioni gestionali, criteri che sono stati via via perfezionati attraverso un triplice passaggio di applicazione e verifica. La verifica è consistita nel sottoporre una serie di risultati estratti dal sistema al giudizio di rilevanza espresso da bibliotecari esperti nei servizi di reference. Questi sono i criteri di base usati:

- 1) le corrispondenze nei campi titolo, autore, soggetto e nei campi che descrivono la natura o la forma di un documento sono più significative che la corrispondenza in generale nel record
- 2) la corrispondenza in più di uno dei campi precedenti è più significativa di una sola corrispondenza
- 3) la corrispondenza dell'intera frase di ricerca in un unico campo del record è più importante della corrispondenza delle singole parole in vari campi
- 4) il peso da attribuire alle corrispondenze del punto 3 è diverso se la corrispondenza avviene nel campo titolo, autore o soggetto: mentre nel campo titolo l'esatta corrispondenza (exact match) è molto più significativa della corrispondenza delle stesse parole ma non nella stessa sequenza o della corrispondenza di alcune singole parole (phrase match e word match), questa differenza è meno significativa per il campo soggetto
- 5) le corrispondenze nei campi principali sono più importanti, per esempio viene attribuito più peso alle corrispondenze presenti nel campo del titolo proprio rispetto al campo del titolo alternativo
- 6) le corrispondenze dei termini immessi dall'utente sono più importanti delle corrispondenze ottenute dalla correzione dell'ortografia o dall'applicazione di tecniche di stemming
- 7) a parità delle altre condizioni i record a livello di collezione sono più rilevanti
- 8) a parità delle altre condizioni documenti che sono presenti in più copie sono da considerarsi più rilevanti (Dellit – Boston 2007 p. 6).

Lo studio individua come significative per impostare i criteri di rilevanza anche altri tipi di corrispondenze che tuttavia non è stato fino ad ora possibile incorporare: le corrispondenze tra l'utente e il titolo, per esempio quando l'età dell'utente corrisponde alla fascia di età cui

il documento è destinato; le informazioni sulla disponibilità in loco del documento; informazioni derivanti da citazioni; informazioni generate dagli utenti come i dati di circolazione, le recensioni, i tag o i suggerimenti degli utenti.

E' interessante notare che molti di coloro che hanno testato l'opac hanno suggerito di mantenere anche la ricerca per campi come forma avanzata di ricerca mostrando di condividere l'opinione diffusa che le due forme di ricerca rappresentino un approccio mutualmente esclusivo: si potrebbe invece pensare all'applicazione dell'ordinamento per rilevanza anche alle ricerche per campi. Inoltre non è detto che le ricerche condotte attraverso la ricerca per campi non si possano eseguire attraverso la ricerca per parole ottenendo gli stessi buoni risultati (Dellit – Boston 2007 p. 8). Va notato anche che nelle fasi di test diversi bibliotecari hanno talvolta espresso giudizi contrastanti sulla rilevanza dei diversi set di risposta mostrando interpretazioni differenti di cosa sia significativo per l'utente: una conferma di quanto detto in precedenza sul concetto rilevanza umana e sui fattori non oggettivi che determinano il giudizio⁵⁶.

Criteri simili a quelli esposti sono stati adottati nell'opac NCSU (North Carolina State University). Anche in questo opac, basato sul software Endeca IAP⁵⁷, gli algoritmi di ranking danno priorità ai risultati che corrispondono esattamente alla query, mentre correzioni ortografiche e troncamenti hanno meno peso; se vengono immessi più termini i risultati che contengono la frase esatta sono considerati più significativi; è stata creata una lista di priorità dei campi che consente di definire più significative le corrispondenze che avvengono in certi campi piuttosto che in altri. In questo sistema è possibile definire per ogni indice di ricerca diversi criteri di ranking, cosa che si comprende immediatamente significativa se si pensa alla differenza che può esserci tra cercare per parola chiave in tutti i campi o cercare per ISBN, ma la maggior parte delle attenzioni si sono concentrate sulla strategia di ranking dei risultati della ricerca libera. L'algoritmo di ranking consiste in sette moduli, i primi cinque che indicizzano i risultati in modo dinamico, gli ultimi due che ordinano in base alla data di pubblicazione e alla maggiore o minore circolazione dei documenti. Anche in questo caso si riconosce che la particolare natura dei dati bibliografici

⁵⁶ Ulteriori informazioni sul rankign applicato dalla biblioteca nazionale australiana si trovano anche in Set of rules for Lucene relevance ranking <<http://1101.nla.gov.au/docs/LuceneRRNotes.html>> citato da Bergamin (2008).

⁵⁷ Vedi <<http://www.endeca.com/products-information-access-platform.htm>>.

richiede un approccio per prove ed errori per arrivare ad una configurazione che dia risultati soddisfacenti (Antelman - Lynema - Pace 2006 p. 129-130).

Un approccio diverso al ranking dei materiali librari è suggerito da Lewandowski (2009 p. 588-591). I criteri generalmente usati nel relevance ranking vengono da lui ricondotti a quattro categorie per ciascuna delle quali individua gli aspetti da considerare nel caso degli opac. L'individuazione e l'applicazione di questi criteri però non esaurisce il lavoro perché altri aspetti problematici nascono dalla provenienza dei dati da diverse fonti: non tutte le risorse della biblioteca sono presenti nel catalogo e devono comunque essere prese in considerazione. Analogamente molte risorse possono essere presenti nella forma di link a fonti esterne (si pensi agli abstract o ai full text) e questo rende difficile ottenere e gestire i dati da usare nel ranking e anche generare un set di risposta significativo, che dovrebbe comunque essere costituito da un mix delle diverse tipologie di documenti. Lewandowski propone anche di considerare il tipo di query: se è generica dovrebbe presentare nel set di risposta principalmente opere di reference, libri di testo, materiale corrente insieme ai principali giornali e database; se è specifica avrebbe più senso mostrare principalmente articoli di giornale. Omette tuttavia di indicare in che modo il sistema dovrebbe rilevare la genericità o specificità della query, elemento relativo che dipende da fattori soggettivi non deducibili dai termini di ricerca immessi su cui si formula invece la rilevanza di sistema. Di seguito sono i criteri di ranking individuati da Lewandowsky (2009 p. 588).

Corrispondenza testuale: come abbiamo già detto il problema principale della corrispondenza testuale nei cataloghi è la scarsità di testo presente nei metadati che conduce a risultati di ranking scadenti. Secondo questo autore il punto più problematico è costituito dalla differente quantità di testo presente nei diversi record che può variare dalla semplice descrizione bibliografica, ai dati arricchiti, al full text: applicare lo stesso algoritmo di ranking a tipi diversi di record non produce buoni risultati. Egli propone di esaminare i tipi di record separatamente e generare diverse liste di ordinamento da integrare poi in un'unica lista di risultati.

Popolarità: la popolarità può essere considerata dal punto di vista del singolo documento o sulla base del gruppo cui il documento appartiene, per esempio le opere dello stesso autore o i libri della stessa collana o dello stesso editore. Può essere elaborata a partire dai dati di circolazione dei documenti, come il numero di prestiti o di downloads. Se sono presenti nel sistema, si possono usare anche le citazioni o i voti espressi dagli utenti. Questi

dati sono tutti indipendenti dalle query e possono essere calcolati in anticipo in modo da non influire negativamente sui tempi di risposta, un fattore determinante per la qualità del sistema. Anche per i dati che sono costantemente aggiornati, come quelli provenienti dagli utenti, si può prevedere un aggiornamento periodico.

Novità: questo criterio è il più applicato negli opac, semplicemente ordinando i documenti per data. Ma la novità o l'attualità di un documento non dipendono esclusivamente dalla sua data di stampa. E' evidente che per un libro che si occupa di informatica l'attualità è un fattore essenziale mentre per un libro di filosofia il fattore tempo potrebbe non essere così determinante e bisognerebbe tenerne conto negli algoritmi di ranking magari attribuendo in questo caso più peso ai criteri di corrispondenza testuale. Si tratta in sostanza di determinare il grado di novità necessario per diversi tipi di documenti: questo potrebbe essere fatto dai dati di circolazione del singolo documento oppure considerando i dati di circolazione di un gruppo di documenti accomunati dall'appartenere alla stessa disciplina o dall'aver lo stesso soggetto.

Località: questo criterio dovrebbe prendere in considerazione la posizione fisica dell'utente e la disponibilità dei documenti presentati nella lista dei risultati: un documento presente nella biblioteca locale dovrebbe essere posizionato più in alto rispetto ad uno disponibile presso una biblioteca più lontana e dovrebbe anche essere valutata la sua prestabilità perché la disponibilità per la sola consultazione in sede potrebbe renderlo poco interessante per alcuni utenti. La localizzazione fisica dell'utente può essere determinata attraverso l'indirizzo IP del suo computer. Si può assumere che utenti che si connettono da posizioni esterne alla biblioteca possano essere maggiormente soddisfatti da documenti in formato elettronico di cui possono fare il download: questa restrizione dovrebbe essere applicata tra i criteri di ranking per questi utenti mentre non avrebbe senso per gli utenti presenti in biblioteca per i quali anche i documenti a stampa dovrebbero entrare nella lista dei risultati (Lewandowski 2009 p. 590). L'applicazione di questo criterio sembra essere particolarmente insidiosa in quanto mescola due distinti passi della ricerca: il momento dell'individuazione dei documenti con l'ottenerli fisicamente. Mescola anche criteri di tipo più oggettivo con criteri di utilità, e quindi legati alla soggettività, senza avere le informazioni sull'utente necessarie per farlo. Assumere che l'utente sia interessato ai soli documenti che può avere immediatamente mi sembra arbitrario e anche inutile quando si può ottenere lo stesso risultato con il raffinamento volontario della ricerca successivo al

ranking dei risultati, magari con una presentazione a faccette tra cui la disponibilità (locale o remota) o il formato dei documenti (elettronico o a stampa).

Lewandowski suggerisce anche altri criteri che possono essere utilizzati negli algoritmi di ranking molti dei quali dovrebbero essere pesati attentamente a seconda del contesto: dare più peso alle monografie o agli articoli di rivista, ai materiali on line o ai libri può essere più o meno significativo a seconda dei casi. Molto significativo potrebbe essere il ranking dei risultati personalizzato in base alla classe di utenza (per esempio distinguendo tra i diversi tipi di studenti, i ricercatori, i professori) rendendo gli utenti consapevoli di quali dati personali siano raccolti a questo fine e quali restrizioni applicate (Lewandowski 2009 p. 590). Questa possibilità porta con sé tutte le problematiche tecniche relative alla profilatura degli utenti: dalla raccolta dei dati dalle navigazioni degli utenti, alla necessità di renderli anonimi⁵⁸, alla loro conservazione e non ultima la difficoltà di combinare e pesare correttamente i dati utenti e i dati relativi alle collezioni in criteri di ordinamento significativi. Problematici sono anche alcuni aspetti generali circa i vantaggi, o gli svantaggi, della profilatura: non si può negare che gli utenti possano voler fare ricerche secondo stili diversi da quelli per i quali sono profilati nel sistema e che la presentazione di risultati selezionati in base a precise caratteristiche personali possa portare alla perdita di quei fortunati incontri casuali definiti serendipità. Potrebbe inoltre avere un senso anche in questo contesto il timore espresso da Cass Sunstein relativamente alla possibilità filtrare tutte le nostre fonti di informazione selezionando solo quelle che sono più affini ai nostri interessi e alle nostre opinioni: l'opportunità di creare il nostro profilo personale offerta dai diversi siti Web di informazione o di scegliere solo particolari canali televisivi specializzati negli argomenti di nostro interesse sembrerebbe un grande vantaggio data la sovrabbondanza di informazioni e di stimoli di cui siamo destinatari ma può rivelarsi negativa perché, evitando di esporci a tutto, selezionando le fonti, limita i nostri contatti alle sole opinioni simili alla nostra, polarizza e radicalizza le posizioni, ci impedisce di partecipare ad una vera conversazione democratica (Sunstein 2003).

⁵⁸ La necessità di rendere anonimi i dati nasce dall'opportunità di tutelare la privacy degli utenti. Questa problematica è discussa in Whitney-Schiff (2006) e Levine (2007). Altri aspetti di valutazione si possono trarre da OCLC (2007).

Nel contesto degli opac e della profilatura per fasce di utenti questo potrebbe voler dire chiudere l'utente nell'ambito disciplinare che ha indicato di suo interesse oppure dargli visione dei soli testi che si adattano al suo livello di specializzazione o ai corsi di studio cui è iscritto e così via. Il vantaggio di ridurre i tempi di selezione e valutazione affidandoli a sistemi automatici potrebbe divenire un limite alle opportunità di entrare in contatto con molti documenti e in sostanza essere un ostacolo alla possibilità di conoscere. E' quindi fondamentale che l'utente sia consapevole del modo in cui sta impostando la ricerca e in questo dovrebbe essere aiutato da un sistema che consenta di scegliere tra diverse opzioni possibili in ogni sessione di ricerca piuttosto che applicare algoritmi di ranking basati sul profilo utente applicati senza che l'utente ne sia avvertito.

Un caso in cui il ranking dei risultati non è il modo più efficace di risolvere il problema della grande dimensione del set di risposta è quello delle opere letterarie o musicali o comunque delle opere pubblicate in diverse edizioni e formati. In un catalogo collettivo, anche di dimensioni medie, l'interrogazione sul titolo di un'opera letteraria o su un autore importante dà come risultato centinaia di occorrenze. L'ordinamento secondo rilevanza in questi casi risulta poco coerente, mescolando ad esempio il testo dell'opera con i suoi commenti, l'autore con i suoi chiosatori o interpreti. La faccettazione dei risultati, che vedremo in seguito, può aiutare nella selezione raggruppando il set di risposta secondo i nomi degli autori, consentendo così di distinguere gli autori dai curatori o dagli interpreti, o secondo la lingua di traduzione o i formati, ma la soluzione migliore sarebbe mostrare i risultati secondo la logica di FRBR⁵⁹ distinguendo l'opera (per esempio *Le avventure di Pinocchio*) dalle opere derivate (il film Disney o il film di Comencini) e suddividendola nelle sue diverse espressioni (il testo in lingua originale, la sua traduzione in inglese, la sua traduzione in francese) quindi nelle differenti manifestazioni (le diverse edizioni delle singole espressioni) fino ad arrivare al singolo documento posseduto dalle diverse biblioteche. Purtroppo gli attuali record bibliografici, anche quelli basati sugli ultimi sviluppi del formato cioè su MARC21⁶⁰, non contengono le informazioni necessarie a realizzare questo tipo di visualizzazione in tutte le sue articolazioni. OCLC, che ha studiato il problema e

⁵⁹ Per una introduzione ai concetti di FRBR si veda Ghilli (2001) oppure si vedano le pagine pagine dell'IFLA <<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>>.

⁶⁰ Vedi <<http://www.loc.gov/marc/>>.

realizzato un prototipo che mostra come potrebbe essere una visualizzazione basata su FRBR⁶¹, ha per ora indirizzato i suoi sforzi nella creazione di raggruppamenti almeno a livello di opera. Questi sono ottenuti attraverso l'applicazione di un algoritmo⁶² che, dopo aver normalizzato nomi degli autori e titoli, estrae una chiave di ricerca autore/titolo per ciascuno dei record presenti nel catalogo, chiavi che vengono poi usate per creare i raggruppamenti a livello delle opere. Questi algoritmi sono utilizzati nel catalogo World Cat e sono disponibili gratuitamente per essere applicati in altri contesti.

3.4 Classificazione a faccette e faceted browsing

La definizione del concetto di faccette applicato alla scienza dell'informazione è generalmente attribuita al bibliotecario e matematico indiano Shiyali Ramamrita Ranganathan (1892-1972). Egli rilevò la rigidità delle classificazioni gerarchiche che hanno l'ambizione di organizzare tutta la conoscenza secondo un albero semantico che si ramifica a partire da alcune discipline di base: in questo tipo di classificazioni ogni documento-concetto è collocato in un unico punto di uno schema predeterminato che in genere è molto ampio e gerarchicamente profondo. Intorno agli anni 30 del 1900 Ranganathan propose come alternativa una classificazione più flessibile, che consentisse di descrivere semanticamente e organizzare i documenti secondo categorie tipiche delle diverse discipline cioè secondo i loro diversi aspetti che chiamò faccette⁶³. Secondo Ranganathan ogni faccetta di qualsiasi disciplina può essere ricondotta a cinque categorie base, ciascuna delle quali rappresenta un elemento del concetto descritto: personalità (l'oggetto centrale di un discorso), materia (i componenti e le proprietà dell'oggetto), energia (le caratteristiche dinamiche dei processi che lo interessano), spazio (i suoi elementi geografici o in genere spaziali) e tempo (le sue fasi cronologiche). Ogni faccetta può assumere valori diversi che vengono chiamati fuochi. Le varie faccette devono poi essere rappresentate insieme secondo un preciso ordine di citazione (Gnoli 2000, Gnoli – Marino - Rosati 2006 p. 22-26). Questa classificazione ha il vantaggio di poter essere declinata (potenzialmente) all'infinito aggiungendo oggetti e concetti man mano che si presentano, superando le rigidità degli schemi enumerativi come

⁶¹ Vedi Fiction Finder <<http://fictionfinder.oclc.org/>>.

⁶² Vedi <http://www.oclc.org/research/activities/past/orprojects/frbralgorithm/frbr_workset_algorithm.pdf>.

⁶³ L'ispirazione a studiare un diverso tipo di classificazione nacque nel 1924 da un gioco matematico. La prima edizione della classificazione è del 1933, la sesta, l'ultima da lui curata, è del 1960.

la Classificazione Decimale Dewey o Library of Congress Classification, che hanno difficoltà a incorporare nuovi concetti e discipline all'interno di una struttura rigidamente definita⁶⁴. La classificazione a faccette, che non è riuscita a rimpiazzare i sistemi gerarchici nelle biblioteche, dove fungono anche da metodo di collocazione, è servita da base teorica e ha trovato applicazione in diversi tesauri specialistici che, senza adottare necessariamente i fattori proposti da Ranganathan, considerano le faccette semplicemente come raggruppamenti di termini della stessa natura, o suddivisioni di campi più ampi (Maisonneuve - Toutilou 2007 p. 13). Nel contesto conoscitivo attuale caratterizzato da una grande mole di informazioni e dalla rapidità dei cambiamenti, in cui un'organizzazione flessibile e multidimensionale dell'informazione è una necessità crescente, il concetto di faccette e classificazione a faccette proposto dalla biblioteconomia classica sta tornando in auge perché consente di realizzare sistemi aperti e adattivi. Le faccette possono essere considerate come un sistema di attributi (metadati) mutuamente esclusivi ciascuno dei quali rappresenta un aspetto o una proprietà persistente dell'oggetto e nel loro insieme lo descrivono in modo esaustivo. I principali tratti distintivi di un sistema a faccette sono così sintetizzabili:

- pluridimensionalità: inversamente ai sistemi tradizionali, nei sistemi a faccette, ogni oggetto è classificato secondo una pluralità di attributi (faccette)
- persistenza: tali attributi/faccette costituiscono proprietà essenziali e persistenti dell'oggetto; in questo modo l'impatto sullo schema di classificazione di eventuali cambiamenti è fortemente ridotto o nullo
- scalarità: è sempre possibile aggiungere una nuova faccetta descrittiva di un nuovo aspetto dell'oggetto
- flessibilità: esiste una pluralità di chiavi di accesso parallele (faccette); ogni oggetto può essere reperito utilizzando un singolo attributo di ricerca alla volta, oppure più attributi insieme in combinazione
- non è necessario conoscere il nome della/e categoria/e in cui l'oggetto è collocato

⁶⁴ Va detto che le ultime edizioni della CCD stanno aumentando la struttura a faccette delle tavole specie in alcuni ambiti come la musica, l'amministrazione pubblica, le scienze della vita (Paradisi 2003). Va inoltre precisato che anche nelle classificazioni a faccette l'inserimento di una nuova faccetta comporta qualche difficoltà perché implica la riclassificazione tutte le entità secondo la nuova faccetta, mentre l'introduzione di nuovi fuochi deve comunque rispettare il principio di reciproca esclusività e quindi presenta le stesse problematiche di inserimento e riclassificazione degli schemi gerarchico enumerativi (Gnoli-Marino-Rosati 2006 p. 72).

- il criterio di ordinamento non è necessariamente gerarchico (dal generale al particolare, dall'intero alla parte etc.) ma può anche essere alfabetico
- è possibile costruire relazioni persistenti fra le diverse faccette, tali da fornire una rappresentazione della conoscenza contenuta nel sistema stesso (Rosati 2003 p. 4).

In ambiente digitale le caratteristiche delle faccette, in particolare la capacità di essere usate come attributi di ricerca sia singolarmente sia in combinazione, per fornire una descrizione più precisa dell'oggetto o per limitare i risultati di una ricerca o per descrivere relazioni tra oggetti, le rendono particolarmente efficaci per il recupero dell'informazione (Rosati 2003). Questo sistema consente, infatti, di rispondere alle interrogazioni di una pluralità di utenti, che hanno competenze e motivazioni diverse, con una flessibilità maggiore di quella consentita da sistemi ordinati gerarchicamente o in cui la combinazione degli elementi avviene in fase di indicizzazione, come nel caso della catalogazione per soggetti.

Dal concetto di classificazione a faccette si giunge oggi a quello di navigazione a faccette (faceted browsing o faceted navigation). Sin dai suoi inizi il Web ha offerto due paradigmi di ricerca: la navigazione attraverso strutture gerarchiche che consentono all'utente di esplorare risorse organizzate restringendo ripetutamente il campo di ricerca, tipica delle directory, contrapposta alla ricerca diretta attraverso parole chiave, resa popolare dai motori di ricerca. La navigazione a faccette è un nuovo approccio che sembra mettere insieme i due precedenti: consente all'utente di combinare la ricerca per parole con l'esplorazione dei diversi aspetti delle informazioni che possono essere esaminati attraverso raffinamenti successivi. Gli scopi di una navigazione a faccette dovrebbero essere proprio dare flessibilità alla ricerca integrando la ricerca diretta per parole chiave con strumenti che sostengano l'esplorazione; consentire un passaggio fluido dalle attività di raffinamento a quelle espansione della ricerca; mantenere gli obiettivi di qualsiasi tipo di ricerca, cioè evitare i set di risposta vuoti e dare all'utente una sensazione di controllo dell'interrogazione e comprensione dell'ambiente e dei risultati (Hearst 2006). I siti commerciali⁶⁵ sono quelli che per primi hanno adottato questo approccio che si sta ora diffondendo anche negli opac delle biblioteche i quali arrivano a riappropriarsi di queste forme di organizzazione e presenta-

⁶⁵ Accanto al solito Amazon (www.amazon.com) sono spesso citati come esempio di faccettazione i siti di ricette gastronomiche (come www.Epicurios.com) o di vini (come www.Wine.com).

zione delle informazioni nate dalla tradizione biblioteconomica. Va precisato che questi sistemi di ricerca, anche quelli introdotti negli opac delle biblioteche, non si basano necessariamente su una classificazione a faccette come quella che abbiamo descritto. Come sottolineato da Gnoli – Marino - Rosati (2006 p. 97), la classificazione a faccette nella sua formulazione originaria associa all'accesso pluridimensionale agli oggetti di una collezione altri aspetti importanti, come l'ordine di citazione delle faccette e una notazione in grado di produrre un ordine utile per il browsing. Inoltre per poter parlare di classificazione ci si deve riferire agli aspetti semantici degli oggetti classificati. Nei sistemi di navigazione a faccette oggi in uso viene adottata una logica multidimensionale simile a quella delle classificazioni a faccette ma non se ne utilizzano le altre caratteristiche, così come le faccette non si riferiscono solo ai contenuti semantici ma anche, e in molti casi esclusivamente, alle proprietà formali di un oggetto, alle sue caratteristiche fisiche. E' bene quindi distinguere tra i diversi tipi di classificazioni allo scopo di cogliere le implicazioni nell'uso dell'una o dell'altra: **classificazioni a faccette pure**: si rifanno in modo completo alla teoria della classificazione classica e ai suoi sviluppi e prevedono l'accesso all'informazione mediante più dimensioni semantiche (faccette), l'ordine conveniente delle faccette, un sistema di notazione utile a garantire tale ordine; **classificazione a faccette spurie**: si rifanno alla teoria delle faccette soltanto o principalmente per ciò che riguarda la logica pluridimensionale di accesso all'informazione trascurando l'ordine di citazione e gli aspetti di notazione; **classificazioni apparenti**: sono modelli di organizzazione dell'informazione di tipo relazionale che offrono un accesso all'informazione attraverso più parametri ma i parametri, o metadati, usati sono prevalentemente di tipo descrittivo e non di tipo semantico (Gnoli – Marino - Rosati 2006 p. 99).

3.5 Le faccette negli opac

L'introduzione della navigazione a faccette negli opac sta avvenendo contestualmente alla presentazione della finestra unica di ricerca attraverso la quale l'utente fa una interrogazione per parole. Come abbiamo visto gli opac stanno progressivamente passando in secondo piano la ricerca per campi e la cosiddetta ricerca avanzata proponendo interfacce semplificate, più orientate agli utenti e meno ai professionisti dell'informazione, che non richiedono alcuna conoscenza della struttura dei dati o della logica booleana ma favoriscono l'espressione delle interrogazioni in linguaggio naturale, interfacce alle quali gli utenti

sono abituati per l'uso massiccio che ne fanno sui motori di ricerca. La navigazione a faccette risponde alla necessità di gestire in modo significativo la grande quantità di dati che si ottiene da questo genere di interrogazioni: riduce il rumore che deriva da ricerche troppo generiche organizzando il set di risposta in raggruppamenti creati dinamicamente sulla base delle caratteristiche dei documenti di quel particolare set di risposta. Questi raggruppamenti, accompagnati da un numero che rappresenta la quantità di risorse collegate, costituiscono una presentazione secondo diversi punti di vista (faccette) dei documenti estratti e hanno il vantaggio di dare una visione sintetica ma comprensibile dei documenti, di consentire il raffinamento della ricerca sulla base di opzioni effettivamente presenti, di suggerire strategie di raffinamento della ricerca alle quali l'utente potrebbe non aver pensato. Le faccette sono categorie usate per caratterizzare un oggetto all'interno di insieme; ogni faccetta ha un nome (per esempio Formato può essere una faccetta significativa se stiamo parlando di documenti, Ingredienti può essere una faccetta se stiamo parlando di ricette di cucina); una faccetta può essere piatta (come il nome di un autore) oppure gerarchica (la faccetta Formato può includere Libri, Articoli, DVD, Dischi). Ad ogni oggetto-documento nel sistema vengono associate una o più etichette relative a diverse faccette (un documento può avere una etichetta di formato, di data, di autore, di soggetto, di disponibilità ecc.). Quando l'utente seleziona un'etichetta di una particolare faccetta il sistema recupera tutti i documenti ai quali quella etichetta è stata assegnata. Se seleziona diverse etichette di differenti faccette il sistema costruisce una query che include le etichette scelte escludendo le altre della stessa faccetta (Hearst 2006). A questo corrisponde una visualizzazione per l'utente: la scelta dell'elemento di una faccetta modifica l'intera visualizzazione dei dati per mostrargli quali elementi soddisfano la sua scelta e quali scelte sono ancora possibili: tutte le faccette e le etichette che vengono mostrate contengono degli elementi, corrispondono a dei documenti, mentre le opzioni che non contengono elementi non vengono più mostrate. Com'è facilmente intuibile, la navigazione a faccette, seducente in linea di principio, fonda il suo successo su una scelta delle faccette accurata e pertinente al contenuto del catalogo (Maisonneuve - Touitou 2007 pag. 15). Gli opac, alla stregua dei siti commerciali, non basano la navigazione a faccette su una preesistente classificazione a faccette, né procedono ad una nuova catalogazione al fine di utilizzare questi strumenti, ma utilizzano i metadati di tipo bibliografico e gestionale già presenti nel sistema per trarne delle categorie (faccette) che possano rappresentare il contenuto del catalogo attraverso

raggruppamenti significativi. Si tratta dunque di vedere quali faccette vengono proposte per la navigazione, come le etichette vengono desunte dai metadati presenti nei sistemi e, non ultimo, la quantità e la coerenza dei metadati⁶⁶. Diversi opac mostrano strategie differenti: mentre alcune faccette sono presenti in tutti gli opac altre sono frutto di scelte specifiche. Inoltre alcuni opac traggono i metadati solo dal catalogo locale mentre altri includono anche dati provenienti dalla ricerca federata o da dati di arricchimento bibliografico provenienti da fonti esterne⁶⁷.

Possiamo dividere le faccette in base alla provenienza dei metadati⁶⁸:

Dati ricavati dalla descrizione bibliografica: autore, data di pubblicazione, collana, titolo.

Dati ricavati da codifiche immesse: lingua del documento, formato, età di lettura.

Dati provenienti dalla catalogazione per soggetti: argomento, nome di persona o ente, luogo, data-periodo storico.

Dati provenienti dal sistema di gestione:

localizzazione (nel caso di opac collettivi indica le biblioteche che posseggono documenti), fonte dei documenti (indica i diversi silos da cui le informazioni sono prelevate sia localmente, come specifiche raccolte, sia esternamente come basi dati diverse), fonte delle etichette (indica se le etichette sono state cercate negli autori, nei soggetti o nei titoli) tipo di utilizzo (indica i documenti disponibili per il prestito o per la consultazione, può indicare anche i documenti effettivamente presenti al momento dell'interrogazione), raccomandazioni (possono essere tratte da consigli degli utenti o dai dati di circolazione indicando in questo caso i documenti più prestati), tag (generati dagli utenti o importati da fonte esterne), top level (indica le etichette che hanno il maggior numero di documenti associati), novità (indica i documenti più recentemente immessi nel sistema).

⁶⁶ Per esempio l'attribuzione del codice di genere o la fascia di età di destinazione è una pratica introdotta recentemente dalle biblioteche, rispetto a metadati standard come il nome dell'autore o la data. I documenti estratti scegliendo queste etichette saranno solo una parte rispetto a tutti i documenti che possiedono quella particolare caratteristica ma non sono stati etichettati. Queste difformità possono generare visioni distorte del contenuto del catalogo.

⁶⁷ Molte biblioteche acquistano riassunti, indici, note sugli autori, recensioni, primi capitoli da Syndetic Solution <<http://www.bowker.com/syndetics/>>; OCLC importa dati da WeRead <<http://weread.com/>> e Amazon <<http://www.amazon.com/>>.

Tutte le faccette, anche quelle che riflettono i tradizionali punti di accesso (autore, titolo e soggetto) necessitano di una notevole manipolazione dei dati per arrivare alla semplicità e coerenza di presentazione che lo stile della faccettazione richiede. Le faccette che vanno oltre gli aspetti descrittivi e si focalizzano sulla disponibilità in tempo reale delle risorse, su come ottenerle, presentano problematiche ancora maggiori. Le informazioni che vengono usate in queste faccette non provengono dai record bibliografici, che sono relativamente stabili e per i quali è sufficiente trasferire i dati con procedure batch notturne, ma dai dati gestionali che sono più dinamici e richiedono particolari procedure di estrazione e manipolazione dei risultati quando l'opac non è integrato col sistema gestionale, cosa che accade sempre più spesso, e quando l'opac è il catalogo collettivo di un insieme di biblioteche⁶⁸. Anche faccette meno dinamiche e apparentemente semplici come Disponibilità (prestabile, consultabile, online) e Localizzazione (lista delle biblioteche che possiedono il documento) richiedono un esame approfondito. Per esempio: le risorse online devono essere posizionate tra le etichette di Disponibilità o di Localizzazione? Se vengono posizionate nella Localizzazione saranno confuse in mezzo ad una lista di biblioteche e inoltre un utente che seleziona una biblioteca per restringere la ricerca, non troverà nel set di risposta nessuna risorsa disponibile online. Viceversa metterle nell'etichetta Disponibilità significa moltiplicare per il numero di biblioteche l'informazione che quella risorsa è disponibile online, anche se logicamente non dovrebbe essere necessario, perché se questo non viene fatto e l'utente opera una restrizione sulla faccetta Localizzazione, quella risorsa non gli apparirà nel set di risposta benché sia effettivamente disponibile per gli utenti di quella biblioteca (Denholm – Kauler – Lavelle – Sokvitne 2009).

⁶⁸ I dati sono ricavati da Maisonneuve-Touitou (2007) e da Tosca Consultant (2008) e prendono ad esempio gli opac delle biblioteche pubbliche di Nashville e di Phoenix, della Biblioteca reale di Danimarca, delle biblioteche pubbliche della Gorgia, della biblioteca pubblica della contea di Athens in Ohaio, della biblioteca del Queens di New York che usano i software Ancore, Endeca, Primo, Evergreen, Koha, Aquabrowser.

⁶⁹ Queste problematiche sono presenti anche nelle presentazioni dei dati più tradizionali con la differenza che la presentazione a faccette richiede di elaborare dati per un gruppo di documenti contemporaneamente aumentando gli aspetti critici. Per un'analisi delle diverse tecniche di accesso ai dati si veda la parte introduttiva del rapporto di Tosca Consultants (2008).

<p>Bibliothèque publique de Nashville (E-U) Encore http://nplncore.library.nashville.org</p> <p>Refine by:</p> <ul style="list-style-type: none"> SEARCH FOUND IN Author (991) Subject (882) Title (838) FORMAT PRINTED MATL (1386) VIDEORECORDING (107) DVDS (98) BOOKS ON TAPE (84) NETLIB EBOOKS (56) COLLECTION Bellevue (101) Bookmobile (16) Borckaus (258) Donelson (121) East (51) LANGUAGE English (1818) Spanish (13) French (8) Italian (4) Japanese (3) PUBLISH DATE 2007 (16) 2006 (36) 2005 (48) 2004 (65) 2003 (76) more 	<p>Bibliothèque publique de Phoenix (E-U) Endeca www.phoenixpubliclibrary.org</p> <p>YOUR SEARCH</p> <p>RECOMMENDED</p> <ul style="list-style-type: none"> Most Borrowed Staff Picks Customer Picks N.Y. Times Best Seller Pearl's Picks <p>FORMAT</p> <ul style="list-style-type: none"> Book DVD VHS Downloadable eBook Music (CD) <p>AGE LEVEL</p> <ul style="list-style-type: none"> Adult Children Teen <p>LOCATION</p> <ul style="list-style-type: none"> Acacia Burton Bair Century Cesar Chavez Cholla <p>LANGUAGES</p> <ul style="list-style-type: none"> English Spanish Italian 	<p>Bibliothèque royale du Danemark Primo www.kb.dk</p> <p>Refine My Results</p> <p>Narrow my results by limiting the search to:</p> <p>Top level:</p> <ul style="list-style-type: none"> Online Resources (347) Available (4,636) New in 2007 (642) <p>Resource Type:</p> <ul style="list-style-type: none"> Books (5,214) Scores (101) Journals (27) <p>Types of loan:</p> <ul style="list-style-type: none"> Items for loan (3,441) Reading room use only (1,469) <p>Topic:</p> <ul style="list-style-type: none"> Shakespeare, William (1,806) Shakespeare, William 1564-1616 (468) Shakespeare, William IV (174) drama (65) Theater (48) Show 15 more <p>Creator:</p> <ul style="list-style-type: none"> Shakespeare, W (1,576) Brandes, G (56) Wells, S (38) Wilson, J (27) Taylor, G (12) Show 15 more <p>Language:</p> <ul style="list-style-type: none"> English (3,554) Danish (600) German (443) 	<p>Bibliothèque publique de Géorgie (E-U) Evergreen http://gapines.org</p> <p>Relevant Subjects</p> <ul style="list-style-type: none"> Shakespeare, William 1564-1616 Gardens Gardens in literature Plants in literature Art, Modern English drama Dramatis, English Caliban (Fictitious character) <p>Relevant Authors</p> <ul style="list-style-type: none"> Ackroyd, Peter Bloom, Harold Garfield, Leon Hales, Michael Hightfill, Philip H. Martineau, Jane McGinn, Collin McManaway, James Glimmer <p>Relevant Series</p> <ul style="list-style-type: none"> Bloom's major dramatists The Tupper lectures on Shakespeare 	<p>Les facettes restent identiques quelle que soit la recherche.</p>
<p>Bibliothèque publique de Queens, New York (E-U) AquaBrowser http://aqua.queenslibrary.org</p> <p>Refine</p> <ul style="list-style-type: none"> Library Catalog (239) Recommended by Queens Library (5) Format Books (1556) Videos and DVDs (329) Books on Cassette or CD (143) Music CDs and Cassettes (92) Music Scores (63) more Author Shakespeare, William (468) SHAKESPEARE (28) Lamb, Charles (21) 1862 more Subject English drama (264) English literature (186) Theater (164) Feature films (328) Dramatists, English (155) 1389 more Language english (483) chinese (24) spanish (9) 	<p>Bibliothèque publique du comté d'Athens, Ohio (E-U) Koha http://search.athenscounty.lib.oh.us</p> <p>Refine Your Search</p> <p>Places</p> <ul style="list-style-type: none"> Great Britain (2) England (1) Southwark (London, Eng.) (1) Denmark (1) <p>Series</p> <ul style="list-style-type: none"> Oxford school Shakespe... (3) The New Folger Libra... (2) The plays of William... (1) Evewitness books. (1) The Shakespeare libr... (1) <p>Show More</p> <p>Branches</p> <ul style="list-style-type: none"> Athens (17) Nelsonville (10) The Plains (9) Wells (Albany) (4) Cookville (1) <p>Authors</p> <ul style="list-style-type: none"> Shakespeare, William... (9) Gill, Roma. (3) Mowat, Barbara A. (2) Worsvine, Paul. (2) Brett, Jeremy. (1) <p>Show More</p> <p>Topics</p> <ul style="list-style-type: none"> Dramatis, English (6) Theater (3) English drama (2) Shakespeare, William... (1) Fathers and daughter... (1) 	<p>Bibliothèque publique du comté d'Athens, Ohio (E-U) Koha http://search.athenscounty.lib.oh.us</p> <p>Refine Your Search</p> <p>Places</p> <ul style="list-style-type: none"> Great Britain (2) England (1) Southwark (London, Eng.) (1) Denmark (1) <p>Series</p> <ul style="list-style-type: none"> Oxford school Shakespe... (3) The New Folger Libra... (2) The plays of William... (1) Evewitness books. (1) The Shakespeare libr... (1) <p>Show More</p> <p>Branches</p> <ul style="list-style-type: none"> Athens (17) Nelsonville (10) The Plains (9) Wells (Albany) (4) Cookville (1) <p>Authors</p> <ul style="list-style-type: none"> Shakespeare, William... (9) Gill, Roma. (3) Mowat, Barbara A. (2) Worsvine, Paul. (2) Brett, Jeremy. (1) <p>Show More</p> <p>Topics</p> <ul style="list-style-type: none"> Dramatis, English (6) Theater (3) English drama (2) Shakespeare, William... (1) Fathers and daughter... (1) 	<p>Les facettes restent identiques quelle que soit la recherche.</p>	
<p>Bibliothèque publique de Géorgie (E-U) Evergreen http://gapines.org</p> <p>Relevant Subjects</p> <ul style="list-style-type: none"> Shakespeare, William 1564-1616 Gardens Gardens in literature Plants in literature Art, Modern English drama Dramatis, English Caliban (Fictitious character) <p>Relevant Authors</p> <ul style="list-style-type: none"> Ackroyd, Peter Bloom, Harold Garfield, Leon Hales, Michael Hightfill, Philip H. Martineau, Jane McGinn, Collin McManaway, James Glimmer <p>Relevant Series</p> <ul style="list-style-type: none"> Bloom's major dramatists The Tupper lectures on Shakespeare 	<p>Bibliothèque publique de Géorgie (E-U) Evergreen http://gapines.org</p> <p>Relevant Subjects</p> <ul style="list-style-type: none"> Shakespeare, William 1564-1616 Gardens Gardens in literature Plants in literature Art, Modern English drama Dramatis, English Caliban (Fictitious character) <p>Relevant Authors</p> <ul style="list-style-type: none"> Ackroyd, Peter Bloom, Harold Garfield, Leon Hales, Michael Hightfill, Philip H. Martineau, Jane McGinn, Collin McManaway, James Glimmer <p>Relevant Series</p> <ul style="list-style-type: none"> Bloom's major dramatists The Tupper lectures on Shakespeare 	<p>Bibliothèque publique de Géorgie (E-U) Evergreen http://gapines.org</p> <p>Relevant Subjects</p> <ul style="list-style-type: none"> Shakespeare, William 1564-1616 Gardens Gardens in literature Plants in literature Art, Modern English drama Dramatis, English Caliban (Fictitious character) <p>Relevant Authors</p> <ul style="list-style-type: none"> Ackroyd, Peter Bloom, Harold Garfield, Leon Hales, Michael Hightfill, Philip H. Martineau, Jane McGinn, Collin McManaway, James Glimmer <p>Relevant Series</p> <ul style="list-style-type: none"> Bloom's major dramatists The Tupper lectures on Shakespeare 	<p>Les facettes restent identiques quelle que soit la recherche.</p>	
<p>Bibliothèque publique de Queens, New York (E-U) AquaBrowser http://aqua.queenslibrary.org</p> <p>Refine</p> <ul style="list-style-type: none"> Library Catalog (239) Recommended by Queens Library (5) Format Books (1556) Videos and DVDs (329) Books on Cassette or CD (143) Music CDs and Cassettes (92) Music Scores (63) more Author Shakespeare, William (468) SHAKESPEARE (28) Lamb, Charles (21) 1862 more Subject English drama (264) English literature (186) Theater (164) Feature films (328) Dramatists, English (155) 1389 more Language english (483) chinese (24) spanish (9) 	<p>Bibliothèque publique de Queens, New York (E-U) AquaBrowser http://aqua.queenslibrary.org</p> <p>Refine</p> <ul style="list-style-type: none"> Library Catalog (239) Recommended by Queens Library (5) Format Books (1556) Videos and DVDs (329) Books on Cassette or CD (143) Music CDs and Cassettes (92) Music Scores (63) more Author Shakespeare, William (468) SHAKESPEARE (28) Lamb, Charles (21) 1862 more Subject English drama (264) English literature (186) Theater (164) Feature films (328) Dramatists, English (155) 1389 more Language english (483) chinese (24) spanish (9) 	<p>Bibliothèque publique de Queens, New York (E-U) AquaBrowser http://aqua.queenslibrary.org</p> <p>Refine</p> <ul style="list-style-type: none"> Library Catalog (239) Recommended by Queens Library (5) Format Books (1556) Videos and DVDs (329) Books on Cassette or CD (143) Music CDs and Cassettes (92) Music Scores (63) more Author Shakespeare, William (468) SHAKESPEARE (28) Lamb, Charles (21) 1862 more Subject English drama (264) English literature (186) Theater (164) Feature films (328) Dramatists, English (155) 1389 more Language english (483) chinese (24) spanish (9) 	<p>Les facettes restent identiques quelle que soit la recherche.</p>	

Esempio di faccette proposte da diversi opac a seguito di una interrogazione con il termine Shakespeare (Maisonneuve-Toutou 2009 p. 18)

I diversi opac posizionano le faccette in ordini di sequenza differenti evidenziando un giudizio non omogeneo sulla loro rilevanza e non tutti presentano il numero delle occorrenze, un dato che è ritenuto invece significativo per dare un quadro del contenuto del catalogo e consentire all'utente di operare le scelte successive con consapevolezza. Alcuni opac utilizzano solo etichette in forma piatta, altri le organizzano in forma gerarchica, per esempio l'etichetta Libri può essere suddivisa in Caratteri grandi e Braille, oppure l'etichetta RegISTRAZIONI sonore può essere divisa in LP, CD, audiolibri, musica, anche se le sotto-etichette non sono necessariamente gerarchiche dal punto di vista semantico. Normalmente le etichette appaiono in ordine decrescente dei documenti associati. Abbastanza comune è l'uso di presentare solo le prime etichette seguite dall'indicazione "more...", "more options", "show more" o simili. Questi aspetti di design dell'interfaccia risultano essere molto significativi per una buona la navigazione a faccette e sono molto studiati, per esempio da Hearst (2006, 2008). Il raffinamento attraverso i tag è in genere collocato sulla pagina in posizione diversa dalle altre etichette se i tag vengono rappresentati attraverso la tag cloud⁷⁰. Il progetto Blacklight dell'università della Virginia⁷¹ consente di presentare faccette diverse in funzione delle raccolte della biblioteca: per esempio la collezione di musica presenta, oltre alle faccette comuni, delle faccette specifiche come Tipo di registrazione, Strumenti, Genere (Sadler 2009 p. 63).

Un modo differente di utilizzare la navigazione a faccette è quello studiato dal Flamenco ProjeT⁷² e adottato, almeno come concetto, anche da alcune biblioteche. Flamenco è l'acronimo di FLEXible information Access using METadata in Novel COmbinations. E' un progetto sviluppato dall'Università di Berkeley che ha lo scopo di studiare interfacce che consentano all'utente di muoversi in uno spazio informativo flessibile senza sentirsi smarrito. In questi sistemi la prima visualizzazione proposta all'utente è l'esposizione di tutti gli oggetti-documenti raggruppati in categorie (faccette) sia di tipo formale che di tipo semantico organizzate gerarchicamente (posizionandosi sopra un'etichetta si possono vedere le sottocategorie in un menù a tendina). Questa visualizzazione consente all'utente di avere

⁷⁰ Vedi per esempio < <http://nplencore.library.nashville.org>>.

⁷¹ Vedi <<http://www.lib.virginia.edu>> e <<http://projectblacklight.org/>>.

⁷² Vedi <<http://flamenco.berkeley.edu/index.html>>.

un quadro di tutte le risorse disponibili e di iniziare la ricerca scegliendo la categoria di suo interesse.

Flamenco Fine Arts Search
 Images from the Collections of the Fine Arts Museums of San Francisco,
 Legion of Honor and de Young Museums. <http://www.thinker.org>

Powered by Flamenco

Save Search History and Settings Return to Search New Search Logout

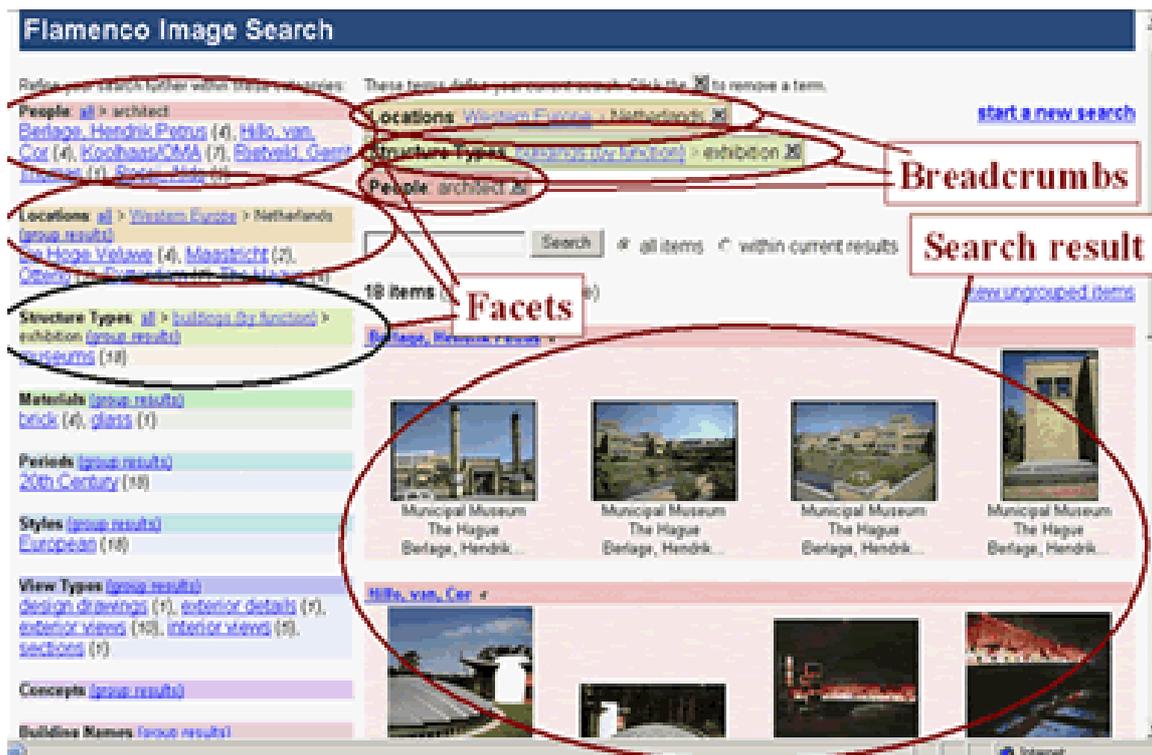
search

Username default Password Log In
 Create a New Account

Show tooltip previews of subcategories

MEDIA Book (309) Ceramic (896) Drawing (1547) Glass (403) Metalwork (1134)	HEAVEN AND EARTH Dawn, Dusk, Night (529) Islands, Deserts, Forests (424) Mountains, Hills, Valleys (2471) Rivers, Lakes, Seas (4098)
LOCATION Africa (101) Asia (945) Australia (5) Central America (57) Europe (17758)	SHAPES AND COLORS Color (4149) Decoration (1680) Metal (256)
OBJECTS Clothing (6018) Containers (2632) Food and Meals (1580) Fuel (453) Lighting (386)	OCCUPATIONS Combatant, Guard (1170) Entertainer (524) Leader (1688)
BUILT_PLACES Bridge (431) Building (2573) Built Open Space (912)	ARTISTS A.C., active 19th century (1) A.H. Heisey and Company (1) Aachen, Hans von, 1552 - 1615 (1) Abbenille, N. Sanson di (1) Abbey, Edwin Austin, 1852 - 1911 (7) Abbiati, Alessandro Paolo, active 1686 (1) Achenbach, Moore S., 1878 - 1963 (1)
ANIMALS AND PLANTS Birds (2569)	Objects (1689) Painting (115) Photograph (333) Print (18206) Sculpture (193)
Middle East (60) North America (3634) Oceania (72) Roman Empire (4) South America (158)	Stone and Rock (18) Storms, Clouds, Floods (1145) Sun, Moon, Stars (1272)
Musical Instruments (634) Timepieces (73) Vehicles (3457) Weapons (1498) Writing Tools (3636)	Scene (6526) Shape (1566)
Dwelling (1528) Part of Building (3159) Road (1204)	Ackerman, James, active 1813 (2) Adam, Georg, 1784 - 1823 (1) Adam, Robert, 1728 - 1792 (1) Adam, Victor, 1801 - 1866 (4) Adams & Co. (1) more...
Mammals, Hoofed (2555)	

La prima selezione comporta la suddivisione in due parti della videata: da una parte sono visibili le categorie disponibili per i raffinamenti successivi, dall'altra vengono presentati i documenti. Ad ogni selezione nelle faccette corrisponde una restrizione della ricerca, un suo affinamento. L'esposizione dei documenti è preceduta da una rappresentazione grafica dei passi di ricerca fatti cioè delle etichette selezionate (breadcrumb): da questa rappresentazione possono essere facilmente rimosse le etichette che rappresentano i raffinamenti applicati estendendo così nuovamente la ricerca. Nell'interfaccia è sempre presente una casella per l'immissione di termini di ricerca che agiscono in combinazione con le faccette selezionate in modo che i due tipi di ricerca sono sempre perfettamente integrati. Questo tipo di utilizzo della navigazione a faccette è più efficace quando viene applicato ad una collezione di documenti relativamente limitata ed omogenea in cui le categorie principali sono poche e facilmente prevedibili dagli utenti (McGrath 2007).



I demo presenti sul sito del Flamenco Project si riferiscono, infatti, alle collezioni del Fine Arts Museums of San Francisco e della UC Berkeley Architecture Visual Resources Library. Nei siti delle biblioteche questo tipo di esplorazione a faccette viene proposto esponendo in prima pagina alcune macrocategorie a partire dalle quali si possono fare raffinamenti successivi attraverso faccette di tipo formale o semantico, specifiche di quella categoria o generiche, fino ad arrivare al documento di interesse. Le scelte operate nelle macrocategorie e nei successivi raffinamenti possono anche essere usate in associazione ad una ricerca per parole che agisce sul set di dati ristretto dalle precedenti selezioni. La visualizzazione della breadcrumb proposta in una forma grafica particolare consente intuitivamente di ampliare nuovamente la ricerca eliminando i raffinamenti che non interessano più in qualsiasi punto del percorso e non tornando indietro passo per passo⁷³. Un modo diverso di applicare lo stesso concetto di navigazione è offerto implicitamente dalla versione beta dell'opac delle NCSU Library⁷⁴ che come risposta ad un'interrogazione vuota,

⁷³ Si veda come esempio il sito della State Library of Tasmania gestito con il software Talis Plus < <http://www.statelibrary.tas.gov.au/>; < <http://www.talis.com/>>

⁷⁴ Vedi <<http://www.lib.ncsu.edu/summon/>>.

quando cioè l'utente non immette alcun termine di ricerca, assume l'etichetta "any" per tutte le faccette disponibili e presenta quindi tutte le risorse disponibili. In questo caso però l'utilizzo della casella "search" successivo alla selezione di una o più faccette lancia la ricerca testuale su tutto il catalogo e non sul set ristretto⁷⁵.

Un aspetto interessante dell'introduzione della navigazione a faccette negli opac è di aver stimolato la ricerca in merito a come estrarre dati significativi dalla catalogazione semantica. Maisonneuve e Touitou (2007 p. 15) rilevano che gli opac di nuova generazione per la prima volta sono riusciti a sfruttare in modo adeguato le suddivisioni geografiche e temporali del soggetto RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié)⁷⁶, il principale strumento di catalogazione per soggetto utilizzato dalle biblioteche francesi. Un esempio di come i dati di tipo semantico siano utilizzati per costruire raggruppamenti di documenti è visibile nel catalogo NCSU⁷⁷ che utilizza il software Endeca: tra le faccette, oltre a quella generale di Soggetto sono presenti una faccetta di suddivisioni geografiche, una di suddivisioni temporali e una di genere che estraggono dati dai descrittori semantici. E' inoltre presente una faccetta riferita alla classificazione (LC Classification). Diversi studi vanno nella direzione di adattare le LCSH (Library of Congress Subject Headings) per consentire la loro navigazione a faccette ma gli ostacoli sono molteplici. In primo luogo bisognerebbe individuare un numero ristretto di categorie di primo livello per consentire un tipo di navigazione come quello che abbiamo appena descritto, cioè senza immettere un termine di ricerca specifico. Considerata la quantità di termini presenti in LCSH, molti dei quali non sono collegati tra loro da relazioni gerarchiche, questo comporterebbe un grandissimo lavoro anche per creare le relazioni sindetiche tra i termini. Una soluzione potrebbe essere usare altri schemi di classificazione, come la CDD, per generare le faccette di livello superiore sulle quali mappare le stringhe LCSH. Un altro aspetto da considerare è la specificità dei termini associata alla mancanza di una struttura

⁷⁵ L'incongruenza di alcuni sistemi di raffinamento attraverso faccette è rilevata da Buizza (2008) che sottolinea come in alcuni sistemi l'utente non sia avvertito del fatto che selezionando un elemento il sistema opera una nuova ricerca piuttosto che raffinare la precedente come accade con Evergreen. Buizza richiama anche l'attenzione su come non sia possibile verificare se la ricerca per parole sia fatta per parola esatta, troncata o per sequenza di caratteri, cosa che restituisce risultati a volte sconcertanti, oppure se in alcuni casi non venga applicato l'operatore or al posto del più consueto and al solo scopo di dare una risposta comunque sia pur di non restituire un set di risposta vuoto.

⁷⁶ Vedi <<http://rameau.bnf.fr/informations/rameauenbref.htm>>.

⁷⁷ Vedi <<http://www.lib.ncsu.edu/>>.

sindetica completa: uno strumento di tipo tesaurale risponderebbe meglio di LCSH alle esigenze della faccettazione. Questa carenza si riflette molto sulla gestione dei termini geografici per i quali non è possibile al momento ricavare dalle stringhe di soggetto le relazioni che consentirebbero una navigazione a faccette gerarchiche e non piatte. Un terzo aspetto riguarda l'organizzazione dei riferimenti incrociati tra termini autorizzati e termini alternativi, che non è gestita correttamente nelle interfacce per parola chiave. Un quarto aspetto riguarda il trattamento dei soggetti composti, non solo da più parole ma da più concetti (per esempio information retrieval e poeti dialettali italiani). Un ulteriore problema è rappresentato dalle faccette implicite: in alcune stringhe LCSH indica espressamente di non utilizzare suddivisioni (per esempio Nazional socialismo non deve essere seguito dalla precisazione geografica Germania, che è implicita nel concetto): per un recupero attraverso parole e per una corretta faccettazione queste stringhe dovrebbero tutte essere rese esplicite. Anche le faccette temporali sono spesso omesse o considerate implicite (McGrath 2007).

Un tentativo pratico di affrontare queste problematiche, pensato soprattutto per risolvere problemi legati alla difficoltà e ai costi della classificazione con LCSH, è il progetto FAST (Faceted Application of Subject Terminology) di OCLC⁷⁸. FAST può essere definito come un vocabolario a faccette post-coordinato per l'ambiente online. Preso atto della vastità dei dati da catalogare, il progetto mette in primo piano la necessità di creare uno schema di metadati per soggetto che sia semplice da imparare e usare, interoperabile, scalabile, facile da mantenere, compatibile con i metadati attualmente esistenti. I soggettari impiegati nella catalogazione tradizionale normalmente sono composti da stringhe di soggetto pre-coordinate, mentre i dizionari controllati usati nei database online sono perlopiù fatti di descrittori semplici che si affidano alla post-coordinazione per restituire i soggetti composti. Dal punto di vista della semplicità e dell'interoperatività semantica la post-coordinazione è il metodo che si sposa meglio con i presupposti e le caratteristiche dell'ambiente online ed è questo il metodo scelto da FAST che lo associa però alla ricchezza del vocabolario controllato offerto da LCSH (Dean 2003). I termini del vocabolario sono stati divisi in otto faccette: di argomento, geografica, formale, cronologica, nomi personali,

⁷⁸ Vedi <www.oclc.org/research/projects/fast>.

nomi di enti, convegni, titoli uniformi. Le faccette sono state individuate basandosi sui tipi di termini che hanno già etichette o sottocampi separati nei record MARC e si possono quindi estrarre facilmente dai record esistenti attraverso procedimenti automatici al fine di recuperare il pregresso garantendo quella compatibilità retrospettiva che è un requisito del progetto. L'authority file di FAST, che sarà reso disponibile con licenza non proprietaria, contiene attualmente oltre 1.600.000 record e può essere interrogato attraverso un prototipo⁷⁹ ma non ci sono ancora applicazioni nell'opac del catalogo World Cat.

Un modo ancora diverso di considerare la faccettazione ma sempre nella prospettiva di valorizzare gli aspetti semantici è quello scelto dal Visual Catalog. Questo catalogo è stato concepito e sviluppato all'università Paris-VIII ed è attualmente in uso come opzione secondaria di visualizzazione nell'opac delle biblioteche universitarie della regione di Artois⁸⁰. Il concetto di fondo di questo prodotto è che il catalogo deve esprimere l'idea che la biblioteca non è una semplice accumulazione di libri: deve essere in grado di mostrare la coerenza tra i testi di quella particolare raccolta, rendere il loro senso complessivo mostrando le relazioni che intercorrono tra le opere, i presupposti logici, la complementarità, il rinvio reciproco di una all'altra. Il catalogo è un punto di convergenza intellettuale e deve mostrare l'attività che i bibliotecari fanno per far emergere queste relazioni. In sostanza deve essere in grado di rendere visibile l'organizzazione intellettuale della biblioteca attraverso uno strumento tecnologico (Papy - Chauvin 2005 p. 10). L'attività dei catalogatori si esprime nel trattamento dei documenti secondo i loro titoli, i loro soggetti e le loro discipline, quindi le tre faccette individuate dal Visual catalog sono Titoli, Soggetti, Classificazione Dewey. L'opac tenta anche di dare una rappresentazione visiva delle opere all'interno delle dieci classi principali della classificazione e delle localizzazioni dei documenti. A seguito di un'interrogazione, in cui si può scegliere il canale di ricerca (titolo, autore, soggetto) e il tipo di ricerca testuale (almeno una parola, tutte le parole, la stringa esatta), si ottengono cinque gruppi di informazioni: le tre faccette Titolo, Soggetto e CDD che mostrano rispettivamente una lista di documenti associati ai termini di ricerca, una lista dei soggetti associati ai titoli estratti e alle CDD corrispondenti (con l'indicazione del

⁷⁹ Vedi <<http://fast.oclc.org>>.

⁸⁰ Vedi <<http://portail.bu.univ-artois.fr/cda/default.aspx?INSTANCE=INCIPIO> oppure direttamente <<http://visualcatalog.univ-artois.fr/vc2/>>.

numero dei documenti corrispondenti), una lista delle classificazioni associate ai titoli e ai soggetti (con l'indicazione del numero dei documenti corrispondenti); una sintesi "grafica metaforica" (Papy - Chauvin 2005 p. 11) che mostra in quali settori disciplinari si collocano le opere estratte (il cerchio che rappresenta ogni disciplina si riempie di colore in relazione alla quantità di documenti associati); un'immagine delle biblioteche dove il documento si trova, che corrisponde ad una faccetta di localizzazione. Ad ogni scelta in ciascuna delle faccette corrisponde una rielaborazione dinamica dei risultati presentati. Questa particolare presentazione dei dati non ha tanto lo scopo di raffinare la ricerca quanto piuttosto di mostrare il contesto in cui si colloca ogni documento scelto.

The screenshot shows the 'Visual... Catalog' interface. At the top, there's a search bar with 'famille' entered and a 'Chercher' button. Below the search bar are several filters: 'Titre' (checked), 'Auteur' (unchecked), 'vedette-matière RAMEAU' (unchecked), and 'Libellé de la classification Dewey' (unchecked). There are also radio buttons for 'au moins l'un des mots', 'tous les mots, quel que soit l'ordre', and 'l'expression exacte'. The main area features a 'Nuvoles de mots' visualization with colored circles representing different disciplines: Philosophie (100), Psychologie (100), Sciences (300), Langues (400), Technologie (600), Littérature Rhetorique (800), Géographie Histoire (900), Arts Sports (700), and Sci. naturelles Mathématiques (500). To the right, there are images of libraries: Bethune, Lens, Arras, Douai, and Lille. Below the visualization, there are three main sections: 'Titre (951)' with a list of books, 'RAMEAU (2358)' with a list of subjects, and 'Classification Dewey' with a list of Dewey numbers. At the bottom, there are search and localization buttons and the text 'Conception & Développement: Fabrice Papy, Etude chromatique: Sophie Chauvin'.

3.6 Nuvoles di parole e grafi

Un modo diverso dalla faccettazione di consentire l'esplorazione delle risorse a partire da una prima interrogazione formulata per parole chiave è rappresentato dalle "nuvoles di parole". In questo caso si tratta piuttosto di un'estensione della ricerca che di un raffinamento. Come abbiamo visto parlando degli opac sociali e dei tag generati dagli utenti, le tags clouds sono un insieme di parole usate per rappresentare concetti o descrivere documenti all'interno di un insieme di informazioni allo scopo di facilitarne il recupero. Ogni parola costituisce un link alle risorse che condividono l'assegnazione di quel tag. Le parole sono visivamente significative perché a quelle in caratteri più grandi o con colori più brillanti sono associate un numero maggiore di risorse o, in alcuni contesti, le risorse più recenti o più popolari. In alcuni opac le nuvoles di parole possono essere costituite sempli-

cemente da una lista di parole accompagnata dal numero che rappresenta i documenti collegati. Anche le words clouds consentono all'utente di avere una visione panoramica dei risultati e di utilizzare questa sintesi per navigare attraverso i risultati di una ricerca mostrandogli informazioni potenzialmente rilevanti che altrimenti potrebbero trovarsi in fondo in una lista di risultati ordinati solo per rilevanza. L'attivazione di un link corrisponde normalmente al lancio di una nuova interrogazione su tutto il catalogo⁸¹. Tuttavia quantità, novità o popolarità sono le uniche informazioni esplicitamente trasmesse dalla rappresentazione visuale attraverso nuvole di parole: non essendo esplicitata in nessun modo la relazione semantica tra termini di ricerca e parole della nuvola proposta, l'utente non può che assumere genericamente che si tratti di relazioni "associative"⁸². In alcuni opac, per esempio in WorldCat, le nuvole di parole non sono presentate parallelamente alla lista dei risultati ma sono presenti nel dettaglio delle singole notizie, in particolare nei dati di arricchimento del catalogo, e costituiscono uno strumento di navigazione a partire da un documento specifico piuttosto che da una interrogazione. I termini proposti sono definiti come "soggetti associati"⁸³.

Analogamente a quanto abbiamo visto per le faccette la generazione di nuvole di parole può avvenire da fonti differenti: per esempio il software Primo⁸⁴ utilizza solo i tag attribuiti dagli utenti, mentre altri prodotti utilizzano diverse fonti secondo due modelli sostanzialmente diversi, uno più semplice, il secondo più sofisticato. Nel primo caso le parole proposte sono quelle che appaiono nelle notizie corrispondenti alla richiesta immessa con l'aggiunta dei termini associati nelle registrazioni di autorità; nel secondo caso viene applicato un algoritmo che analizza le correlazioni esistenti tra le parole usate all'interno del catalogo. Poiché si tratta di un'elaborazione pesante viene svolta di notte su una base dati d'appoggio e non al volo al momento dell'interrogazione. La ricchezza e la diversità delle relazioni tra i termini immessi e la nuvola di parole proposta dipende dalla pertinenza

⁸¹ Un'eccezione a questo modo di intendere le words clouds è la scelta opera dal software Encore <<http://encoreforlibraries.com/>> che usa i termini della tag cloud come raffinamento della ricerca. Si veda per esempio l'opac della Biblioteca di Nashville <<http://nplencore.library.nashville.org>>.

⁸² Uno studio sull'introduzione delle tag cloud in PubMed, una base dati specializzata in medicina, rileva che proprio l'impossibilità di indicare il tipo di relazione tra i termini rende questo strumento poco utile nei casi in cui si tratti di recuperare documenti specifici, restando tuttavia positiva l'impressione generale sulla navigazione espressa dagli utenti sottoposti al test (Kuo – Hentrich - Good - Wilkinson 2007).

⁸³ Vedi per esempio <<http://www.worldcat.org/wcidentities/lccn-n79-21285>>.

⁸⁴ Vedi <<http://www.exlibrisgroup.com/category/PrimoOverview>>

dell'algoritmo e dalla dimensione del catalogo (Tosca Consultant 2008 p. 40). Secondo Breeding (Breeding 2007 p. 44), citato da Maisonneuve - Touitou (2007 p. 17) si possono avere dei buoni risultati con almeno 150.000 record mentre negli altri casi si può ricorrere ad interrogazioni su altri cataloghi che utilizzano lo stesso software oppure optare per l'acquisto di dati per l'arricchimento dei record bibliografici.

Una forma particolare di nuvola di parole è quella rappresentata in forma di grafo nei prodotti basati su AquaBrowser⁸⁵. In questo caso la rappresentazione, di cui si può vedere un esempio nell'immagine che segue, cerca di esprimere qualcosa in più delle relazioni che esistono tra i termini di ricerca e le parole proposte per la navigazione: il termine di ricerca appare al centro di un grafo; la distanza più o meno grande delle parole dal centro suggerisce una relazione più o meno stretta tra i termini; il colore della parola indica il tipo di relazione. I termini possono avere una relazione associativa, provenire da un tesaurus o da un vocabolario controllato, essere una traduzione del termine di ricerca, essere una sua variante ortografica o un suo sinonimo.

The screenshot shows the Queens Library website interface. At the top, there is a navigation menu with categories: Books, Music, Movies, Digital Media, Websites, Articles, and Events. A search bar contains the text "digital libraries". Below the search bar, the results are displayed, sorted by relevance. Two book entries are visible:

- Recommended by Queens Library:** The contributors cover the basics of evaluating digital libraries, but then explore how DLs make a difference in people's lives and their social worlds. Topics include documents a...
Digital library use
 social practice in design and evaluation
 Ann P. Bishop, Nancy A. van House, Barbara Pfeil Buttenfield
 Viewing digital libraries as sociotechnical systems networks of people and technology interacting with society.
 Call number: 025.0028 D
 Year: 2003
 Found: digital (4) ; libraries (3) ; Library (1)
 Series: Digital libraries and electronic publishing
 Subject: digital libraries, information technology
- Recommended by Queens Library:** A compilation of Roy Tennant's Library Journal columns. Tennant offers valuable commentary and analyses of technology and public policy issues. His column helps the reader to know...
Managing the digital library
 Roy Tennant
 Call number: 025.0028 T
 Year: 2004
 Found: digital (3) ; libraries (2) ; Library (2) ; Digitization (1)
 Subject: digital libraries

On the left side of the page, there is a word cloud visualization. The central term is "digital libraries". Other terms include "school", "management", "media", "history", "copyright", "guide", "encyclopedia", "catalogue", "image", "law", "congress", "system", "photography", "people", "science", "bookshop", "libraries", "digital", and "biblioteca". The words are color-coded and sized according to their relationship to the search term.

⁸⁵ Vedi <<http://www.serialssolutions.com/aquabrowser/>>.

I termini associati sono estratti da fonti prescelte e rielaborate attraverso una serie di algoritmi parametrizzabili (si possono selezionare i campi del record, il numero di occorrenze, la vicinanza al termine ricercato etc.). Per le traduzioni dei termini AquaBrowser utilizza come default un dizionario di lingue europee che riconduce ciascun termine dall'esperanto e dall'esperanto riparte per proporre la traduzione del termine stesso ma, in sede di configurazione, possono essere scelti altri dizionari, anche dizionari tecnici da usare in un ambiente specialistico. Per le relazioni gerarchiche utilizza strutture semantiche (tesauri, soggettari, vocabolari controllati) integrabili nel sistema in un qualsiasi formato elettronico (file di authority, file XML etc.) (Serial Solution s.d). Posizionandosi su un termine si vedono anche le relazioni che quel termine ha con gli altri nella nuvola e non solo con quello posto al centro: viene rappresentata così la rete dei collegamenti tra tutti i termini e si può avere un'anticipazione, benché parziale, dell'esito di un eventuale rilancio della ricerca. Il grafo di termini si "rigenera" ad ogni nuova ricerca e ad ogni nuova esplorazione del catalogo così come si modifica la lista di rilevanza dei documenti estratti (Kaiser - Hodge, 2005). AquaBrowser porta alle estreme conseguenze il concetto che il catalogo deve sempre dare una risposta, associandolo alle tecniche di stemming: interrogando il catalogo Hollis col termine ZZZ si ottengono nel grafo di risposta 5 nuovi termini (ayz, azz, ayez, yuz,ioz) tratti da un tesoro e in grado di consentire un rilancio significativo della ricerca⁸⁶.

Le nuvole di parole generate attraverso algoritmi, come quelle che AquaBrowser usa per mostrare le relazioni di tipo associativo, sono un tentativo diverso dal ranking dei risultati per dedurre la rilevanza per l'utente. Gli algoritmi di ranking applicano un concetto di rilevanza di sistema che non considera in alcun modo il significato semantico dei termini immessi per la ricerca. Specialmente nei casi in cui l'utente immetta un solo termine di ricerca alcuni documenti significativi per l'utente possono finire in fondo alla lista dei risultati in dipendenza dal contesto informativo in cui si trovano piuttosto che dal significato del termine e dalla sua rilevanza per l'utente. Per esempio se si immette il termine jaguar in un motore di ricerca questo non sarà in grado di distinguere se si cercano documenti

⁸⁶ Vedi <<http://discovery.lib.harvard.edu/>>. Ma interrogazioni simili su un catalogo meno ricco di quello della Harvard University danno risultati poco significativi. Si veda per esempio il catalogo Co.Bi.Re <<http://www.aquabrowser.it/COBIRE3/default.ashx>>.

sull'animale o sulle automobili: se la ricerca si sta svolgendo in una collezione che contiene più documenti sugli animali che sulle auto questi saranno considerati più rilevanti e saranno presentati in alto nella lista dei risultati mentre i documenti che riguardano le auto finiranno in coda. Questo meccanismo costituisce un limite delle presentazioni basate sul relevance ranking: costringe l'utente a scorrere la lista dei risultati fino alla fine o a rifare la ricerca in modo più preciso. Le nuvole di parole dovrebbero sopperire a queste limitazioni, dovrebbero aiutare l'utente a trovare più facilmente qualcosa che sia diverso dal significato principale del termine e nello stesso tempo proporgli una visione semantica dei documenti che sono in relazione alla sua query che gli consenta di manifestare il suo giudizio di rilevanza. Per individuare termini in relazione semantica vengono applicate delle tecniche di analisi statistica e semantica dei documenti che consentono al sistema di distinguere automaticamente i possibili diversi significati di una query. Si tratta di algoritmi che calcolano la rete di co-occorrenza dei termini (Veling - Weerd 1999).

3.7 Ranking e faccettazione nel giudizio degli utenti

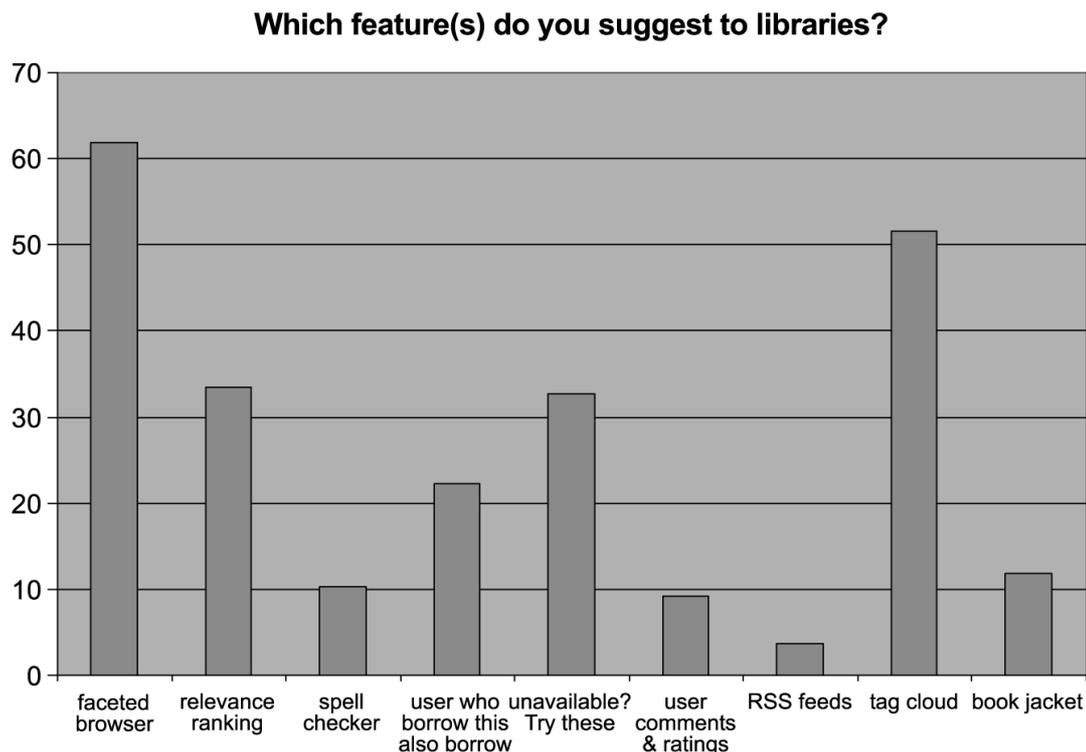
Gli opac di nuova generazione hanno adottato funzionalità di ricerca e interfacce innovative sulla scorta di considerazioni in merito all'evoluzione del contesto informativo e tecnologico e delle esigenze degli utenti manifestate in studi specifici o desunte dai comportamenti generali degli utenti sul Web. Le scelte operate, introduzione della ricerca per parole, del relevance ranking, navigazione a faccette, sono giustificate da una lunga tradizione di studi di IR e da un numero imponente di applicazioni commerciali. Resta tuttavia per le biblioteche la necessità di monitorare l'uso che di questi strumenti viene fatto allo scopo di migliorare ulteriormente e costantemente gli opac e quindi l'esperienza di ricerca degli utenti (non disgiunta dall'urgenza di dimostrare agli stakeholders l'uso che viene fatto delle risorse (Lown – Hemminger 2009). La letteratura su questo argomento è molto limitata, probabilmente perché gli opac di nuova generazione sono un prodotto relativamente nuovo (Tam – Cox – Bussey 2009 p. 354): alcune ricerche hanno cercato di individuare quali fossero le preferenze degli utenti in vista dell'adozione di nuovi opac mentre altre sono focalizzate sulla valutazione successiva all'installazione di nuovi prodotti.

Nello studio svolto nelle biblioteche dell'Università del Wisconsin nel 2008, di cui abbiamo parlato in precedenza (University Of Wisconsin-Madison Libraries 2008), gli utenti hanno indicato che la maggiore difficoltà che incontrano nello svolgimento delle proprie

ricerche consiste nell'individuazione dei materiali rilevanti all'interno dei risultati di una ricerca (42,5% delle risposte). Coerentemente, nell'indicare le caratteristiche che vorrebbero presenti in un nuovo opac hanno scelto l'ordinamento per rilevanza nel 60,8% dei casi ponendo questa funzionalità tra quelle maggiormente preferite, preceduta solo dalla funzionalità "Did you mean?" cioè dal suggerimento di parole alternative in caso di fallimento della ricerca (votata dal 63,5 %). Nessuna domanda indagava l'interesse per la faccettazione dei risultati.

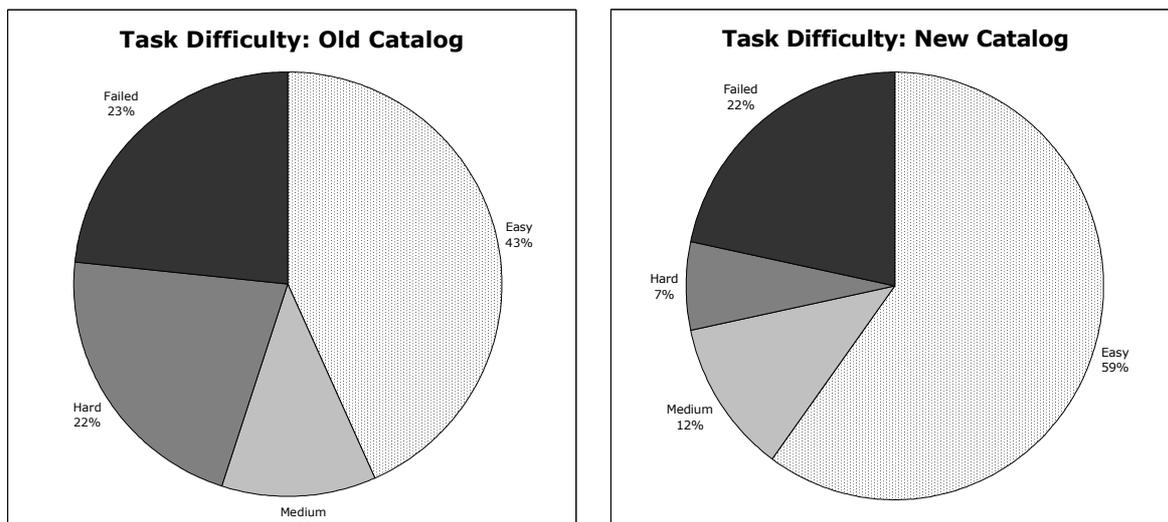
Secondo un'indagine svolta in vista dell'adozione di un nuovo software presso l'Università nazionale di Singapore nel 2007, condotta attraverso la somministrazione di questionari online e cartacei agli studenti e ai docenti, la faccettazione dei risultati è stata la prima delle caratteristiche preferite dagli utenti, seguita dalla possibilità della ricerca simultanea su più cataloghi e database, la ricerca federata di cui parleremo in seguito. In questo caso non è stata rivolta alcuna domanda specifica sul ranking dei risultati che tuttavia è stato scelto dalle biblioteche come prima opzione di ordinamento nel nuovo software adottato (Lim 2008).

Anche nel caso dell'Università di Sheffield, secondo lo studio descritto in precedenza (Tam – Cox – Bussey 2009), la possibilità di esplorare il set di risposta attraverso la navigazione delle faccette è stata l'opzione preferita dagli intervistati seguita dall'ordinamento per rilevanza dei risultati. Gli studenti hanno dimostrato di avere familiarità con questo dispositivo, per averlo usato nei siti commerciali, e di apprezzarlo per la capacità di dare una visione globale dei risultati e di restringere velocemente il set di risposta. Hanno inoltre riconosciuto che la navigazione a faccette svolge la stessa funzione della tradizionale ricerca avanzata ma la considerano molto più semplice da usare. Rispetto al relevance ranking il giudizio degli utenti è stato molto positivo comparato al tradizionale ordinamento alfabetico o per data. La metà degli intervistati esprime inoltre un interesse a capire come effettivamente il motore di ricerca calcoli questa rilevanza meravigliandosi che effettivamente riesca a proporre ciò essi giudicano rilevante (Tam – Cox – Bussey 2009 p. 360). Nell'immagine che segue sono rappresentati i giudizi espressi su tutte le caratteristiche innovative proposte (Tam – Cox – Bussey 2009 p. 367).



Un esempio di valutazione post installazione è quello offerto dalla North Carolina State University (NCSU) che nel gennaio 2006 ha pubblicato il suo opac di nuova generazione basato sul software Endeca, oggetto da allora di continui miglioramenti e introduzione di nuove funzionalità. Nonostante il nuovo catalogo fosse manifestamente superiore al precedente, considerando che le caratteristiche che si sarebbero introdotte erano sostanzialmente diverse da quelle in uso precedentemente, la biblioteca aveva individuato sin dalle fasi di progettazione diversi aspetti da sottoporre a valutazione. Un primo elemento di valutazione significativo, rilevato attraverso l'analisi dei log dei primi due mesi, è che la ricerca attraverso canali specifici sui record di autorità (che in un primo momento era rimasta presente nell'interfaccia di ricerca) è diminuita del 45 % a fronte di un aumento del 230 % della ricerca per keyword (che però diversamente dal vecchio sistema era impostata come default, scelta che contribuisce in modo imponderabile ad alzare la percentuale d'uso). Complessivamente il 55 % delle ricerche effettuate sono risultate essere semplici ricerche per parola chiave, nel 30 % dei casi c'è un raffinamento della ricerca attraverso le faccette, nel restante 15 % si tratta di una esplorazione del catalogo senza immissione di alcun termine di ricerca (Antelman - Lynema - Pace 2006 p. 133). Gli stessi dati rilevati a maggio 2007 indicano rispettivamente 67, 25 e 8% (Antelman 2007 slide 19) con un aumento della

ricerca semplice e una diminuzione dell'uso delle faccette di raffinamento. Un test di usabilità è stato svolto con 10 studenti i quali hanno svolto, 5 sul vecchio catalogo e 5 sul nuovo, le stesse ricerche (4 su documenti conosciuti e 6 per argomento). Il nuovo catalogo ha richiesto un tempo di ricerca significativamente inferiore. Come si vede dall'immagine che segue gli utenti inoltre giudicano significativamente più facile effettuare delle ricerche sul nuovo catalogo anche se resta alto il numero di ricerche fallite, forse perché gli utenti continuano a restringere la ricerca a "parole di soggetto" anziché lanciarla "anywhere". Considerando solo le ricerche per argomento il successo del nuovo catalogo è molto superiore: 22% di fallimenti contro 34% del vecchio (Antelman - Lynema - Pace 2006 p. 135) dimostrando la migliore qualità dell'opac in una tipologia di interrogazione da sempre considerata problematica.



L'uso delle faccette è cambiato nel corso del tempo e anche il loro posizionamento è stato modificato per adeguarsi alle preferenze degli utenti. Nel test tutti i 5 utenti che hanno usato il nuovo opac hanno capito intuitivamente l'uso delle faccette come metodo per restringere il set di risposta ma solo 3 le hanno effettivamente usate, forse condizionati dalle precedenti esperienze negli opac. Nel 2007, secondo dati ricavati dai file di log, le faccette più usate sono risultate essere: soggetto 26 %, classificazione 21 %, formato 10 %, novità 10 %, biblioteche 10 %, soggetto-genere 6 %, autore 6%, luogo 4 %, lingua 3 %, periodo storico 2 %, disponibilità 2 %. Questi numeri non sono direttamente corrispondenti al posizionamento della faccetta (Pennell 2007). Un nuovo test con gli utenti effettuato nel

2008 evidenzia ancora qualche difficoltà nell'uso delle faccette. Emerge che gli utenti le usano occasionalmente e in modo poco coerente: spesso non le guardano e anche quando ne cercano una specifica faticano a trovarla, forse per la presenza di troppo testo nella colonna; capiscono poco la differenza tra faccetta di soggetto e di classificazione (call number); non capiscono che quella di Call number corrisponde ad una esplorazione dello scaffale secondo l'ordinamento fisico dei documenti; non usano quasi mai il raffinamento per luogo o data e preferiscono rilanciare la ricerca utilizzando parole chiave più specifiche ottenendo in modo diverso lo stesso risultato (Littletree 2008). Gli utenti sottoposti al test hanno giudicato l'ordinamento per rilevanza dei risultati il principale miglioramento del catalogo. L'analisi dei log conferma che gli utenti scorrono solo la prima pagina dei risultati per stabilire la pertinenza dei documenti: se non vi trovano nulla normalmente immettono una nuova ricerca con un numero superiore di parole mentre solo il 13% passa a visualizzare la seconda pagina di risultati (Antelman - Lynema - Pace 2006 p. 136). Nel tentativo di verificare la corrispondenza tra la rilevanza indicata dal sistema e la rilevanza umana, almeno per quanto concerne l'argomento (rilevanza topica), è stato svolto un test specifico. Sono state effettuate 100 ricerche per argomento nel vecchio e nel nuovo opac usando l'opzione "Keyword anywhere": circa il 40 % dei primi risultati del vecchio opac sono stati giudicati rilevanti dal ricercatore mentre lo sono stati il 68 % di quelli ottenuti dal nuovo opac, con un miglioramento del 70%. Assumendo che il primo risultato nel nuovo opac sia rilevante si è anche verificata la sua posizione nella lista dei risultati proposta nel vecchio opac: è risultato che il documento può trovarsi tra la posizione 1 e la 4126, con più di un terzo dei casi in cui il documento è presentato dopo la seconda schermata di risultati e quindi ben oltre il numero di documenti che l'utente è normalmente disponibile ad esaminare⁸⁷.

⁸⁷ E' interessante notare che il catalogo testato, nonostante i successi, non è ancora considerato un catalogo di terza generazione secondo i criteri stabiliti da Hildret, di cui abbiamo parlato, in quanto non sono sufficientemente sviluppate 3 aree: espressione delle query in linguaggio naturale, espansione ed estensione del catalogo, metodi di relevance feedback (Antelman - Lynema - Pace 2006 p. 137).

4. OPEN DATA, LINKED DATA

4.1 Il Web arricchisce il catalogo

Tra le caratteristiche degli opac di nuova generazione vi è la capacità di gestire dati provenienti da diverse fonti per consentire all'utente di semplificare il suo processo di ricerca interrogando attraverso un'unica interfaccia numerose basi di dati. I dati possono essere costituiti da varie raccolte della biblioteca, per esempio raccolte di fotografie o stampe o altri documenti catalogati su database separati, oppure da documenti provenienti da diverse biblioteche per costituire un catalogo collettivo, oppure essere dati registrati su fonti esterne alla biblioteca raggiungibili via Web. Quest'ultima opzione è particolarmente in uso nelle biblioteche universitarie e delle istituzioni di ricerca che sottoscrivono abbonamenti di accesso a banche dati di periodici o di altri documenti digitali. Quando i dati sono accessibili su fonti per le quali la biblioteca ha la disponibilità dell'esportazione, questi vengono normalmente ricopiati nella base dati interna e l'opac ha la capacità di gestire diversi formati di metadati mettendo in corrispondenza i formati dei dati esterni con i formati interni. Questo trattamento consente di utilizzare tutti i dati come un unico insieme e di organizzare la visualizzazione unitaria nell'opac applicando il relevance ranking o la faccettazione dei risultati. Nel caso in cui i dati non possano essere trasferiti, come avviene in tutti i casi di banche dati di tipo editoriale a pagamento, gli opac mettono in campo i moduli per la ricerca federata e i link resolver. La ricerca federata (Ridi 2007 p. 201, Marshall – Herman – Rajan 2006) consiste nella ricerca simultanea e in tempo reale su molteplici database online o su molteplici risorse Web, selezionate dalla biblioteca, attraverso un'interfaccia unica. Questi dispositivi, a seguito dell'interrogazione immessa dall'utente, si fanno carico di preparare per ogni fonte esterna un'interrogazione organizzata secondo le esigenze della base dati che vanno ad interrogare e gestiscono la presentazione dei diversi set di risposta raggruppandoli in un unico insieme, per quanto possibile deduplicato, per il quale sono in genere previste diverse possibilità di ordinamento e raffinamento: cliccando sul record di interesse avviene il trasferimento sul database nativo per una completa esplorazione della risorsa. I link resolver sono server in grado di gestire il protocollo OpenUrl per lo scambio di metadati finalizzato alla gestione di servizi di linking contestualizzato

(context-sensitive)⁸⁸. Lo schema OpenUrl (NISO 2004, Bucchioni 2004) è stato messo a punto tra il 1998 e il 2000 con lo scopo di favorire i collegamenti citazionali (reference linking) in ambiente informativo accademico su Web. Quando, utilizzando un server configurato per la ricezione di OpenUrl, si interroga una base dati configurata per l'invio di OpenUrl, i due sistemi si riconoscono e, accanto ad ogni link, compare un pulsante che attiva il servizio. L'information provider invia una stringa di metadati che il link resolver interpreta e, sulla base dell'identificazione della fonte del collegamento e dei metadati descrittivi associati, è in grado di accertare quali sono i diversi accessi disponibili per quella risorsa, per poi sottoporli tutti all'utente, che effettuerà la sua scelta in base alle proprie esigenze. I link resolver consentono quindi di gestire i servizi connessi ad ogni singolo documento indicando il tipo di disponibilità (se è accessibile l'informazione citazionale, l'abstract, il full text, ecc.), attivando un servizio di document delivery oppure rilanciando una ricerca sul Web, mostrando le azioni possibili in funzione dei diritti posseduti dall'utente che sta svolgendo la ricerca. Nel caso della ricerca federata l'utente si trova di fatto in un ambiente diverso dall'opac vero e proprio: l'esplorazione dei dati locali e dei dati esterni non sono integrate ma solo giustapposte. Questo comporta la compresenza in un'unica interfaccia di logiche di ricerca diverse, cosa che crea una certa complessità per l'utente: egli dovrebbe essere messo in condizione di capire in quali casi è opportuno interrogare la base dati locale piuttosto che i dati esterni sapendo che le modalità di presentazione dei dati e le possibilità di utilizzo sono diverse (Tosca Consultant 2008 p. 16). Il tema dell'utilizzo di dati provenienti da diverse fonti va molto al di là della semplice schematizzazione qui proposta. In una visione più ampia, fatta propria specie dalle biblioteche delle istituzioni di formazione e ricerca e dagli ambienti decisamente orientati alla gestione di risorse digitali, questi dati dovrebbero essere usati in modo tale da creare una rete di informazioni che possa inserirsi senza soluzione di continuità nel flusso di lavoro degli utenti, sostenendo non solo le fasi di ricerca ma anche di uso e manipolazione dei dati⁸⁹. Questa visione coinvolge diversi aspetti problematici che vanno dall'apertura dei dati alla loro interoperabilità, alla loro conservazione a lungo termine, dalla definizione di

⁸⁸ Per una definizione più puntuale del linking citazionale e dei link resolver si veda Bucchioni-Spinelli (2007).

⁸⁹ Si veda come esempio di questo approccio il rapporto finale del progetto OLE (Open Library Environment Project) <<http://oleproject.org/final-report-on-ole-project/>>.

nuovi modelli di servizio all'adozione di nuove pratiche di collaborazione tra le diverse comunità di interesse (utenti, bibliotecari, istituzioni, editori) (Beit - Arie 2009). Già la semplice possibilità di svolgere in diverse forme una ricerca estesa costituisce comunque un notevole miglioramento dell'opac che va sia nella direzione di ampliare il contenuto del catalogo sia nella direzione di creare per l'utente un ambiente dinamico e interconnesso più in sintonia con il mondo delle informazioni attuale. In questa stessa prospettiva possono essere considerate le attività che mirano ad arricchire le singole registrazioni bibliografiche: alla presentazione della pura descrizione bibliografica vengono aggiunti dati che hanno lo scopo di favorire la selezione puntuale del documento sia attraverso l'impatto visivo che si ottiene mostrando la copertina, sia favorendo la valutazione della pertinenza attraverso l'esposizione dell'indice o del primo capitolo o di recensioni. Ulteriori arricchimenti possono essere costituiti da link ad altri contenuti di tipo bio-bibliografico. I dati di arricchimento possono essere ricopiati nella base dati interna ma più spesso sono prelevati dinamicamente da fonti esterne sia in forma gratuita sia a seguito del pagamento di un servizio di fornitura. L'idea di sfruttare dati non posseduti direttamente per migliorare i propri servizi viene alle biblioteche (Engard 2009) dalle pratiche del Web 2.0 (O'Reilly 2005 a, 2005 b) che dell'uso e remixing di dati aperti, definito come mashup, fanno un proprio cavallo di battaglia. Il mashup è un'applicazione Web che usa contenuti provenienti da due o più fonti per creare un nuovo servizio Web all'interno di una singola interfaccia grafica (Fitcher 2009 p. 3). Per creare un mashup è essenziale essere in grado di riunire informazioni strutturate provenienti da siti Web (oppure estrarre i contenuti e convertirli in un formato strutturato). L'esempio più usato per spiegare il mashup è l'utilizzo delle mappe di Google nei contesti più diversi: quando nel sito di un'azienda o di una istituzione o in qualsiasi altro sito, si mette insieme l'indirizzo o l'immagine di un edificio con la mappa che ne indica la localizzazione presa da Google maps si è realizzato un mashup. Condizione per il riuso dei dati, per la loro combinazione in contesti diversi da quello in cui sono stati creati, è che i dati siano liberamente accessibili senza necessità di negoziazione: una organizzazione rende i dati disponibili e descrive il modo in cui vi si può accedere. L'accesso può avvenire sotto forma di RSS feed o di altri formati XML, ma principalmente attraverso API (Application Programming Interface) cioè attraverso programmi che indicano come formulare un'interrogazione per ottenere i dati. Le API indicano la logica su cui il servizio si basa, le sue risorse e le funzioni aperte all'uso in altri ambienti, cioè consen-

tono ad un programma informatico di accedere e manipolare i dati disponibili su un servizio Web (Biancu 2009 p. 19). Molte applicazioni consentono di fare uso delle API semplicemente copiando e incollando nel proprio sito Web o nel proprio blog pezzetti di codice generati automaticamente (Biancu 2009 p. 31). In altri casi è richiesta una maggiore competenza di programmazione ma in sostanza i dati sono disponibili per chiunque abbia abbastanza fantasia da usarli nel proprio contesto anche in modi diversi da quello per cui sono stati originariamente creati (Fitcher 2009 p. 7-8). Naturalmente è possibile, e di fatto avviene, che alcuni siti, pur dotati di API per dare accesso ai propri dati, non le mettano liberamente disponibili a terzi ma solo dietro il pagamento di un compenso. Anche tra le biblioteche è molto diffuso l'uso del mashup per indicare la propria posizione geografica utilizzando Google maps oppure le tecniche di mashup vengono usate, come abbiamo detto, per arricchire i dati bibliografici facendo ricorso a dati liberi provenienti da Amazon o da Wikipedia o a dati a pagamento: quando l'utente interroga il catalogo e identifica un volume di suo interesse il sistema interroga una di queste fonti esterne e, se il dato è disponibile, incorpora nella propria interfaccia l'immagine di copertina o l'indice del volume o la recensione. Va detto per inciso che il mashup di dati aperti ha notevoli implicazioni legate all'affidabilità dei dati, alla loro persistenza, alla gestione di eventuali diritti d'autore e non ultimo alla tutela della privacy (Fitcher 2009 p. 14-15) ma non c'è dubbio che consenta di creare un servizio più accattivante e porti dentro la staticità dell'opac elementi di dinamicità e di interconnessione con la rete Web molto apprezzati dagli utenti.

4.2 Il catalogo arricchisce il Web

L'arricchimento dei record bibliografici attraverso il tipo di mashup che abbiamo descritto e la ricerca federata su diversi repository di dati esterni alla biblioteca seguono la logica di far convergere verso la biblioteca una maggiore quantità e varietà di dati in modo che l'utente continui a trovare negli strumenti e nelle risorse che essa mette a disposizione un punto privilegiato di ricerca e accesso ai documenti e la biblioteca conservi un ruolo che ha sempre avuto nella storia. Come abbiamo visto parlando del nuovo contesto informativo e della net generation, questo ruolo della biblioteca è messo oggi in secondo piano dall'immagine e dalla centralità che la rete internet ha acquisito nella vita delle persone anche se non è completamente oscurato: la funzione delle biblioteche come "ultimo miglio", cioè come ultimo segmento del processo di ricerca che consiste nella fornitura fisica

del documento all'utente, continua ad essere riconosciuta ed apprezzata (Weston 2007 p. 241). In una logica completamente diversa, da attuare in modo complementare alla prima, si pongono le indicazioni di coloro che, come Breeding (2006), propongono che la biblioteca diventi un punto di destinazione nelle ricerche degli utenti e non un punto di partenza: i contenuti e i servizi della biblioteca dovrebbero essere disseminati ovunque sul Web, presentarsi attraverso interfacce diverse dai cataloghi e dai siti istituzionali delle biblioteche in modo che gli utenti, svolgendo le proprie ricerche o le proprie attività liberamente sul Web, possano incontrarli e farne uso secondo le proprie necessità e modalità. Usa gli stessi termini in modo diverso, ma esprime lo stesso concetto Miller:

Rather than trying, repeatedly, to become the destination site for users, we should recognise that library services are rarely an end in themselves, and that the library makes most sense in the context of the task that users find themselves attempting to complete. We should not make our users have to consciously decide to 'do library stuff now'. Instead, we should integrate into the workflows that they already undertake, and we should implement methods that make library content visible and relevant to those who might never have thought to turn to a library for anything more than a warm place to check their e-mail (Miller 2006)⁹⁰.

Non solo le biblioteche debbono fare uso di dati altrui per migliorare i propri servizi o crearne di nuovi ma devono anche mettere a disposizione i propri dati in forme e luoghi diversi per essere visibili anche fuori del proprio ristretto contesto.

Modern approaches to thinking about provision of library data and services online create opportunities for numerous applications beyond the traditionally defined library management system. By adhering to standards from the wider Web community, by considering the library system as an interlocking set of functional components rather than a monolithic black box, and by taking a bold new approach to defining the ways in which information from and about libraries are 'owned' and exposed to others, we make it straightforward for information from the library to find its way to other places online... Rather than being locked inside the library system, data can add value to the experience of users wherever they are, whether it is Google, Amazon, the institutional portal, or one of the social net-

⁹⁰ Piuttosto che cercare ripetutamente di diventare il sito di destinazione degli utenti, dovremmo probabilmente riconoscere che i servizi della biblioteca raramente sono in se stessi un fine e che le biblioteche hanno molto più senso nel contesto dei compiti che gli utenti stanno cercando di svolgere. Non dovremmo mettere i nostri utenti nella situazione di decidere consciamente di "fare cose di biblioteca". Dovremmo invece integrarci nel flusso di lavoro che hanno già intrapreso e dovremmo sviluppare metodi che rendano i contenuti della biblioteca visibili e rilevanti per coloro che potrebbero non aver mai pensato di rivolgersi ad una biblioteca per qualcosa di più che un posto caldo dove controllare le proprie e-mail.

working sites such as MySpace or Facebook. By unlocking data and the services that make use of it, the possibilities are literally endless ... (Miller 2007)⁹¹.

Ma in un modo ancora diverso e più significativo si tratta di riconoscere che i dati bibliografici provenienti dalla lunga tradizione di lavoro e competenza delle biblioteche possono essere significativi non solo per singoli utenti individualmente dentro e fuori dalle biblioteche ma per intere comunità di utenti nel Web ed è necessario creare le condizioni e mettere in campo i dispositivi che possano favorire la loro apertura e disponibilità in modo che queste comunità possano farne uso nei modi che ritengono più utili alle proprie attività.

Tra le istituzioni che hanno intrapreso questa strada esemplare è il caso di OCLC. OCLC (Online Computer Library Center) è un'organizzazione internazionale che gestisce il più grande opac collettivo mondiale, WorldCat: 130 milioni di record, 1,4 miliardi di dati sul possesso di 71 mila biblioteche di 120 paesi, 20 milioni di documenti diversi dai libri, pubblicazioni di oltre 220 paesi in 470 lingue. Negli ultimi anni il catalogo è stato significativamente incrementato con le risorse di molti cataloghi collettivi regionali e nazionali ma anche acquisendo dati da editori e aggregatori commerciali. Inoltre i dati sono stati arricchiti da contributi degli utenti. Per consentire all'utente di navigare e selezionare questa enorme massa di documenti sono stati approntati dispositivi di ranking e faccettazione dei risultati, recensioni e valutazioni provenienti da fonti esterne come Amazon e WeRead, stime di popolarità e livello di lettura basate sul data mining (vengono valutate la quantità e il tipo di biblioteche che possiedono il documento), raccomandazioni fornite da WeRead e dagli utenti (Hickey - Gatenby 2009 p. 1-2). Utilizzando i dati del catalogo associati al Virtual International Authority File (VIAF)⁹² OCLC ha creato WorldCat Identities⁹³ un data base che contiene informazioni bio-bibliografiche su autori ed enti che vengono utilizzate come arricchimento dei record nel momento in cui un utente svolge una

⁹¹ Approcci moderni al modo di pensare la fornitura di dati e servizi librari online creano la possibilità per numerose applicazioni che vanno al di là dei sistemi di gestione di biblioteca tradizionalmente definiti. Aderendo agli standard della vasta comunità del Web, considerando i sistemi di biblioteca come in insieme di componenti funzionali interconnessi piuttosto che come una scatola nera monolitica e adottando un audace nuovo approccio per definire i modi in cui le informazioni sulle biblioteche e dalle biblioteche sono possedute ed esposte agli altri, rendiamo semplice per le informazioni dalla biblioteca trovare la loro via verso altri luoghi online. Piuttosto che restare chiusi dentro i sistemi di biblioteca, i dati possono aggiungere valore alle esperienze degli utenti ovunque si trovino, sia Google o Amazon, i portali istituzionali, o uno dei siti di social networking come MySpace o Facebook. Aprendo i dati e i servizi che ne fanno uso, le possibilità sono letteralmente infinite...

⁹² Vedi <www.viaf.org>.

⁹³ Vedi <<http://www.worldcat.org/identities/>>

ricerca in WorldCat. Le pagine per gli autori, di cui si vede un esempio nell'immagine che segue, comprendono indicazioni sulle loro opere, le principali opere di critica, persone collegate, varianti del nome, immagini delle copertine dei libri, soggetti e classificazioni collegate, statistiche sul posseduto, una time line delle pubblicazioni, link a risorse esterne significative.

The screenshot shows the WorldCat Identities page for Luciano Pavarotti. The page is divided into several sections:

- Overview:** Shows 1,249 works, 3,191 publications in 21 languages, and 93,291 library holdings. It lists genres (Opera, Song), subjects (Tenor, Singer), and roles (Performer, Vocalist).
- Publication Timeline:** A bar chart showing the number of publications from 1980 to 2010. The chart shows a significant increase in publications starting around 1990, peaking around 2000, and then declining.
- Audience Level:** A slider showing the audience level from 0 (Kids) to 1 (Special). The current level is 0.04, with a range from 0.02 for 'Let's Talk...' to 0.59 for 'La Dile al...J'.
- Related Identities:** A list of related composers and performers, including Verdi, Sutherland, Domingo, Bonynge, Carreras, Freni, Ghislanzoni, Donizetti, National Philharmonic Orchestra, and Puccini.
- Useful Links:** Links to the Library of Congress Authority File (English), Deutsche Nationalbibliothek Authority File (German), and Wikipedia Luciano Pavarotti.
- Associated Subjects:** A list of related subjects, including 'The Ring and I: the unceremonious tale of Luciano Pavarotti's rise to fame by his manager, friend, and sometime adversary by Herbert H. Gold', 'Pavarotti: life with Luciano by Adam Pavarotti', 'Luciano Pavarotti: the myth of the tenor by Jürgen Hoffing', 'Grandissimo Pavarotti by Martin Mayer', 'Meeting Luciano: a memoir by Anna Exler Smith', 'Pavarotti, my own story by Luciano Pavarotti', 'Carreras, Domingo, Pavarotti in concert', 'Pavarotti, my world by Luciano Pavarotti', 'La Bohème by Giacomo Puccini', 'Rigoletto by Giuseppe Verdi', and 'Turandot by Giacomo Puccini'.

Utilizzando i dati che servono per la gestione del posseduto, WorldCat ha realizzato il WorldCat Registry che raccoglie informazioni su 107 mila istituzioni collegate. Interrogando il database si possono avere indirizzi, contatti, oltre che collegamenti alle home page, ai cataloghi e ai servizi⁹⁴. Da questa posizione di massimo accentramento OCLC è ora in grado di mettere a disposizione i dati, variamente collegati e ricombinati in modo significativo, per le biblioteche, per una vasta platea di utenti sul Web, per i motori di

⁹⁴ Vedi <<http://www.worldcat.org/registry/Institutions/>>.

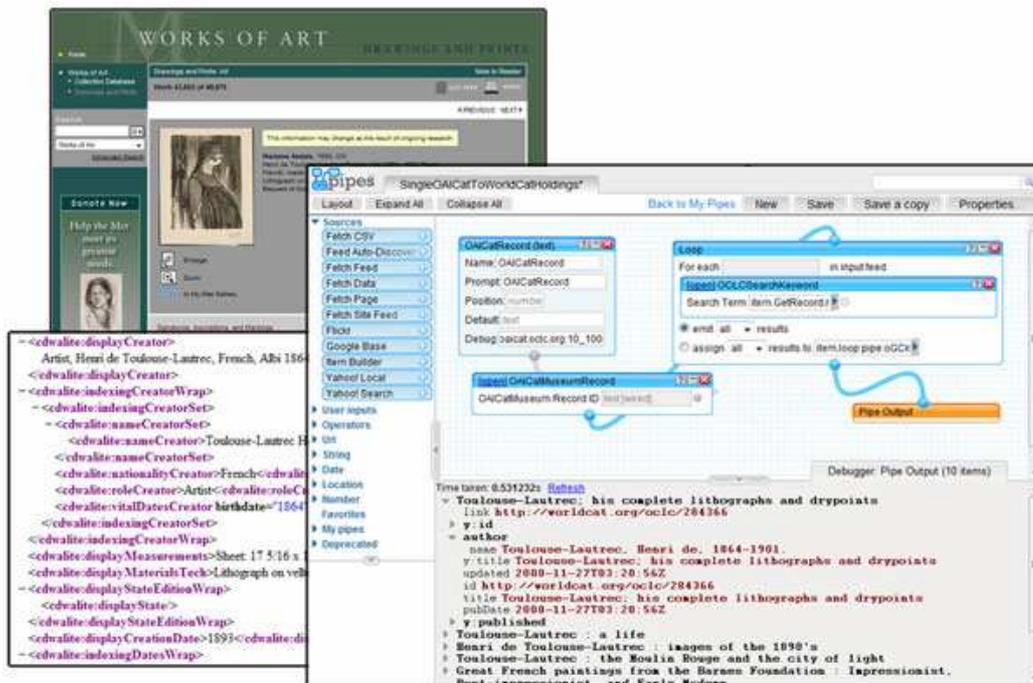
ricerca, per particolari comunità come quelle legate all'editoria e alla distribuzione libraria. OCLC sta infatti cercando di rendere i propri metadati disponibili per il mashup attraverso diversi strumenti:

- il catalogo può essere interrogato da dispositivi come gli iPhone;
- sono stati creati dei widget che consentono di fare una ricerca libera sul catalogo restituendo i risultati in una visualizzazione ristretta che consente però di fare lo scroll di tutti i risultati fino al livello di dettaglio del singolo documento; altri widget consentono di svolgere una ricerca preimpostata su termini particolari (per esempio sul tema di interesse del sito su cui il widget è posizionato);
- sono state create delle API che consentono vari tipi di interazione coi dati:
 - permettono di interrogare il catalogo ed estrarre liste di record bibliografici o singoli record relativi al possesso di particolari biblioteche; di ricevere informazioni sulle biblioteche che possiedono i documenti; di creare link diretti alle biblioteche che possiedono particolari documenti;
 - a fronte dell'immissione di un numero ISBN, consentono di ricevere una lista di numeri ISBN associati al primo un modo da mostrare altre edizioni dell'opera di cui si è immesso il numero standard;
 - a fronte dell'immissione di un numero ISSN, consentono di ottenere informazioni sul seriale corrispondente inclusi i titoli alternativi e i record che ricostruiscono la storia del periodico (continuazione di, continua con)
 - consentono di interrogare il WorldCat registry e ricevere informazioni di base o dettagliate sulle istituzioni aderenti a OCLC;
- sono stati creati dei plug-in che interrogano il catalogo nella forma di un'Applicazione di Facebook, di un'estensione di ricerca del browser Firefox, a partire da un link nella Toolbar di Google⁹⁵.

Per dare solo un esempio del possibile riuso dei dati e delle applicazioni messi a disposizione da OCLC si pensi a questo prototipo creato dal Metropolitan Museum of Art di New York: la collezione online delle opere possedute dal Museo è costruita mettendo insieme dati di tipo museale in formato di scambio combinati a dati bibliografici attraverso il tool

⁹⁵ Vedi <www.worldcat.org> in particolare la sezione "Use WorldCat tools".

Yahoo pipes⁹⁶. I dati bibliografici sono record prelevati con le Search API di WorldCat, informazioni sui libri dell'autore o sull'autore presi da WorldCat Identities, termini collegati per estendere la ricerca presi da OCLC Terminology Service⁹⁷.



Anche altre istituzioni negli Stati Uniti e in altri paesi del mondo stanno lavorando sul tema dell'apertura dei dati, sperimentando in diversi settori. Il gruppo di lavoro della Library of Congress sul futuro del controllo bibliografico si esprime in questo modo

Library bibliographic data will move from the closed database model to the open Web-based model wherein records are addressable by programs and are in formats that can be easily integrated into Web services and computer applications. This will enable libraries to make better use of networked data resources and to take advantage of the relationships that exist (or could be made to exist) among various data sources on the Web (On 2008 p. 26)⁹⁸.

⁹⁶ Yahoo Pipes è uno strumento che consente di aggregare, manipolare e fare mashup di contenuti sul Web usando comandi relativamente semplici <<http://pipes.yahoo.com/pipes/>>.

⁹⁷ Per una lista di esemplificazioni del possibile uso di questi strumenti software si veda la Application gallery all'indirizzo <<http://www.oclc.org/us/en/registry/applicationgallery/default.htm>>. Per la descrizione del Terminology Services vedi <<http://www.oclc.org/research/activities/termservices/default.htm>>.

⁹⁸ I dati bibliografici della biblioteca passeranno da un modello di database chiuso al modello aperto Web-based in cui i record sono indirizzabili da programmi e sono in formati che possono facilmente essere integrati in servizi Web e applicazioni per computer. Questo consentirà alle biblioteche di fare miglior uso di risorse collegate in rete e di usufruire delle relazioni che esistono (o possono essere create) tra varie fonti di dati sul Web.

Ci sono certamente molti modi per realizzare quest'apertura ma sicuramente i dati dovranno trasformare la loro attuale struttura essenzialmente basata su una descrizione testuale per diventare un insieme di elementi ai quali possano essere applicate elaborazioni automatiche. Questi dati dovranno essere compatibili con la direzione principale intrapresa dalla tecnologia cioè basarsi sul World Wide Web. Una delle possibili direzioni per le biblioteche è unirsi alla nuvola dei Linked Data, un insieme crescente di dati sul Web che molti vedono come un'iniziativa portatrice di grandi promesse per il futuro delle risorse informative (Coyle 2010 b p. 14). Questo ci introduce al concetto di Web semantico, la visione e le tecnologie che sono alla base dei Linked Data.

4.3 Il Web semantico

Il concetto di Web semantico⁹⁹, presentato come novità alla World Wide Web Conference nel 1994 (Coyle 2008 b p. 263), è stato introdotto al grande pubblico nel 2001 da un articolo di Berners-Lee, Hendler e Lassala su *Scientific American* (Berners-Lee – Hendler – Lassila, 2001). Vi veniva proposta una futuribile visione del Web popolato da “bots”, insetti con quattro zampe e un'antenna che sfrecciano sul Web raccogliendo e analizzando dati e consultando “ontologie” rappresentate da oggetti simili a libri. Questi bots sarebbero stati in grado di prendere un appuntamento dal dottore e segnarlo nel calendario del fratello incaricato di accompagnare la madre all'appuntamento (Coyle 2009). In sostanza programmi di ricerca intelligenti (software agents) sarebbero stati in grado di dedurre sofisticate inferenze da metadati legati a informazioni basate sul web e operare di conseguenza (Robu – Robu – Thirion 2006 p. 199). Ancora oggi non esiste una definizione formale di Web semantico in quanto la complessità e varietà delle applicazioni che si riferiscono ad esso cresce continuamente in diversi ambiti di sviluppo (integrazione di dati, analisi e rappresentazione della conoscenza, servizi di catalogazione, miglioramento di algoritmi e metodi di ricerca, social network ecc.) ciascuno dei quali enfatizza aspetti differenti delle tecnologie del Web semantico. Possiamo dire che il Web semantico si occupa essenzialmente di due cose: creare formati comuni per l'integrazione e la combinazione di dati estratti da diverse fonti e creare linguaggi per registrare come i dati siano correlati al mon-

⁹⁹ Per le informazioni sul Semantic Web si fa riferimento, oltre agli articoli citati, alla relativa sezione del W3C < <http://www.w3.org/2001/sw/>> e siti collegati.

do degli oggetti reali (W3C 1994 a). L'adozione di questi formati e linguaggi consentirà alle persone, o alle macchine, di partire da un database e muoversi tra un insieme infinito di database che sono connessi tra loro non da reti fisiche ma dal fatto di trattare della stessa cosa, ottenendo rapidamente le informazioni desiderate, indipendentemente dal linguaggio utilizzato per cercarle.

Il Web semantico, che si definisce il Web dei dati al posto del Web dei documenti, può essere considerato come un'estensione del Web attuale. Il World Wide Web è stato sviluppato inizialmente come un modo per consentire agli studiosi di scambiarsi tra loro documenti in modo rapido ed economico. L'idea dei link è analoga a quella della citazione in un articolo accademico: da un punto in un articolo rinvia ad un altro articolo. Il legame tra i due testi non è chiarito dal link ma dipende dall'interpretazione del contesto verbale in cui si trova: può essere la citazione di un passaggio o il riferimento ad una pubblicazione precedente da cui l'autore ha tratto ispirazione: niente nel link stesso spiega il significato del link (Coyle 2009). Va inoltre notato che il link va solo in una direzione: un documento non è a conoscenza dei link che sono puntati verso di lui (Coyle 2010 a p. 18). Nel Web oggi ci sono moltissime pagine che non hanno niente in comune con un documento scientifico e il suo contenuto, il Web si è evoluto in qualcosa di molto più eterogeneo per forma e contenuto. La creazione sul Web di opere di reference come dizionari, enciclopedie, mappe ha consentito di creare link di tipo diverso che non sono più tra documento e documento ma tra termini e concetti. Per esempio all'interno di un documento ci si può riferire ad una persona e creare un link tra quel nome e la corrispondente voce in Wikipedia. Queste capacità hanno fatto comprendere che i documenti hanno al loro interno dei dati che possono essere trattati come informazioni in se stesse (Coyle 2009). L'oggetto dei motori di ricerca tuttavia resta il testo e la ricerca sulle parole del testo è incompleta perché non è in grado di risolvere ambiguità, polisemia, sinonimia e tutte le problematiche legate al linguaggio: serve qualcosa che consenta di trovare le informazioni dentro i testi. Si deve passare da un Web dei documenti ad un Web dei dati che esistono dentro i documenti o, detto in altri termini, dal linguaggio umano contenuto nei documenti alle informazioni sottostanti e bisogna rendere quelle informazioni usabili dalle macchine. Nelle parole di Berners-Lee, Hendler, and Lassala:

To date, the World Wide Web has developed most rapidly as a medium of documents for people rather than of information that can be manipulated automatically. By augmenting

Web pages with data targeted at computers and by adding documents solely for computers, we will transform the Web into the Semantic Web (Berners-Lee – Hendler – Lassila, 2001).

La tabella che segue riassume le differenze tra il Web attuale e la sua evoluzione¹⁰⁰

	Web dei documenti	Web semantico
Analogia	Filesystem globale	Database globale
Destinato a	Uso umano	Machine prima, umani dopo
Oggetti primari	Documenti	Cose (o descrizioni di cose)
Link tra	Documenti (o parti di documenti)	Cose
Strutturazione degli oggetti	Abbastanza bassa	Alta
Contenuto semantico e link	Implicito	Esplicito

L'idea di base del Web semantico è quella di connettere, cercare e fare uso di dati dentro i documenti. Questo richiede non solo di scoprire le informazioni ma anche di capire come creare link significativi tra questi nuclei di dati: non basta dire “A ha relazione con B” bisogna anche dire di che tipo di relazione si tratti e inoltre questa relazione deve essere tanto comprensibile agli esseri umani quanto utilizzabile dalle macchine per prendere decisioni operative (Coyle 2009). Comunicare il senso di una relazione tra esseri umani è abbastanza semplice: se dico che *Calvino* è autore di *Il barone rampante* tutti comprendono quello che voglio dire perché colgono il contesto delle mie parole, sanno che *Calvino* è una persona, che *Il barone rampante* è un'opera letteraria e sanno che cosa significa “essere autore di”. Per rendere la stessa frase comprensibile da un computer perché possa manipolarla e usarla dobbiamo spiegare tutto il contesto e farlo in un modo che lo possa recepire. Prendiamo la frase “*Herman Melvill* fu uno scrittore americano di metà '800 e scrisse numerose opere letterarie tra le quali la più famosa è *Moby Dick*”. Contiene una serie di fatti che nessun sistema di recupero delle informazioni è in grado di dedurre dal testo: *Herman Melvill* era americano, era uno scrittore, non è più vivente, ha scritto *Moby Dick*, *Moby Dick* è un libro, il libro *Moby Dick* è famoso. Insegnare ad un computer a comprendere semplici deduzioni è stata la meta delle discipline dell'intelligenza artificiale per oltre mezzo secolo ma si sono raggiunti scarsi risultati. Nella visione del Web semantico, perché le macchine possano avere questa comprensione è necessario che alcune parti del testo siano codificate in modi particolari: si tratta di creare metadati che siano usabili dai computer. Per fare questo il Web semantico si serve di tre elementi chiave: identità, relazioni e

¹⁰⁰ La tabella è liberamente rielaborata da Heath (2009).

regole. Va precisato che il termine “semantico” non deve essere inteso nel senso linguistico di “relativo al significato delle parole” ma nel modo in cui viene usato in matematica, nei linguaggi formali: interpretazione di una formula o definizione delle possibili azioni di calcolo che un linguaggio può fare secondo le sue regole (Coyle 2010 a p. 19).

Identità: Per il funzionamento dei Linked data l’identificazione dei concetti e delle relazioni è un punto fondamentale. Gli esseri umani hanno una grande capacità di tollerare l’ambiguità del linguaggio: se parlando con un amico questi mi nomina la Georgia dal contesto della conversazione sono in grado di capire se intende lo stato americano o quello caucasico. Perché la differenza sia compresa dal computer invece, ciascuna di queste due entità deve avere un proprio identificatore univoco sia perché il computer è poco in grado di capire il contesto, sia perché questo rende possibile utilizzare questo dato perfettamente identificato fuori dal contesto, cioè fuori dal documento che lo contiene, come dato atomico nel Web. Nel Web semantico le identità sono stabilite attraverso uno tipo specifico di identificatore, l’Uniform Resource Identifiers (URI) e in particolare vengono create nella forma di un Uniform Resource Location (URL), quindi in molti casi l’identificatore coinciderà con l’identificazione della localizzazione. Un insieme di identificatori viene definito nel Web semantico “ontologia” o anche vocabolario. Le ontologie possono essere espresse anche in forme comprensibili agli esseri umani, ci sarà un URI che punta a un documento che le descrive in forma leggibile, e possono anche avere informazioni aggiuntive per consentire l’elaborazione automatica, in altre parole possono contenere relazioni con altri termini all’interno dello stesso vocabolario.

Relazioni: Nell’esempio di Melville e Moby Dick abbiamo una persona e un libro, la loro relazione è che la persona ha scritto il libro. Come possiamo renderla comprensibile ad una macchina? Possiamo definire la relazione dandogli un’identità, dire che è una relazione “X”. Questo non consentirà al computer di comprendere in senso umano la relazione ma potrà riconoscere l’identificatore “X” e ogni volta che lo incontra capire che è la stessa relazione perché ha lo stesso identificatore.

Regole: In un ambiente automatico la definizione dei dati non è sufficiente, è necessario avere delle regole che governino le inferenze tra i dati. Ogni applicazione può creare le proprie regole ma queste devono essere interpretabili dai computer.

Vediamo di seguito l’esempio, tratto da Coyle (2009), di un semplice vocabolario di ambito biblioteconomico e delle sue regole (espresso in forma leggibile dagli umani e non da

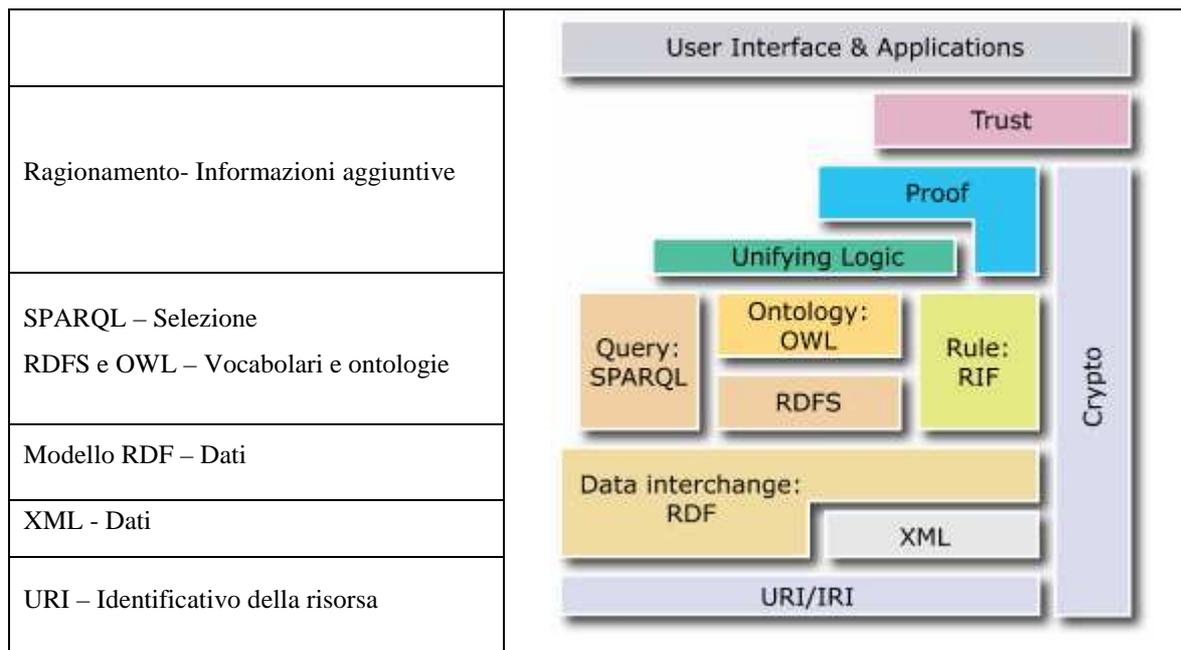
una macchina). Per ogni termine abbiamo una parola, un identificatore che potrà essere usato in una applicazione, una definizione; viene definita una semplice relazione, seguita da un insieme di regole.

identificatori	Term 1 : Series Identifier: http://www.example.com/publishingTerms/3279 Definition: A group of documents published in an order over time
	Term 2 : Book Identifier: http://www.example.com/publishingTerms/101 Definition: An independently published document
relazioni	Relationship: isMemberOf Identifier: http://www.example.com/publishingTerms/73 Definition: Belonging to a set
regole	Book can be “isMemberOf” Series
	Series cannot be “isMemberOf” Book
	Series can be defined as the sum of all Books with relationship isMemberOf
	Aggiungendo ulteriori informazioni alla definizione di series e book si possono definire anche queste regole
	Series can be ordered by: series number
	Series can be ordered by: publication date

Con la definizione dei termini e delle regole e di questa particolare relazione “fa parte di” i libri che fanno parte di una collana possono essere identificati e mostrati in uno degli ordinamenti possibili ovunque nel Web da parte di una qualsiasi comunità di utenti, vale a dire fuori dai sistemi chiusi, tra i quali gli opac, in cui questi termini e queste relazioni sono oggi usati in molteplici schemi di metadati tipici di specifiche comunità (bibliotecari, editori, aggregatori). Questa evoluzione del Web dovrebbe, infatti, fornire la struttura che consente, con l’adozione di tecnologie comuni, di scambiare e riutilizzare dati attraverso diverse applicazioni, imprese e comunità per creare migliori applicazioni e servizi. Per esempio potrebbe essere usata nel campo dell’integrazione dei dati, per mettere insieme più facilmente dati provenienti da diverse fonti e in differenti formati; nel campo della classificazione ed esplorazione delle risorse potrebbe servire a sviluppare motori di ricerca per specifici domini; nel campo della catalogazione essere adottata per descrivere contenuti e relazioni; potrebbe essere usata nella gestione dei diritti di proprietà dei documenti e delle informazioni.

4.4 Componenti del Web semantico

La definizione di identità, relazioni e regole, che abbiamo introdotto per poter almeno figurarci in modo semplicistico il senso di quello che potrebbe essere il Web semantico, è solo una parte del complesso insieme di concetti e tecnologie che costituisce l'architettura del Web semantico. Questa viene di solito definita come Semantic stack o Semantic layer cake e rappresentata con l'immagine che segue.



L'architettura complessiva è costituita da sette livelli e diversi elementi ognuno dei quali usa le tecnologie rappresentate nei livelli inferiori. I livelli più bassi contengono le tecnologie che sono già oggi usate nel Web e che forniscono le basi al Web semantico: sono anch'esse rappresentate nello schema a riprova del fatto che il Web semantico sia un'estensione, e non una cosa diversa, del Web come lo conosciamo oggi.

1) Il primo livello, livello dell'URI (Uniform Resource Identifier) che include URL (Uniform resource Location) e URN (Uniform Resource Name)¹⁰¹, serve a dare il nome alle cose e alle risorse informative e a fornire i mezzi per accedere alle informazioni che li riguardano nel Web (Thompson 2008, Dizionario 2006 p. 638). L'assunto di base è che ciascuna risorsa Web, compresi i concetti del mondo reale (persone, organizzazioni, luoghi

¹⁰¹ Vedi <<http://www.w3.org/Addressing/>>, <<http://www.w3.org/TR/uri-clarification/>>.

e cose), devono avere un proprio URI. Il Web semantico ha bisogno di identificatori univoci perché le manipolazioni delle risorse che avvengono ai livelli superiori siano validabili (Coyle 2010 a p. 21). Un sistema di indirizzamento globale standardizzato è la base del web semantico: estendendo il concetto di URI ad ogni cosa possa avere una presenza nel web si incrementa il numero delle entità che possono essere rappresentate e rese oggetto di inferenze ed elaborazioni eseguite dalle macchine (Campbell – Fast 2004 p. 385).

2) L'XML (Extensible Markup Language)¹⁰² è un linguaggio di marcatura che consente la creazione di documenti composti da dati strutturati. Il livello dell'XML riguarda i marcatori di contenuto e la sintassi di validazione da usare nei vocabolari e nei datatypes definiti dagli utenti. Viene usato per la serializzazione delle dichiarazioni RDF (Yang - Yi Lee - Xu 2009 p. 2).

3) L'RDF (Resource Description Framework)¹⁰³ è un linguaggio per la codifica, lo scambio e il riutilizzo di metadati strutturati e consente l'interoperabilità tra applicazioni (W3C 1994 b, Dizionario 2006 p. 506). E' stato creato per rappresentare informazioni su risorse presenti nel Web, in particolare per rappresentare metadati sulle risorse, come il titolo, l'autore, le date di modifica di una pagina, informazioni di copyright e di licenze, le informazioni sulla disponibilità di risorse condivise. Poiché adotta un concetto esteso di risorsa Web, RDF può essere usato per rappresentare qualsiasi cosa possa essere identificata nel Web ed inoltre è pensato per situazioni in cui l'informazione debba essere elaborata dalle applicazioni piuttosto che essere solo mostrata alle persone. RDF fornisce un modello comune per esprimere le informazioni in modo che queste possano essere scambiate tra le applicazioni senza perdite di significato: questo fa sì che le informazioni possano essere rese disponibili per applicazioni diverse da quelle che le hanno create. RDF si basa sull'idea di identificare le cose in modo univoco attraverso un URI e descriverle nei termini di semplici proprietà e valori di quelle proprietà. I concetti principali su cui si basa RDF sono quelli di risorsa, classe, proprietà e valore. Risorsa è un termine neutro per indicare un soggetto sul web. Il Web semantico non pone limiti a quello che può essere descritto: ogni cosa e ogni concetto possono essere definiti, anche una relazione può essere una risorsa.

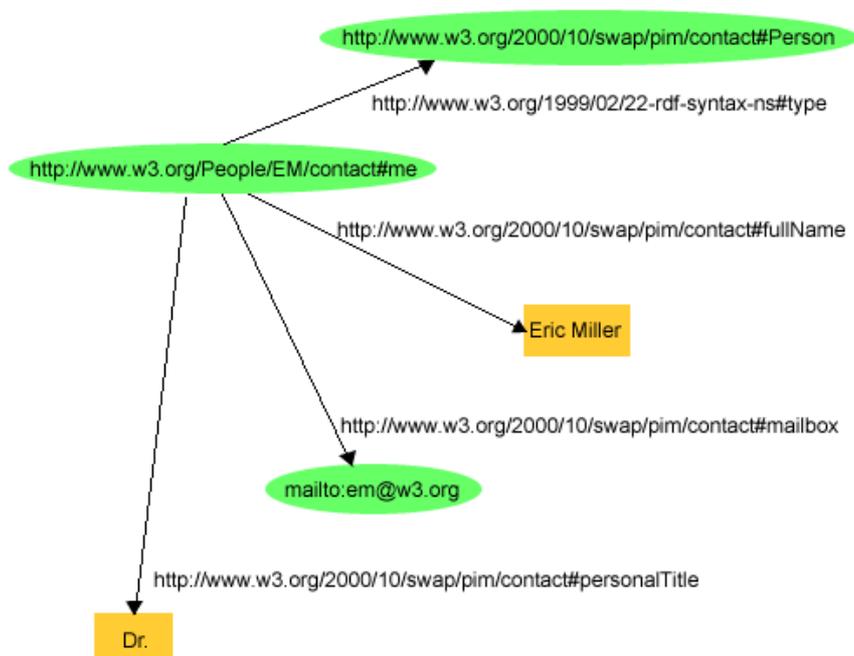
¹⁰² Vedi <<http://www.w3.org/XML/>>

¹⁰³ Vedi <<http://www.w3.org/RDF/>>.

Con classe ci si riferisce ad un raggruppamento di risorse simili che hanno caratteristiche comuni che le rendono parte di uno stesso insieme: è simile al concetto di classe usato nelle tassonomie scientifiche ma qui le classi non hanno un valore universale, possono cioè essere definite classi che sono utili solo nel contesto dei metadati che si stanno utilizzando (per esempio la classe delle cose che io possiedo può includere auto, computer, cane). Le classi possono essere divise in sottoclassi (Coyle 2010 a p. 20-21, Robu – Robu – Thirion 2006 p. 200). Le proprietà sono le relazioni che legano tra loro risorse e valori e sono anch'esse identificate da URI. Un valore, invece, è un tipo di dato primitivo, è il contenuto effettivo dell'elemento, che può essere una stringa, un numero, una data, oppure essere l'URI di una risorsa collegata al soggetto. L'unità base per rappresentare un'informazione in RDF è la dichiarazione (statement): una dichiarazione è una tripla Soggetto – Predicato – Oggetto, dove il soggetto è una risorsa, il predicato è una proprietà e l'oggetto è un valore. Prendiamo per esempio lo statement (Manola – Miller 2004 parte 1):

"there is a Person identified by <http://www.w3.org/People/EM/contact#me>, whose name is Eric Miller, whose email address is em@w3.org, and whose title is Dr."

Può essere rappresentato da un grafo sui cui nodi ci sono risorse o tipi primitivi (valori) e i cui archi rappresentano le proprietà, come nell'immagine che segue:



In questo esempio RDF ha rappresentato:

- un individuo, Eric Miller, identificato da <http://www.w3.org/People/EM/contact#me>

- un tipo di cosa, persona, identificata da <http://www.w3.org/2000/10/swap/pim/contact#Person>
- una proprietà di quella cosa, la sua casella di posta, identificata da <http://www.w3.org/2000/10/swap/pim/contact#mailbox>
- un valore di quella proprietà, <mailto:em@w3.org> (il valore della proprietà Casella di posta)

I valori, come “Eric Miller” e “Dr.” possono essere registrati sotto forma di stringhe, numeri, date. RDF fornisce anche una sintassi per rappresentare (serializzare), registrare e scambiare questi dati in formato XML (RDF/XML)¹⁰⁴. Come l’HTML, l’RDF/XML è elaborabile dal computer e, usando l’URI, può linkare parti di informazioni ovunque sul Web ma diversamente dai normali ipertesti, gli RDF URIs possono riferirsi a qualsiasi cosa identificabile comprese quelle che normalmente sul Web non sarebbe possibile recuperare, come la persona Eric Miller. Inoltre le proprietà stesse di RDF sono identificate da un URI allo scopo di individuare precisamente la relazione che esiste tra le cose linkate.

4) Il quarto livello comprende il linguaggio di interrogazione SPARQL, lo Schema RDF, le ontologie e il RIF.

Il linguaggio SPARQL (SPARQL Protocol and RDF Query Language)¹⁰⁵ è un linguaggio di interrogazione che consente di formulare query su risorse che siano registrate nativamente come RDF oppure viste come RDF attraverso middleware cioè attraverso programmi informatici che fungono da intermediari (Yang-Yi Lee- Xu 2009 p. 3). Definisce il linguaggio di interrogazione (attraverso quattro forme di base Select, Construct, Ask, Describe), il formato dei risultati (che possono essere accessibili da API, serializzati in XML o in grafi RDF o altro), il protocollo di accesso ai dati.

Lo schema RDF (RDFS)¹⁰⁶ consente alle comunità di utenti di definire propri vocabolari cioè di indicare quali termini intendono usare nei propri statement, in particolare fornisce gli strumenti per indicare quali tipi o classi di risorse appartengono al proprio dominio e quali proprietà vengono usate nella descrizione delle risorse. Per esempio coloro che sono interessati alle risorse di tipo bibliografico avranno interesse a definire classi come “libro”

¹⁰⁴ Vedi <<http://www.w3.org/TR/REC-rdf-syntax/>>.

¹⁰⁵ Vedi <<http://www.w3.org/TR/rdf-sparql-query/>>.

¹⁰⁶ Vedi <<http://www.w3.org/TR/rdf-schema/>>.

o “articolo di rivista” e ad usare proprietà come “autore”, “titolo”, “soggetto” per descriverle.

Questi vocabolari devono essere costruiti secondo un modello particolare indicato nello schema RDF (il quale a sua volta è costruito secondo lo stesso schema). Possiamo quindi dire che lo schema RDF è un vocabolario per descrivere classi e proprietà di un set specialistico: non dice quali caratteristiche devono avere le classi e le proprietà da usare in specifiche applicazioni, dice che devono essere descritte nel modo previsto dallo schema (Manola – Miller 2004 parte 5).

Le ontologie sono il tentativo di formare uno schema concettuale esaustivo e rigoroso nell’ambito di un particolare dominio (Robu – Robu – Thirion 2006 p. 200-201, Sure - Studer, 2005 p. 191-192). Si tratta di strutture gerarchiche che contengono tutte le entità rilevanti, le relazioni esistenti fra esse, le regole, gli assiomi e i vincoli di uno specifico dominio. Possono essere molto complesse, fino a migliaia di termini, oppure descrivere pochi concetti. Il linguaggio OWL (Web Ontology Language)¹⁰⁷ estende lo schema RDFS consentendo di realizzare costrutti più complessi per descrivere la semantica delle dichiarazioni RDF. Consente di esplicitare restrizioni di valori, unicità, cardinalità, di rappresentare relazioni complesse come l’equivalenza tra classi, di descrivere caratteristiche delle proprietà come la transitività (Robu – Robu – Thirion 2006 p. 202). Diverse comunità di utenti stanno già sviluppando le proprie ontologie. Il linguaggio SKOS (Simple Knowledge Organization System)¹⁰⁸ ha lo scopo di favorire una più facile pubblicazione di vocabolari controllati e strutturati (tesauri, schemi di classificazione e soggettazione, tassonomie) per il Web semantico basati su RDF e RDFS. Tra i vocabolari già in uso da parte di diverse applicazioni si possono citare: FOAF (Friend of Friend)¹⁰⁹ un’ontologia leggibile dai programmi informatici che descrive le persone, le loro attività, le loro relazioni con persone e oggetti; SIOC (Semantically-Interlinked Online Communities)¹¹⁰ un’ontologia che rappresenta informazioni contenute nei gruppi sociali del Web (blog, forum, mailing list); Dublin

¹⁰⁷ Vedi < <http://www.w3.org/TR/owl-features/>>, <<http://www.w3.org/2004/OWL/>>.

¹⁰⁸ Vedi < <http://www.w3.org/TR/skos-reference/>>.

¹⁰⁹ Vedi <<http://www.foaf-project.org/>>.

¹¹⁰ Vedi <<http://sioc-project.org/>>.

Core¹¹¹ un insieme di elementi di metadati utili per la descrizione di risorse informative digitali (video, registrazioni sonore, immagini, testi, multimediali)¹¹².

Il RIF (Rule Interchange Format)¹¹³ è in corso di sviluppo da parte di un apposito gruppo del W3C. Nel conteso del Web semantico con il termine “regole” ci si riferisce ad elementi di programmazione logica che in genere si presentano nella forma “se condizione allora conseguenza”. Un esempio molto semplice può chiarire il concetto: integrando dati provenienti da diverse fonti posso avere riferimenti a persone, i loro nomi, homepage, indirizzo e-mail ecc. I dati non dicono tuttavia sotto quale condizione due persone devono essere considerate identiche: deve essere espressa una condizione che consenta di affermarlo, una regola che dica “se due persone hanno lo stesso nome, la stessa homepage, lo stesso indirizzo e-mail, allora sono identiche” (Kifer - Boley 2009, W3C 1994 c). Questa condizione può essere espressa in molti modi diversi. Esistono vari sistemi di regole molto diversi gli uni dagli altri, usati da comunità diverse in contesti differenti e non è possibile arrivare ad un unico linguaggio di regole che li comprenda tutti: è possibile tuttavia definire un nucleo minimo che possa essere compreso da tutti i sistemi di regole. Il gruppo di lavoro sta studiando un linguaggio base, e le sue estensioni, che insieme consentano alle regole di essere tradotte nei diversi linguaggi di regole esistenti e sia così trasferito ai diversi sistemi di regole, dopo aver trovato il consenso delle comunità interessate.

5-7) Gli ultimi livelli del Semantic stack, quello logico, cioè del ragionamento deduttivo nel Web semantico, della validazione e della fiducia sono ancora in corso di sviluppo anche se la struttura e le tecnologie sono già state definite negli ultimi 10 anni (Yang - Yi Lee - Xu 2009 p. 5). La possibilità di verificare e assicurare che le dichiarazioni semantiche provengano da una fonte affidabile è molto importante e potrà essere conseguita dall'utilizzo di firme digitali adeguate al formato RDF. Lo studio di meccanismi di autenticazione è parte delle attività del W3C e si sta concretizzando nel protocollo POWDER (Protocol for Web Description Resources)¹¹⁴ (Bermès 2009 p. 10). La crittografia è un aspetto che sottende al processo di validazione. La fiducia nelle dichiarazioni derivate

¹¹¹ Vedi <<http://dublincore.org/>>.

¹¹² Una lista di ontologie con il rinvio ai rispettivi siti si può vedere all'indirizzo <http://umbel.org/lod_constellation.html>.

¹¹³ Vedi <<http://www.w3.org/TR/rif-core/>>.

¹¹⁴ Vedi <<http://www.w3.org/TR/2009/NOTE-powder-primer-20090901/>>.

dovrebbe essere sostenuta sia dalla possibilità di verificare le fonti da cui provengono le premesse sia dall'affidarsi alla correttezza della logica formale nel processo di derivazione di nuove informazioni. L'ultimo livello riguarda lo sviluppo di interfacce che consentano agli esseri umani di usare le applicazioni del Web semantico.

4.5 Linking Open Data initiative e Linked data

La Linking Open Data initiative nasce intorno al 2007 all'interno di un gruppo di interesse del W3C¹¹⁵ dedicato alla diffusione delle soluzioni e delle tecnologie del Web semantico (SWEO Semantic Web Education and Outreach interest group)¹¹⁶ (Bermès 2009 p. 1). Le possibilità aperte dalle tecnologie del Web semantico sono molto sofisticate, vanno dal campo delle ontologie a quello del ragionamento automatico e della programmazione logica, e per molto tempo sono rimaste riservate al campo della ricerca. Si voleva quindi cercare di far passare il Web semantico dall'essere una "visione" all'essere una realtà educando la comunità del Web alle sue tecnologie e potenzialità e mostrando in breve tempo realizzazioni di successo. Fu quindi lanciata la Linking Open Data initiative per promuovere l'idea del Web semantico come database globale di dati interconnessi comprensibili da macchine che su questi dati sono in grado di creare applicazioni. La Linking Open Data initiative ha sviluppato linee guida per la costruzione di questo tipo di Web e incoraggiato una varietà di istituzioni a pubblicare contenuti non coperti da licenze nella forma di triple RDF e a linkarli nel Web attraverso URI. L'approccio pratico adottato, di puntare ad una semplice conversione dei dati in formato RDF e alla loro immediata pubblicazione piuttosto che alla realizzazione di complesse ontologie, ha fatto sì che l'iniziativa potesse essere pensata come un sottoinsieme del Web semantico di più facile adozione. Nel 2008 il gruppo SWEO ha concluso i suoi lavori ma l'iniziativa aveva raggiunto una sua diffusione tanto da poter continuare a vivere col nome di Linked Data (Bermès 2009 p. 2). I Linked Data sono un modo di creare tipi di link tra dati provenienti da diverse fonti: possono essere diversi database gestiti da differenti organizzazioni in luoghi geografici distanti oppure semplicemente sistemi eterogenei all'interno della stessa organizzazione che fino ad ora non erano interoperabili a livello dei dati. Nella visione di questo progetto

¹¹⁵ Il World Wide Web Consortium è una comunità internazionale che si occupa di sviluppare standard per il WEB <http://www.w3.org/>.

l'interoperabilità non è limitata alla realizzazione di specifiche API che danno accesso a silos di dati altrimenti isolati ma è basata sulla connessione dei dati ad un livello di granularità più basso. Mentre l'unità base del Web ipertestuale sono i documenti HTML (HyperText Markup Language) collegati da link non qualificati, i Linked Data si basano su documenti che contengono dati in formato RDF. I Linked Data non si limitano a collegare questi documenti ma usano il linguaggio RDF per fare delle dichiarazioni che creano link con cose reali. Più che parlare di Web dei dati di dovrebbe più precisamente dire "Web delle cose del mondo descritto attraverso dati sul Web" (Bizer – Heath - Berners-Lee 2009 p. 2). Con Linked Data ci si riferisce a dati pubblicati sul Web in modo che siano leggibili alle macchine, abbiano un significato definito esplicitamente, siano linkati ad altri dataset e possano a loro volta essere linkati: si tratta di un modo di pubblicare dati che incoraggia il riuso, riduce la ridondanza, massimizza le interconnessioni reali e potenziali tra i dati, utilizza gli effetti di rete per aggiungere valore ai dati (Heath 2009 p. 13). Lo scopo del progetto Linked Data è rendere disponibili grandi quantità di dati di tipo RDF come insiemi collegati e non come singole unità. Non si richiede di rispettare nessun modello nei dati e nelle loro proprietà se non l'essere espressi secondo le triple RDF e non ci si aspetta che tutti coloro che forniscono informazioni concordino sullo stesso URI: i vari set di dati sono legati da link ma mantengono la loro indipendenza e non è necessario che ogni set locale abbia conoscenza della forma e del modello degli altri data set. Quando ci si riferisce alla stessa entità del mondo reale con due URI diversi in due contesti diversi questi possono essere collegati da alias. Ogni tripla RDF è autosufficiente e non importa quali altre dichiarazioni possono essere fatte sulla stessa risorsa. Per fare un paragone con il tipo di interoperabilità cui le biblioteche sono abituate, è come se fosse possibile linkare un campo MARC di un record di un catalogo con un elemento Dublin Core dentro il record di un altro catalogo e usare questi link per manipolare insieme entrambi i record senza bisogno di nessuna mappatura o crosswalk dei metadati, scambio o import di dati, sviluppo di software ad hoc (Bermès 2009 p. 2). Nonostante questa apertura, è considerata una buona pratica riutilizzare i termini di vocabolari RDF ben conosciuti come FOAF, SIOC, Dublin Core ecc.. e solo se questi non contengono i termini necessari procedere con la creazione di nuovi termini (Bizer – Heath - Berners-Lee 2009 p. 4). Dal punto di vista tecnologico i

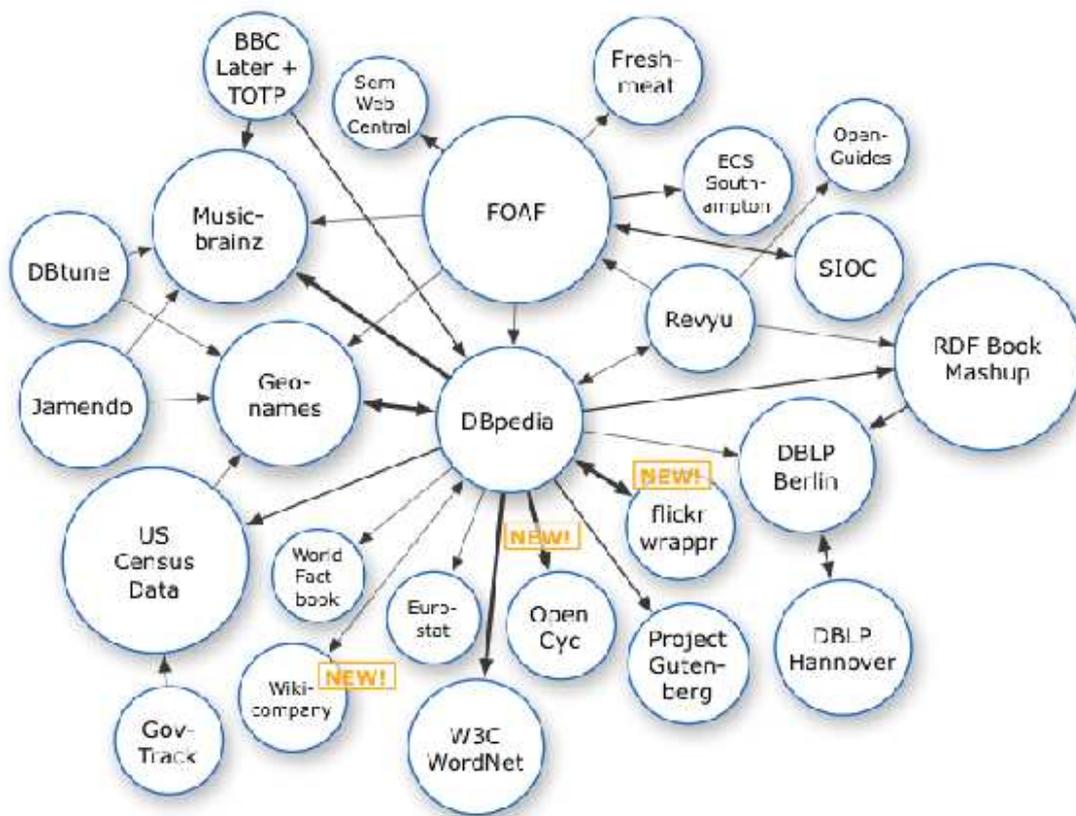
¹¹⁶ Vedi <<http://www.w3.org/2001/sw/sweo/>>.

Linked Data si basano su due tecnologie fondamentali del Web: l'URI per l'identificazione delle risorse e l'HTTP per il loro recupero associati a RDF come formato del contenuto. I principi base dell'architettura dei Linked data sono stati descritti da Barbers-Lee (2006):

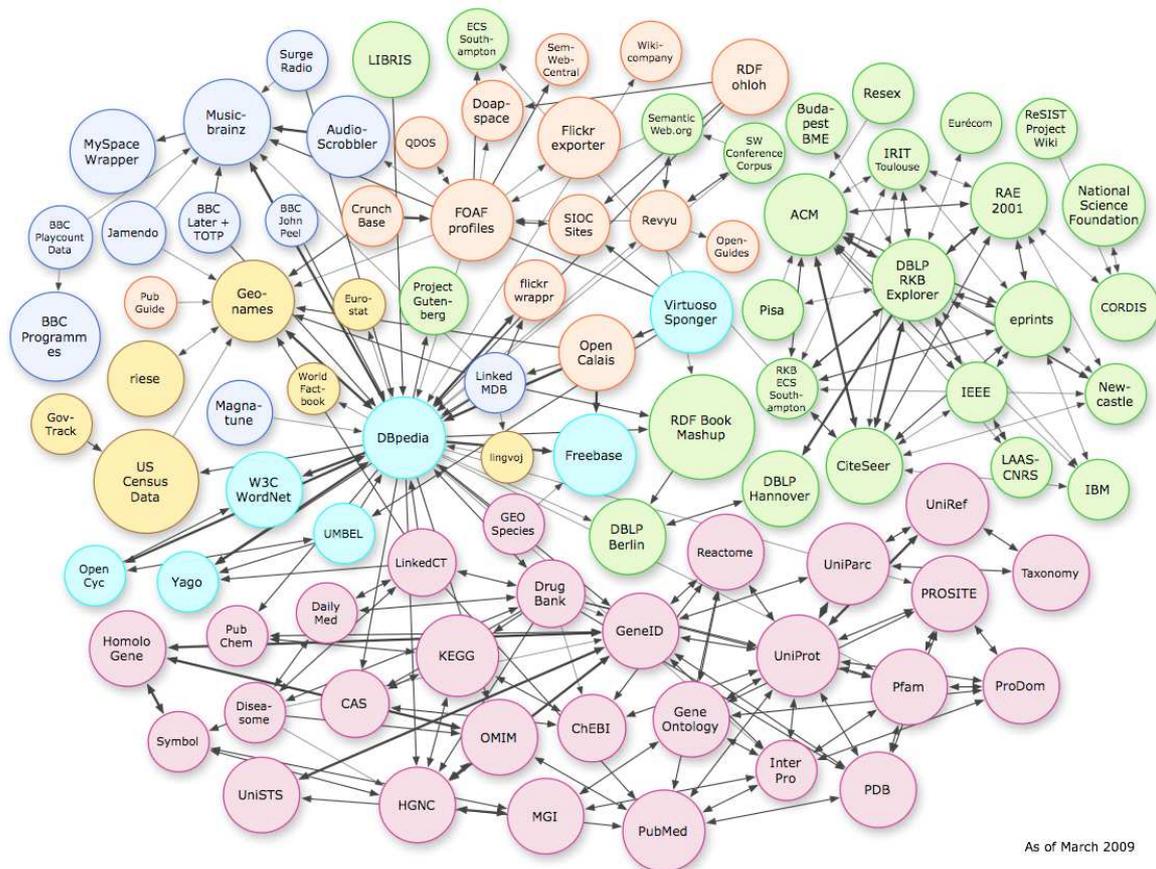
- usare gli URI come nomi delle cose (nomi globalmente unici, proprietà distribuita)
- usare HTTP URI perché le persone possano cercare quei nomi
- quando qualcuno cerca un URI dare informazioni utili usando gli standard (rdf, sparql)
- includere link ad altri URI in modo che possano scoprire più cose

Se questi principi non vengono osservati si perde l'opportunità di rendere i dati interconnessi, si limita il modo in cui possono essere usati in modi diversi da quelli attesi, mentre è proprio questo riuso delle informazioni che dà valore al Web.

Dal 2006 molte comunità hanno iniziato a creare i propri insiemi di dati nella forma di Linked data: la crescita della Linked Data cloud è molto veloce come si vede dalle rappresentazioni che seguono e attualmente ha raggiunto dimensioni tali da essere difficilmente rappresentabile.



set di Linked Data nel 2007



set di Linked Data nel 2009

Nelle rappresentazioni la dimensione dei cerchi indica il numero di triple RDF presenti nel dataset mentre la direzione delle frecce indica il dataset che contiene i link: se la freccia è bidirezionale significa che i link sono reciproci. Lo spessore indica il numero di link mentre il colore distingue i vari tipi di risorse. Il colore verde mare che ha come centro DBpedia rappresenta le risorse di tipo generale. DBpedia¹¹⁷ estrae informazioni strutturate da Wikipedia e le rende disponibili come triple RFD. È un progetto iniziato dall'Università di Berlino che ha dato grande impulso allo sviluppo di altri dataset in molte comunità ed è linkato da molte altre risorse. Il colore rosa rappresenta le risorse di tipo sociale, quelle legate ai social network. Il colore giallo rappresenta risorse di tipo geografico e statistico

¹¹⁷ Vedi < <http://dbpedia.org/About>>.

tra le quali la conversione in RDF di Eurostat¹¹⁸, la banca dati statistica della comunità europea, e GeoNames¹¹⁹ che contiene i dati di milioni di località geografiche di tutto il mondo ai quali si collegano moltissimi altri set di dati specialistici. Il colore azzurro rappresenta risorse multimediali mentre quello viola comprende le risorse di medicina e scienze della vita. Quelli in verde sono i dati di tipo bibliografico tra i quali il catalogo collettivo delle biblioteche svedesi, il primo catalogo reso disponibile come insieme di Linked Data (Malmsten 2009).

Questi insiemi di dati promuovono la realizzazione di nuovi tipi di applicazioni. Ci sono browser¹²⁰ per esplorarli che consentono all'utente di partire da una risorsa e navigare verso le risorse collegate non attraverso i link HTML ma seguendo le triple RDF. Ci sono motori di ricerca che seguono i link e consentono di interrogare dati aggregati come se si trovasse in un database locale: ne esistono sia di orientati agli essere umani sia più orientati alle applicazioni¹²¹. Il Web dei dati apre nuove possibilità non solo per le applicazioni generali ma anche per applicazioni in specifici domini. Per esempio DBpedia Mobile è un browser di linked data da usare su iPhone o altri dispositivi mobili che consente ad un utente di esplorare la città: basandosi sulla posizione dell'iPhone stabilita dal GPS invia foto collegate via Linked data estratte da Flickr e informazioni tratte da DBpedia su luoghi, edifici, risorse culturali ecc. presenti in quella zona e consente anche all'utente di pubblicare le sue foto e la sua posizione come Linked Data (Bermès 2009 p. 7, Becker – Bizer 2008).

Diversamente dai mashup del Web 2.0 realizzati attraverso le API, le applicazioni Linked Data operano in uno spazio globale non limitato e questo consente loro di migliorarsi, di dare una risposta più completa, man mano che nuove risorse compaiono nel Web. Le API per accedere ai dati usano un vasto insieme di meccanismi, rappresentano i dati in diversi formati e agiscono su un insieme prestabilito di fonti mentre i Linked Data usano un set standardizzato di tecnologie e non hanno limiti quanto all'insieme di dati cui fare riferimento. Inoltre, e qui sta la principale differenza concettuale, le API non assegnano identificatori globali ai dati quindi non consentono di creare legami tra i diversi insiemi di dati e di contribuire alla creazione di uno spazio globale (Bizer – Heath - Berners-Lee 2009 p. 2).

¹¹⁸ Vedi < <http://riese.joanneum.at/>>.

¹¹⁹ Vedi < <http://www.geonames.org/>>.

¹²⁰ Vedi per esempio Open Linking Data explorer <<http://linkeddata.uriburner.com/ode/>>.

Avendo questo spazio globale come base, le proposte più sofisticate del Web semantico, come quella degli agenti intelligenti, potranno forse trovare realizzazione. Non mancano tuttavia aree problematiche sulle quali la comunità del Web semantico sta ragionando e lavorando e che qui si accennano:

Interfacce utenti e paradigma di interazione: non è ancora chiaro come l'utente possa essere messo in grado di interagire con applicazioni costruite sul raggruppamento dinamico di tante risorse, come possa include o escludere fonti e mantenere una visione comprensibile attraverso appropriate interfacce.

Schemi di mappatura e fusione di dati: una volta che risorse provenienti da diverse fonti sono state recuperate è necessario integrarle e per farlo è necessario mappare i diversi vocabolari e risolvere i conflitti che si creano per poter unificare i dati che riguardano la stessa entità.

Mantenimento dei link: la persistenza dei link è fondamentale. Oggi gli RDF link sono aggiornati solo sporadicamente.

Licenze: è necessario costruire schemi che consentano di pubblicare le specifiche condizioni sotto le quali i dati possono essere riutati e ripubblicati rispettando i diritti connessi.

Fiducia, qualità e rilevanza: si tratta di trovare modi per recuperare le risorse più rilevanti, affidabili, di qualità e appropriate per gli utenti in questo nuovo contesto analogamente a quanto si fa con gli algoritmi di relevance ranking.

Privacy: l'integrazione globale dei dati mentre offre notevoli benefici espone anche a molteplici opportunità di violare la privacy che devono essere controllate da una combinazione di strumenti tecnici e legali (Bizer – Heath - Berners-Lee 2009 p. 20).

4.6 Web semantico, Linked Data e biblioteche

Le biblioteche hanno una lunga tradizione nella creazione e nello scambio di metadati. Possiamo dire che sono state tra le prime istituzioni sociali a capire la funzione e il valore dei metadati: i metadati bibliografici agiscono come inventario del posseduto, aiutano nel reperimento dei documenti e sono usati come surrogati dei documenti quando non sia possibile accedere direttamente agli scaffali per esaminarli (Coyle 2010 a p. 9). Sono usati

¹²¹ Vedi per esempio Falcons < <http://iws.seu.edu.cn/services/falcons/conceptsearch/index.jsp>>; vedi per esempio Sindice < <http://sindice.com/>>.

non solo per soddisfare le necessità degli utenti ma anche nelle funzioni gestionali e amministrative delle biblioteche. Anche rispetto al tema dell'interoperabilità, cioè della capacità di scambiare e riutilizzare i dati (Ridi 2010 p. 22), le biblioteche hanno da lungo tempo colto la forza di codificazioni e formati di scambio condivisi. L'adozione di regole comuni di catalogazione è stata il fondamento della possibilità di far circolare e riusare i dati a diversi livelli fino a sostenere il controllo universale delle pubblicazioni (UBC)¹²²: i principi base della catalogazione sono condivisi internazionalmente anche se possono trovare applicazioni diverse nelle regole nazionali; sono internazionali gli standard descrittivi ISBD (International Standard Bibliographic Description)¹²³ e possono essere considerate internazionali anche le Anglo American Cataloguing rules per l'ampia adozione che hanno avuto in tutti i paesi anglofoni e le nuove regole RDA (Resource Description & Access)¹²⁴ che oggi le sostituiscono. Un formato internazionale di metadati catalografici è stato realizzato con l'Unimarc (Universal Marc Format) oggi nella più recente versione di MARC21¹²⁵. La tradizione di scambio e circolazione dei cataloghi risale all'epoca dei cataloghi in volume che, riprodotti in più copie, venivano dati o venduti ad altre biblioteche e che possiamo forse considerare la prima forma di accesso remoto al catalogo (Coyle 2010 a p. 11), ma più recentemente, e con effetti sostanzialmente diversi, i dati sono stati resi disponibili e interoperabili nelle reti a partire dall'uso del protocollo Z39.50¹²⁶ per arrivare ai protocolli oggi in uso nel Web come SRU/SRW¹²⁷, OAI-PMH¹²⁸ o ai servizi e alle API messe a disposizione sul Web. Da questo punto di vista si potrebbe considerare l'adesione al Web semantico solo come l'utilizzo di una diversa tecnologia.

¹²² Il controllo universale delle pubblicazioni è uno dei progetti principali dell'IFLA (International Federation of Library Associations and Institutions). Avviato nel 1974 con lo scopo di catalogare la produzione intellettuale mondiale si basa sulla responsabilità delle biblioteche nazionali nel documentare la produzione dei singoli paesi e sull'adozione di standard per la condivisione e lo scambio delle registrazioni bibliografiche a livello internazionale.

¹²³ Vedi <<http://www.ifla.org/en/isbd-rg>>.

¹²⁴ Vedi <<http://www.rdatoolkit.org/>>.

¹²⁵ Il formato Marc fu studiato a partire dal 1963 all'interno della Library of Congress per facilitare il trattamento informatico dei dati catalografici e la loro esportazione in diverse biblioteche. Si è poi sviluppato in diverse varianti nazionali fino a consolidarsi in una versione internazionale nel 1977 con l'Unimarc (Metitieri – Ridi 2005 p. 16). Vedi <<http://www.loc.gov/marc/>>.

¹²⁶ Vedi <<http://www.loc.gov/z3950/agency/>>.

¹²⁷ Vedi <<http://www.loc.gov/standards/sru/>>.

¹²⁸ Vedi <<http://www.openarchives.org/OAI/openarchivesprotocol.html>>.

I Linked Data tuttavia sono qualcosa di diverso perché consentono di creare una rete di conoscenze basate su standard condivisi da ampie comunità e abbastanza flessibili da poter essere usate da una molteplicità di utenti (Bermès 2009 p. 1). Fino ad oggi i dati bibliografici sono stati creati e scambiati solo all'interno della comunità bibliotecaria e questo nel mondo di oggi è diventato un limite. Molti dei protocolli e degli standard che le biblioteche usano sono stati pensati e realizzati per scambiare dati tra biblioteche e non coinvolgono, o solo molto parzialmente, altri attori presenti nel ciclo della produzione e distribuzione editoriale così come non sono adatti a comunicare con altre istituzioni culturali quali i musei o gli archivi, per restare solo negli ambiti più affini. Inoltre, nonostante il costante aggiornamento che i metadati bibliografici hanno avuto, oggi l'ambiente informativo è così cambiato da richiedere un cambiamento più sostanziale: non si tratta più di creare dati da inserire nei cataloghi, silos chiusi accessibili solo attraverso canali prestabiliti, ma di creare dati che possano essere utilizzati fuori dai cataloghi, nel Web, e non solo di creare dati in formato testuale destinati agli esseri umani ma di creare dati di cui possano fruire i computer. Secondo Coyle attualmente il web non è il semplicemente il luogo in cui i dati vengono scambiati ma anche lo spazio in cui si creano nuove informazioni e le biblioteche devono essere parte di questo ambiente

If we are to serve our users, then we need to deliver library services to users via the Web. But delivery over the network is not enough; our services must not only be on the Web, but need to be of the Web. The services can not just pass through, but must live and interact on the Web. With Web-based data, we can use the vast information resources there to enhance our data by creating relationships between library data and information resources. This will not only increase opportunities for users to discover the library and its resources, but will also increase the value of the data by allowing its use in a wide variety of contexts (Coyle 2010 b p. 6)¹²⁹.

Il cambiamento da affrontare è dunque rispetto alle esigenze di ricerca degli utenti che hanno nel web il loro punto di riferimento e rispetto alle necessità di scambio fra comunità diverse e in un ambiente tecnologico sostanzialmente differente ma più in generale riguar-

¹²⁹ Se dobbiamo servire i nostri utenti allora dobbiamo fornire i servizi bibliotecari agli utenti via Web. Ma fornire i servizi in rete non è abbastanza: i nostri servizi devono non solo essere nel Web ma del Web. I servizi non devono solo attraversare il Web ma devono vivere e interagire nel Web. Con dati web-based possiamo usare le vaste risorse che ci sono per migliorare i nostri dati creando relazioni tra i dati bibliotecari e le risorse informative. Questo non solo aumenterà le opportunità degli utenti di scoprire le biblioteche e le loro risorse, ma, permettendo il loro uso in un'ampia varietà di contesti, farà crescere il valore dei dati.

da la necessità di contribuire alla creazione di contenuti significativi di cui altre comunità possano fare uso. I Linked data possono essere un modo di realizzare questa prospettiva:

Linked Data gives libraries an exciting opportunity to make new connections between our collections and the world. It can extend discovery platforms to explore library data, broaden benchmark services, and support the role of libraries as creators and publishers. From bibliographic records to digital library collections, library data is also valuable and interesting to communities outside of our own. Opening up and sharing it will lead to greater innovation as others mashup and extend it in ways we never thought possible (Bradley 2009)¹³⁰.

Linked Data is rather about creating opportunities: opportunities for the data to be discovered and re-used in a variety of ways and by a diversity of users, opportunities for the libraries to build exciting interfaces to contextualize their collections and assets, opportunities for our community to mash-up with other domains and with the Web at large (Bermès 2009 p. 8)¹³¹.

In concreto il ruolo che le biblioteche possono assumere in questo momento nel Web semantico è stato individuato nella realizzazione di queste attività (Miller 2004 slide 26):

- Esporre le collezioni, usare le tecnologie del web semantico per rendere i contenuti disponibili
- Adattare al web tesauri, mappature, servizi
- Condividere le lezioni apprese
- Persistenza.

Persistenza. La persistenza è ciò di cui hanno bisogno le biblioteche e i creatori del Web semantico per rendere i legami semantici una realtà (Yang -Yi Lee - Xu 2009 p. 6). Il tema della consistenza e persistenza nella creazione degli identificatori e degli URI è uno dei principali dei Linked Data. Esistono linee guida date per la creazione di “cool URIs” (Berners-Lee 1998, Sauermann - Cyganiak 2008) ma le biblioteche devono studiare il modo di rendere persistenti i propri URI nel Web perché siano di qualità e possano essere riusati (Bermès 2009 p. 8-9). Per capire quali caratteristiche deve avere un identificatore appro-

¹³⁰ I Linked Data danno alle biblioteche un'eccitante occasione di fare nuove connessioni tra le nostre collezioni e il mondo. Possono estendere la piattaforma di ricerca che esplora i dati delle biblioteche, ampliare i servizi di riferimento e sostenere il ruolo di autore ed editore delle biblioteche. Dai record bibliografici alle collezioni digitali, i dati delle biblioteche sono di valore e di interesse per comunità fuori dalla nostra. Aprirli e condividerli porterà maggiore innovazione poiché altri li mescoleranno ed estenderanno in modi che non abbiamo mai ritenuto possibili.

priato dobbiamo considerare diversi aspetti: cosa identifica, chi crea l'identificatore, quale impegno viene preso per il suo mantenimento, in quale contesto questo identificatore è univoco, quanto è affidabile è l'identificatore. Nel contesto del Web semantico gli identificatori devono essere costruiti in modo da potersi muovere in tutto il web senza perdere la propria identità e di questa responsabilità le agenzie che li creano devono farsi carico (Coyle 2010 a p. 23-24).

Condividere le lezioni apprese. Il concetto di condivisione fa riferimento alla lunga tradizione che la comunità bibliotecaria ha nello sviluppare termini e concetti per descrivere contenuti che le ha consentito di acquisire una competenza che oggi può trasferire ad altri (Yang -Yi Lee - Xu 2009 p. 6). Vocabolari controllati, tassonomie, sistemi di classificazione, recupero dell'informazione, gestione di metadati, gestione della conoscenza sono parte integrante della professione. Molte delle problematiche che vengono oggi affrontate nelle ricerche sul Web semantico sono state studiate per anni dai biblioteconi ben prima della nascita del Web stesso. Per esempio l'approccio alle classi e alle proprietà di RDF/RDFS, che consente un modo flessibile di etichettare e recuperare fonti informative, ricorda la classificazione a faccette proposta da Ranganathan negli anni 1930, benché l'attuale protocollo abbia il vantaggio di fondarsi sulla logica formale (Robu – Robu – Thirion 2006 p. 204). Più in generale si può dire che la comunità del Web semantico e quella delle biblioteche condividono lo stesso scopo: dare un nome ai concetti e alle entità e tenere insieme le diverse forme che questi nomi possono assumere. Gli strumenti sviluppati in questa direzione dalle biblioteche sono sofisticati e avanzati: se saranno tradotti nelle tecnologie del Web semantico potranno contribuire a realizzare la visione di Berners- Lee (Harper – Tillet 2007 p. 1).

Adattare al web tesauri, mappature, servizi. Si tratta di trasformare i tesauri, i vocabolari controllati e i relativi servizi in formati che possano interagire con altri servizi e applicazioni Web dopo aver creato gli strumenti che consentono di farlo. Trasformarli secondo gli standard del web semantico, come il linguaggio OWL (Web Ontology Language) offrirà

¹³¹ Linked Data riguarda piuttosto il creare opportunità: opportunità per i dati di essere scoperti e riutilizzati in diversi modi e da una varietà di utenti, opportunità per le biblioteche di costruire interfacce stimolanti per contestualizzare le proprie collezioni e le proprie risorse, opportunità per la nostra comunità di mescolarsi con altri domini e con il Web nel suo complesso.

innumerevoli possibilità di usarli in modi diversi: consentirà per esempio di integrare le funzionalità di ricerca (searching e browsing delle risorse, ricerche sugli autori o su argomenti collegati ad un particolare concetto) in molti tipi di strumenti diversi dai cataloghi delle biblioteche. Gli strumenti creati per codificare le ontologie come SKOS (Simple Knowledge Organization System), alla cui realizzazione ha contribuito anche la Library of Congress, consentono di trasformare la struttura e i contenuti dei tradizionali sistemi di organizzazione della conoscenza in senso semantico. Attraverso SKOS i tesauri e i sistemi di classificazione possono essere re-ingegnerizzati come ontologie formali, trasformando la loro struttura di concetti in un insieme più sfaccettato ed elaborato di classi e proprietà identificate da URI (Biagetti 2009 p. 577). Il compito del W3C è creare gli strumenti per realizzare queste trasformazioni: esula invece dal suo scopo occuparsi del contenuto dei tesauri, stabilire se siano validi e affidabili. L'utilità di questi tesauri tuttavia dipende certamente dalla loro validità e affidabilità e in questo senso il ruolo delle biblioteche viene riconosciuto come estremamente rilevante: le biblioteche hanno una lunga storia nel fornire servizi che fanno uso di vocabolari controllati e se questi fossero trasformati secondo i nuovi standard sarebbe di grande valore per tutto il Web semantico. Da qui l'interesse a coinvolgere la comunità bibliotecaria in questo processo. Le biblioteche inoltre hanno da offrire la posizione di fiducia di cui godono presso il pubblico e che sarebbe estremamente utile nel costruire uno degli ultimi livelli del Semantic Stack, quello appunto della fiducia (Harper – Tillet 2007 p. 2). Le biblioteche potrebbero aumentare la fiducia nello spazio informativo rappresentato dai Linked Data fornendo propri dataset di qualità e aggiungendo metadati in altri dataset dimostrando così di ritenerli affidabili. Anche riutilizzare dataset messi a disposizione da altri nel proprio ambiente, negli opac, nelle biblioteche digitali, nei siti delle biblioteche, darebbe alle comunità di utenti, ai ricercatori, agli studenti, al pubblico in generale che fanno riferimento alle biblioteche, una dimostrazione del valore aggiunto e del possibile impiego dei Linked Data (Bermès 2009 p. 10).

Il lavoro a questo livello è portato avanti da grandi biblioteche e organizzazioni bibliotecarie come la Library of Congress (LC) e OCLC o in gruppi interessati a specifiche discipline, come la National Library of Medicine, ed è orientato alla conversione del formato MARC e delle regole di catalogazione in formati XML e RDF. Viene condotto in collabo-

razione con il W3C e con DCMI (Public Core Metadata Initiative)¹³². Finora si è concretizzato nella traduzione dei vocabolari controllati¹³³ della LC e di altre istituzioni in ontologie OWL e SKOS con l'assegnazione di URI secondo i principi dei Linked Data e con la loro registrazione in registri di metadati come il Dublin Core Metadata Registry¹³⁴. Appartiene a quest'area di lavoro anche la conversione in triple RDF, con l'assegnazione di specifici URI per ogni singolo elemento, dei modelli di dati definiti in FRBR¹³⁵, RDA¹³⁶, Dublin Core¹³⁷ e la loro iscrizione nei registri di metadati (Yang -Yi Lee - Xu 2009 p. 6-7).

Esporre le collezioni, usare le tecnologie del web semantico per rendere i contenuti disponibili¹³⁸. Esporre le collezioni significa consentire la ricerca semantica delle collezioni librarie e dei contenuti scientifici su internet (Yang -Yi Lee - Xu 2009 p. 6). Come esempi del lavoro in questo livello possono essere citate tutte le iniziative che consentono la ricerca, l'esportazione e il riuso di dati bibliografici o di dati di autorità come il progetto WorldCat Identities e FAST di OCLC, di cui abbiamo già parlato, oppure il progetto europeo MACS (Multilingual Access to Subject)¹³⁹ che coinvolge le biblioteche nazionali di Svizzera, Francia, Germania e la British Library. Ad un livello successivo di intervento si collocheranno le iniziative volte a creare i collegamenti di questi dati di autorità con altri dati, per esempio alla DBpedia (Yang -Yi Lee - Xu 2009 p. 8). Molto significativa in questo ambito è la pubblicazione dell'intero catalogo collettivo delle biblioteche svedesi come Linked Data. Mentre le biblioteche americane sono ancora caute e ritengono sia necessario un maggiore approfondimento prima di passare a realizzazioni concrete in questo senso, altre biblioteche come la British Library, la biblioteca nazionale canadese e

¹³² Vedi < <http://dublincore.org/>>.

¹³³ Tra questi Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), Library of Congress Subject Headings (LCSH), Medical Subject Headings (MeSH), National Library of Medicine Classification (NLM), The Getty Thesaurus of Geographic Names (TGN), Universal Decimal Classification (UDC) (Harper – Tillet 2007 p. 9).

¹³⁴ Vedi < <http://dcmi.kc.tsukuba.ac.jp/dcregistry/>>.

¹³⁵ Vedi <<http://vocab.org/frbr/core.html>> e < <http://vocab.org/frbr/extended.html>> .

¹³⁶ Vedi < <http://dublincore.org/dcmirdataskgroup/>>, <<http://www.rdvocab.org/>>.

¹³⁷ Vedi < <http://dublincore.org/dcmirdataskgroup/>>.

¹³⁸ In un articolo che esamina le implicazioni pratiche del ruolo delle biblioteche nel Web semantico così come esposto da Miller, Joint (2008 p. 183) evidenzia che i bibliotecari potrebbero a ragione ironizzare sul fatto che non sia una priorità esporre maggiormente le collezioni perché queste sono ancora costituite da documenti cartacei e quindi sono geograficamente e fisicamente circoscritte: in sostanza non c'è richiesta di opac di tipo Web semantico. Altrettanto si potrebbe dire riguardo alle collezioni digitali full text per le quali i bibliotecari spendono più tempo nell'usare sistemi di autenticazione che ne restringono l'esposizione di quanto ne passano a sognare che siano libere e interoperabili .

¹³⁹ Vedi <<https://macs.hoppie.nl/pub/topic/about-macs>> .

altre biblioteche come la British Library, la biblioteca nazionale canadese e quella austriaca stanno lavorando in questa direzione (Yang -Yi Lee - Xu 2009 p. 15). La biblioteca nazionale svedese è stata la prima a concludere il lavoro e a pubblicare il suo nuovo catalogo LIBRIS¹⁴⁰.

4.7 LIBRIS il catalogo delle biblioteche Svedesi

Il catalogo LIBRIS è creato collettivamente da circa 200 biblioteche, principalmente universitarie ma anche di archivi, musei e da alcune biblioteche pubbliche, ed è ospitato e gestito dalla Royal Library¹⁴¹. Contiene circa 6,5 milioni di record bibliografici e di autorità. L'idea di rendere il catalogo nella forma di Linked Data è nata dalla voglia di confrontarsi in modo pratico con le idee del Web semantico e verificare la loro applicabilità ed è stata sostenuta dalla convinzione che i Linked Data possano realmente connettere le biblioteche al Web. I responsabili del progetto ritengono che aprire i dati a comunità esterne al settore bibliotecario crei interesse nei confronti delle biblioteche e che condividere strumenti e tecniche sia un modo per comunicare più facilmente con professionisti in altri settori. Nel procedere ad una revisione dell'opac in uso, tra il 2007 e il 2008, era anche emerso un effettivo interesse per i dati bibliografici da parte di persone esterne che non avevano esperienza con i protocolli di scambio in uso tra le biblioteche come Z39.50 e MARC. L'idea di considerare come risorsa non solo i documenti contenuti nel catalogo ma il catalogo stesso, e quindi la volontà di renderlo accessibile, così come l'idea che fosse necessario renderlo leggibile non solo dagli esseri umani ma anche dalle macchine e dai web robots sono diventate rapidamente naturali. Rendere il catalogo accessibile avrebbe potuto essere fatto anche con metodi più tradizionali e in effetti questi metodi sono stati adottati in parallelo: è stato predisposto un web server basato su HTTP che può esportare dati in formato MARC-XML, Dublin Core, Json, RIS e MODS. Nel caso di altre necessità si sarebbe tuttavia dovuto passare il tempo a costruire continuamente nuove API e per questo si è scelta la strada più innovativa. Inoltre, cosa di grande importanza, l'approccio tradizionale consente di utilizzare solo i dati presenti nei record MARC originali, perdendo

¹⁴⁰ Vedi <<http://libris.kb.se/>>.

¹⁴¹ Vedi <<http://www.kb.se/english/>>.

tutte le altre informazioni in forma di link che normalmente vengono presentate agli utenti che consultano il catalogo.

Il nuovo opac è stato realizzato anche con un'attenzione particolare all'utente e con un metodo di sviluppo aperto e collaborativo che mira all'indipendenza da fornitori esterni. Piuttosto che attendere di realizzare la struttura perfetta per i record nel web, che non si sa ancora come potrebbe essere, gli sviluppatori del progetto hanno preferito provare e “sporcarsi le mani facendo” (Söderbäck – Malmsten 2008 p. 20), produrre qualcosa che in seguito si potrà migliorare ma su cui intanto si può ragionare.

Il primo passo del lavoro è stato l'esportazione dei dati dalle strutture proprietarie degli ILS in formato MARCXML. Il secondo passo è stato l'assegnazione degli identificatori alle risorse. I dati sono stati trasformati in triple RDF, usando le ontologie SKOS, Dublin Core, Bibliontology¹⁴² e BDPedia, con una specifica applicazione. La scelta di usare ontologie fuori dallo stretto dominio delle biblioteche è stata dettata dal desiderio di entrare in contatto con comunità diverse. Per i nomi delle persone e degli enti nei record di autorità sono state usate rispettivamente le classi Persona e Organizzazioni del FOAF, per i record che descrivono intestazioni di soggetto e tesauri è stata usata la classe Concept di SKOS mappando Marc21 su SKOS, per i record bibliografici sono state usate classi individuate in Bibliontology e in Dublin Core. Per esprimere le relazioni tra record di autorità e bibliografici sono stati usati elementi Dublin core (dc:subject e dc:creator) mentre le relazioni tra dati di autorità per soggetto sono state definite usando SKOS (skos:broader e skos:narrower). Per consentire una migliore esplorazione dei dati è stato predisposto un server SPARQL. Per migliorare la qualità dei dati sono stati creati un gran numero di link, la maggior parte interni al dataset Libris, circa 12 mila a LCSH.INFO (dato che le intestazioni di soggetto in Libris sono mappate su LCSH si è ritenuto utili mostrarle nella forma di link a quel vocabolario), alcune decine tra gli autori e le corrispondenti voci in DBPedia (allo scopo di dimostrare il desiderio di andare oltre la comunità delle biblioteche).

La creazione di questi legami è considerata essenziale in quanto è proprio da questi link che ci si aspettano i maggiori benefici per la prospettiva futura di dedurre inferenze significative man mano che la nuvola dei Linked Data crescerà. Anche pubblicare i dati nella forma di RDF senza legami sarebbe utile per la singola biblioteca, non fosse altro che per la possibilità di utilizzare SPARQL che si è dimostrato un linguaggio di interrogazione

la possibilità di utilizzare SPARQL che si è dimostrato un linguaggio di interrogazione molto efficace, ma la creazione di legami consente di fare uso delle attività intellettuali di altri e di garantire un sicuro miglioramento del proprio catalogo contribuendo contemporaneamente al miglioramento del Web in generale. I dati di autorità, in particolare, sono risorse importanti che possono essere utilizzate molto al di là della semplice ricerca negli opac delle singole biblioteche: legarli alla nuvola dei Linked data può aprire interessanti prospettive di visibilità e suscitare interesse verso le attività delle biblioteche (Söderbäck – Malmsten 2008, Malmsten 2008, 2009).

¹⁴² Vedi < <http://bibliontology.com/> >

CONCLUSIONE

I cataloghi delle biblioteche si trovano di fronte ad un ulteriore passaggio della loro evoluzione. Il nuovo contesto informativo, rappresentato da una molteplicità di fonti online e da una generazione di utenti nata con internet o comunque avvezza alle nuove tecnologie, ha fatto loro perdere la posizione di preminenza che occupavano quale punto di fornitura e accesso ai documenti e alle informazioni e li ha costretti a superare una tradizione di isolamento, autosufficienza, autoreferenzialità per confrontarsi con un mondo in cui informazioni, dati, applicazioni, servizi e utenti hanno come punto di riferimento il Web, la sua filosofia, il suoi strumenti.

L'esigenza delle biblioteche e dei loro cataloghi di restare visibili in questo nuovo ambiente e di contribuire in modo significativo al suo sviluppo mettendo a disposizione l'enorme massa di documenti e di informazioni sui documenti che hanno fin qui raccolto ed elaborato, fa sì che esse stiano cercando strategie per rompere le barriere che fino ad oggi hanno tenuto i dati bibliografici separati e poco accessibili. Una possibilità in questo senso sembra essere offerta dalla visione e dalle tecnologie del Web semantico. Dopo aver adottato all'interno degli opac tutti i dispositivi che semplificano la ricerca, migliorano i risultati, arricchiscono i contenuti le biblioteche si apprestano dunque a portare i propri dati fuori dal catalogo adottando modelli e tecnologie condivisi da ampie comunità di utenti che cambiano significativamente l'approccio ai dati bibliografici e l'idea stessa di catalogo.

Non è detto che questa sia l'unica strada possibile ma al momento sembra essere un buon modo per espandere i servizi di ricerca ed esplorazione dei dati che possono essere offerti agli utenti (Coyle 2010 b p. 5). L'inclusione dei dati delle biblioteche nella nuvola dei Linked Data non significa tuttavia la fine dei cataloghi delle biblioteche o la loro sostituzione. Le funzioni legate alla gestione continueranno ad essere parte dei database istituzionali delle biblioteche: acquisizione, controllo, circolazione dei materiali sono funzioni che continueranno ad essere svolte attraverso questi strumenti. Il Web semantico può invece essere considerato come un'estensione dei servizi informativi forniti dalle biblioteche che espanderà il numero delle persone che potranno fruire di quei servizi: i dati delle biblioteche, trasformati in Linked Data e diventati parte di un più vasto universo informativo potranno raggiungere gli utenti nel web ed essere usati in combinazione con i molti altri dati che già vi si trovano (Coyle 2010 a p. 29).

BIBLIOGRAFIA

ABRAM, Stephen – LUTHER, Judy [2004] *Born with the Chip*, “*Library Journal*”, 129(2004), 8, p. 34-37.

<<http://www.libraryjournal.com/article/CA411572.html>>

ANTELMAN, Kristin [2007] *Empowering the library search experience: Endeca at NCSU Libraries*, Maryland Library Association, May 3, 2007. Presentazione PPT.

<www.lib.ncsu.edu/staff/kaantelm/MLA050407.ppt>

ANTELMAN, Kristin – LYNEMA, Emily – PACE, Andrew K. [2006] *Toward a Twenty-First Century Library Catalog*, “*Information Technology And Libraries*”, 2006, September, p. 128-139.

<<http://eprints.rclis.org/7332/>>

BADE, David [2007] *Relevance ranking is not relevance ranking or, when the user is not the user, the search results are not search results*, “*Online Information Review*”, 31(2007), 6, p. 831-844.

BATES, Marcia J. [1989] *The Design of Browsing and Berrypicking Techniques for the Online Search Interface*, “*Online Review*”, 13(October 1989), p. 407-424.

<<http://www.gseis.ucla.edu/faculty/bates/berrypicking.html>>

BEALL, Jeffrey [2008] *The Weaknesses of Full-Text Searching*, “*The Journal of Academic Librarianship*”, 34(2008), 5(September), p. 438-444.

BECKER, Christian - BIZER, Christian [2008] *DBpedia mobile: A LocationEnabled Linked Data Browser*, Proceedings of the Semantic Web Challenge at ISWC, 2008.

<<http://events.linkeddata.org/ldow2008/papers/13-becker-bizer-dbpedia-mobile.pdf>>

BEIT-ARIE, Oren [2009] *Toward Service Oriented Librarianship*, NISO Forum: Library Resource Management Systems, October 8, 2009, Boston, MA, presentazione ppt.

<<http://www.niso.org/news/events/2009/lrms09/agenda/>>

BERGAMIN, Giovanni [2008] *Opac: migliorare l'esperienza degli utenti*, “*Bibliotime*”, 11(2008), 1(marzo).

<<http://www2.spbo.unibo.it/bibliotime/num-xi-1/bergamin.htm>>

BERMÈS, Emmanuelle [2009] *Linking Open Data: a case for releasing library data on the Semantic Web*, Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009.

<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/EmmanuelleBermes.pdf>>

BERNERS-LEE, Tim [1998] *Cool URIs don't change*, <<http://www.w3.org/Provider/Style/URI>>.

BERNERS-LEE, Tim – HENDLER, James – LASSILA, Ora [2001] *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*, “*Scientific American Magazine*”, May 2001.

BERNERS-LEE, Tim [2006] *Linked data*, Ultimo aggiornamento giugno 2009.

<<http://www.w3.org/DesignIssues/LinkedData.html>>

BIAGETTI, Maria Teresa [2009] *Biblioteche digitali e Web semantico*, “*Bollettino AIB*”, 49(2009), 4(dicembre), p. 577-582.

BIANCU, Bonaria [2009] *Behind the Scenes: Some Technical Details on Mashups*, in Engard, Nicole: *Library mashup. Exploring new way to deliver library data*, edited by Nicole C. Engard, London, Facet, 2009.

BIZER, Christian – HEATH, Tom - BERNERS-LEE, Tim [2009] *Linked Data. The Story So Far*, “*International Journal on Semantic Web and Information Systems*”, 5(2009),3 – Special issue on Linked Data.
preprint a: <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>

BLYBERG, John [2008] *Getting Started with SOPAC*, 09/24/2008.
<<http://thesocialopac.net/getting-started>>

BRADLEY, Fiona [2009] *Discovering Linked Data*, “*Library Journal*”, 4/15/2009.
<<http://www.libraryjournal.com/article/CA6650835.html>>

BREEDING, Marshall [2006] *Trends in Library Automation: Meeting the challenges of a new generation of library users*, OCLC Office of Research Distinguished Seminar Series, November 29, 2006, presentazione ppt.
<<http://www.oclc.org/research/dss/ppt/breeding.ppt>>

BREEDING, Marshall [2007] *Next-Generation Library Catalogs* “*Library Technology Reports*”, 43 (2007), 4 (Jul/Aug).

BREEDING, Marshall [2009] *Investing in The Future: Automation Marketplace 2009*, “*Library Journal*”, 4/1/2009.
<<http://www.libraryjournal.com/article/CA6645868.html>>

BREEDING, Marshall [2010] *Automation Marketplace 2010: New Models, Core Systems*, “*Library Journal*”, 4/1/2010.
<<http://www.libraryjournal.com/article/CA6723662.html>>

BUCCHIONI, Cinzia [2004] *L'evoluzione dello standard OpenURL*, “*Bibliotime*”, 7(2004), 3(novembre).
<<http://didattica.spbo.unibo.it/bibliotime/num-vii-3/bucchion.htm>>

BUCCHIONI, Cinzia - SPINELLI, Serafina [2007] *Teoria e pratica del linking citazionale*, “*Biblioteche oggi*”, 25 (2007), 8, pag. 70-86.
<<http://www.bibliotecheoggi.it/2007/20070807001.pdf>>

BUIZZA, Pino [2008] *Gli opac: funzionalità e limiti nel mondo del web*, “*Bibliotime*”, 11(2008), 1(marzo).
<<http://www2.spbo.unibo.it/bibliotime/num-xi-1/buizza.htm>>

CALHOUN, Karen [2006] *The Changing Nature of the Catalog and its Integration with Other Discovery Tools. Prepared for the Library of Congress. Final report, March 17 2006*.
<<http://www.loc.gov/catdir/calhoun-report-final.pdf>>

CAMPBELL, D. Grant – FAST, Karl V. [2004] *Academic Libraries and the Semantic Web: What the Future May Hold for Research-Supporting Library Catalogues*, “*The Journal of Academic Librarianship*”, 30(2004), 5, p. 382–390.

CAMPBELL, D. Grant – FAST, Karl V. [2006] *From Pace Layering to Resilience Theory: The Complex Implications of Tagging for Information Architecture*, Conference paper presentato ad IA summit 2006.

CASE, Donald O. [2008] *Looking for information. A survey on Research on Information Seeking, Needs, and Behavior*, 2. ed. United Kingdom, Emerald, 2008.

CHRISTENSEN, Anne [2009] *Next generation catalogs: What do users think? Conclusions from the beluga project in Hamburg*, Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009
<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/AnnChristensen.pdf>>

COLLABORATIVE and Social Tagging Networks [2008] di Tonkin e altri, “*Ariadne*”, 54(January 2008).

COOPER, Michael D. [2001] *Usage Patterns of a Web-based Library Catalog*, “*Journal of the American Society for Information Science & Technology*”, 52(2001), 2, p. 137-148.

COYLE, Karen [2007 a] *The Libray Catalog in a 2.0 World*, “*The Journal of Academic Librarianship*”, 33(2007), 2, p. 289-291.

COYLE, Karen [2007 b] *The Libray Catalog: Some Possible Future*, “*The Journal of Academic Librarianship*”, 33(2007), 3, p.414-416.

COYLE, Karen [2008 a] *Machine indexing*, “*The Journal of Academic Librarianship*”, 34(2008), 6, p. 530-531.

COYLE, Karen [2008 b] *Meaning, Thecnology, and the Semantic Web*, “*The Journal of Academic Librarianship*”, 34(2008), 3, p. 263-264.

COYLE, Karen [2009] *Making connections*, “*NetConnect*”, 4/15/2009.
<<http://www.libraryjournal.com/article/CA6650836.html>>

COYLE, Karen [2010 a] *Understanding the Semantic Web: Bibliographic Data and Metadata*, “*Library Technology Reports*”, 46(2010), 1(January), p. 5-31.

COYLE, Karen [2010 b] *RDA Vocabularies for a Twenty-First-Century Data Environment*, “*Library Technology Reports*”, 46(2010), 2(February-March), p. 5-39.

DATEMA, Jay [2007] *Tag, You’re It*, “*Libray Journa*”l, 132(2007), 1, p. 41.

DE GENNARO, Richard [1982] *Library automation & networking perspectives on three decades*, “*Library journal*”, 108 (1983), 7(April), p. 629-635.

DEAN, Rebecca J. [2003] *FAST : La messa a punto di voci semplificate per i metadati*.
<http://www.sba.unifi.it/ac/relazioni/dean_ita.pdf>
<<http://www.oclc.org/research/activities/fast/default.htm>>

DELLIT, Alison - BOSTON, Tony [2007] *Relevance ranking of results from MARC-based catalogues: from guidelines to implementation exploiting structured metadata*, National Library of Australia Staff Papers.
<<http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1052/1321>>

DEMPSEY, Lorcan [2006] *The Library Catalogue in the New Discovery Environment: Some Thoughts*, "Ariadne", 48(July 2006).

<<http://www.ariadne.ac.uk/issue48/dempsey/>>

DENHOLM, Carmel – KAULER, Leto – LAVELLE, Jan – SOKVITNE, Lloyd [2009] *Making the new OPAC seamless: dealing with the transition from "finding" to "getting"*, "Library Hi Tech", 27(2009), 1, p. 13-29.

DERVIN, Brenda [1976] *Strategies for dealing with human information needs: Information or communication?*, "Journal of Broadcasting", 20(1976), 3, p. 324-351.

DERVIN, Brenda - NILAN, Michael. S. [1986] *Information needs and uses: A conceptual and methodological review*, "Annual Review of Information Science and Technology", 21(1986), p. 3–33.

DIZIONARIO [2006] *Dizionario di informatica*, Segrate, Mondadori informatica, 2006.

ECKERT, Kai – HÄNGER, Christian – NIEMANN, Christof [2009] *Tagging and Automation: Challenges and Opportunities for Academic Libraries*, "Library Hi Tech", 27(2009), 4, p. 557-569.

ENGARD, Nicole [2009] *Library mashup. Exploring new way to deliver library data*, edited by Nicole C. Engard, London, Facet, 2009.

FAST, Karl V. - CAMPBELL, D. Grant [2004], "I still like Google": *university student perceptions of searching OPACs and the web*, in Schamber, L. and Barry, C.L., *Proceedings of the 67th ASIS&T Annual Meeting*, 41(2004), 1, p. 138-146.

FICHTER, Darlene [2009] *What is a Mashup?* In Engard, Nicole: *Library mashup. Exploring new way to deliver library data*, edited by Nicole C. Engard, London, Facet, 2009.

GAMBARI, Stefano – GUERRINI, Mauro [2002] *Definire e catalogare le risorse elettroniche*, Milano, Editrice Bibliografica, 2002.

GHILLI, Carlo [2001] *Introduzione a FRBR*, Milano, Bibliografica, 2001.

GNOLI, Claudio [2000] *Il tavolino di Ranganathan*, "Bibliotime", 11(2008), 1 (marzo).

<<http://www2.spbo.unibo.it/bibliotime/num-xi-1/gnoli.htm>>

GNOLI, Claudio – MARINO, Vittorio – ROSATI, Luca [2006] *Organizzare la conoscenza. Dalle biblioteche all'architettura delle informazioni per il Web*, Milano, Hops Tecniche nuove, 2006.

GOLDER, Scott A. - HUBERMAN, Bernardo A. [2006] *Usage patterns of collaborative tagging systems*, "Journal of Information Science", 32(2006), p. 198-208.

HADRO, Josh [2008] *Darien Library's Open Source SOPAC 2.0 Emphasizes Patron*, "Library Journal", 8/29/2008.

<<http://www.libraryjournal.com/article/CA6591377.html>>

HAMMOND, Tony - HANNAY, Timo - LUND, Ben - SCOTT, Joanna [2005] *Social Bookmarking Tools (I). A General Review*, "D-Lib Magazine", 11(2005), 4(April).

<<http://www.dlib.org/dlib/april05/hammond/04hammond.html>>

HANLON, G. – ROBERTSON, P. [2009] *Web2 in Scottish Libraries*. Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009.

HARPER, Corey A. - TILLET, Barbara B. [2007] *Library of Congress controlled vocabularies and their application to the Semantic Web*, "Cataloguing & Classification Quarterly", 43(2007), 3-4, p. 47-68. Citato da:
<https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/3269/ccq_sem_web.pdf;jsessionid=A8AAF6E4A4AB6D26F32EEC5D75A205B6?sequence=1>.

HARRIS, Christofer [2007] *Catalog Manifesto* in "Infomancy", 14 giugno 2007. Intervento su weblog. <<http://schoolof.info/infomancy/?p=388>>

HEARST, Marti [2006] *Design Recommendations for Hierarchical Faceted Search Interfaces*, ACM SIGIR Workshop on Faceted Search, August, 2006.
<<http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf>>

HEARST, Marti [2008] *Uls for Faceted Navigation. Recent Advances and Remaining Open*, Workshop on Computer Interaction and Information Retrieval, HCIR 2008, October 2008, Redmond, WA.
<<http://flamenco.berkeley.edu/papers/faceted-workshop06.pdf>>

HEATH, Tom [2009] *An Introduction to Linked Data*, Slides presentate il 13/14 febbraio 2009 ad Austin, Texas.
<<http://tomheath.com/slides/2009-02-austin-linkeddata-tutorial.pdf>>

HICKEY, Thom – GATENBY, Janifer [2009] *Opening library data for web scale use and re-mixing*, Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009
<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/HickeyThom.pdf>>

HILDRETH, Charles R. [1982] *On line public access Catalogs: The user Interface*, Dublin (Ohio), OCLC, 1982.

HILDRETH, Charles R. [1989] *The online catalogue. Developments and directions*, London, The library Association, 1989.

HILDRETH, Charles R. [1991] *Advancing toward the E³ OPAC. The Imperative and the Path* in "Think Tank on the present and future of the Online Catalog: proceedings". Occasional Paper n. 9. Chicago, American Library Association, Reference and Adult Services Division, 1991, p. 17-38.

HILDRETH, Charles R. [1994] *Opac history*, in "BULB", March 11, 1994.
<<http://bubl.ac.uk/archive/lis/opac/opachi21.htm>>

HILDRETH, Charles R. [1995] *Online Catalog Design Models: Are We Moving in the Right Direction? A Report Submitted to the Council on Library Resources*, August, 1995.
<<http://myweb.cwpost.liu.edu/childret/clr-opac.html>>

JISC – SCONUL [2008] *Library Management Systems Study. An Evaluation and horizon scan of the current library management systems and related systems landscape for UK higher education*, Sero Consulting Ltd with Glenaffric Ltd and Ken Chad Consulting Ltd, March 2008.
<<http://www.jisc.ac.uk/media/documents/programmes/resourcediscovery/lmsstudy.pdf>>

JOINT, Nicholas [2008] *The practitioner librarian and the semantic web*, "Library Review", 57(2008), 3, p. 178-186.

KAIZER, Jasper – HODGE, Antony [2005] *AquaBrowser Library: Search, Discover, Refine*, "Library Hi Tech News", 10(2005), p. 9-12.

KIFER, Michael – BOLEY, Harold [2009] *RIF Overview*, W3C working Draft 1 October 2009.
<<http://www.w3.org/TR/rif-overview/>>

KIPP, M.E.I. - CAMPBELL, D.G. [2006], *Patterns and inconsistencies in collaborative tagging systems: an examination of tagging practices*, in "Proceedings of the American Society for Information Science and Technology", Austin, Texas.
<<http://dlist.sir.arizona.edu/1704/01/KippCampbellASIST.pdf>>

KUO, Byron Y.L. - HENTRICH, Thomas - GOOD, Benjamin M. - WILKINSON, Mark D. [2007] *Tag clouds for summarizing web search results*, International World Wide Web Conference archive, Proceedings of the 16th international conference on World Wide Web, poster session, 2007, p. 1203 – 1204.
<<http://www.ra.ethz.ch/CDstore/www2007/www2007.org/posters/poster1046.pdf>>

LAWSON, Karen G. [2009] *Mining Social Tagging Data for Enhanced Subject Access for Readers and Researchers*, "The Journal of Academic Librarianship", 35(2009), 6, p. 574-582.

LEVINE, Jenny [2007] *Presence in the 2.0 World*, "Library Technology Reports", 43 (2007), 5(Sep/Oct), p. 7-8.

LEWANDOWSKI, Dirk [2009] *Ranking library materials*, "Library Hi Tech", 27(2009), 4, p. 584-593.

LI, Dezhu [1985] *On-Line public access catalog display formats* in *Encyclopedia of library and information science*, editors Allen Kent and Harold Lancour, [poi] executive editor Allen Kent, administrative editor Carolyn M. Hall, New York - Basel, Marcel Dekker, 1968-2003, Vol. 38 (1985), p. 325-332.

LIM, Lynette [2008] *LINC+: Our leap towards a Web 2.0 OPAC interface*, National University of Singapore.
<http://www.las.org.sg/pa_ly.pdf>

LITTLETREE, Sandy [2008] *Endeca Catalog Usability Test. Findings Report*, February 2008.
<<http://www.lib.ncsu.edu/endeca/publications.html>>

LOWN, Cory - HEMMINGER, Brad [2009] *Extracting User Interaction Information from the Transaction Logs of a Faceted Navigation OPAC*, "Code4Lib", 7(2009/06/26)
<http://journal.code4lib.org/articles/1633>

MAISONNEUVE, Marc – TOUITOU, Cécile [2007] *Une nouvelle famille d'Opac. Navigation à facettes et nuages de mots*, "BBF", 52(2007), 6, p. 12-19
<<http://bbf.enssib.fr/consulter/bbf-2007-06-0012-002>>

MALMSTEN, Martin [2008] *Making a Library Catalogue Part of the Semantic Web*, International Conference on Dublin Core and Metadata Applications, Berlin, 2008, Proceedings.
<<http://dcpapers.dublincore.org/ojs/pubs/article/view/927/923>>

MALMSTEN, Martin [2009] *Exposing Library Data as Linked Data*, Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009.

<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/MartinMalmsten.pdf>>

MANN, Thomas [2006] *The Changing Nature of the Catalog and its Integration with Other Discovery Tools. Final Report. March 17, 2006*. Prepared for the Library of Congress by Karen Calhoun. A critical Review. Prepared for AFSCME 2910 The Library of Congress Professional Guild. April 3, 2006. <<http://www.guild2910.org/AFSCMECalhounReviewREV.pdf>>

Traduzione italiana in “*Bollettino Aib*”, 46, (2006), 3, p. 186-206.

MANOLA, Frank – MILLER, Eric [2004] *RDF Primer*, W3C Recommendation 10 February 2004.

<<http://www.w3.org/TR/2004/REC-rdf-primer-20040210/>>

MARCHITELLI, Andrea – PIAZZINI, Tessa [2008] *OPAC, SOPAC e Social networking: cataloghi di biblioteca 2.0?*, “*Biblioteche Oggi*”, 26(2008), 2, p. 82-92.

MARKEY, Karen [2007] *The Online Library Catalog. Paradise lost and Paradise Regained?*, “*D-Lib Magazine*”, 13(2007), 1-2(January-February 2007).

<<http://dlib.org/dlib/january07/markey/01markey.html>>

MARSHALL, Peg – HERMAN, Shawn – RAJAN, Sri [2006] *In Search of More Meaningful Search*, “*Serials Review*”, 32(2006), p. 172–180.

MATHES, Adam [2004] *Folksonomies, cooperative classification and communication through shared metadata*, Computer Mediated Communication-LIS590CMC, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, Urbana, FL.

<www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>

MATTHEWS, Joseph R – LAWRENCE, Gary S. - FERGUSON, Douglas K [1983] *Using online catalogs : a nationwide survey : a report of a study sponsored by the Council on Library Resources*, New York, Neal-Schuman, 1983.

McGRATH, Kelley [2007] *Facet-Based Search and Navigation With LCSH: Problems and Opportunities*, “*Code4Lib*”, 1(2007/12/7).

<<http://journal.code4lib.org/articles/23>>

MENDES, Luiz H. – QUINONEZ-SKINNER, Jennie – SKAGGS, Daniel [2009] *Subjecting the catalog to tagging*, “*Library Hi Tech*”, 27(2009), 1, p. 30-41

METITIERI, Fabio [2009] *L'OPAC collaborativo tra socialità e folksonomia*, “*Biblioteche Oggi*”, 27(2009), 2, p. 7-14.

METITIERI, Fabio – RIDI, Riccardo [2005] *Biblioteche in rete. Istruzioni per l'uso*, 3.ed., Roma-Bari, Laterza, 2005.

oppure <<http://www.laterza.it/bibliotecheinrete/index.htm>> con un aggiornamento a giugno 2008.

MILLER, Eric [2004] *The Semantic Web and Digital Libraries*. Keynote presentation from The International Conference on Dublin Core and Metadata Applications, 2004. Shanghai, China, 11-14 October 2006. PowerPoint presentation.

<<http://dc2004.library.sh.cn/english/prog/ppt/talk.ppt>>

MILLER, Paul [2006] *Coming Together around Library 2.0. A Focus for Discussion and a Call to Arms*, “*D-Lib Magazine*”, 12 (April 2006), n. 4.
<<http://www.dlib.org/dlib/april06/miller/04miller.html>>

MILLER, Paul [2007] *What Happens When We Mash The Library*, “*Ariadne*”, 50(2007 January).
<<http://www.ariadne.ac.uk/issue50/miller/>>

NIELSEN, Jakob [2005] *Mental Models for Search Are Getting Firmer*, in “*Jakob Nielsen’s Alertbox*”, May 9 2005.
<<http://www.useit.com/alertbox/20050509.html>>

NISO [2004] *ANSI/NISO Z39.88. The OpenURL Framework for Context-Sensitive Services*, 2004 (ultima modifica gennaio 2010)
<http://www.niso.org/kst/reports/standards?step=2&project_key=d5320409c5160be4697dc046613f71b9a773cd9e>

NISO [2005] *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, ANSI/NISO Z39.19-2005*, by the American National Standards Institute, NISO Press, Bethesda, Maryland, U.S.A., 2005.
<http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a>

O'REILLY, Tim [2005 a] *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*, in “*O'Reilly*”, 9/30/2005.
<<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>
traduzione italiana <<http://www.xyz.reply.it/web20/>>

O'REILLY, Tim [2005 b] *Web 2.0: Compact Definition?*, “*O'Reilly Radar*”, blog post, October 1, 2005 <http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html>

OCLC [2004] *The 2003 OCLC Environmental Scan: Pattern Recognition* Dublin (Ohio), OCLC Online Computer Library Center, 2004.
<<http://www.oclc.org/reports/escan/>>

OCLC [2005] *Perceptions of libraries and information resources: a report to the OCLC membership*, Dublin (Ohio), OCLC Online Computer Library Center, 2005.
<http://www.oclc.org/reports/pdfs/Percept_all.pdf>

OCLC [2007] *Sharing, Privacy and Trust in Our Networked World. A Report to the OCLC*, Dublin, Ohio, OCLC Online Computer Library Center, 2007.
<<http://www.oclc.org/reports/pdfs/sharing.pdf>>

OCLC [2009] *Online catalogs: What users and librarians want. An OCLC report*, Dublin, Ohio, OCLC Online Computer Library Center, 2009.
<<http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf>>

ON [2008] *On The Record. Report of the Library of Congress Working Group on the Future of Bibliographic Control*, Washington, DC: Library of Congress, Jan. 9, 2008.
<<http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>>

PAPY, Fabrice - CHAUVIN, Sophie [2005] *Au-delà de la transfiguration du catalogue Le Visual... Catalog*, “*BBF*”, 50(2005), 4, p. 5-12.

<<http://bbf.enssib.fr/consulter/bbf-2005-04-0005-001>>

PARADISI, Federica [2003] *Classificazione Dewey fra tradizione e innovazione*, "Bibliotime", 6(2003), 1(marzo)
<<http://didattica.spbo.unibo.it/bibliotime/num-vi-1/paradisi.htm>>

PENNELL, Charley [2007] *A new kind of catalog*, North Carolina Library Association Annual Meeting, Hickory, NC, October, 2007. Presentazione PPT.
<<http://www.lib.ncsu.edu/endeca/presentations.html>>

PHILLIPSON, Mark [2006] *Dear Penntags*, post nel blog Clayfox, june 14th, 2006.
<<http://www.clayfox.com/2006/06/14/dear-penntags/>>

PRENSKY, Marc [2001 a] *Digital Natives Digital Immigrants*, "On the Horizon", 9(2001), 5(October).

PRENSKY, Marc [2001 b] *Digital Natives Digital Immigrants: Part II: Do They Really Think Differently?*, "On the Horizon", 9(2001),6(December).

QUINTARELLI, Emanuele [2005] *Folksonomies: power to the people*, paper presented at the ISKO Italy-UniMIB meeting, Milan, June 24, 2005.
<<http://www.iskoi.org/doc/folksonomies.htm>>

RIDI, Riccardo [2007] *La Biblioteca come ipertesto. Verso l'integrazione dei servizi e dei documenti*, Milano, Editrice Bibliografica, 2007.

RIDI, Riccardo [2008] *Strumenti e strategie per la ricerca di informazioni WWW*, Versione 3.10 (2008-03-03)
<<http://www.burioni.it/forum/ridi-mot.htm>>

RIDI, Riccardo [2010] *Il mondo dei documenti. Cosa sono, come valutarli e organizzarli*, Roma-Bari, Laterza, 2010.

ROBU, Ioana - ROBU, Valentin – THIRION, Benoit [2006] *An introduction to the Semantic Web for health sciences librarians*, "Journal of the Medical Library Association", 94(2006), 2(April), p. 198- 205.

ROSATI, Luca [2003] *La classificazione a faccette fra Knowledge Management e Information Architecture (parte I)*, IT Consult, 2003
<http://www.itconsult.it/knowledge/articoli/pdf/itc_rosati_faccette_e_KM.pdf>

SADLER, Elizabeth (Bess) [2009] *Project Blacklight: a next generation library catalog at a first generation university*, "Library Hi Tech", 27(2009), 1, p. 57-67.

SANTORO, Michele [2007] *Questa sera si cataloga a soggetto. Breve analisi delle folksonomies in prospettiva bibliotecaria*, "Bibliotime",10(2007),2(luglio).
<<http://didattica.spbo.unibo.it/bibliotime/num-x-2/santoro.htm>>

SARACEVIC, Tefko [2007 a] *Relevance: a review of the literature and a framework for thinking on the notion, Information Science. Part II: Nature and manifestations of relevance*, "Journal of the American Society for Information Science and Technology", 58(2007),13, p. 1915-1933.

SARACEVIC, Tefko [2007 b] *Relevance: a review of the literature and a framework for thinking*

on the notion, *Information Science. Part III: Behavior and Effects of Relevance*, "Journal of the American Society for Information Science and Technology", 58(2007), 13, p. 2126-2144.

SAUERMANN, Leo - CYGANIAK, Richard [2008] *Cool URIs for the Semantic Web*, (prima versione novembre 2006).

<<http://www.w3.org/TR/cooluris/>>

ŠAUPERL, Alenka – SAYE, Jerry D. [2009] *Have we made any progress? Catalogues of the future revisited*, "Journal of Documentation", 65(2009), 3, p. 500-514.

SERIALS SOLUTIONS [s.d.] *AquaBrowser library. Search, discover, refine. Breve descrizione del prodotto.*

<<http://www.serialssolutions.com/assets/publications/Brochure-AquaBrowser.pdf>>

SU, Shiao-Feng [1994] *Dialogue with an Opac: how visionary was Swanson in 1964?*, "Library Quarterly", 64(1994), 2(April), p. 130-161.

SHERMAN, Chris [2003] *ComScore Launches Search Engines Tracking System.*

<<http://searchenginewatch.com/searchday/article.php/2197801>>

SHIRKY, Clay [2005 a] *Ontology is overrated: categories, links, tags* in *Clay Shirky's Writings About the Internet.*

<http://shirky.com/writings/ontology_overrated.html>

SHIRKY, Clay [2005 b] *Folksonomy + controlled vocabularies*, weblog post

<http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php>

SHIRKY, Clay [2008] *Here comes everybody: the power of organizing without organizations*, New York, Penguin Press, 2008.

Traduzione Italiana: *Uno per uno, tutti per tutti. Il potere di organizzare senza organizzazione*, Torino, Codice edizioni, 2009.

SINHA, Rashmi [2005] *A cognitive analysis of tagging*, Rashmi's Blog: Thoughts about People, Technology and Running a company, <<http://rashmishinha.com/2005/09/27/acognitive-analysis-of-tagging/>>

SMITH, Gene [2008] *Tagging. People-Powered Metadata for the Social Web*, Berkeley CA, New Riders, 2008.

SÖDERBÄCK, Anders – MALMSTEN, Martin [2008] *LIBRIS. Linked Library Data*, "Nodalities Magazine", 5(Nov/Dec 2008), p. 20-21.

<http://www.talis.com/nodalities/pdf/nodalities_issue5.pdf>

SPALDING, Tim [2007] *When tags work and when they don't: Amazon and LibraryThing* post in "Thingology Blog", Tuesday, February 20, 2007.

<<http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php>>

SPALDING, Tim [2010] *Next next generation catalogs*, post in NGC4LIB, March 27, 2010, 01:48:24.

<<http://listserv.nd.edu/cgi-bin/wa?A2=ind1003&L=NGC4LIB&T=0&F=&S=&P=103530>>

SPITERI, Louise F. [2007] *Structure and form of folksonomy tags: The road to the public library catalogue* in "Webology", Volume 4, Number 2, June, 2007.

<<http://www.webology.ir/2007/v4n2/a41.html>>

SUNSTEIN, Cass R. [2003] *Republic.com. Cittadini informati o consumatori di informazioni?*, Bologna, Il mulino, 2003. Traduzione italiana di *Republic.com*, Princeton, N.J., Princeton University Press, 2001.

SURE, York - STUDER, Rudi [2005] *Semantic Web technologies for digital libraries*, "Library Management", 26(2005), 4/5, p. 190-195.

TAM, Winnie – COX, Andrew M. – BUSSEY, Andy [2009] *Student user preferences for features of next-generation OPACs. A case study of University of Sheffield international students in "Program: electronic library and information systems"*, Vol. 43 No. 4, 2009, pp. 349-374.

THOMAS, Marliese – CAUDLE, Dana M. – SCHMITZ, Cecilia M. [2009] *To tag or not to tag?* in "Library Hi Tech", 2009, v. 27, n. 3 p. 411-434.

THOMPSON, Henry S. [2008] *What's a URI and why does it matter?*, School of Informatics, University of Edinburgh World Wide Web Consortium, 4 December 2008.
<<http://www.ltg.ed.ac.uk/~ht/WhatAreURIs/>>

TOSCA CONSULTANT [2008] *Le catalogue de la bibliothèque à l'heure du Web 2.0. Étude des opacs de nouvelle génération, Étude réalisée par Marc Maisonneuve*, Paris, ADBS, 2008.

UNIVERSITY COLLEGE LONDON [2008] *Information behaviour of the researcher of the future*, 11 gennaio 2008,
<www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf>

UNIVERSITY OF CALIFORNIA BIBLIOGRAPHY SERVICES TASK FORCE [2005], *Rethinking how we provide bibliographic services for the University of California*.
<<http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf> >

UNIVERSITY OF WISCONSIN-MADISON LIBRARIES [2008] *Resource discovery exploratory task force final report*,
<<http://staff.library.wisc.edu/rdetf/RDETF-final-report.pdf>>

VANDER WAL, Thomas [2007] *Folksonomy in Vanderwal.net*
<www.vanderwal.net/folksonomy.html>

VELING, Anne – WEERD, Peter van der [1999] *Conceptual grouping in word co-occurrence networks*, Technical paper, IJCAI-99, the Sixteenth International Joint Conference on Artificial Intelligence, Stockholm, Sweden.
<<http://www.veling.nl/anne/work/grouping/grouping.pdf>>

W3C [1994 a] *W3C Semantic Web Frequently Asked Questions, What is the Semantic Web?*, (ultimo aggiornamento novembre 2009).
<<http://www.w3.org/RDF/FAQ>>

W3C [1994 b] *W3C Semantic Web Frequently Asked Questions, What is RDF?* (ultimo aggiornamento novembre 2009).
<<http://www.w3.org/2001/sw/SW-FAQ#whrdf>>

W3C [1994 c] *W3C Semantic Web Frequently Asked Questions, What are rules on the Semantic Web?* (ultimo aggiornamento novembre 2009).

< <http://www.w3.org/2001/sw/SW-FAQ#whrules> >

WENZLER, John [2007] *LibraryThing and the Library Catalog: Adding Collective Intelligence to the OPAC. A Workshop on Next Generation Libraries*, San Francisco State University Carl Nitig, September 7, 2007.

<<http://online.sfsu.edu/~jwenzler/research/LTFL.pdf>>

WESTON, Paul Gabriele [2002] *Il catalogo elettronico. Dalla biblioteca cartacea alla biblioteca digitale*. Roma, Carocci editore, 2002.

WESTON, Paul Gabriele [2007] *La gestione elettronica delle biblioteche* in *Biblioteconomia: principi e questioni* a cura di Giovanni Solimine e Paul Gabriele Weston, Roma, Carocci, 2007, p. 221-256.

WHITNEY, Colleen – SCHIFF, Lisa [2006] *The Melvyl Recommender Project. Developing Library Recommendation Services*, “*D-Lib Magazine*”, 12(2006), 12(December).

<<http://www.dlib.org/dlib/december06/whitney/12whitney.html>>

WISNIEWSKI, Jeff [2009] *Next-Gen OPACs: No Time Like the Present*, “*Online Magazine*”, 33(2009), 5(Sep-Oct), p.54-57

<www.onlinemag.net>

WONTOROWSKA, Marta [2009] *Readers 2.0. What kind of library do they need?* In 17th annual BOBCATSSS symposium (Bobcatsss 2009), Porto (Portugal), 28-30 January 2009.

<<http://eprints.rclis.org/15911/>>

YEE, Martha M. [2006] *Beyond the Opac: Future Directions for Web-Based Catalogues*, [Relazione presentata al seminario Beyond the OPAC : future directions for Web-based catalogues tenuto nel settembre 2006 a Perth, Western Australia].

<<http://www.nla.gov.au/lis/stdnrds/grps/acoc/papers2006.html>>

YANG, Sharon - YI LEE, Yan – XU, Amanda, [2009] *The Semantic Web and Libraries in the United States: Experimentation and Achievements*, Conference paper, Satellite Meetings IFLA 2009, Emerging trends in technology: libraries between Web 2.0, semantic web and search technology, Florence, 19-20 August 2009

<<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/SharonYang.pdf>>

YEE, Martha M. - LAYNE, Sara Shatford [1996] *Online public access catalogs* in “*Encyclopedia of library and information science*, editors Allen Kent and Harold Lancour, [poi] executive editor Allen Kent, administrative editor Carolyn M. Hall, New York - Basel, Marcel Dekker, 1968-2003,” Vol. 58 (1996), p. 149-238.

ZANI, Maurizio [2006] *Granularità: un percorso di analisi*, “*Digitalia*”, 2006, 2, p. 60-128.

http://digitalia.sbn.it/upload/documenti/digitalia20062_ZANI.pdf

Gli URL sono stati verificati l’ultima volta il 25 maggio 2010.

Le traduzioni sono della candidata.

La tesi è stata discussa il 29/06/2010. Emendata di alcuni refusi, è in linea da luglio 2010 su ESB FORUM <<http://www.burioni.it/forum/>>.