

I destini incrociati di strumenti di indicizzazione, dizionari, enciclopedie, motori di ricerca

ANNA LUCARELLI

Biblioteca nazionale centrale di Firenze
anna.lucarelli@bncf.firenze.sbn.it

*In verità, non c'è al mondo oggetto librario più fascinoso,
seducente, innamorativo di una Enciclopedia
(Giorgio Manganelli, *Il rumore sottile della prosa*)*

*Tuttavia non si può negare che nel definire come tale
un uovo prevalga la testimonianza dei sensi,
mentre per definire come tale un mammifero
prevalga la conoscenza delle classificazioni
(Umberto Eco, *Kant e l'ornitorinco*)*

Fra i nuovi alfabeti della biblioteca, ce n'è uno che è per antonomasia fatto di paradigmi in continuo mutamento. È quello dell'indicizzazione per soggetto, un'attività che si basa, appunto, sull'impiego di linguaggi, e che affida ai suoi elementi primari, alle parole, alla loro capacità comunicativa, l'obiettivo ambizioso di facilitare l'acquisizione della conoscenza sui contenuti concettuali dei documenti. Ma quale conoscenza, e soprattutto quali documenti? Ha ancora senso, nell'universo digitale in cui siamo immersi, occuparsi di indicizzazione tradizionale e di linguaggi con un'alta dose di formalizzazione? È vero, questi stanno evolvendo rapidamente e proprio in funzione dei mutamenti dei meccanismi di accesso al sapere attivati dal web, ma comportano comunque una certa distanza dalla lingua naturale che sempre più, invece, è al centro degli interessi, ad esempio, degli sviluppatori dei motori di ricerca. Si pensi a Powerset, la tecnologia di Bing, il motore di Microsoft che intende competere con Google proprio per la sua capacità di accogliere nella ricerca formule espressive direttamente legate all'uso corrente e diretto dei parlanti.

E ancora, in tempi di drammatiche emergenze economiche e di inevitabili contrazioni di risorse destinate alla catalogazione, è davvero necessario allestire sistemi

di recupero dell'informazione basati sull'attività intellettuale di umani? Chiedersi, come ha fatto la Library of Congress qualche anno fa tramite un noto rapporto, se siano migliori (anche in funzione dei costi che comportano) i linguaggi pre-coordinati o quelli post-coordinati non è più sufficiente. Il punto è se l'attività di indicizzazione continui ad avere una funzione culturale insostituibile nei processi di acquisizione della conoscenza, se abbia un ruolo anche in un panorama di *information retrieval* totalmente innovato da tecnologie in continuo sviluppo. Siamo ovviamente consapevoli delle intrinseche diversità che caratterizzano (caratterizzano?) OPAC e motori di ricerca, ma nello stesso tempo sappiamo quanto questi ultimi siano divenuti la principale "interfaccia" impiegata per il ricupero di informazioni, anche di tipo bibliografico, dalla maggior parte degli utenti-lettori-ricercatori di varia tipologia e composizione socio-culturale. Eppure, anche grandi progetti come quello di Google sulla digitalizzazione di massa di collezioni bibliografiche sembrano non valorizzare la componente dell'indicizzazione,¹ continuando a confidare sulle capacità offerte dalle parole del testo. A questo proposito, può anche essere interessante ricordare che i meccanismi di Google non sono affatto soltanto automatici, come saremmo portati a credere. Da un articolo di "Business Insider", sembra che la company paghi 14,5 dollari l'ora affinché mano umana controlli i risultati prodotti dal suo famoso algoritmo.² Certo è che i linguaggi di indicizzazione tradizionali, pur mantenendo le loro specificità, dovranno trovare una maggiore interazione con altri strumenti di accesso al sapere e alla conoscenza. Non seguendo percorsi isolati e obsoleti e rafforzando la loro capacità di mettere a disposizione dati di qualità, i linguaggi di indicizzazione non solo non vengono annullati da altri strumen-

ti per la ricerca e l'accesso alla conoscenza ma, come ormai riconosciuto da esperti e tecnici, rappresentano anzi architetture a cui quest'ultimi si possono agganciare. Per non parlare di quanto vocabolari controllati, classificazioni ecc., stiano diventando componenti fondamentali del *semantic web*.

Come dicevo, l'indicizzazione negli ultimi tempi si è trovata immersa in percorsi nuovi perché non solo è mutato il concetto di documento, di risorsa indicizzabile, ma sono mutate, in generale, le modalità di acquisizione della conoscenza e di recupero dell'informazione. Tutto questo all'interno di un processo sociale di progressiva, anche se lenta, alfabetizzazione digitale. Alle esigenze di *information literacy* vanno incontro servizi di *social web*, OPAC 2.0 (invero poco diffusi nelle biblioteche italiane), pagine Facebook, ed anche *discovery tools* che aggregano dati da fonti diverse e consentono di offrire servizi immediati. Su questo fronte si stanno sviluppando iniziative che promuovono standardizzazione e integrazione con strumenti tradizionali. Sui *discovery tools*, ad esempio, il NISO (National Information Standards Organization) ha costituito un gruppo di lavoro per sviluppare standard e raccomandare buone pratiche per i servizi bibliotecari di nuova generazione.³ Insomma, c'è un accordo consolidato sul fatto che le biblioteche debbano mettere a disposizione degli utenti le loro risorse in maniera innovativa. E trovo interessante che la ricerca si stia concentrando su modalità che aiutino a stabilire, fra oggetti del sapere, relazioni inattese, collegamenti di varia natura, non necessariamente legami rigidi, simbiosi.

Il movimento *Open access* invita a facilitare, tramite standard, l'interoperabilità fra gli archivi e, in una visione nuova del concetto di pubblicazione, a concepire la ricerca in riferimento ai dati piuttosto che ai testi o ai documenti. Nello stesso tempo, ci si chiede cosa connoti un documento, un documento scientifico in particolare, per essere considerato tale, e quali siano le sue misure qualitative; perché questo incide sulle scelte e sui "processi di conservazione, valutazione, e disseminazione della ricerca" (quali documenti digitali conservare, quali dati, chi potrà auto-valutarli, scartarli), ma incide fortemente anche sugli strumenti di recupero dell'informazione.⁴

Certo è che il web consente di collegare strutture concettuali anche molto diverse fra loro, facendole colloquiare con modalità impensabili fino a non molto tempo fa. Non a caso i temi più attuali della riflessione biblioteconomica riguardano proprio la possibilità di mappare e rendere interoperabili strutture che sinora,

in un universo pre-digitale, hanno avuto vite separate. Questo è particolarmente stimolante perché consente di ipotizzare e realizzare connessioni fra strumenti che hanno avuto origine in mondi diversi: ad esempio, fra quelli di indicizzazione (di ambito tipicamente bibliotecario) e quelli di tipo enciclopedico e lessicografico. Si tratta di una parentela solo apparentemente nuova, che riporta al ricordo degli antichi legami che secoli fa si svilupparono fra il mondo della cultura, dell'erudizione, della filosofia e quello delle biblioteche, facilitando gli esordi delle scienze bibliografiche. Si pensi a Leibniz e alle esperienze dell'enciclopedismo sei e settecentesco. Rapporti che sembravano essersi dispersi, dissolti e che invece tendono singolarmente a ristabilirsi grazie a internet e alle possibilità della biblioteca digitale.⁵

Trovo interessante chiedersi se esista una *koinè*, un linguaggio omologante, che uniformi e raccordi questi diversi mondi, quello degli strumenti di indicizzazione e quello di altri strumenti che lavorano con le parole e con i significati, come appunto le enciclopedie, i dizionari, i lessici, senza ovviamente perdere di vista le differenze che li caratterizzano. È suggestiva la metafora che Umberto Eco, ispirato da Deleuze e Guattari, echeggia nel suo *Vertigine della lista* per descrivere le caratteristiche logiche e strutturali delle enciclopedie moderne (altro si dovrebbe dire di quelle antiche e medievali). Queste tentano di mettere a disposizione il sistema complesso delle conoscenze alla maniera di "rizomi"; infatti, a differenza delle radici, che hanno natura gerarchica, connettono "un punto qualsiasi con qualsiasi altro punto", senza un inizio e una fine.⁶ Un'immagine che sembra condurci senza esitazioni alla filosofia dei *linked data*...

Che l'interoperabilità sia diventato uno dei temi centrali della biblioteconomia è dimostrato da molti progetti in corso come, ad esempio, da Europeana che ha stipulato convenzioni per mettere a disposizione milioni di metadati in applicazione del nuovo Europeana data model (EDM) per la *linked open data cloud*.⁷

In questo contesto di forte dinamicità, gli esperti di catalogazione cercano di orientarsi in direzioni nuove. L'IFLA Cataloguing Section, alla Conferenza 2012 di Helsinki, oltre che di regole e standard che possano supportare infrastrutture maggiormente adatte ai bisogni degli utenti ("Inspired moments in cataloguing"), si è occupata proprio di metadati integrati da fonti diverse e di interazione fra OPAC e altri strumenti di ricerca.⁸ Parallelamente, la sezione Classification and Indexing ha indagato architetture nuove per l'accesso semantico, strumenti per navigare nei *repositories*, metadati da dif-

fondere in contesti non tradizionali.⁹ A Tallin, nel Satellite meeting, si è parlato di come muoversi fra metodi tradizionali e innovativi di indicizzazione, di cooperazione con musei e archivi, di sinergie con editori; si è discusso di web semantico, di coinvolgimento degli utenti nella creazione di metadati per soggetto e del loro riuso da parte di *communities* non bibliotecarie.¹⁰

In questo panorama variegato, è allora indispensabile che chi si occupa di indicizzazione allarghi la propria visuale verso altre esperienze, ad esempio verso quanto si sta facendo nel campo delle cosiddette “ontologie” (strutture concettuali ritenute particolarmente adatte al web),¹¹ e dei motori di ricerca, ormai dichiaratamente proiettati anche sul fronte del web semantico e tuttavia ancora deludenti sul piano delle metodologie e tecniche relative all’indicizzazione e recupero per soggetto. La “Ricerca avanzata” di Google segue logiche di tipo testuale;¹² lo stesso avviene in Google books, Google Livres, Google libri, e sappiamo quale bassa pertinenza si ottenga nel caso di ricerche su concetti complessi o in presenza di polisemie.

Situazione analoga riguarda altri motori. Anche Bing non rivela grossi passi avanti, neppure tramite la “Ricerca avanzata”; non suggerisce sinonimi, pur offrendo la possibilità di effettuare “Ricerche correlate”. Rispetto a Google, gestisce meglio le polisemie e così, ad esempio, partendo dal termine *Gru* possiamo scegliere, nell’ambito delle “Ricerche correlate”, se intendiamo cercare risorse sul macchinario o l’animale.¹³

Ma è pur vero che Google sta comunque lanciandosi sul fronte della semantica a partire da altre direzioni, ad esempio, con l’ appena inaugurato *Knowledge Graph*,¹⁴ come viene detto: “a knowledge base used by Google to enhance its search engine’s search results with semantic-search information gathered from a wide variety of sources. Knowledge Graph display was added to Google’s search engine in 2012, starting in the United States, having been announced on May 16, 2012”.¹⁵

Un’iniziativa interessante di cui seguiremo gli sviluppi, benché non ancora implementata nella versione italiana del motore di ricerca.



L’interfaccia di *Google libri* con i campi di ricerca per parole contenute ma non per soggetto

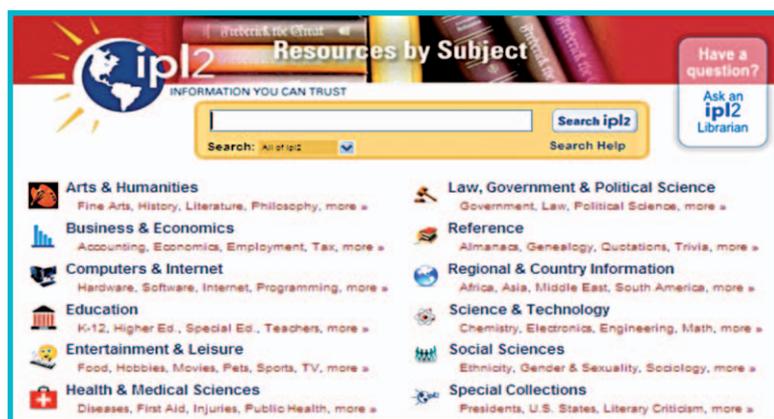


Il Grafico della conoscenza recentemente rilasciato dalla versione statunitense di Google

Quelli citati non sono ovviamente gli unici motori del web; in Internet Archive,¹⁶ la “Ricerca avanzata” prevede dei “subjects”,¹⁷ metamotori più o meno specializzati per aree disciplinari¹⁸ si prefiggono lo scopo di favorire il recupero di risorse qualificate come, ad esempio, nel nostro ambiente, Internet Public Library2 (IPL2).¹⁹ Ma in genere, nonostante vogliano illuderci di offrire accessi sulla base di soggetti (“Resources by Subject”), in realtà organizzano contenuti in base a categorie, oppure a classificazioni.²⁰

In questo quadro, che non ha certamente la pretesa di essere esaustivo, come stanno evolvendo i linguaggi di indicizzazione tradizionali, i più diffusi, come *Library of Congress subject headings* (LCSH), RAMEAU, le tedesche *Regeln für den Schlagwortkatalog* (RSWK), il *Nuovo soggettario*?

Il dato interessante è che, pur con le loro differenze, tutti si stanno orientando verso l’interoperabilità. Quelli europei stanno inglobando equivalenze con LCSH (“il principe dei soggettari”, come lo definisce Carlo Revelli),²¹ quasi tutti stanno convertendo i pro-



La pagina di Internet Public Library 2 dedicata al recupero di risorse in base ai contenuti

pri thesauri in formati che ne agevolano l'uso e sono disponibili in SKOS (Simple Knowledge Organization System),²² *Nuovo soggettario* compreso. L'uso del formato SKOS/RDF facilita notevolmente non solo l'interazione fra i vari strumenti di indicizzazione, ma anche quella con strutture di altro tipo, rendendoli fruibili come servizi web in ambiti non necessariamente bibliotecari. Per queste ragioni, che non sono le sole pur apparendomi le più macroscopiche, chi si occupa di indicizzazione dovrà lavorare – maggiormente di quanto non stia ancora facendo – a contatto con altri esperti di recupero dell'informazione, di linguaggi, di parole, di significati come, ad esempio, con coloro che operano nel settore degli studi di linguistica, di linguistica computazionale, in quello scientifico ed editoriale delle enciclopedie online e degli strumenti lessicografici.

Numerose esperienze italiane e straniere stanno già dimostrando che il mondo dei linguisti e quello dei bibliotecari che allestiscono e curano strumenti di indicizzazione affrontano problematiche affini. Sempre più i linguisti si occupano di “ontologie”: si pensi al progetto di Tullio De Mauro “Senso comune”, che intende produrre una “base di conoscenza per la lingua italiana”.²³ Realizzano ricerche avanzate nel settore dell'*information technology* e dell'indicizzazione automatica, tema di cui si occupa anche la Biblioteca nazionale centrale di Firenze che sperimenta l'impiego del *Nuovo soggettario* per l'indicizzazione semiautomatica di risorse digitali. Così i bibliotecari non possono non guardare con interesse alle esperienze innovative del Laboratorio di documentazione dell'Università della Calabria, al Centro di linguistica computazionale di Salerno, alla Fondazione Kessler di Bolzano e ovviamente all'Istituto di linguistica computazionale del CNR di Pisa.²⁴ Ma potrei citare anche l'Istituto per il lessico intellettuale europeo e

storia delle idee (ILIESI), che si occupa di terminologia e pubblica indici, concordanze, appronta banche dati e “strumenti di informatica umanistica applicati alla terminologia tecnica”, e ancora associazioni come Ass.I.Term, interessata ai linguaggi speciali e settoriali. Insomma con questa realtà attiva, ricca di iniziative, il mondo delle biblioteche dovrebbe maggiormente dialogare e avviare progetti comuni.

Come sappiamo, una delle tipologie editoriali che il web ha contribuito maggiormente a diffondere e rinnovare, potenziandone l'offerta ad accesso aperto (che non signifi-

ca sempre gratuito), è proprio quello delle enciclopedie e dei dizionari; attivarne il colloquio con gli strumenti di indicizzazione creerà un valore aggiunto per gli utenti delle biblioteche e dei motori di ricerca, ammesso che abbia ancora un senso operare questa distinzione. Il progetto con cui, già qualche anno fa, le biblioteche universitarie di Washington hanno reso disponibili tramite Wikipedia le proprie collezioni digitali,²⁵ è stato un antesignano in questo senso, una sorta di profezia se pensiamo, ad esempio, ai legami che OCLC sta ora ampiamente consolidando con Wikipedia e i wikipedisti, come abbiamo potuto constatare dalle iniziative diffuse in rete proprio negli ultimi mesi.²⁶

Dunque viene finalmente condiviso che le enciclopedie, per antonomasia strumenti di approfondimento della conoscenza, abbiano un forte legame con quanto realizzato e sviluppato dai bibliotecari. Insieme ai dizionari, da cui si differenziano non sempre in modo rigido, hanno avuto una straordinaria evoluzione grazie ai formati digitali. Come già accennato, i rapporti fra il mondo dell'enciclopedismo, della lessicografia e quello delle biblioteche risalgono in realtà ad epoche lontane, a tempi in cui le discipline del libro, emancipandosi dall'erudizione in senso stretto, muovevano i primi passi. La lessicografia, nel Medioevo e nell'età monastica, è stata alle origini degli “indici”. Ci diceva Luigi Balsamo nella sua bellissima opera di storia della bibliografia, che anche le enciclopedie e i dizionari, con i cataloghi di vendita dei librai, sono da sempre “assieme ai repertori bibliografici parte essenziale dei meccanismi della memoria collettiva”; a partire dal Settecento hanno avuto una funzione di stimolo alla lettura, sono stati fonti di notizie sui libri, libri su determinati argomenti, pur non volendovisi sostituire, perché come avvertiva D'Alembert in persona: “[le enciclopedie] possono tutt'al più giova-

re a dare qualche lume a coloro i quali, senza tale strumento, non avrebbero avuto il coraggio di procurarseli: ma non potranno mai sostituire i libri per coloro che vogliono istruirsi”.²⁷

Questo ruolo delle enciclopedie come “ponte” fra la descrizione dei temi e degli oggetti del sapere e le risorse bibliografiche e documentarie ad essi collegate, dunque, è sempre esistito. Ce l’avevano le enciclopedie cartacee (con la bibliografia di riferimento posta a chiusura dei lemmi); ce l’hanno, a maggior ragione, quelle digitali che sempre più frequentemente collegano i lemmi ad altre risorse che riguardano i medesimi concetti, segnalando ulteriori percorsi di ricerca. Così, ad esempio, ai lettori della voce *Social Epistemology* della *Stanford Encyclopedia of Philosophy*, fa molto piacere trovare a fine pagina, fra le “Related entries”, notizie su risorse online affini e altre linee di ricerca.²⁸

Il legame fra le enciclopedie, i dizionari e il mondo dei libri può potenziarsi grazie all’interazione con i thesauri, i più efficaci “attrezzi” impiegabili nell’indicizzazione di contenuti concettuali; questi possono colloquiare con i cataloghi delle biblioteche dove sono descritte le manifestazioni che veicolano le opere. Si tratta di strumenti diversi, ma dalla valorizzazione dei loro “destini incrociati” possono derivare grandi benefici per il mondo della ricerca.

Non è certo un caso che, da quando sono online, le enciclopedie e i dizionari abbiano mutuato funzionalità tipiche degli strumenti di indicizzazione. Offrono ai lettori percorsi di ricerca fra argomenti affini; introducono una sorta di classificazioni e di “categorie”,²⁹ creano relazioni semantiche fra i lemmi e la possibilità di navigare fra loro, attivano collegamenti con risorse do-

documentarie. Tutte funzionalità tipiche dei thesauri, anche se in questi ben più formalizzate.

Wikipedia, ormai da tempo in progressivo miglioramento qualitativo, si sta sviluppando proprio in questa direzione. Ad esempio, gestisce la sinonimia con un sistema di re-indirizzamento da forme non preferite: *Figli dei fiori* rinvia a *Hippy*.³⁰ Alla voce *Thesaurus*³¹ prevede una disambiguazione: “Se stai cercando il trattatista italiano, vedi Emanuele Tesauro”; propone il significato in ambiti disciplinari diversi, sulla base di una sorta di analisi a faccette; fornisce una bibliografia (anche se cita opere il cui tema di base non è propriamente quello dei thesauri...); presenta voci correlate che rammentano le relazioni associative dei nostri vocabolari controllati, nonostante da nessuna parte indichi con quali logiche o algoritmi si stabiliscano le correlazioni. In fondo alla pagina dedicata al soggetto *Thesaurus*, presenta quelle che vengono dette “categorie”. Una di queste, guarda caso, è proprio la Lessicografia,³² la cui pagina è peraltro strutturata differentemente da quelle dei concetti comuni. Dunque *Wikipedia* prevede una sorta di struttura classificatoria, un sistema di “categorie” e “categorie nascoste” che hanno anche lo scopo di organizzare il lavoro dei collaboratori. A loro, ma non solo, viene spiegato cosa si debba intendere per categorie,³³ come crearle, a chi rivolgersi in caso di dubbio, come attribuirle, cosa fare nel caso in cui il concetto appartenga a più di una, come creare sottocategorie o strutture ad albero in cui siano gerarchizzate, cosa fare di quelle “orfane” ecc. Analogo impegno sul versante delle relazioni semantiche rivela l’*Encyclopaedia Britannica*³⁴ che, partendo da una qualsiasi ricerca, prevede una funzione “Thesaurus” in cui il termine è presentato all’interno di una rete

The screenshot shows the Britannica website interface. At the top, there is a search bar with the text 'library' and a 'Go' button. Below the search bar, the search results are displayed. On the left side, there is a navigation menu with links for 'Encyclopedia', 'Year in Review', 'Dictionary', 'Thesaurus', 'Kids Encyclopedia', 'Magazines', 'Ebooks & Primary Sources', and 'Web Search'. The main content area shows an advertisement for 'Web Based Library System' and a definition for the word 'library' from Merriam-Webster. The definition includes the entry word, function (noun), meaning (1 a place where books, periodicals, and records are kept for use but not for sale), synonyms (archive), and related words (stacks). There is also a note about the collection of comic books.

Il lemma *Library* dell’Enciclopedia Britannica online, nell’inglobata funzione Thesaurus integrata con il Merriam-Webster Dictionary

Il lemma *Spread* dell'Enciclopedia Treccani online, con *Approfondimenti* e *Argomenti correlati* linkabili

di sinonimi, termini correlati, antonimi ecc. estratti dal *Merriam-Webster Online Dictionary* ad esso collegato. Digitando *Library* nel campo della ricerca, si accede anche a un browser da cui si vedono voci collegate, e pagine su specifiche biblioteche. Queste funzioni di navigabilità sono attivate nelle varie versioni della *Britannica* (conci- sa, estesa, accademica ecc.).

Trovo bellissimo che in un'enciclopedia sia possibile transitare sia al corrispondente lemma di un dizionario (funzionalità prevista anche dalla nostra *Treccani* online) sia ad uno strumento thesaurale. Recentissima, peraltro, la notizia della *partnership* fra la *Britannica* e Bing, un accordo in base al quale il servizio online di Microsoft includerà, nella pagina dei risultati della ricerca su un certo argomento, informazioni provenienti dalla prestigiosa enciclopedia;³⁵ Bing rafforza così anche la competizione con Google che a sua volta, come sappiamo, ha consolidato sempre più i suoi legami con Wikipedia. Appare invece meno spregiudicata, almeno al momento, la politica della gloriosa *Brockhaus Enzyklopädie* tede-

sca, “Das Wissen der Welt”, come si auto-definisce,³⁶ peraltro più riluttante a rendersi disponibile gratuitamente in rete (al primo annuncio di una sua versione digitale del 2008 hanno fatto seguito varie vicissitudini editoriali che ne hanno intralciato il percorso) e comunque accessibile con modalità meno dirette.

Analoghe funzioni di navigabilità semantica vengono proposte anche da enciclopedie online tematiche o specialistiche, dove troviamo macrostrutture classificatorie, applicazioni di analisi a faccette, sistemi di “see also”,³⁷ sebbene ancora una volta intesi molto liberamente e senza alcuna deferenza verso quegli standard sui thesauri elaborati dall'ISO, riferimento indiscusso di noi bibliotecari.

Una struttura semantica articolata è prevista anche in *Treccani.it*, il portale da cui è disponibile una selezione dell'enciclopedia e del vocabolario prodotti dall'Istituto dell'Enciclopedia italiana.³⁸ Contiene lemmi afferenti a tutte le discipline, talvolta molto specialistici e presentati anche in lingue diverse dall'italiana nel caso siano attestati nell'uso in questa forma. *Spread*, tanto per fare un esempio, alla descrizione del significato aggiunge la categoria “Temi generali” all'interno della classe dell'Economia e indica “Approfondimenti” e “Argomenti correlati”, ovviamente navigabili.

Ma *Treccani.it* fa qualcosa in più rispetto ad altre enciclopedie online, perché non solo prevede funzionalità simili a quelle dei thesauri, ma con uno di essi interagisce concretamente: migliaia di collegamenti attivi la collegano al Thesaurus del *Nuovo soggettario*, ormai ricco di oltre 46.000 termini.³⁹

Lo strumento, realizzato dalla Biblioteca nazionale centrale di Firenze, ha fatto molta strada dalla pubblicazione del prototipo; si inserisce in un progetto aperto, perché aperti sono i prodotti che mette a disposizione da anni, anni importanti in cui si sono evoluti il Thesaurus, il *Manuale applicativo*, la rete delle istituzioni che collaborano al progetto. Le biblioteche italiane lo stanno via via adottando, mentre le *Linee guida sui siti web delle Pubbliche Amministrazioni* (in versione finale da novembre del 2011) lo citano come uno degli strumenti di riferimento per l'organizzazione di contenuti concettuali.⁴⁰

Come mostra l'esempio del termine *Biopolitica*, dal campo "Fonte", tramite la freccetta che precede la dicitura Treccani.it, si può transitare nel corrispondente lemma dell'enciclopedia. Il coinvolgimento del *Nuovo soggettario* con altri sistemi di organizzazione della conoscenza (i cosiddetti KOS), è fortemente esplicitato proprio nel campo "Fonte" dei termini. Le sigle citate fanno riferimento agli oltre quattrocento fra repertori, banche dati, enciclopedie, dizionari impiegati per i controlli semantici, un vero e proprio apparato di *reference* in continua crescita, oggetto di un costante lavoro di verifica ed aggiornamento.⁴¹

In questa ottica di colloquio con altri linguaggi di indicizzazione e altri strumenti, sempre più termini del *Nuovo soggettario* stanno acquisendo le equivalenti *Library of Congress subject headings* e attivano link con altri thesauri generali o specialistici (ad esempio, con quello della Biblioteca Rostoni dell'Università di Castellanza, con Eurovoc, Agrovoc ecc.).⁴² Oltre ai collegamenti con la Treccani online sono in via di implementazione quelli con Wikipedia e, come si può visualizzare da termini come *Badanti* o *Mobbing*,⁴³ con il servizio *Parole nuove* dell'Accademia della Crusca, recente partner del *Nuovo soggettario*. In questo caso, i collegamenti funzionano già anche in direzione inversa: i lettori possono navigare dal lemma previsto come neologismo dall'Accademia della Crusca, ad esempio *Cartolarizzazione*, al corrispondente del *Nuovo soggettario* (cliccando sul relativo logo posto in fondo alla pagina) e da lì, tramite il tasto "Noti-

Il termine *Biopolitica* del Nuovo soggettario, linkato alla Treccani online, agli equivalenti della Library of Congress (LCSH) e di RAMEAU

Il lemma *Cartolarizzazione*, fra i neologismi dell'Accademia della Crusca, linkato al corrispondente termine del Nuovo soggettario

zie bibliografiche", alle risorse su quel tema descritte nell'OPAC della BNCF.

È evidente che di questo genere di connessioni possono beneficiare varie tipologie di utenti: servono ad approfondire conoscenze, ad espandere percorsi di ricerca. Dunque anche il settore della lessicografia online si sta evolvendo e gli studi che la riguardano mostrano ancora una volta un'incredibile convergenza con le tematiche affrontate dai bibliotecari che si occupano di indicizzazione.⁴⁴ Interessante la tedesca *Internetlexikografie*,⁴⁵

The screenshot shows the Zanichelli online dictionary interface. At the top, there are logos for Zanichelli, Loescher, CINECA, and META. The search bar contains the text 'Sfumature' and has buttons for 'cerca' and 'ricerca avanzata'. Below the search bar, there is a list of 'Documenti trovati: 965' with various words like 'accordo', 'accudire', 'acquisto', etc. The main content area is titled 'FLESSIONE' and shows the lemma 'allegria' with its phonetic transcription and a list of synonyms. The synonyms include 'contentezza', 'esultanza', and 'euforia'. The interface also includes a navigation menu on the left and a search history section at the bottom.

Il lemma *Allegria* del Dizionario Zingarelli 2012 online, navigabile con altri lemmi citati fra le *Sfumature*

una rete a cui partecipa anche l'EURAC di Bolzano,⁴⁶ e non possiamo ignorare il progetto WordNet, elaborato all'Università di Princeton, un dizionario per la lingua inglese disponibile gratuitamente, dotato di strutture relazionali del tutto simili a quelle dei thesauri,⁴⁷ che sul versante italiano prende il nome di MultiWordNet.⁴⁸ Come sappiamo, è nutrita anche l'offerta online di vocabolari della lingua italiana, di dizionari di sinonimi, dizionari analogici, benché non tutti con le stesse funzionalità, non tutti ad accesso gratuito, non tutti di qualità, diversi fra loro e tuttavia con l'attitudine comune a stabilire connessioni semantiche e collegamenti fra termini e significati. Lo *Zingarelli 2012* online (disponibile su abbonamento), ad esempio, all'indicazione dei sinonimi aggiunge anche cosiddette "Sfumature", di cui arricchisce un gran numero di lemmi.

Non tutti i dizionari online prevedono queste funzionalità; in quelli bilingue le strutture relazionali sono più rare. I numerosi *Larousse*⁴⁹ non riportano alcun tipo di reti semantiche e il massimo che ci può capitare, ad esempio, è di scorgere nella pagina del lemma italiano/tedesco relativo a *Termosifone*,⁵⁰ banner di aziende produttrici di caldaie, radiatori, stufe, camini: oggetti correlati dalla pubblicità...

Insomma, è certamente con le esperienze e i prodotti editoriali di maggiore livello e qualità che potremmo iniziare a condividere i nostri percorsi, pur nella consapevolezza delle finalità e dei contesti diversi in cui ci muoviamo e con l'obiettivo di far decollare una cooperazione sinora inesistente o al massimo troppo debole. Non ci sono dubbi che chi si adopera per allestire thesauri deve rispettare standard, disambiguare significati, mettere a fuoco differenze fra termini/concetti/, fra senso e significati, con parametri diversi da quelli impiegati da linguisti, lessicografi o curatori di enciclopedie. Ma è pur vero che i vantaggi di una fattiva collaborazione sarebbero così consistenti da spiazzare le ragioni di chi teme contaminazioni o inappropriata invasioni di campo. Arricchirebbero di tessere importanti il caleidoscopico mosaico in cui gli strumenti di indicizzazione stanno ridefinendosi come uno snodo importante fra molti possibili linguaggi, rivestendo un ruolo fondamentale nel mondo del *Linked data*. E sappiamo che è qui il futuro del Web, quello di Tim Berners-Lee, un Web di dati e non solo di documenti, come un recente convegno (fiorentino nella sede ma internazionale nei contenuti), ha suggestivamente confermato.⁵¹

NOTE

L'articolo sviluppa e aggiorna la relazione tenuta dall'autrice in occasione del convegno "I nuovi alfabeti della biblioteca" (Milano, Palazzo delle Stelline, 15-16 marzo 2012).

¹ Sull'accordo fra Google e il Ministero per i beni e le attività culturali per la digitalizzazione delle opere delle biblioteche italiane: <http://www.beniculturali.it/mibac/export/MiBAC/sito-MiBAC/Contenuti/Ministero/Accordi/Altri/visualizza_asset.html_1672918906.html>. Il coinvolgimento della Biblioteca nazionale centrale di Firenze al progetto è stato sintetizzato recentemente in: PIERANTONIO METELLI, *La digitalizzazione di massa: Google e ProQuest, in 1861/2011: l'Italia unita e la sua biblioteca*. [Catalogo della Mostra tenuta a Firenze nel 2011-2012], Firenze, Polistampa, 2011, p. 205-206.

² "Google pays outside contractors \$14.50 an hour to rate search results for relevance, then feeds this data back into its search algorithms. But to get the job, testers have to pass a couple of difficult exams" (<http://articles.businessinsider.com/2012-01-20/tech/30646009_1_search-algorithms-google-search-results>).

³ "Questo gruppo di lavoro, interno a NISO [...] vuole trovare strategie per riunire i fornitori di servizi e di contenuti per sviluppare uno standard comune e per rendere questi strumenti veramente efficaci per l'utente. Inoltre, i componenti del gruppo auspicano di trovare un "linguaggio comune" sulle collezioni indicizzate, sui termini d'uso, sulla disponibilità dei contenuti indicizzati, etc., dal momento che non sempre è chiaro quali risorse sono disponibili e come, se i metadati sono ricavati direttamente dal testo o se invece provengono da aggregatori" (<<http://bib04.caspar.it/cibernewsletter/?p=14584>>).

⁴ I temi sono emersi allo scorso CERN Workshop, promosso da Open Archive Iniziative: PAOLA CASTELLUCCI, *Tempo e massa: nuova energia nella comunicazione scientifica*, "Bollettino AIB", 2011, n. 3, p. 237-244.

⁵ Ho avviato una riflessione su questo tema in ANNA LUCARELLI, *Enciclopedie e Thesauri, "essenze" e "proprietà". Una lista semiseria, in Books seem to me to be pestilent things. Studi in onore di Piero Innocenti per i suoi 65 anni*, Manziana, Vecchiarelli, 2011, p. 975-978.

⁶ Cfr. UMBERTO ECO, *Vertigine della lista*, Milano, Bompiani, 2009, p. 231-241.

⁷ Cfr. MARIA TERESA NATALE, *Europeana data Exchange Agreement*, "Bollettino AIB", 2011, n. 3, p. 246-247.

⁸ <<http://conference.ifla.org/ifla78/session-80>>.

⁹ <<http://conference.ifla.org/ifla78/session-117>>.

¹⁰ <<http://www.nlib.ee/index.php?id=17763>>.

¹¹ Per una panoramica sui lavori in questo ambito: *Le ontologie*, a cura di Maria Teresa Biagetti, "Aida informazioni", 2010, n. 1-2.

¹² <http://www.google.com/advanced_search>.

¹³ <<http://it.bing.com/search?q=gru&q=gru&sk=n&f=1&sk=gru&sc=8-3&sp=-1&sk=&q=1&FORM=AXRE>>.

¹⁴ <<http://www.google.com/insidesearch/features/search/knowledge.html>>.

¹⁵ <http://en.wikipedia.org/wiki/Knowledge_Graph>.

¹⁶ Come si auto-definisce nella sua home page: "The Internet Archive, a 501(c)(3) non-profit, is building a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public" (<<http://www.archive.org/>>).

¹⁷ <<http://www.archive.org/advancedsearch.php>>.

¹⁸ Interessante la rassegna che si trova in TERESA NUMERICO – DOMENICO FIORMONTE – FRANCESCA TOMASI, *L'umanista digitale*, Bologna, Il Mulino, 2010, p. 185-190.

¹⁹ <<http://www.ipl.org/>>.

²⁰ <<http://www.ipl.org/div/subject/>>.

²¹ Cfr. CARLO REVELLI, *Pareri sui soggetti*, "Biblioteche oggi", 29 (2011), 1, p. 52-58: 54.

²² <<http://www.w3.org/2004/02/skos/>>.

²³ Come è indicato nel sito, "oltre a svolgere le funzioni di un normale dizionario, Senso Comune consente di classificare ciascuna accezione dal punto di vista concettuale, di specificare esempi d'uso, sotto-accezioni, relazioni lessicali e, in futuro, tematiche e argomentali, attraverso una piattaforma di acquisizione di facile uso" (<http://www.sensocomune.it/portale/index.php?option=com_content&task=view&id=14&Itemid=70>).

²⁴ Uno dei progetti del CNR più interessanti in questo senso è "Intraweb semantico: gestione avanzata dell'informazione in organizzazioni complesse", di cui è responsabile Vito Pirrelli. Come si legge: "Nel corso del 2008 le attività del modulo sono state in gran parte rivolte alla valutazione e sperimentazione della portabilità degli strumenti di analisi documentale esistenti verso altri potenziali usi e applicazioni, all'estensione degli stessi strumenti ad altre lingue, nonché all'individuazione di ulteriori obiettivi a breve termine e all'acquisizione di nuovi finanziamenti. Il dimostratore realizzato implementa funzionalità avanzate di indicizzazione e navigazione semantica di ampie basi documentali, disponibili come servizi a richiesta all'interno di un sistema cooperativo accessibile in rete, con finalità di i) gestione di conoscenza distribuita non strutturata, ii) data mining e iii) semantic scouting" (<<http://www.cnr.it/istituti/Ricerca.html?cds=048>>).

²⁵ FABIO METITIERI, *Spiacenti, la conoscenza enciclopedica si è trasferita online*, "Biblioteche oggi", gennaio-febbraio 2009, p. 15-22, <<http://www.bibliotecheoggi.it/content/20090101501.pdf>>.

²⁶ Cfr., ad esempio, *OCLC Research to Host "Wikipedia and Libraries: What's the Connection?"*, webinar on 31 July, <<http://www.oclc.org/research/news/2012/07-06a.html>>. A Susanna Giaccari il merito di avere recentemente stimolato la riflessione sul tema nell'ambito di AIB-CUR.

²⁷ Entrambe le citazioni sono tratte da LUIGI BALSAMO, *La Bibliografia. Storia di una tradizione*, Firenze, Sansoni, 1984, p. 98.

²⁸ <<http://plato.stanford.edu/entries/epistemology-social/>>.

²⁹ La presenza di strutture categoriali in Wikipedia era stata segnalata qualche anno fa da: SERENA SPINELLI, *I thesauri nelle biblioteche universitarie oggi*, in *I Thesauri tra cataloghi e web: riflessioni*

sul tema in occasione della presentazione del *Thesaurus italiano infanzia e adolescenza*, Firenze, Istituto degli Innocenti, 6 febbraio 2009, <<http://www.iskoi.org/doc/thesauri.htm>>.

³⁰ <http://it.wikipedia.org/w/index.php?title=Figli_dei_fiori&redirect=no>.

³¹ <<http://it.wikipedia.org/wiki/Thesaurus>>.

³² <<http://it.wikipedia.org/wiki/Categoria:Lessicografia>>.

³³ <<http://it.wikipedia.org/wiki/Aiuto:Categorie>>.

³⁴ <<http://www.britannica.com/>>.

³⁵ Cfr., ad esempio, <http://www.bing.com/community/site_blogs/b/search/archive/2012/06/07/bing-introduces-new-britannica-online-encyclopedia-answers.aspx>.

³⁶ <<http://www.brockhaus.de/>>.

³⁷ Solo per fare qualche esempio, nella *Internet Encyclopedia of Philosophy* il lemma *Epistemology* (<http://www.iep.utm.edu/epistemo/>) a fine pagina prevede il link a *Categories* (<http://www.iep.utm.edu/category/m-and-e/epistemology/>) da cui si accede a una macrostruttura entro cui il concetto viene “sfaccettato”. Anche nell’enciclopedia americana di medicina *MEDPEDIA*, in fondo al lemma *Aphasia* (<http://wiki.medpedia.com/Aphasia>), troviamo sia i *See also*, che le solite *Categories*.

³⁸ <<http://www.treccani.it/>>.

³⁹ <<http://thes.bncf.firenze.sbn.it/ricerca.php>>. Sulla storia del progetto e i suoi stati di avanzamento: ANNA LUCARELLI, *Nuovo soggettario. Un servizio per le biblioteche italiane e il mondo della ricerca*, “Accademie & Biblioteche d’Italia”, n.s., a. V, luglio-dicembre, n. 3-4, 2010, p. 69-81.

⁴⁰ <http://www.digitpa.gov.it/sites/default/files/LINEE%20GUIDA%20SITI%20WEB%20DELLE%20PA%202011_finale.pdf>. Il paragrafo dedicato a *Classificazione e semantica* è il 5.1.

⁴¹ Cfr. ELISABETTA FRANCONI, *Fare reference con le “Fonti” del Nuovo soggettario*, relazione al seminario *Comunicare significato: Information literacy e thesauri*, Firenze, Biblioteca nazionale centrale, 12 aprile 2012 (Atti non pubblicati); per il programma: <<http://www.aib.it/attivita/2012/20449-comunicare-significato-information-literacy-e-thesauri/>>.

⁴² Su questo GIOVANNI BERGAMIN – ANNA LUCARELLI, *Il Nuovo soggettario come servizio per il mondo dei linked data*, paper presentato al seminario *Global Interoperability and Linked Data in Libraries*, University of Florence, Aula Magna, June, 18-19-2012 (in corso di pubblicazione).

⁴³ <<http://thes.bncf.firenze.sbn.it/termine.php?id=35195>>; <<http://thes.bncf.firenze.sbn.it/termine.php?id=20405>>.

⁴⁴ Si veda, ad esempio, <<http://multimedia.ids-mannheim.de/mediawiki/web/index.php/Publications>>.

⁴⁵ <<http://www.internetlexikografie.de>>.

⁴⁶ <<http://www.eurac.edu/it/research/projects/ProjectDetails.html?pid=9619>>.

⁴⁷ <<http://wordnet.princeton.edu/>> (per informazioni in italiano, <<http://it.wikipedia.org/wiki/Wordnet#mw-head#mw-head>>).

⁴⁸ <<http://multiwordnet.fbk.eu/english/home.php>>.

⁴⁹ <<http://www.larousse.com/it/dizionari>>.

⁵⁰ <<http://www.larousse.com/it/dizionari/italiano-tedesco/termosifone/45696>>.

⁵¹ *Global Interoperability and Linked Data in Libraries*, University of Florence, Aula Magna, June, 18-19-2012: <<http://www.linkedheritage.org/linkeddataseminar/>>.

DOI: 10.3302/0392-8586-201209-005-1

ABSTRACT

In an information retrieval world with continuous technological innovations and ever-growing web functionalities, subject indexing tools continue to have an irreplaceable cultural role in the information and knowledge acquisition process. Their role has been changing and is following common and crossing paths with web search engines and other online resources. Many Italian and international experiences show that linguists, lexicographers, encyclopedists and librarians who are involved in indexing have to face similar issues. Thanks to the Net and to linked data, tools from different sources and with different backgrounds are becoming interoperable and communicating, with common advantage.

Libraries offer disambiguating semantic structures to web search engines, according to the logic behind authority files. Thesauri do not only communicate with other thesauri but also with dictionaries and authoritative encyclopedias. Therefore, since quality data are made available to the Net, indexing languages are becoming more and more important components of the “semantic web”.