

L'accesso per soggetto: dal catalogo a schede al catalogo online, con uno sguardo all'Intelligenza Artificiale

di Alberto Cheti

Ad Anna Lucarelli

Quello che mi stupisce di più è che i catalogatori continuano a creare intestazioni, scrupolosamente, onestamente, nonostante il fatto che non siano utilizzate nei sistemi per biblioteche nel modo previsto, e non lo siano più da quando i primi sistemi sono stati sviluppati più di 30 anni fa. Le intestazioni sono foraggio per la ricerca mediante parole chiave, ma non più di quanto lo sarebbe un semplice insieme di tag. Le intestazioni non svolgono mai la funzione organizzativa per la quale erano state concepite.

Essendomi occupata della transizione dai cataloghi a schede ai cataloghi online, ho iniziato a scavare nella letteratura biblioteconomica e ho trovato l'oro [...]. Ciò che vi si scopre è al tempo stesso sbalorditivo e scoraggiante [...]: in termini di modelli di catalogo, sono stati fatti pochissimi progressi e continuiamo a dedicare più tempo all'organizzazione di dati bibliografici atomistici ignorando l'accesso ai soggetti.

(Karen Coyle)

Introduzione

Nell'estate del 2023 mi capitò di rileggere due contributi di Karen Coyle sulla transizione dal catalogo a schede al catalogo online, nei quali ha un posto di rilievo il tema dell'accesso per soggetto alle risorse bibliografiche. La sua disamina giunge alla conclusione che i vari modelli di catalogo sviluppatisi nel tempo hanno apportato ben pochi progressi in questo ambito. Non solo, buona parte del potenziale informativo contenuto negli strumenti con cui si costruiscono i metadati semantici si perde nel catalogo e resta così inutilizzato dagli utenti. Un bel paradosso oggi nell'epoca delle tecniche di *semantic search*, che utilizzano gli stessi elementi di semanticità che sono incorporati nei linguaggi di indicizzazione (il significato di parole e

ALBERTO CHETI, e-mail: chetialberto@infinito.it.

Ultima consultazione siti web: 22 dicembre 2024.



frasi, la sua contestualizzazione, le correlazioni tra i termini), mirando a fornire risposte più pertinenti e rilevanti, che è anche una delle funzioni proprie dei cataloghi.

Mi parve allora indispensabile tornare a quella transizione attraverso le parole di alcuni studiosi, italiani e stranieri, che intorno alla metà degli anni Ottanta fecero sentire la loro voce sulle opportunità offerte dalle nuove tecnologie, sul modo di utilizzarle al meglio in questo ambito dell'attività bibliotecaria, sulle aspettative di miglioramento degli strumenti di catalogazione semantica, sulla progettazione di un nuovo modello di catalogo. Di questo si occupano appunto i primi due paragrafi di questo contributo.

A novembre dello stesso anno, il Congresso dell'AIB affrontò il tema dell'intelligenza artificiale e delle sue applicazioni in biblioteca, con molti sguardi ottimistici e qualche preoccupazione, con i punti di forza, i punti di debolezza, le opportunità e i rischi. Uno degli ambiti di applicazione auspicati è quello della descrizione delle risorse bibliografiche, della sommarizzazione dei testi e della creazione di metadati semantici. In generale, raccolti dal Congresso la raccomandazione di affrontare queste sfide in modo intelligente e critico, con ingegno e sapienza, ricercando in ogni ambito di applicazione la complementarità tra compiti umani e compiti affidati a un sistema di intelligenza artificiale, valutando per ciascun ambito i confini entro i quali l'IA può operare con successo, assumendo come 'determinanti' delle scelte strategiche le competenze professionali bibliotecarie, le risorse occorrenti, le necessità degli utenti.

Ecco profilarsi una nuova transizione, su cui lo sguardo si apre fuggevolmente nell'ultimo paragrafo, con qualche spunto di riflessione sulla natura dei nostri dati e, conseguentemente, sui vincoli e le restrizioni di cui tenere conto nell'applicare ad essi un sistema di intelligenza artificiale.

Questo contributo si muove, dunque, tra due transizioni, con uno sguardo al passato e uno al futuro, entrambi limitati a un unico campo visivo, l'accesso per soggetto alle risorse bibliografiche. I due paragrafi intermedi fanno da ponte, interrogandosi sulla natura dei metadati semantici, sul linguaggio con cui sono costruiti, sul loro rapporto con la tecnologia, sulla loro gestione nei cataloghi, sui principi che possono tornare utili per progettare un efficace accesso per soggetto. In questa parte del testo, si toccano anche alcune nozioni della teoria dell'informazione – *probabilità, varietà, economia, ridondanza* – per le loro implicazioni nell'ambito dell'indicizzazione e della ricerca semantica.

Da tempo sono in atto progetti di trasformazione in *linked data* delle informazioni bibliografiche contenute nei cataloghi, implementando *data model* che si conformano a LRM e utilizzando il sistema di codifica RDF. Di tali modelli, che incidono sull'accessibilità e sullo scambio dei dati nell'ambiente Web, non si occupa questo contributo. Abbattere i muri che separano i cataloghi di biblioteca dal più vasto mondo del Web, uscire dal giardino recintato del catalogo, che può apparire agli utenti di oggi limitato e antiquato – come invita a fare Karen Coyle –, è un compito certamente necessario. Ma non cambia i problemi dell'origine, della natura e della qualità dei dati semantici, del loro significato pragmatico e della loro disponibilità per l'utente nella ricerca. Questi problemi sono anche il filo conduttore delle pagine seguenti.

Il fermento della transizione

Oggi, dopo decenni di progressi della tecnologia dell'informazione, interni ed esterni al mondo delle biblioteche, e di fronte alle sfide dell'intelligenza artificiale, perché occuparci del passaggio dal tradizionale catalogo a schede a quello online? Che utilità può avere, a quasi quarant'anni di distanza? Ho trovato una risposta a queste

domande in due contributi di Karen Coyle: *Catalogs and context* (2016) e *Creating the catalog, before and after FRBR* (2017). Scrive Coyle:

There is a great deal of talk today about the future of the library catalog. There is also ongoing work on developing a 'next generation' library bibliographic data format, possibly based on the model presented by the IFLA study group on the Functional Requirements for Bibliographic Records. There is a tendency, however, to treat the current state of data and catalogs as a starting point. The historical context is, however, much richer¹.

Morale: se guardassimo ai nostri cataloghi da una prospettiva storica, ci renderemo conto che il loro stato attuale non è solo un punto di partenza per ulteriori sviluppi, ma anche il punto di arrivo di un processo del quale possiamo valutare i risultati raggiunti. L'espressione «the historical context is, however, much richer» assume così un valore euristico: è un invito a ripercorrere le principali tappe dello sviluppo dei cataloghi e del loro design, chiedendoci in che misura le biblioteche, in questo ambito, siano state al passo con le nuove tecnologie; soprattutto, se in questa transizione si sia verificata la perdita di qualche importante indicazione riguardante i nostri dati; e, alla fine, se da questa indagine emergano domande utili per il futuro dei cataloghi di biblioteca.

È su questo passaggio che si focalizza l'attenzione di Karen Coyle. Volendo ridurre a estrema sintesi la sua disamina, potremmo riepilgarne la conclusione in questo modo: ciò che di più importante abbiamo perduto sono le indicazioni di contesto dei dati, il contesto nelle sue diverse sfaccettature, come significato, ordinamento, organizzazione, ricerca. Contesti perduti o, per meglio dire, nascosti: infatti, la maggior parte di quelle indicazioni sono tuttora presenti implicitamente nei dati descritti nei cataloghi, ma non sono pienamente visibili e utilizzabili dagli utenti². Per dirlo ancora con le sue parole:

Presumably, libraries are an organized knowledge space. If they weren't the books would just be thrown onto the nearest shelf, and subject cataloging would not exist. However, if this organization isn't both visible and comprehended by users, we are, firstly, not getting the return on our cataloging investment and secondly, users are not getting the full benefit of the library³.

L'avvento dei cataloghi online segna un tempo «of great excitement in information systems design», scrive Marcia Bates⁴. Riandare a quel tempo consente, dunque, di

1 Karen Coyle, *Creating the catalog, before and after FRBR*, September 12th 2017, <<https://kcoyle.net/mexico.html>>.

2 «C'è un catalogo 'occulto', non visibile agli utenti perché non è ricercabile dall'OPAC»: questo catalogo è fatto di elementi che nell'OPAC «sono spesso carenti o negati agli utenti». Cfr. Pino Buizza, *Gli opac: funzionalità e limiti nel mondo del web*, «Bibliotime», 11 (2008), n. 1, <<https://www.aib.it/aib/sezioni/emr/bibtime/num-xi-1/buizza.htm>>.

3 Karen Coyle, *User tasks, Step one*, «Coyle's InFormation», August 31st 2016, <<https://kcoyle.blogspot.com/2016/>>.

4 Marcia J. Bates, *Subject access in online catalogs: a design model*, «Journal of the American society for information science», 37 (1986), n. 6, p. 357-376: p. 357.

cogliere gli effetti che si prevedeva potessero avere i sistemi online sulla natura e l'uso dei cataloghi, le aspettative di arricchimento e facilitazione dell'accesso alle risorse di biblioteca, le implicazioni sul piano progettuale del cambio di prospettiva da cui analizzare l'accesso per soggetto – spostando l'attenzione dal punto di vista del sistema (*system-oriented view*) al punto di vista dell'utente (*user-oriented view*) –, i principi ai quali ancorare un nuovo design del catalogo.

Scavando nella letteratura biblioteconomica, nel corso della sua indagine sulla transizione dal catalogo a schede al catalogo online, Karen Coyle ritrova questo fermento nelle lezioni di Pauline Cochrane del 1984 *Modern subject access in the online age*, pubblicate in sei parti su *American libraries*:

In 1984, Pauline Atherton Cochrane, one of the great thinkers in library land, organized a six-part 'continuing education' to bring librarians up to date on the thinking regarding the transition to new technology⁵.

È da queste lezioni che Coyle prende lo spunto da cui partire per esaminare e valutare il passaggio dal catalogo a schede al catalogo online. In esse, infatti, vi è delineato lo spazio di intersezione di una molteplicità di elementi, che costituiscono, oggi come allora, l'ambiente in cui ha luogo l'accesso per soggetto: l'utente, la tecnologia, la progettazione del sistema informativo e delle interfacce, il loro impatto sui modelli e le regole di catalogazione e indicizzazione. Dice tutto questo, in sintesi e in modo figurato, il disegno che campeggia nella pagina iniziale della prima lezione: un elefante (che rappresenta l'accesso per soggetto), infastidito da alcuni uomini dagli occhi bendati, ciascuno dei quali tocca una parte differente dell'animale, facendosene una percezione limitata; solo dal confronto delle diverse percezioni, ognuno di loro potrà iniziare a conoscere l'intero animale. La nota favola è utilizzata dalla Cochrane come metafora dell'intento che vuole perseguire con le sue lezioni: mostrare la complessità delle operazioni comprese in quello che chiamiamo 'accesso per soggetto' e confrontare i diversi punti di vista al fine di ottenerne una visione d'insieme, necessaria per utilizzare al meglio i progressi tecnologici, cogliendone le sfide, le opportunità e i rischi.

Opportunità:

Loading catalog and index records onto a computer changes them into something other than unit records, because the computer allows access in ways never dreamed of in a card catalog or printed index⁶.

Rischi:

Features of the old catalog that helped make searches more accurate, such as guide cards and references, are missed by the computer searcher who often gets too many or too few search results. Expectations have been raised by new systems, but system design needs some overhauling⁷.

⁵ Karen Coyle, *Catalogs and context*, June 2016, <<https://kcoyle.net/catcon.html>>.

⁶ Pauline A. Cochrane, *Modern subject access in the online age: lesson 3*, «*American libraries*», 15 (1984), n. 4, p. 250-255: p. 250.

⁷ *Ibidem*.

Le aspettative di miglioramento puntano a un design del sistema che garantisca modalità di ricerca caratterizzate da «a redundant but carefully controlled subject access system»⁸. Entra in gioco, a fornire un ricco, vario e controllato apparato di punti di accesso, la cosiddetta struttura sindetica del linguaggio di indicizzazione, costituita dalle relazioni semantiche basate sul significato dei termini (sinonimiche, gerarchiche, associative), da valorizzare nel catalogo online più di quanto non lo fossero nel catalogo a schede. Quali termini per ogni opera indicizzata? Quali altri termini, oltre quelli impiegati per indicizzare il/i soggetti di un'opera? Tutti i termini che compongono ogni stringa di soggetto assegnata all'opera, ciascuno con il collegamento alla stringa di cui fa parte e con i propri 'vedi' e 'vedi anche' originati dalla struttura sindetica.

Sono queste le principali risposte ai problemi incontrati dagli utenti nell'accesso per soggetto, che Cochrane sintetizza così:

The most severe problem appears to be 'finding related words'. The lack of cross reference structure in the file and the disuse of the printed LCSH beside the catalogs are now seen by many to be the focal points for improving subject [...]. As more and more index publishers and libraries use computers to print catalogs or produce online catalogs, the use of pre-formed strings of subject headings (main heading plus subdivisions) with access only by first word has also come into question. This feature, which we listed among the otherwise good features of traditional catalogs and printed indexes, has received the most criticism and is easiest for any computer-based system to 'fix'⁹.

Queste sono anche le principali questioni che si pone Karen Coyle nel suo post *Catalogs and context* del 2016, in particolare, a caratteri cubitali: «What should happen between the search box and the bibliographic display?»¹⁰; o, citando Jeffrey Griffith in *Lesson 5* di Cochrane: «Why should a user ever enter a search term that not provide a link to the syndetic apparatus and a suggestion about how to proceed?»¹¹.

Rendere trasparente all'utente l'apparato sindetico, integrandolo nel catalogo online, farlo funzionare automaticamente nel percorso di ricerca mediante meccanismi chiari e facilmente comprensibili di *selection/switching*, ricorrendo anche a una sorta di «inference engine», grazie all'intelligenza artificiale, sono suggerimenti che ricorrono negli interventi di esperti ospitati da Cochrane nelle sue lezioni. Non sono gli unici. Se ne possono ricavare altri: la ricerca *full text* in linguaggio naturale, da combinarsi con la ricerca per termini assegnati, l'uso di parole chiave e operatori booleani, l'ottimizzazione del dialogo interattivo tra utente e computer, ecc.

⁸ Da un articolo di Marcia J. Bates, *Factors affecting subject catalog search success*, «Journal of the American society for information science», 28 (1977), n. 3, p. 161-169 citato da Cochrane in *Lesson 3* cit., p. 251.

⁹ Pauline A. Cochrane, *Modern subject access in the online age: Lesson 2*, «American libraries», 15 (1984), n. 3, p. 145-149: p. 147.

¹⁰ K. Coyle, *Catalogs and context* cit.

¹¹ Jeffrey C. Griffith, in Pauline A. Cochrane, *Modern subject access in the online age: Lesson 5*, «American libraries», 15 (1984), n. 6, p. 438-443: p. 441.

Qui, tuttavia, si sottolineano soprattutto quei dispositivi che hanno maggiore impatto sul linguaggio di indicizzazione. Quest'ultimo è, infatti, l'altro ambito che offriva ampi margini di miglioramento. Due sviluppi interconnessi: la progettazione del sistema in funzione dell'utente, secondo le nuove possibilità offerte dalla tecnologia, e la modernizzazione del linguaggio di indicizzazione.

In primo luogo, se le nuove strategie di ricerca assegnano all'apparato sintetico un ruolo importante nell'accesso per soggetto, molto più di quanto non avvenisse nel catalogo a schede, allora forse è necessario configurare questo apparato in modo rigoroso, comprensibile e ricco. In questa direzione vanno, per esempio, le molte sollecitazioni a riorganizzare le LCSH in forma di thesaurus, tra cui quella di Charles Hildreth, che prospetta non un cambiamento formale, di facciata, ma una vera e propria ristrutturazione:

However, before a new authority record format can be defined, the call for restructuring LCSH into a hierarchical thesaurus – with broader and narrower terms – must be answered.

Many opportunities for improving our subject vocabulary system would arise during the hierarchical restructuring of LCSH. Conceptual gaps would be identified and new subject terms added. New lead-in terms (*see* references) could be added. Consistency could be achieved in the forms of subject terms and sub-heading patterns. The resulting syndetic structure would provide the related terms and subject vocabulary guidance users of today's online library catalogs are demanding¹².

In secondo luogo, se la tecnologia offre la possibilità di accedere a una stringa di soggetto da ogni termine che ne fa parte, allora anche l'apparato sintattico ha l'occasione propizia per una auspicata modernizzazione, che consenta di superare le strutture rigide delle tradizionali intestazioni di soggetto, in cui i termini non hanno lo stesso valore (voce principale/suddivisioni) né lo stesso grado di controllo e che non assicurano un coerente e prevedibile ordine di citazione. L'auspicio, almeno nelle parole di Toni Petersen, direttore di *Art & architecture thesaurus*, è quello di andare verso una più netta distinzione tra due tipi, differenti ma interdipendenti, di relazioni concettuali: le relazioni semantiche e quelle sintattiche.

Proper definition of the elements of the subject authority record is of primary importance. Subject access is achieved by identifying wanted terminology, combining the terminology into a indexing system, and then connecting the indexing to desired information. Current Library of Congress practice confuses these two elements by treating subject terms and accompanying subdivisions as one unit on an authority record. It is important for future uses of authority records that indexing terms and their authority apparatus be clearly separated from indexing applications like subdivisions¹³.

12 Charles Hildreth, *LCSH needs hierarchical restructuring*. In Pauline A. Cochrane, *Modern subject access in the online age: Lesson 6, «American libraries»*, 15 (1984), n. 7, p. 527-529: p. 529.

13 Toni Petersen, *Subject authority record still confusing*. In Pauline A. Cochrane, *Modern subject access in the online age: Lesson 3 cit.*, p. 255.

Si trattava, in sostanza, di applicare alla soggettazione tradizionale il modello di linguaggio analitico-sintetico, derivante dalle classificazioni a faccette, la cui principale caratteristica è appunto la distinzione fra relazioni paradigmatiche e relazioni sintagmatiche, fra il piano della semantica e quello della sintassi¹⁴.

Ad arricchire i punti di accesso per soggetto nel catalogo online non ci sono solo l'apparato sindetico e la combinazione/scomposizione dei termini che formano una stringa di soggetto. Vi contribuisce anche la possibilità di assegnare all'opera una molteplicità e varietà di soggetti, maggiore di quanto non avvenisse nel catalogo a schede, «developed under conditions of severe resource limitations», superando l'approccio tradizionale caratterizzato da un livello minimo di ridondanza, passando da un regime di 'austerità' a un regime di 'abbondanza'. Infatti, mentre una proliferazione di più di uno o due soggetti per opera avrebbe avuto l'effetto di accrescere le dimensioni del catalogo a schede in misura tale da comprometterne la gestione, «with automation, we have the opportunity to introduce many access points to a given book». La voce di Marcia Bates non è l'unica: «important is to break from the restrictions of the past in assigning subject terms and access points to a document» (Petersen); «what is need is an enrichment and expansion of the subject content in our MARC records [...] based on in-depth, part-by-part indexing of new works» (Hildreth)¹⁵. Sia Petersen che Hildreth aggiungono un punto chiave: questo arricchimento dovrebbe essere il risultato dell'attività cooperativa di tutta la comunità bibliotecaria.

Dunque, anche l'analisi concettuale dei documenti è interessata dal cambiamento indotto dalla tecnologia, così come l'apparato sindetico e la struttura delle stringhe di soggetto. In definitiva, tutti gli ambiti della catalogazione semantica vi sono coinvolti. Tuttavia, la nozione di 'accesso per soggetto' non è riducibile a quelle di catalogazione o indicizzazione, le quali, per restare nella metafora, possono rappresentare la coda o la proboscide dell'elefante, ma non l'intero animale.

Si torna così all'impostazione iniziale. L'ambito in cui si colloca l'accesso per soggetto è un'area di intersezione di una serie di componenti: i bisogni informativi degli utenti, il modello di recupero dell'informazione loro familiare, la biblioteca e le reti di biblioteche, con le loro attività (e prodotti) di catalogazione e indicizzazione, l'infrastruttura informatica. Il nucleo essenziale di quest'area è l'interfaccia utente-sistema: un'interfaccia dinamica, un luogo di dialogo, di comunicazione, di interazione¹⁶. Quest'area presenta due caratteristiche:

14 Il modello analitico-sintetico aveva avuto in PRECIS (Preserved Context Index System) un'applicazione coerente e avanzata sia sul piano concettuale che tecnologico. Il sistema, ideato da Derek Austin, era stato adottato fin dal 1971 dalla *British national bibliography*, dove continuerà a essere operativo fino al 1991, sostituito prima da COMPASS, una versione semplificata del PRECIS, poi nel 1996 da LCSH, che già da tempo aveva assunto la forma di thesaurus («disguised as a thesaurus», per usare l'espressione critica di Mary Dykstra (*LC subject headings disguised as a thesaurus*, «Library journal», 113 (1988), n. 4, p. 42-46). Dal PRECIS a LCSH: vista da una prospettiva storica, la vicenda è singolare ed emblematica di una certa evoluzione del rapporto tra vocabolario e tecnologia nella transizione al catalogo online.

15 M. J. Bates, *Factors affecting subject catalog search success* cit.; T. Petersen, *Subject authority record still confusing* cit.; C. Hildreth, *LCSH needs hierarchical restructuring* cit.

16 All'interfaccia utente nel catalogo in linea è dedicata l'opera di Charles R. Hildreth, *Online public access catalogs: the user interface*. Dublin (Ohio): OCLC, 1982: «Choosing an online catalog is choosing to enter the world of 'human-computer interaction', with all that world's new opportunities, chal-

a) la complessità, dovuta ai diversi elementi che la compongono e alle loro interrelazioni;
 b) la priorità del punto di vista orientato all'utente, nella progettazione dell'interfaccia e del processo di ricerca per soggetto.

Per esemplificare queste caratteristiche, possiamo fare riferimento all'apparato sindetico del Thesaurus del *Nuovo soggettario*. Dal punto di vista del sistema, è uno strumento che guida l'indicizzatore nella selezione dei termini appropriati a esprimere un soggetto. Tutto concorre a questo scopo: definizioni, note d'ambito, note di orientamento, relazioni, equivalenti in altre lingue, fonti, ecc. Dal punto di vista dell'utente, questi stessi dispositivi sono strumenti per la ricerca. La prospettiva è capovolta. Dalla prospettiva dell'utente, per esempio, sinonimi, quasi-sinonimi, varianti grafiche, varianti storiche, ecc., servono ad ampliare il vocabolario di accesso e a facilitare la corrispondenza tra i termini espressi nella *query* e i termini impiegati per descrivere il soggetto. Considerazioni analoghe valgono per le relazioni gerarchiche e associative, per le note di orientamento, ecc., che possono guidare l'utente ad affinare la ricerca iniziale, modificarla, ampliarla, restringerla e a scegliere ulteriori punti di accesso.

Di nuovo, è una questione di design: con quali modalità e procedure si progetta l'accesso per soggetto, facendo sì che il suo apparato sia più facile da usare e più rispondente alle abitudini di ricerca dell'utente? Come integrare nel catalogo online la struttura sindetica e quella sintattica del linguaggio di indicizzazione? Quali elementi del linguaggio si prestano a essere utilizzati per lo *switching* automatico oppure per il *browsing* e/o la *selection* da parte dell'utente? A questo proposito, prima di lasciare le lezioni di Cochrane, vale la pena ascoltare l'auspicio, attuale ancora oggi, di Jeffrey Griffith:

We know what we need to improve subject access in our online catalogs, because we have not yet built into our online systems much of the structure for subject access that is already present in subject cataloging. That structure is internal and known by the person analyzing the work; it needs to be external and known by the person seeking the work¹⁷.

«Costruiamo strumenti informativi secondo le nuove possibilità»

È l'invito di Luigi Crocetti alla comunità bibliotecaria italiana in occasione del 33° Congresso AIB (Sirmione 1986): una specie di viatico per la transizione dal catalogo a schede al catalogo online, e non solo. Questo invito scaturisce da un ragionamento il cui nucleo centrale è contenuto in un passaggio della sua introduzione ai lavori congressuali, nel quale possiamo avvertire, distinto, l'eco delle parole di Marcia Bates

lenges, and alas, problems» (p. 33); «The telling difference between an online catalog and a card catalog does not lie in the area of bibliographic description, control, or record storage, but rather in the way the user interacts with and is assisted by the online catalog. [...] As interactive computer systems, online catalogs add an entirely new dimension to catalog use [...]. In terms of process, this new dimension is referred to as human-computer interaction or man-machine communication. In structural terms it is known as the user-system interface or man-machine interface component of the computer system or environment» (ibidem); «Implicit in our general definition of 'interface' is the notion that the user, whether an intermediary or a direct user, is an integral component of the system. This view of the user as an active, creative and productive participant within the system itself [...] has major consequences for the design of interactive systems, as well as for our evaluation of systems currently in use» (p. 34).

17 J. C. Griffith, in P. A. Cochrane, *Modern subject access in the online age: Lesson 5* cit., p. 441.

sulla condizione di «resource limitations» dei cataloghi tradizionali. La citazione non è breve, ma temo che a riassumerla si corra il rischio di sottovalutarne il peso:

I bibliotecari si rendono conto che non è cambiato l'insieme delle loro idee, ma sta cambiando – e in maniera rapidissima – il tessuto connettivo che le stringe l'una all'altra. Il processo che aveva portato a costruire una serie di procedure e di norme concatenate si fondava in gran parte sul riconoscimento di stati di fatto, sul riconoscimento di necessità, per così dire, 'economiche'. Ora, fa parte del concetto di 'economia' e dell'aggettivo 'economico' la nozione di 'limitatezza delle risorse'. [...] In molti dei settori della nostra attività le nuove tecnologie non dirò che cancellano quella nozione, ma certamente ne attenuano grandemente il peso. Molte delle nostre procedure e molte delle nostre norme non hanno più ragione di essere.

Ed ecco un compito fondamentale del bibliotecario: adeguare la mente alle nuove condizioni, discriminare tra procedure e norme che rispondono a reali necessità conoscitive e informative e norme, procedure o abitudini che sono derivate soltanto da situazioni di fatto, da considerazioni di carattere pratico, di carattere empirico, come quelle norme che sono state autorevolmente definite 'illogiche'¹⁸.

Con queste parole Crocetti suggerisce un metodo e indica una direzione di lavoro. Che cosa implica adeguare la mente alle nuove condizioni? Innanzitutto, un lavoro di discriminazione tra procedure/norme che rispondono a «reali necessità conoscitive e informative» e procedure/norme «che non hanno più ragione di essere», perché illogiche, empiriche, dettate da situazioni di fatto, da condizioni di limitazione delle risorse. Ne consegue un duplice lavoro: da un lato, l'eliminazione di illogicità e incompatibilità, rinuncia ad abitudini e idiosincrasie; dall'altro, la ridefinizione delle procedure/norme in funzione degli scopi che esse vogliono raggiungere, la loro spiegazione in termini di ragioni, di motivazioni, l'esplicitazione del loro significato pragmatico¹⁹. Un lavoro «faticoso, remunerante, bellissimo», da cui scaturirà una nuova concatenazione di procedure/norme adeguata alle «reali necessità» e alle possibilità offerte dalle tecnologie. Schematizzando, possiamo descrivere il processo di una nuova concatenazione con questa sequenza: possibilità tecnologiche, impatto delle possibilità tecnologiche sulla catena mezzi-fini, adeguamento delle procedure/norme e degli strumenti alle possibilità tecnologiche.

Esempi di questa concatenazione si sono incontrati nel precedente paragrafo, relativamente a tutte le attività finalizzate all'accesso per soggetto. Tra quelli scelti da Crocetti troviamo la molteplicità dei punti di accesso («non ci si può ostinare a

18 Luigi Crocetti, *Introduzione e conclusione del 33° Congresso (Sirmione 1986)*. In *Id.*, *Le biblioteche di Luigi Crocetti: saggi, recensioni, paperoles (1963-2007)*, a cura di Laura Desideri e Alberto Petruccianni; presentazione di Stefano Parise. Roma: Associazione italiana biblioteche, 2014, p. 214-219; p. 215-216.

19 Sul legame tra la regola e lo scopo ritornerà più volte Crocetti. Qui mi limito a richiamare il suo ammonimento in occasione del Seminario "L'automazione delle biblioteche nel Veneto: tra gli anni '90 e il nuovo millennio" (Venezia, Palazzo Querini Stampalia, 29-30 gennaio 1999): «La biblioteconomia fornisce gli strumenti per maneggiare la cultura: è una disciplina tecnica, ma guai se i suoi principî cominciano a svilupparsi autonomamente, a dedursi l'uno dall'altro, senza tener conto dei cambiamenti culturali e delle necessità del pubblico» (L. Crocetti, *La tradizione culturale italiana del Novecento*. In *Id.*, *Le biblioteche di Luigi Crocetti cit.*, p. 501-507; p. 507).

disquisizioni di sottigliezza rara quando è possibile che tutto, o quasi tutto, diventi accesso; questa parte dei codici di catalogazione appare ormai arcaica») e l'opportunità, grazie alla tecnologia, di tenere sempre aggiornato l'indice di un catalogo classificato, allineandolo ai cambiamenti che si verificano nello schema di classificazione. Crocetti si riferisce alle classi della DDC, ma il discorso è estendibile anche al catalogo per soggetti: il linguaggio cambia nel tempo e con esso inevitabilmente anche il vocabolario di indicizzazione; i termini impiegati precedentemente per esprimere un determinato concetto e non più in uso ('variabili storiche') possono essere allineati automaticamente ai termini equivalenti in uso, grazie alla tecnologia, senza bisogno di correggere le relative stringhe di soggetto, ottenendo da qualsiasi termine di partenza un identico risultato in quanto a richiamo e precisione.

Aggiungo un altro esempio, tratto anche questo dal versante dell'indicizzazione per soggetto. L'automazione dei cataloghi ha reso possibile l'accesso a una stringa di soggetto da tutti i termini significativi della stringa, indipendentemente dalla loro posizione nell'ordine della sequenza e dal loro ruolo sintattico. Il beneficio che ne è derivato è una moltiplicazione dei punti di accesso a un soggetto. Non è l'unico: da un termine si può accedere a tutte le stringhe di soggetto in cui il termine è presente. Nel primo caso, a trarne beneficio è il grado di richiamo; nel secondo, il grado di precisione. Ne consegue un nuovo design delle procedure riguardanti l'accesso e il *browsing*. Ma «un'innovazione puramente tecnica, o presentabile come tale, raramente resta, in biblioteca, confinata nel suo ambito originario»²⁰. Infatti, il cambiamento concettuale derivante da quella innovazione tecnica ha investito il linguaggio di indicizzazione in tutti i suoi aspetti. Quando, prima dell'automazione, l'accesso a un soggetto era possibile solo dal primo termine della stringa, questa condizione di fatto, dovuta ai limiti del catalogo cartaceo, costringeva a scegliere uno dei concetti del soggetto come più importante, più significativo, più rappresentativo, assegnandogli una funzione predominante ('voce principale') e collocando il termine corrispondente in prima posizione nella stringa. Gli altri concetti venivano relegati in una posizione subalterna ('suddivisioni'), non direttamente accessibili, ma reperibili solo a partire dalla voce principale. Del resto, non si poteva fare altrimenti, data una stringa lineare in un catalogo a schede. Venuto meno questo limite, tutti i concetti di un soggetto hanno acquisito pari importanza sia ai fini dell'accesso che della descrizione del soggetto. Questo ha consentito, da un lato, di elaborare nuove norme per la costruzione delle stringhe di soggetto, in funzione di una maggiore uniformità, precisione, flessibilità nella descrizione dei soggetti; dall'altro, di considerare ogni termine della stringa come un termine di indicizzazione a pieno titolo, acquistando la titolarità di quelle funzioni di controllo e di organizzazione proprie di un vocabolario di indicizzazione. Ne consegue che a un soggetto l'utente potrà accedere non solo da tutti i termini di una stringa, ma anche dai loro sinonimi, varianti storiche, iponimi, iperonimi, ecc. Alla fine di questo processo troveremo la separazione del linguaggio nei due ambiti, distinti ma complementari, della semantica (apparato sintetico) e della sintassi (strutture sintattiche). Questo approccio, inoltre, ha rimescolato le carte nel rapporto tra pre- e post-coordinazione: nella forma sintetica di pre-coordinazione, l'utente può beneficiare dei vantaggi di entrambe. Infine, volendo estendere fino all'oggi questa concatenazione, potremmo comprenderci il requisito della 'granularità' dei dati, funzionale, oltre che all'accesso ai dati, anche a una loro rappresentazione secondo i modelli descrittivi del Web.

20 L. Crocetti, *Il nuovo in biblioteca*. In Id., *Le biblioteche di Luigi Crocetti* cit., p. 24-28: p. 25.

Come mostra l'esempio, la concatenazione non è solo di procedure e norme, ma anche di mezzi e fini. La ricerca dell'accordo tra mezzi e fini è il presupposto, la premessa del ragionamento di Crocetti: non basta affermare che le tecnologie sono un mezzo per raggiungere determinati fini, uno strumento per realizzare determinate idee; occorre aggiungere che i fini, le idee devono essere in armonia col mezzo²¹. Crocetti accompagna questa raccomandazione con la metafora dell'arma:

A me sembra che il risultato generale che si può trarre dai nostri lavori è che noi abbiamo ora davanti una nuova arma per fare alcune cose che abbiamo sempre voluto fare, cui abbiamo sempre aspirato anche se non riuscivamo a realizzarle; di questo ci rendiamo conto, ma un'arma è solo un'arma²².

Un'arma 'a doppio taglio'. Quando Crocetti avverte che si stia profilando un affievolirsi di quella consonanza, allora la sua critica si fa severa. «I bibliotecari sono in torto [...] quando si pongono in maniera sbagliata di fronte all'automazione», ossia quando «non comprendono i meccanismi (forse meglio dire i percorsi) e gli scopi della ricerca e dello studio»²³. Tanto da fargli dire come l'applicazione dell'elaboratore alle biblioteche «abbia sospinto indietro, abbia costretto ad arretrare conoscenze e metodi», contrariamente a quanto avvenuto in altri campi e discipline²⁴. D'altro canto:

Abbiamo ora strumenti potentissimi che nessun bibliotecario delle età passate ha avuto a disposizione; e mi sembra che questi strumenti stiano insegnando molto, almeno ai bibliotecari delle generazioni più giovani. Quando si ha a disposizione un cannone, invece di una pistola, è inevitabile chiedersi se sia conveniente adoperarlo contro le zanzare o se non convenga indirizzarne la mira a bersagli più degni e anche, diciamo pure, più ambiziosi²⁵.

Mezzi e fini hanno, inoltre, una duplice estensione: i mezzi non sono solo la tecnologia, l'automazione, ma anche la cooperazione tra le biblioteche; i fini non sono solo le idee, ma anche la loro realizzazione nei servizi. Su questi due temi, cari a Crocetti, non c'è bisogno di soffermarci.

Se la riflessione di Crocetti offriva alla transizione dal catalogo a schede al catalogo online una guida metodologica e deontologica, sul piano tecnico dell'accesso per soggetto un'idea del fermento della transizione in Italia, a metà degli anni Ottan-

21 Questa correlazione tra fini e mezzi è descritta efficacemente da Weston: «Le finalità debbono essere tenute distinte dai mezzi utilizzati per il loro conseguimento: ciò è tanto più importante quanto maggiori e più rapidi sono i cambiamenti a carattere tecnologico, perché vi è il rischio concreto che, se da un lato è la disponibilità del mezzo a dettare le finalità, dall'altro lato non si dia sufficiente importanza alla necessità di reinterpretare e riaffermare le finalità alla luce dei cambiamenti in atto, finendo per usufruire solo in parte e male delle potenzialità dei nuovi strumenti». Cfr.: Paul Gabriele Weston, *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*, postfazione di Giovanni Solimine. Roma: Carocci, 2002, p. 29.

22 L. Crocetti, *Introduzione e conclusione del 33° Congresso (Sirmione 1986)* cit., p. 217.

23 L. Crocetti, *Biblioteche e archivi*. In Id., *Le biblioteche di Luigi Crocetti* cit., p. 64-72: p. 69.

24 *Ivi*, p. 71.

25 L. Crocetti, *La tradizione culturale italiana del Novecento* cit., p. 507.

ta, si può ricavare dagli atti del Convegno "Indicizzazione per soggetto e automazione" (Trieste, 21-22 ottobre 1985), nel quale, scrivono i curatori,

si è cercato di dare spazio al più ampio numero possibile di proposte di soluzione, anche perché spesso il catalogatore, abituato a una metodologia tradizionale, non sempre vede chiaramente le possibilità e i vantaggi offerti al suo lavoro dalle nuove tecnologie informatiche²⁶.

Vi si ritrovano argomentazioni e proposte in gran parte analoghe a quelle che abbiamo già incontrato nelle lezioni della Cochrane. Non le ripeterò qui. Piuttosto, cercherò di indicare schematicamente alcuni punti fermi enucleabili dal Convegno, unendo i quali si possa ottenere il profilo di uno strumento o più strumenti progettati «secondo le nuove possibilità»:

1. Design del catalogo

- il catalogo serve agli utenti, permettendo loro di trovare le informazioni che cercano nel modo più semplice e più rapido possibile;
- la tecnologia può aiutarci a evolvere dal catalogo orientato al bibliotecario al catalogo orientato all'utente;
- l'interfaccia utente creata dal software dovrebbe disporre di meccanismi come lo *switching*, il *browsing* e il *dialogue*, attraverso i quali l'utente sia messo in grado di utilizzare correttamente gli strumenti a disposizione, di orientarsi nella scelta, di esplorare le risorse informative del catalogo;
- la creazione dell'interfaccia utente potrà avvalersi delle tecniche dell'intelligenza artificiale;

2. Accesso e ricerca per soggetto

- l'accesso avviene mediante parole chiave, intese come le parole in linguaggio naturale impiegate dall'utente come chiavi di ricerca;
- tra le parole dell'utente e i termini di indicizzazione opera automaticamente un dizionario formato da sinonimi, quasi-sinonimi, varianti grafiche, varianti storiche, ecc., il cui scopo è quello di aumentare le probabilità di corrispondenza tra le parole chiave e i termini di indicizzazione, attraverso la 'traduzione' delle prime nei secondi, facilitando così l'accesso alle risorse;
- i punti precedenti valgono anche in caso di accesso alle classi di un catalogo classificato;
- prima di recuperare le risorse, l'utente ha la possibilità di proseguire la ricerca ampliando, restringendo o modificando l'area concettuale di interesse;
- e/o di accedere dal termine a una lista di stringhe di soggetto oppure alle classi di uno schema di classificazione, entrambe collegate al concetto su cui s'indirizza la ricerca dell'utente, con possibilità di scorrimento della lista alfabetica o della catena gerarchica;

3. Sistema di indicizzazione

- analisi concettuale 1: pur non disconoscendo il criterio della sommarizzazione specifica, si auspica un incremento del numero dei soggetti per documento, sulla base di un'analisi più approfondita, che non si limita al contenuto concettuale del docu-

26 *Il recupero dell'informazione: atti del Convegno - Esposizione bibliografica "Indicizzazione per soggetto e automazione", Trieste, 21-22 ottobre 1985*, a cura di Adriano Dugulin, Antonia Ida Fontana, Annamaria Zecchia. Milano: Editrice Bibliografica, 1986, p. 15.

mento nel suo insieme, ma si focalizza anche su parti e aspetti (distinzione tra libro/unità bibliografica e testo/opera/unità letteraria);

- analisi concettuale 2: l'utilizzo di fonti supplementari sul contenuto di soggetto dei documenti, come titoli, sommari, abstract, ecc., consente di aggiungere alla 'indicizzazione assegnata' una 'indicizzazione derivata' direttamente da queste fonti;
- analisi concettuale 3: il riferimento allo standard ISO 5963:1985²⁷ può contribuire a promuovere procedure uniformi, che prescindono in larga misura dallo specifico sistema di indicizzazione adottato;
- thesaurus 1: lo strumento di indicizzazione più idoneo (anche se non l'unico) a rispondere ai suddetti requisiti di accesso e di ricerca per soggetto è il thesaurus: un thesaurus 'generale' del Servizio bibliotecario nazionale, costruito in cooperazione tra le biblioteche sotto la responsabilità dell'agenzia bibliografica nazionale;
- thesaurus 2: le funzioni del thesaurus: a) controllo terminologico per la scelta dei descrittori; b) preparazione del dizionario dei termini di accesso; c) impiego di un unico vocabolario normalizzato nell'ambito di linguaggi diversi (soggettazione e classificazione), favorendo lo sviluppo della cooperazione; d) struttura generale di riferimento per lo sviluppo di thesauri specializzati;
- thesaurus 3: riferimento allo standard ISO 2788:1986²⁸ come base per la normalizzazione del vocabolario;
- un'attenzione particolare è da riservare alle classificazioni, come strumento di ricerca nel catalogo online, e al sistema PRECIS, come modello di analisi e indicizzazione;
- entrambi, lo standard ISO 2788:1986 e PRECIS, presuppongono un modello di linguaggio cosiddetto 'analitico-sintetico', basato sulla distinzione tra relazioni paradigmatiche e relazioni sintagmatiche.

Questi sono i punti, certo non esaustivi, da cui partire per rinnovare procedure e norme e su cui progettare un nuovo modello di catalogo. Non esaustivi né analiticamente descritti: in questa specie di inventario si perdono, infatti, le singole voci del Convegno, le loro differenze, le argomentazioni, gli approfondimenti, i dettagli, la distinzione tra proposte ed esperienze in corso di realizzazione, ecc.

Un solo esempio, relativo all'intelligenza artificiale: ben due relazioni sono dedicate a questo tema, una sulla sommarizzazione automatica dei testi, l'altra sulla creazione di un'interfaccia intelligente²⁹.

27 International Organization for Standardization, ISO 5963:1985 – *Documentation. Methods for examining documents, determining their subjects, and selecting indexing terms*. Geneva: ISO, 1985, current version, reviewed and confirmed in 2020 (versione italiana: Ente nazionale italiano di unificazione, ISO 5963:1989 – *Documentazione. Metodi per l'analisi dei documenti, la determinazione del loro soggetto e la selezione dei termini di indicizzazione*. Milano: UNI, 1989).

28 International Organization for Standardization, ISO 2788:1986 – *Documentation. Guidelines for the establishment and development of monolingual thesauri*. Geneva: ISO, 1986; versione corrente, in due parti: International Organization for Standardization, ISO 25964-1:2011 – *Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*. Geneva: ISO, 2011 (rivista e confermata nel 2017); Id., ISO 25964-2:2013 – *Information and documentation. Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. Geneva: ISO, 2013 (rivista e confermata nel 2018).

29 Danilo Fum, *SUSY: un sistema per la sommarizzazione automatica di testi in linguaggio naturale*. In: *Il recupero dell'informazione cit.*, p. 91-113; Carlo Tasso, *IR-NLI: un'interfaccia intelligente a sistemi di documentazione online*. In: *Il recupero dell'informazione cit.*, p. 114-126.

Dovremmo, inoltre, allargare un po' lo sguardo per cogliere il fermento della transizione in un pullulare di interventi ed esperienze di quegli anni, a partire dallo stesso Congresso di Sirmione³⁰; o proiettarlo in avanti, per seguire l'onda lunga della transizione. Faccio anche qui un solo esempio, relativo alla costruzione delle stringhe di soggetto, tema rimasto un po' in ombra nel Convegno (a parte la relazione introduttiva di Carlo Revelli³¹). Qualche anno dopo, il GRIS elaborerà un progetto di ricerca proprio sulla sintassi dell'indicizzazione per soggetto, dal quale prese avvio il lavoro che ha portato alla realizzazione della *Guida all'indicizzazione per soggetto*³².

Vocabolario e tecnologia

Due caratteristiche si possono evidenziare nell'inventario di proposte e aspettative elencate nel paragrafo precedente, esprimibili mediante due coppie di nozioni: 'parole e contesti', 'vocabolario e tecnologia'.

Le parole usate dagli utenti e i termini impiegati dall'indicizzatore attivano una varietà di contesti determinata dalle loro relazioni con altri termini, sia sul piano paradigmatico che su quello sintagmatico. I contesti erano presenti anche nel catalogo a schede, ma le opportunità offerte dall'automazione generavano l'aspettativa di una loro valorizzazione a vantaggio dell'utente. I meccanismi interattivi del catalogo online (*switching, browsing, dialogue*) avrebbero reso disponibile per l'utente ciascun tipo di relazione, favorendo la corrispondenza tra parole chiave e termini di indicizzazione e la modulazione della ricerca in funzione del richiamo e della precisione.

Possiamo affermare oggi che quelle aspettative si siano realizzate? In che misura? Si è verificata un'effettiva valorizzazione degli apparati sintetici e delle strutture sintattiche a vantaggio degli utenti? Abbiamo forse ritenuto che l'uso legittimo e appropriato delle parole chiave da parte dell'utente li rendesse non necessari per la ricerca? Non è piuttosto vero il contrario? Se le principali prerogative del catalogo elettronico sono ravvisabili nell'*intertestualità* e nella *navigazione*³³, perché ben poco si è fatto per integrare nel catalogo i collegamenti dizionari, enciclopedici e multilinguistici di cui sono dotati i termini del Thesaurus del *Nuovo soggettario*? Gli strumenti di indicizzazione, nonostante le innovazioni e i miglioramenti apportati, non sono forse rimasti primariamente istruzioni per gli indicizzatori piuttosto che dispositivi per la ricerca, proprio quando la tecnologia avrebbe potuto aiutarci a evolvere dal catalogo orientato al bibliotecario al catalogo orientato all'utente?

30 *Il futuro delle biblioteche: atti del 33° Congresso nazionale dell'Associazione italiana biblioteche, Sirmione, 8-11 maggio 1986*, a cura di Giuseppe Origgi e Gianni Stefanini. Roma: Associazione italiana biblioteche, 1987. Si veda anche: «Bollettino d'informazioni AIB», 26 (1986), n. 4; Michael S. Malinconico, *Intervista a Michael Malinconico*, «Bollettino d'informazioni AIB», 27 (1987), n. 1.

31 Carlo Revelli, *Il catalogo per soggetti e le aspettative dei bibliotecari nei confronti dell'automazione*. In: *Il recupero dell'informazione* cit., p. 27-53.

32 Associazione italiana biblioteche, GRIS – Gruppo di ricerca sull'indicizzazione per soggetto, *Guida all'indicizzazione per soggetto*. Roma: Associazione italiana biblioteche, 1996, rist. 2001. Il progetto sulla sintassi fu presentato al Seminario della Commissione nazionale AIB "Informazione e documentazione", Roma, 4 giugno 1990: Alberto Cheti, *Problemi di sintassi dell'indicizzazione per soggetto*. In: *Informazione e documentazione*, a cura di Vilma Alberani. Roma: Associazione italiana biblioteche, 1990, p. 24-32.

33 P. G. Weston, *Il catalogo elettronico* cit., p. 135 e seguenti; Carlo Bianchini, *Futuri scenari: RDA, REL-CAT e la granularità dei cataloghi*, «Bollettino AIB», 50 (2010), n. 3, p. 219-238.

Domande analoghe, e altre ancora, concludono il post di Karen Coyle su *Catalogs and context*. Nell'altro post *Creating the catalog, before and after FRBR*, Coyle afferma che «it's all technology (except the users)» e che «the aspect of design that is not technology is the human users. These users and their needs should be the primary focus of any standards». E aggiunge, a proposito dei nuovi standard:

Both FRBR and RDA make statements that they are 'technology neutral' – but they are not. FRBR made use of the concepts of database management technology of the 1980's and 1990's and that influenced its design. [...] Unless modelers are aware of the technology behind their designs, and the assumptions they are using for development, they risk being technology ignorant³⁴.

In realtà, questa «neutralità dal punto di vista tecnologico» non si comprende se non si fa intervenire, nell'ambito dell'analisi per soggetto, un'altra componente, il vocabolario. Il modello che sta dietro a FRBR non è un modello tecnologico, ma di analisi concettuale. Per quanto riguarda in particolare l'indicizzazione per soggetto, il modello che a partire dagli anni '30 del secolo scorso informa, più o meno direttamente, i linguaggi di indicizzazione è conosciuto come 'modello analitico-sintetico'.

Semmai ce ne fosse bisogno, per dimostrare questa 'neutralità', si potrebbe ricorrere a una prova documentale: il noto diagramma con cui si distinguono le relazioni paradigmatiche e sintagmatiche, che è alla base del modello analitico-sintetico, lo troviamo riproposto nella prima edizione (1986) e nell'ultima (2011) dello standard ISO sui thesauri, e prima ancora nel manuale del PRECIS (1973), poi nella Guida GRIS (1996) e nella Guida al *Nuovo soggettario* (2021), per citare solo alcuni riferimenti. Dietro questo schema ci sono almeno settant'anni di ricerca sul linguaggio di indicizzazione, a partire da quella del Classification Research Group (CRG) negli anni '50 e '60: se, da un lato, la teoria dei 'livelli integrativi' forniva le basi per analizzare e organizzare i concetti sul piano paradigmatico, dall'altro, la teoria dei 'sistemi' suggeriva un metodo per analizzare e ordinare i concetti di un soggetto sul piano sintagmatico. In Italia, poi, veniamo da un'esperienza quarantennale di costruzione di un sistema di regole in questo campo: dagli studi, le proposte, le sperimentazioni degli anni '80 alla *Guida GRIS* e al *Nuovo soggettario*: un lavoro collettivo, collaborativo, un processo caratterizzato da una sorprendente continuità di intenti, di riferimenti scientifici e culturali, di approcci metodologici.

Oppure, potremmo limitarci a un'osservazione banale. Talvolta, ci si riferisce alla pre-coordinazione, ossia alla costruzione delle stringhe di soggetto, come a un retaggio del catalogo tradizionale, a una metodologia tecnologicamente inadeguata, orientata al bibliotecario, a fronte del mondo tutto post-coordinato e orientato all'utente di Internet³⁵. Ebbene, non facciamo in tempo a digitare su Google la parola chiave

34 K. Coyle, *Creating the catalog, before and after FRBR* cit.

35 Cfr. P. G. Weston, *Il catalogo elettronico* cit. (in particolare, il Cap. 5: *Le prospettive del catalogo elettronico e la biblioteca digitale*). Weston riconduce la distinzione tra pre- e post-coordinazione, in *information retrieval*, all'opposizione tra due coppie di nozioni: 'termini di un vocabolario controllato/correlazione dei termini in fase di indicizzazione' (pre-coordinazione) e 'termini ricavati direttamente dal documento/correlazione in fase di ricerca' (post-coordinazione). In realtà, è solo il secondo elemento di ciascuna coppia che distingue i due metodi. Inoltre, come osserverò nel prossimo paragrafo, il metodo della pre-coordinazione non contrasta con la necessità della 'varietà' delle «opzioni a disposizione del lettore e soprattutto l'attivazione di tutta la rete di collegamenti reciproci» (p. 137), giustamente auspicata da Weston.

che subito ci vengono proposte (*Google Suggest*) parole chiave composte da più termini (*keyword long tail*), che esprimono argomenti più specifici e tendono ad avere minori risultati. Lo stesso avviene in fondo alla pagina dei risultati (SERP, *Search Engine Results Pages*). Non interessa qui stabilire l'origine di questo tipo di parole chiave, ma sottolinearne l'affinità (per natura, funzione e risultato) con le stringhe di soggetto: sequenze di più parole, che individuano temi più specifici e che tendono a risultati con minore grado di richiamo. La 'coda lunga' (*long tail*) è la rappresentazione grafica del rapporto inverso tra richiamo e precisione, che abbiamo studiato sui testi di biblioteconomia³⁶. È il caso di tenerlo presente quando ci si appella al *google like*.

Dunque, nel caso delle stringhe di soggetto, non si tratta necessariamente di un dispositivo antiquato, tecnologicamente vecchio stile, e la sua affinità con quanto suggeriscono i motori di ricerca ha basi logiche e linguistiche, non certo tecnologiche.

In conclusione, il significato del modello analitico-sintetico, nel quale si riflettono le necessità degli utenti, prescinde dalla tecnologia, anche se non la sua implementazione. Potremmo rifare la storia del 'catalogo per soggetti' seguendo nel tempo i modi in cui la tecnologia ne ha supportato il design secondo questo modello: dal tradizionale catalogo per soggetti all'indice di un catalogo classificato, dal *subject index* in volume della *British national bibliography* ai primi cataloghi automatizzati e alle diverse generazioni di OPAC sviluppatesi in questi ultimi decenni.

E oggi? Per restare nell'ambito delle stringhe di soggetto, dobbiamo chiederci se le modalità con cui sono utilizzate nei cataloghi ne valorizzano il ruolo nel processo di ricerca³⁷; se le applicazioni tecnologiche forniscano, a partire da un termine di indicizzazione, una lista di stringhe ben formata (quali stringhe), ben ordinata (in

36 Si confronti la curva richiamo-precisione che compare a p. 42 in Antony C. Foskett, *Il soggetto*. Milano: Editrice bibliografica, 2001 con quelle presenti in qualsiasi manuale SEO (per esempio, Andrea Minini, *La Long Tail o coda lunga*, <<https://www.andreaminini.com/seo/la-long-tail-o-coda-lunga>>).

37 Un esempio di conflitto tra vocabolario e tecnologia è dato dal modo in cui è effettuato il collegamento tra un termine del Thesaurus del *Nuovo soggettario* e le relative stringhe di soggetto presenti nell'OPAC. Ogni termine del Thesaurus è definito precisamente nel suo significato da un ricco apparato sindetico, che non lascia spazio all'ambiguità; note di orientamento rinviano, quando opportuno, ad altri significati del termine rappresentati da termini alternativi; varianti e note storiche riequilibrano il vocabolario sul piano diacronico; le fonti forniscono il significato dizionario e l'uso enciclopedico del termine; equivalenti in altre lingue collegano il termine ad altri termini presenti in strumenti di indicizzazione di altri Paesi, ecc. Potremmo dire che c'è una grande *ridondanza*, che è l'altra faccia dell'*economia* di un vocabolario controllato. Qui s'incontra un primo punto critico: a che serve questa ridondanza se non a facilitare la comunicazione tra utente e catalogo, a rendere più probabile la corrispondenza tra parole chiave e termini di indicizzazione, a ridurre il 'rumore'? Questa ridondanza sarebbe davvero uno spreco se non si rispecchiasse nel catalogo. Nel Thesaurus è data la possibilità di accedere dal termine alle stringhe di soggetto, in alternativa alla visualizzazione diretta delle opere, per selezionare contesti sintagmatici più precisi, più pertinenti alla ricerca. Ci si aspetta che questa espansione sia coerente con il significato del termine. L'informazione cercata, infatti, non riguarda il suo significato, la cui ridondanza è massima, ma i soggetti specifici nei quali quel concetto è trattato, non la sua interpretazione semantica ma la sua contestualizzazione sintattica. Per esempio, partendo dal termine del Thesaurus 'fede' dovrebbero essere recuperate le stringhe di soggetto che trattano della fede in contesti specifici: la fede nei suoi rapporti con la cultura, la filosofia o la ragione; il tema della fede nella Bibbia, in un determinato Vangelo o nei documenti conciliari; la concezione della fede di Dante, di Hegel o di Rosmini; la dottrina pontificia della fede, la sua interpretazione psicanalitica o la sua trattazione dal punto di vista della teologia cristiana, ecc. Resterebbe poi da stabilire in

quale sequenza), facilmente consultabile (con quale interfaccia)³⁸, tale da adempiere efficacemente alle funzioni del *browsing*, offrendo una visione d'insieme e, al tempo stesso, articolata del campo di indagine³⁹.

Scriva Antonio Scolari:

Anche l'approccio e l'uso dei soggetti da parte degli utenti è con tutta probabilità da ripensare, proprio perché si chiede alle biblioteche e ci si aspetta dalle biblioteche, che i loro dati siano da questo punto di vista certificati, rispetto a quelli della rete. Ma forse è anche il concetto di accesso semantico che va espanso:

quale ordine presentarle nella lista, in modo da facilitarne il *browsing*. La tecnologia applicata attualmente, invece, effettua una ricerca per parole e sequenze di caratteri, recuperando da 'federe' anche stringhe contenenti termini di indicizzazione, parole o parti di nomi propri – come 'fedeltà', 'federalismo', 'buona fede', 'confederazione', 'federale', 'Federico', ecc. – incoerenti con il significato di 'fede', scoraggiando così la consultazione dell'elenco delle stringhe di soggetto (che in maggioranza risultano non pertinenti) e vanificando il loro ruolo nell'atto del *browsing*. Quella procedura mira probabilmente a evitare risposte nulle, ma è del tutto inappropriata in questo contesto. A ciò si aggiunge la limitazione contenuta nell'avvertenza che compare nell'OPAC in testa alla lista delle stringhe: «È possibile ricercare un solo termine di ricerca. Per ricerche con più termini vai alla Ricerca avanzata», che, essendo quest'ultima una ricerca di documenti, non consente di affinare la ricerca di stringhe, che servirebbe in caso di un numero elevato di risultati pertinenti. Naturalmente, non è il Thesaurus a essere difettivo, ma il suo collegamento all'OPAC.

38 Le forme di presentazione delle stringhe di soggetto devono essere efficaci dal punto di vista della comprensibilità, come suggerisce Andrea Fabbrizzi: «La comprensibilità delle descrizioni e degli indici che permettono l'organizzazione e il recupero dei dati catalografici, troppo spesso trascurata dai bibliotecari, non è un dettaglio: al contrario, è la condizione indispensabile per rendere realmente utile il lavoro dei catalogatori» (Andrea Fabbrizzi, *I contesti dell'indicizzazione*. In: *Viaggi a bordo di una parola: scritti sull'indicizzazione semantica in onore di Alberto Cheti*, a cura di Anna Lucarelli, Alberto Petrucciani, Elisabetta Viti, presentazione di Rosa Maiello. Roma: Associazione italiana biblioteche, 2019, p. 77-89: p. 80). Cfr. anche Agnese Galeffi, *Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica*, «AIB studi», 57 (2017), n. 2, p. 239-252.

39 Il processo di *browsing* è così descritto da Marcia Bates: «browsing can be seen to contain four elements, iterated indefinitely, until the overall episode ends: 1. glimpsing a field of vision; 2. selecting or sampling a physical or representational object from the field; 3. examining the object; and 4. physically or conceptually acquiring the examined object, or abandoning it». Cfr. Marcia J. Bates, *What is browsing really? A model drawing from behavioural science research*, «Information research», 12 (2007), n. 4, <<http://InformationR.net/ir/12-4/paper330.html>>. Essenziale, dunque, è la possibilità di acquisire un campo visivo che consenta di prendere visione di una serie di oggetti potenzialmente interessanti, di esaminarli e compiere poi i passi successivi. Una lista di stringhe di soggetto incoerenti e male organizzate non è in grado di soddisfare i requisiti del *browsing*. Analogamente, Thomas Mann considera una lista di stringhe di soggetto, visualizzata mediante la modalità di ricerca *Subject Browse* (nel catalogo della Library of Congress), come una 'mappa concettuale' che «permette di farsi un'idea non solo della 'configurazione' dell'argomento, ma anche della 'profondità' delle sue diverse parti»; la lista dovrebbe fornire agli studiosi «un quadro complessivo della configurazione e della profondità della letteratura che li riguarda», «un quadro sia comprensivo che comprensibile». Cfr. Thomas Mann, *Il catalogo e gli altri strumenti di ricerca: un punto di vista dalla Library of Congress*, «Bollettino AIB», 46 (2006), n. 3, p. 186-206: p. 194-195; nelle pagine successive, la visualizzazione di una lista di stringhe di soggetto è confrontata, in termini di benefici, con quella di una manciata di parole chiave da mettere in una 'finestrella'.

le stringhe di soggetti delle registrazioni catalografiche o le esoteriche numerologiche classificatorie non sono più sufficienti a rispondere alle esigenze degli utenti, quando si hanno a disposizione, indici, *abstracts*, *full-text* e altrettanto vale per i metodi di ricerca per soggetto molto limitati di cui disponiamo⁴⁰.

Ebbene, questa esigenza di espandere i metodi di accesso semantico, aggiungendo alla 'indicizzazione assegnata' una 'indicizzazione derivata', l'abbiamo avvertita fin dagli albori del passaggio dal catalogo a schede al catalogo online, ma le criticità si sono manifestate proprio nel progettare un modello di catalogo in grado di valorizzare entrambi questi metodi a vantaggio dell'utente. È una criticità che chiama in causa il rapporto tra vocabolario e tecnologia.

Ma il punto qui è anche un altro, preliminarmente all'esigenza di espansione dell'accesso semantico e anch'esso collegato al rapporto tra vocabolario e tecnologia. Prima di dare atto che i dati creati tradizionalmente dalle biblioteche (termini di indicizzazione, stringhe di soggetto e numeri di classificazione) non sono più sufficienti a rispondere alle esigenze degli utenti, sarebbe logico chiedersi se siano sufficienti le condizioni di operatività dei dati nei nostri cataloghi; se il 'mansionario' o 'le regole di ingaggio' siano adeguate alle loro potenzialità nel garantire l'accesso semantico; se i modi in cui vi sono presentati appaiano rispondenti alla varietà, profondità, finalità che li caratterizzano in origine. Saremmo costretti a prendere atto che i dati semantici nei nostri cataloghi non ci sono; o meglio, prendendo spunto da Maja Žumer⁴¹, sono sì in pianta organica, ma con scarsa produttività, poiché li teniamo spesso in vacanza, invece di farli lavorare di più⁴².

Insomma, se si vuole che i dati delle biblioteche siano 'certificati', rispetto a quelli della rete, come auspica Antonio Scolari, siano valorizzati dai bibliotecari e apprezzati dagli utenti occorre che linguaggio (condiviso dalla comunità bibliotecaria) e tecnologia siano allineati nelle finalità e nelle procedure e insieme concorrano alla progettazione di un efficace ed efficiente modello di catalogo. Se, come stigmatizza Karen Coyle, i bibliotecari non possono rischiare di essere ignoranti in materia di tecnologia, tanto meno è ammissibile l'ignoranza del linguaggio impiegato nei loro strumenti professionali; o, come direbbe Crocetti, «il bibliotecario deve sapere che cosa la macchina e il programma gli consentono di fare; ma deve preoccuparsi soprattutto di ciò che mette dentro»⁴³. Sempre che il bibliotecario ritenga di avere ancora qualcosa di utile per l'utente da mettere dentro.

Negli ultimi due decenni, l'attenzione si è spostata sulle caratteristiche formali e sulle esigenze di comunicazione dei dati, indotte dai progressi tecnologici e dal Web (granularità, atomizzazione, autonomia, aggregabilità, leggibilità secondo gli

40 Antonio Scolari, *OPAC & dintorni: essere o non essere nella rete*, «Digitalia», 5 (2010), n. 2, p. 29-40: p. 37.

41 Maja Žumer, *Il parere di Maja Žumer*, in *Old wine, new bottle? Principi e metodi per una reale innovazione nelle prospettive LIS. Il parere di Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer*, «AIB studi», 55 (2015), n. 3, p. 385-403: p. 400.

42 Sui dati semantici e la ricerca per soggetto cfr. Andrea Fabbrizzi, *Indici semantici e ricerca nel catalogo elettronico*, <<http://www.iskoi.org/doc/catalogo.htm>>; anche Emanuela Casson; Andrea Fabbrizzi; Aida Slavic, *Subject search in italian OPACs: an opportunity in waiting?*. In: *Subject access: preparing for the future*, edited by Patrice Landry [et al.]. Berlin: De Gruyter Saur, 2011, p. 37-50.

43 L. Crocetti, *Biblioteche e archivi cit.*, p. 70.

standard descrittivi del Web, RDF, *linked open data*, ecc.; apertura, interoperabilità, riutilizzabilità, fruibilità, ecc.). A queste caratteristiche si è accompagnata la definizione dei requisiti dell'interfaccia (amichevolezza, semplicità, interattività, ecc.) e delle modalità di ricerca (*google like*, interrogazione unica tramite un'interfaccia semplice, serendipità, ecc.)⁴⁴.

Occorre, tuttavia, riprendere le domande, più volte riproposte in queste pagine, sull'utilizzo dei dati semantici nei nostri cataloghi per facilitare l'accesso, orientare la ricerca e modularne i risultati in termini di richiamo e precisione, in modo che la ricerca dell'utente non si fermi all'uso di una parola chiave. Analoghe riflessioni suscitano le questioni legate all'interfaccia e alle modalità di ricerca. Semplicità dell'interfaccia e semplicità dei dati non sono caratteristiche concettualmente isomorfe, ossia non c'è corrispondenza tra la (ragionevole e legittima) semplicità delle modalità di accesso ai dati mediante parole chiave e la (presunta necessità di) semplicità della struttura dei dati ridotti a parole chiave. Anzi, al contrario, è la complessità dei dati, con i loro significati pragmatici, che richiede e permette l'uso di un'interfaccia semplice, amichevole, interattiva. Non c'è, dunque, motivo di invocare la semplificazione dei dati, se essa non fa un buon servizio all'utente. Così come non c'è isomorfismo tra termini di indicizzazione e parole chiave: sebbene possano essere isotopici, essendo entrambi strumenti metatestuali indicanti potenzialmente lo stesso *topic*, tuttavia non sono isomorfi, in quanto alla 'semplicità' della parola chiave si contrappone la 'complessità' del termine di indicizzazione, che ha la configurazione di una fitta rete di termini tra loro variamente collegati.

Il tema della complessità ci riporta alla lezione di Pauline Cochrane e alla necessità di una riprogettazione profonda del catalogo, come già allora si auspicava e per la quale possono rivelarsi ancora utili i *design principles* forniti da Marcia Bates.

Principi di design

Nel suo saggio *Subject access in online catalogs: a design model* del 1986, Marcia Bates propone tre *design principles* per la progettazione dell'accesso per soggetto in un catalogo online: *uncertainty* (l'indicizzazione per soggetto è indeterminata e probabilistica), *variety* (varietà nell'indicizzazione dei documenti e varietà nella formulazione della ricerca), *complexity* (complessità del processo di ricerca, soprattutto nelle fasi di accesso, orientamento e navigazione). Ciascuno di questi principi si ispira ad altrettante teorie: il 'principio di indeterminazione' di Heisenberg, la 'legge della varietà necessaria' di Ashby, il 'principio di ridondanza' di Shannon, il 'principio del lato di un fienile' a indicare, quest'ultimo, l'ampiezza del bersaglio da colpire sufficiente perché l'utente possa entrare nel sistema e interagire con esso.

Ciascun principio, nella sua applicazione all'ambito dell'indicizzazione e dell'accesso per soggetto, ne tocca prevalentemente (anche se non esclusivamente) un aspetto: l'indeterminazione ha implicazioni soprattutto per l'analisi concettuale dei documenti e l'individuazione dei soggetti; la varietà e la ridondanza per la struttura sintetica; mentre la complessità, e le relative raccomandazioni, riguardano il processo di ricerca. L'indeterminazione è la premessa iniziale, il presupposto per generare varietà e ridondanza, sia sul versante dell'indicizzazione che della ricerca, e per

⁴⁴ Per un ampio, articolato e ragionato panorama, ricco di riferimenti bibliografici, cfr. Roberto Raiele, *Vecchi paradigmi e nuove interfacce: la ricerca di un equilibrato sviluppo degli strumenti di mediazione dell'informazione*, «AIB studi», 55 (2015), n. 1, p. 35-55 (*Prima parte*); «AIB studi», 55 (2015) n. 2, p. 197-214 (*Seconda parte*).

attivare meccanismi che facilitino l'accesso al catalogo, ne consentano l'esplorazione attraverso una fitta rete semantica, forniscano all'utente tutta l'informazione disponibile per una scelta consapevole.

Principi e loro implicazioni per l'indicizzazione e la ricerca per soggetto sono illustrati da Marcia Bates in modo articolato e multidisciplinare. Non è possibile darne conto qui in modo esaustivo. Mi limiterò a cogliere alcuni punti essenziali, utili ai fini di un bilancio e di una riflessione critica sulle prospettive future dell'accesso per soggetto.

Il primo dei *design principles* ai quali fa ricorso Bates è il 'principio di indeterminazione' di Heisenberg, il quale stabilisce che non è possibile, in un dato momento, conoscere contemporaneamente, con precisione assoluta, i valori di due variabili, la posizione e la velocità di una particella in movimento. Ne consegue l'impossibilità di predire esattamente la sua traiettoria, essendo possibile soltanto stabilire una probabilità che la particella vada a finire in un determinato punto. Nella sua accezione generale: non abbiamo più previsioni esatte ma delle probabilità che qualcosa succeda⁴⁵.

Trasportato nel campo della scienza dell'informazione, il principio suona così: per quanto si possa presumere di essere capaci di formulare, sulla base di un «ideal indexing system or language», una «perfect description or set of descriptions for each document», così da assicurare «the best possible match with users' needs as expressed in queries», questa corrispondenza tra descrizione del soggetto di un documento e richiesta di un utente resta almeno in parte indeterminata e probabilistica⁴⁶.

Analogamente alle condizioni che generano incertezza nel campo delle misure fisiche (l'atto di misurare modifica la cosa misurata), anche nel campo dell'indicizzazione il comportamento dell'indicizzatore genera inevitabilmente incertezza, dovuta soprattutto a due fattori: l'impossibilità di definire che cos'è il soggetto di un documento e, quindi, di istruire gli indicizzatori su come identificare un soggetto quando esaminano il documento; l'incoerenza degli indicizzatori, che consiste nella scelta, da parte di indicizzatori diversi o di uno stesso indicizzatore in tempi diversi, di differenti temi, aspetti o altri elementi del contenuto di un medesimo documento. Si potrebbe imputare questa incoerenza a inesperienza o incompetenza; ma, nota Bates, in essa si riflette piuttosto quella «variability in associations» documentata dagli studi psicologici, che osservano la diversità delle associazioni mentali che gli individui fanno in risposta a uno stesso stimolo: è, dunque, un tratto umano fondamentale. E conclude:

Let us assume, then, that there is a certain indeterminacy associated with information description and retrieval that is rooted in the nature of the human mind. Given a document, it is impossible (and is likely to remain so) to predict

⁴⁵ Marcia Bates prende spunto da un articolo di Victor Rosenberg, *The scientific premises of information science*, «Journal of the American Society for Information Science», 25 (1974), n. 4, p. 263-269. Werner Karl Heisenberg enunciò il principio di indeterminazione in un articolo pubblicato su «Zeitschrift für Physik» nel 1927. Cfr. Riccardo Ridi, *La piramide dell'informazione e il realismo strutturale*, «ALB studi», 61 (2021), n. 2, p. 233-255: in particolare, p. 240 e seguenti.

⁴⁶ M. J. Bates, *Subject access in online catalogs* cit., p. 360. Questa visione trova riscontro nell'interpretazione, sul piano epistemologico, dell'analisi concettuale dei documenti ai fini dell'indicizzazione come un processo abduttivo; rientra, dunque, nel campo delle previsioni probabilistiche. Cfr. Alfredo Serrai, *Indici, logica e linguaggio: problemi di catalogazione semantica*. Roma: Consiglio nazionale delle ricerche, 1974, in particolare p. 51-53; anche Alberto Cheti, *I processi cognitivi nell'analisi concettuale dei documenti: una postilla tra biblioteconomia e linguistica*, «ALB studi», 56 (2016), n. 1 p. 7-39.

exactly what description a trained indexer will give to it. Given a topic of interest to a searcher, it is impossible to predict what specific aspect of the topic a searcher will pursue and which specific terms or phrasings of terms the searcher will use. Rather than viewing this human characteristic as an obstacle to overcome, let us instead work with this complexity and variety, and design information systems that accept and use this trait⁴⁷.

Dunque, la risposta all'incertezza e all'indeterminazione che caratterizzano i processi di indicizzazione e ricerca per soggetto sta nell'accettare e valorizzare la varietà dei temi, dei significati che possono essere associati al contenuto concettuale di un documento, sia nella fase di indicizzazione che in quella di ricerca, progettando un sistema che ne faciliti l'esplorazione.

Marcia Bates esamina alcune strategie che tenderebbero a ridurre la varietà, aumentando conseguentemente l'incertezza. La prima è la 'pre-coordinazione', associata all'indicizzazione del documento nel suo insieme: un'unica stringa che esprime il tema complessivo del documento.

Library of Congress practice is precoordinate, so the subject concepts appear in that single long heading instead of in separate entries. Additional headings are assigned only if there is no one heading available which reflects the contents of the entire document⁴⁸.

La seconda è la 'specificità', intesa come limitazione dell'indicizzazione di un documento al suo soggetto specifico, con esclusione della pratica di *upward posting*, ossia di indicizzazione a un soggetto più generale. Questa «low-redundancy», insita nella scelta del solo soggetto specifico, dovrebbe essere compensata dalla rete sindetica, che orienta il ricercatore verso temi più generali o più specifici; ma questa rete, osserva Bates, è limitata in origine, ossia nel linguaggio di indicizzazione, e quasi del tutto assente tra le funzionalità dei cataloghi online.

Alle argomentazioni di Bates, sviluppate nel testo del suo saggio in modo molto più articolato di quanto risulti da questa sintesi, aggiungo le seguenti osservazioni:

1. Accettazione della varietà dell'informazione, ossia dei significati associabili a un documento. Qual è la varietà accettabile? In che modo gli indicizzatori possono produrre questa varietà? A quali condizioni la loro incoerenza è da interpretare come condivisibilità di differenti descrizioni del contenuto concettuale di un documento? Servono a questo scopo criteri di pertinenza e rilevanza di vario tipo⁴⁹: testuali

⁴⁷ M. J. Bates, *Subject access in online catalogs* cit., p. 361.

⁴⁸ *Ivi*, p. 363.

⁴⁹ Con 'pertinenza' s'intende qui l'attinenza (*aboutness*) di un'opera (o di una richiesta di informazione) a un determinato tema (*topic*); con 'rilevanza' l'importanza (potenziale informativo, interesse, utilità, adeguatezza al livello cognitivo, grado di novità della trattazione, ecc.) di un tema in relazione all'utente o a un determinato contesto (storico, epistemologico, sociale, culturale), ai fini del recupero dell'opera. In sede di individuazione dei soggetti di un'opera, queste nozioni sono tenute insieme da una domanda del tipo: in base a quali criteri si possono fare presupposizioni sulla rilevanza di un tema di pertinenza dell'opera? In *The notion of relevance in information science: Everybody knows what relevance is. But, what is it really?* (Williston, Vermont: Morgan & Claypool, 2017) di Tefko Sara-

(tema di base/temi particolari/altri elementi di interesse riferibili al contenuto di un'opera; aspetti, punti di vista, ecc.); contestuali (potenziale informativo, intertestualità, variazioni diacroniche nell'uso o nell'interesse suscitato da un'opera, ecc.); strutturali (livelli diversi di analisi di un documento: singola opera indipendente, raccolta di opere indipendenti, aggregato di opera/opere indipendenti e opere dipendenti); relativi alla forma, al supporto, al contenuto intellettuale o artistico (libri, opuscoli, periodici, documenti archivistici, ecc.; documenti a stampa o digitali; testi, fotografie, carte geografiche, opere d'arte, ecc.)⁵⁰. I criteri costruiti su queste caratteristiche non aspirano a una corrispondenza deterministica tra un tema selezionato in un documento e una richiesta di informazione formulata dall'utente; bensì a ridurre l'incertezza e accrescerne la probabilità, producendo varietà in modo condivisibile da parte della comunità bibliotecaria.

2. Pre-coordinazione. Bates associa la pre-coordinazione alla descrizione di un unico soggetto del documento, costituito dal suo tema complessivo. Non c'è dubbio che questa associazione abbia prevalso nella pratica tradizionale dell'indicizzazione per soggetto. Tuttavia, le due nozioni hanno una natura distinta: la pre-coordinazione ha una natura sintattica (combinazione in un'unica stringa dei concetti che descrivono un tema, sia esso il tema complessivo o qualsiasi altro tema selezionato); il tema complessivo (detto anche 'tema di base') ha una natura semantica (è il tema che esprime il contenuto dell'opera nel suo insieme, indipendentemente dalla forma pre- o post-coordinata con cui è rappresentato nel linguaggio di indicizzazione). La pre-coordinazione fa riferimento alla 'coestensione' come proprietà di una stringa di esprimere nel modo più completo i contenuti di un tema; il tema complessivo all' 'esaustività'

cevic, il più autorevole esperto di *relevance*, la nozione di 'pertinenza' data sopra è riconducibile a quella di *topical or subject relevance*, la nozione di 'rilevanza' a quelle di *cognitive relevance or pertinence* e *usefulness or situational relevance*. Al di là di queste precisazioni, si rinvia all'opera di Saracevic per l'ampia, articolata, ricca disamina del concetto di rilevanza e delle sue 'manifestazioni'. Cfr. anche Maria Teresa Biagetti, *Nuove funzionalità degli OPAC e relevance ranking*, «Bollettino AIB», 50 (2010), n. 4, p. 339-356, in cui un paragrafo è dedicato alla nozione di *relevance*, sulla base delle analisi e riflessioni di Saracevic, Yunjie Xu e Zhiwei Chen, Birger Hiørland e Robert Fugmann. Segnalo anche l'ultimo paragrafo del saggio di Biagetti su «nuove prospettive per la ricerca semantica attraverso gli OPAC», il cui intento principale è quello di promuovere «la varietà degli argomenti presentati da un documento», «adottando un'ottica che privilegia la prospettiva della pertinenza [nel senso della 'rilevanza cognitiva' di Saracevic o della 'rilevanza' *tout court*, come intesa sopra] invece di limitarsi alla rilevanza topica (*aboutness*)». Del tutto condivisibili sono anche l'obiettivo e l'impostazione generale della sua argomentazione: «spostare l'attenzione dal livello dell'organizzazione dei risultati delle ricerche (*ranking*) al livello dell'indicizzazione semantica, studiare strategie di miglioramento che arricchiscano le funzioni di ricerca semantica espandendone le potenzialità intrinseche, ed elaborare interventi che rendano possibile utilizzare attraverso le interfacce gli esiti di un'indicizzazione più approfondita e sfaccettata» (p. 353). Entrambi questi aspetti sono anche al centro della discussione in queste pagine. Sul *ranking* e, più in generale, sui problemi che presentano gli OPAC nel rendere disponibili i dati semantici cfr. anche Pino Buizza, *Gli opac: funzionalità e limiti nel mondo del web cit.*

50 Per una panoramica sulla nozione di soggetto, sui criteri di pertinenza e rilevanza e sul processo di analisi, cfr. Biblioteca nazionale centrale di Firenze (BNCF), *Nuovo soggettario. Guida al sistema italiano di indicizzazione per soggetto*, seconda edizione interamente rivista e aggiornata. Roma: Associazione italiana biblioteche; Firenze: Biblioteca nazionale centrale di Firenze, 2021, <https://thes.bnfc.firenze.sbn.it/index_eng.html>, in particolare par. B, C e D.1.

come proprietà dell'enunciato di soggetto di esprimere nel modo più completo i contenuti di un'opera. È in questo senso che va intesa la nozione di 'stringa unica coestesa' proposta dalla *Guida GRIS*⁵¹. Dunque, la pre-coordinazione non limita necessariamente la varietà dei soggetti, mentre ne accresce la precisione. D'altra parte, in entrambi i casi (pre- o post-) è sempre garantito l'accesso ai singoli concetti di un soggetto.

3. Specificità. Sull'efficacia del 'soggetto specifico', accompagnato dalla possibilità di esplorare tutta la sua catena gerarchica, non vi sono dubbi. I benefici sono a vantaggio tanto dell'indicizzatore che dell'utente. In estrema sintesi, il soggetto specifico è il tema che realizza in massimo grado la 'pertinenza', ossia l'attinenza di un'opera a un determinato tema. Qui dobbiamo intendere per 'soggetto specifico' il tema dell'opera ('soggetto') espresso al livello di specificità con cui è trattato nel documento ('specifico'). Un'opera che tratta di 'composizione musicale' è anche un'opera di 'musica', ma solo il primo è il soggetto specifico da selezionare, essendo l'accesso dal secondo assicurato dall'apparato sintetico. Alla base di questa scelta c'è una relazione di implicazione. Tuttavia, nel caso in cui un tema specifico, con elevato grado di specificità e in un settore altamente specializzato, abbia scarsa probabilità di essere reperito attraverso la rete delle relazioni semantiche, allora è conveniente utilizzare una sorta di *upward posting*, in aggiunta all'indicizzazione del soggetto specifico. Analoghe considerazioni valgono anche in direzione opposta. Un'opera che ha come soggetto specifico la 'musica' tratterà presumibilmente anche della 'composizione musicale', che ne è un'attività fondamentale: è il primo, dunque, il soggetto da selezionare. Anche qui vale la relazione di implicazione. Tuttavia, si possono dare casi in cui un singolo tema e il tema più generale che lo comprende sono selezionati e indicizzati entrambi. Questa circostanza può verificarsi per il rilievo che il tema più specifico ha nell'opera, pur essendo compreso nella trattazione di un tema più generale; oppure, per la sua rilevanza contestuale (es., il tema che ha determinato la 'fortuna' dell'opera). Inoltre, quando si ritiene che un tema incluso nel tema di base di un'opera, per la sua specificità e 'singolarità/unicità', abbia un basso grado di implicazione nel tema di base e, quindi, una scarsa probabilità di essere inferito a partire da quest'ultimo, allora è opportuno che il tema specifico abbia una segnalazione autonoma nell'indice (una sorta di *downward posting*), che ne favorisca il recupero⁵². Insomma, anche sull'asse, diciamo così, 'gerarchico' dei temi trattati in un'opera si può produrre varietà senza che venga meno il metodo della 'sommarrizzazione specifica'.

A questa varietà di informazioni prodotta dal sistema informativo dovrebbe corrispondere altrettanta varietà prodotta dal sistema di ricerca, affinché l'interazione tra i due sistemi abbia successo. È a questo proposito che Marcia Bates fa intervenire il principio della varietà di Ashby: un sistema per funzionare con successo dovrebbe generare, in risposta agli input dell'ambiente, una varietà adeguata a ristabilire un equilibrio omeostatico tra i due sistemi⁵³.

51 Associazione italiana biblioteche. GRIS – Gruppo di ricerca sull'indicizzazione per soggetto, *Guida all'indicizzazione per soggetto*. Roma: Associazione italiana biblioteche, 1996 (ristampa con correzioni 2001), in particolare p. 10-11.

52 Cfr. Alberto Cheti, *Il criterio della 'distanza semantica' nell'analisi concettuale dei documenti*, «Bibelot», 23 (2017), n. 2, p. 43-48, <<https://riviste.aib.it/index.php/bibelot/article/view/11679>>.

53 W. Ross Ashby, *An introduction to cybernetics*. London: Chapman & Hall, 1956 (trad. italiana: *Introduzione alla cibernetica*, a cura di Mauro Nasti. Torino: Einaudi, 1971).

We cannot get the full, desired information unless the variety of our search formulation is as great as the variety in the information. If, as I have argued in the previous section, indexers produce great variety in their indexing, then in order to cope successfully the searcher must produce an equal variety in formulating a search on any given topic⁵⁴.

Come promuovere questa varietà per il ricercatore? In sintesi, si dovrebbero adottare due strategie complementari: a) arricchire l'apparato sintetico, ampliando sia le espressioni in linguaggio naturale che descrivono uno stesso tema per consentire al ricercatore interessato a quel tema di entrare più facilmente nel sistema (i 'vedi'), sia le relazioni semantiche tra i temi da esplorare da parte del ricercatore (i 'vedi anche'); b) integrare nel catalogo l'apparato sintetico, sfruttando tutto il suo potenziale ai fini del recupero dell'informazione.

Attraverso questa seconda strategia, il catalogo per soggetto viene a configurarsi come un canale di comunicazione, al quale si può applicare la nozione di 'ridondanza' di Shannon⁵⁵: da una parte, la 'codifica di sorgente', con l'eliminazione della ridondanza, dall'altra, la 'codifica di canale' con l'aggiunta della ridondanza necessaria a facilitare la trasmissione dell'informazione⁵⁶.

54 M. J. Bates, *Subject access in online catalogs* cit., p. 361.

55 La teoria dell'informazione di Claude E. Shannon è stata formulata nell'articolo *A mathematical theory of communication*, pubblicato in «Bell system technical journal» nel 1948 (ristampato in Claude E. Shannon; Warren Weaver, *The mathematical theory of communication*. Urbana, Illinois: The University of Illinois press, 1964).

56 Si domanda Isabella Chiari a proposito del linguaggio: «'ridondante' rispetto a cosa?» (*Ridondanza e linguaggio. Un principio costitutivo delle lingue*. Roma: Carocci, 2002, p. 193). Nell'ambito dell'accesso per soggetto mediante il catalogo, la ridondanza è intesa come una proprietà del codice di trasmissione rispetto al codice linguistico impiegato nell'indicizzazione delle risorse. Il linguaggio di indicizzazione si caratterizza come un codice fortemente economico, che non ammette più forme per una stessa funzione (eliminazione della sinonimia) né più funzioni per una stessa forma (eliminazione dell'omonimia). Queste restrizioni servono ad assicurare un'informazione completa e chiara. Gli stessi elementi (forme diverse per una stessa funzione e funzioni diverse per una stessa forma) dovrebbero essere reintrodotti nel processo di trasmissione come antidoto al rumore, ossia per prevenire eventuali ostacoli al recupero di un'informazione il più possibile esaustiva e pertinente. Questo tipo di ridondanza può essere definita come 'ridondanza paradigmatica', poiché riguarda i rapporti semantici tra i termini, ed è associata alla varietà. Un altro tipo di ridondanza ('ridondanza sintagmatica') si manifesta nella combinazione dei termini in sequenze (stringhe di soggetto) ed è associata alla prevedibilità: l'adozione di un ordine di citazione standard impone restrizioni alla formazione di sequenze valide, rendendole in questo modo più facilmente prevedibili. Per accrescere questa prevedibilità, occorre adottare nel catalogo forme di presentazione e di ordinamento delle stringhe di soggetto che ne favoriscano la comprensione e il *browsing*. Possiamo adattare al contesto dell'accesso per soggetto mediante il catalogo queste parole di Shannon: «The redundancy must be introduced in the proper way to combat the particular noise structure involved. However, any redundancy in the source will usually help if it is utilized at the receiving point. In particular, if the source already has a certain redundancy and no attempt is made to eliminate it in matching to the channel, this redundancy will help combat noise» (C. E. Shannon; W. Weaver, *The mathematical theory of communication* cit., p. 76). La «particular noise structure involved» nel catalogo è determinata dai fattori che incidono sulla probabilità che l'informazione prodotta dal sistema di indicizzazione corrisponda alla richiesta del ricercatore. Questa probabilità è inversamente proporzionale alla quantità di forme che è stata ridotta nell'indicizzazione delle opere (economia) e direttamente proporzionale alla quantità di forme messa a disposizione dell'utente nella ricerca (ridondanza). In altre parole, la corrispondenza tra descrizione del soggetto

A subject catalog is a communication channel as well. The receiver (searcher) needs information about books, which is conveyed by catalog entries. This communication channel functions in a particular way, however. The messages (catalog entries) are labeled by subject headings, and the message is not transferred until and unless there is a match between subject heading and term used by the searcher. To increase the chances of matching, there should be a number of these labels for each book⁵⁷.

Entrambi i processi di codifica sono impiegati nel controllo del vocabolario, anche se il secondo è ampiamente sottoutilizzato: è questa la premessa fondamentale dell'articolo di Bates. Nelle sue parole ritorna la preoccupazione di Coyle incontrata all'inizio:

Most searchers do not realize the great variety that exists in the target information. They use a reasonable term for the topic and stop there. The system must not only help the searcher generate the variety, but also first show the searcher that the variety is there⁵⁸.

A tale scopo, occorre progettare un sistema di ricerca «complex and subtle», che consenta al ricercatore di entrare nel sistema informativo nel modo più facile⁵⁹ e di interagire con esso, esplorandone la rete semantica.

A partire da queste premesse, Bates suggerisce una serie di tecniche e dispositivi che possono essere sintetizzati nella nozione di «a front-end system mind, or FSM», che ha in «an end-user thesaurus» la principale componente:

The FSM is a dense semantic network. [...] It is called front-end because it is the part of the system the searcher encounters first, and while the FSM can and will be used throughout the search, its heaviest use is expected to come at the beginning and early stages of the search. It is called a system mind because it reflects the thinking and organization imposed on the data by systems designers and catalogers⁶⁰.

e *query* ha maggiore probabilità di verificarsi se la varietà ridotta per creare la prima è utilizzata nel catalogo per accrescere la varietà della seconda. Dovremmo allora chiederci: qual è la misura della ridondanza presente nella fonte che può essere utilizzata in fase di ricezione e quanta di questa ridondanza è resa effettivamente disponibile per l'utente attraverso il catalogo? Sul catalogo e la comunicazione, cfr. anche Mauro Guerrini, *Il processo di comunicazione e il catalogo*. In *Id.*, *Riflessioni su principi, standard, regole e applicazioni: saggi di storia, teoria e tecnica della catalogazione*. Udine: Forum, 1999.

57 M. J. Bates, *Subject access in online catalogs* cit., p. 363.

58 *Ivi*, p. 362.

59 La facilità di accesso è enfatizzata nell'immagine del «side-of-a-barn»: il ricercatore deve solo colpire il lato di un fienile; ossia, qualsiasi parola o frase ragionevole nella propria lingua deve essere sufficiente al ricercatore per avviare la ricerca e ottenere le informazioni che la orientino nelle fasi successive.

60 M. J. Bates, *Subject access in online catalogs* cit., p. 368-369; anche Marcia J. Bates, *Rethinking subject cataloging in the online environment*, «Library resources & technical services», 33 (1989), n. 4, p. 400-412, <<https://pages.gseis.ucla.edu/faculty/bates/rethinkingcataloging.html>>. Il *browsing*, la *variety* nelle sue diverse forme, il *side of the barn principle*, il *front-end vocabulary* sono richiamati da Marcia Bates anche nelle raccomandazioni elaborate in occasione della "Library of Congress Bicentennial conference on bibliographic control for the New Millennium" *Improving user access to library catalog and portal information: Final report (Version 3)*, June 1st, 2003, <<https://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>>.

Un tale sistema può essere considerato un «sistema cognitivo», nell'accezione proposta da Riccardo Ridi, in base alla quale si può definire 'sistema cognitivo' ogni 'sistema informazionale' sufficientemente complesso, ossia «abbastanza dinamico e autonomo da essere in grado di ricevere, riconoscere, memorizzare, organizzare, contestualizzare, interpretare, trasformare, elaborare e trasmettere dati»⁶¹. E, sempre sulla falsariga del ragionamento di Ridi, il processo informativo, compreso quello che avviene attraverso il catalogo, può essere considerato come «un processo di 'semantizzazione a due stadi'»: nel primo stadio, i dati sono contestualizzati dal soggetto che li sceglie, li organizza e li trasmette a scopi comunicativi, secondo un codice, ossia un insieme di regole e procedure; nel secondo, è il sistema cognitivo ricevente che deve riconoscere il codice («the thinking and organization», nelle parole di Marcia Bates), effettuare un'ulteriore contestualizzazione dei dati, scegliere «fra le varie informazioni che il dato gli sta trasmettendo, quella che diventerà una conoscenza»:

Solo al termine dell'intero processo, quando viene integrata in un sistema cognitivo, diventando una delle sue conoscenze, l'informazione trasmessa dai dati diventa davvero significativa e quindi si può dire che sia completamente emerso il significato (o, per meglio dire: uno dei possibili significati) dei dati da cui il processo è scaturito⁶².

Contesto, contestualizzazione dei dati: riecheggia qui un'altra delle osservazioni iniziali di Karen Coyle: la perdita delle indicazioni di contesto dei dati, nel passaggio dal catalogo a schede al catalogo online, il contesto nelle sue diverse sfaccettature, come significato, ordinamento, organizzazione, ricerca.

Le considerazioni svolte fin qui si riferiscono in gran parte al modello tradizionale di recupero dell'informazione, basato sui dati creati dalle biblioteche attraverso le attività di classificazione e indicizzazione e sull'incontro, la corrispondenza (*match*) tra una rappresentazione del soggetto del documento e una richiesta che rappresenta un bisogno informativo. Lo scopo è valorizzare il loro potenziale in funzione della ricerca dell'utente. Pare ovvio convenire che la definizione delle strategie di ricerca che utilizzano questi dati sia un compito precipuo delle biblioteche. Tuttavia, seguendo ancora le indicazioni di Marcia Bates, questa fonte non è l'unica e diverse strategie di ricerca (*footnote chasing, citation searching, journal run, area scanning, index or catalog subject search, author subject search*) dovrebbero essere integrate in un modello di ricerca da lei chiamato '*berrypicking*'⁶³:

⁶¹ Riccardo Ridi, *La piramide dell'informazione: una proposta. Prima parte*, «AIB studi», 60 (2020), n. 2, p. 219-267: p. 235.

⁶² *Ivi*, p. 249.

⁶³ Marcia J. Bates, *The design of browsing and berrypicking techniques for the online search interface*, «Online review», 13 (1989), n. 5, p. 407-424 (anche in Marcia J. Bates, *Information users and information system design, Selected Works*, vol. 3, 2016, p. 195-216, <<https://pages.gseis.ucla.edu/faculty/bates/articles/berrypicking.pdf>>). Sulle strategie di ricerca dell'informazione, anche con riferimento a quelle proposte da Marcia Bates, si trovano interessanti e agili spunti nel blog di Luca Rosati, a partire dal suo *Sul web come uomini delle caverne? Dalla raccolta delle bacche alla navigazione*, <<https://www.lucarosati.it/blog/raccolta-delle-bacche>>. Un'applicazione della strategia di *area scanning*, che valorizza ai fini della ricerca l'impiego del sistema di classificazione, è stata realizzata da Andrea Fabbrizzi per la Biblioteca di scienze sociali dell'Università di Firenze. Cfr. Andrea Fabbrizzi, *An atlas of classification. Signage between open shelves, the Web and the catalogue*, «JLIS.it», 5 (2014), n. 2, p. 101-122, <<https://www.jlis.it/index.php/jlis/article/view/224>>.

non una singola ricerca puntuale, basata unicamente su un improbabile *match* tra un termine d'indice e una parola chiave della *query*, ma una ricerca che si evolve, che si modifica, approvvigionandosi da una varietà di fonti e utilizzando differenti tecniche di ricerca, a ciascuna delle quali è associabile una particolare configurazione del *browsing*.

Ridefinire l'indicizzazione nell'era dell'IA

Se il sistema catalogo è un sistema complesso, che implica un duplice processo di semantizzazione/contestualizzazione, tanto da essere assimilabile a un sistema cognitivo; se la fitta rete semantica che lo caratterizza deve essere trasparente per l'utente ed esplorabile; se l'apparato di indicizzazione/accesso deve «stimolare, intrigue, and give pleasure in the hunt», come un gioco⁶⁴; se le sue potenzialità ai fini del recupero dell'informazione sono state ampiamente sottoutilizzate fino a oggi, come stigmatizzano Bates e Coyle; se una varietà di strategie di ricerca potrebbero essere integrate nel catalogo, così da favorire una ricerca 'evolutiva', non sarebbe ragionevole guardare a un possibile utilizzo dell'intelligenza artificiale per questi compiti?

Se un sistema catalogo così fatto presuppone, a monte, un'analisi accurata delle opere da indicizzare, tale da produrre una varietà di soggetti che tenga conto di molteplici fattori di rilevanza, sia testuali che contestuali; se l'analisi dei testi, l'individuazione e la selezione dei temi, i livelli di sommarizzazione richiedono conoscenze storiche, culturali, linguistiche, bibliografiche, oltre a quelle del linguaggio e degli strumenti di indicizzazione; se, tradizionalmente, si ricorre anche a 'fonti esterne' per effettuare le valutazioni di rilevanza, come non pensare ai possibili vantaggi che un algoritmo in grado di sintetizzare grandi quantità di dati può offrire al catalogatore nell'effettuare l'analisi concettuale dei documenti? Un'intelligenza generalista, tipo ChatGPT, potrebbe fungere da 'fonte esterna'? Con quali risultati? Un sistema basato su *Large Language Model* (LLM) ha effettivamente accesso diretto alle opere da analizzare? Se sì, la sommarizzazione di testi eseguita dal sistema è funzionale all'analisi concettuale dei documenti? Che tipo di addestramento è necessario perché il sistema possa produrre i risultati attesi? Quali insiemi di dati di alta qualità e ben curati sono indispensabili, in aggiunta ai cosiddetti *big data*? È possibile e utile attivare, per questi compiti, un 'gioco di interazione/collaborazione' tra processi cognitivi (criteri semantici di pertinenza e rilevanza) e processi statistici (associazioni probabilistiche di parole)? Come far sì che questo gioco possa svolgersi in un contesto cooperativo che coinvolga l'intera comunità bibliotecaria?

Assistere l'utente nella ricerca e il bibliotecario nella catalogazione e nell'indicizzazione per soggetto vengono segnalati come due possibili campi di applicazione dell'IA⁶⁵, entrambi richiamati anche negli interventi al Congresso nazionale AIB del

64 M. J. Bates, *Subject access in online catalogs* cit., p. 363.

65 Cfr. *Artificial Intelligence and machine learning in libraries*, edited by Jason Griffey, «Library technology reports», 55 (2019), n. 1, <<https://journals.ala.org/index.php/ltr/issue/view/issue/709/471>>; Richard Brzustowicz, *From ChatGPT to CatGPT: the implications of Artificial Intelligence on library cataloging*, «Information technology and libraries», 42 (2023), n. 3, <<https://ital.corejournals.org/index.php/ital/article/view/16295/11874>>; Brady D. Lund; Ting Wang, *Chatting about ChatGPT: how may AI and GPT impact academia and libraries?*, «Library Hi Tech news», 40 (2023), n. 4; Jenny Bodenhamer, *The reliability and usability of ChatGPT for library metadata*, <<https://openresearch.okstate.edu/server/api/core/bitstreams/c6187052-e971-48bf-86fo-538b2d36568e/content>>; Christopher Cox; Elias Tzoc, *ChatGPT: implications for academic libraries*, «College & research libraries news», 84 (2023), n. 3, <<https://crln.acrl.org/index.php/crlnews/article/view/25821/33770>>.

2023. A proposito del secondo campo, scrive Gino Roncaglia: «Le intelligenze artificiali generative potranno inoltre lavorare alla produzione automatica di metadati in contesti come gli archivi, le biblioteche, i musei»⁶⁶. Con quali vincoli e restrizioni?

Si tratta ancora una volta di una questione di design, come è capitato più volte di sottolineare in questo excursus e, forse più che in passato, il design richiede una forte integrazione tra il lavoro intellettuale e una tecnologia potenzialmente capace di svolgere efficacemente determinati compiti, di dare risposte pertinenti a domande intelligenti. Senza il primo, non avremmo né domande intelligenti, né risposte pertinenti; neppure sapremmo interpretare una risposta del sistema come pertinente o meno. In altre parole, non avremmo nessun controllo sulle sue prestazioni⁶⁷. Si può, dunque, prevedere che il sistema si comporti in modo efficace quanto più la comunità professionale saprà valorizzare e sviluppare il proprio lavoro intellettuale e, quanto più farà questo, tanto più creerà le condizioni adatte allo sviluppo di un buon sistema di IA. Ciò che durante la transizione dal catalogo a schede al catalogo online è capitato qualche volta di leggere – ossia, che la scarsa qualità di norme e strumenti professionali sarebbe stata compensata dalle potenzialità della tecnologia – non è più valido *a priori* oggi, nell'era delle nuove generazioni di IA, sebbene si debba riconoscere *a posteriori* che non lo sia stato neppure nel passato.

«Il futuro prevedibile dell'IA dipenderà dalle nostre capacità di design e ingegno» (Luciano Floridi). «Il futuro ha bisogno di saggezza» (Helga Nowotny). Se il design e l'ingegno si applicano alla «natura dei dati e dei problemi perché i primi consentono all'IA di funzionare e i secondi delineano i confini entro i quali l'IA può operare con successo»⁶⁸, la saggezza «è un *ethos* di pratiche condivise che viene continuamente coltivato, aggiustato e rifinito al fine di essere esercitato se e quando la situazione lo richiede»⁶⁹. Ingegno e saggezza sono richiesti, dunque, anche nell'affrontare la questione della natura dei dati e dei processi dell'*information retrieval*, in funzione dell'applicazione dell'IA come pratica condivisa, controllata e sorvegliata.

Solo qualche spunto di riflessione. C'è una parola chiave che più di ogni altra definisce la natura dei nostri dati: 'garanzia letteraria'. Alla base di ogni operazione di indicizzazione e ricerca per soggetto ci sono i documenti, le opere. Il bisogno di

66 Gino Roncaglia, *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*. Bari; Roma: Laterza, 2023, p. 119. Sul libro di Roncaglia si veda Maurizio Vivarelli, *L'architetto e l'oracolo. Una nuova alleanza per l'organizzazione della conoscenza?*, «AIB studi», 64 (2024), n. 2, p. 261-268.

67 Scrive Luciano Floridi: «Ci saranno sempre più contesti in cui potremmo non doverci trovare più 'nel o sul processo' (vale a dire, essere parte del processo stesso o almeno in condizione di controllarlo), se possiamo delegare i nostri compiti all'IA» (*Etica dell'intelligenza artificiale: sviluppi, opportunità, sfide*. Milano: Raffaello Cortina, 2022, p. 285). Queste parole possiamo intenderle come un invito a riflettere sull'importanza di individuare e definire in modo corretto i processi. Che cosa fa sì che un insieme di azioni (con gli attori coinvolti e le risorse collegate) costituiscano un processo unitario? Se assumiamo come parametro di riferimento il 'bisogno' (in ogni campo, compreso il bisogno di informazioni) alla cui soddisfazione sono finalizzate le azioni, allora appare difficile immaginare che gli attori umani possano non essere parte del processo o non averne il controllo e la responsabilità. Quelli affidati all'IA potranno essere compiti che stanno dentro un processo unitario e che concorrono a realizzarne i risultati.

68 L. Floridi, *Etica dell'intelligenza artificiale* cit., p. 83.

69 Helga Nowotny, *Le macchine di Dio. Gli algoritmi predittivi e l'illusione del controllo*. Roma: Luiss university press, 2022, p. 128.

informazione che dà origine al processo di *information retrieval* non consiste in una domanda chiusa che può essere soddisfatta da una risposta (tipo *question answering*), ma in una domanda aperta in cerca di informazioni che può essere soddisfatta solo attraverso il reperimento di documenti pertinenti e rilevanti:

the presumption that an AI system with an integrated knowledge base would be superior to a file of documents misses the point that individual document texts have their own value as accounts of the information their authors want to convey. Thus eviscerating documents to construct knowledge bases, and in the process sanitising their language, loses the critical element of who said what, which includes *how* they said it⁷⁰.

Il 'che cosa', da una parte, il 'chi' e il 'come', dall'altra, costituiscono la duplice dimensione della conoscenza di un fenomeno, quale si riflette nelle opere: quella oggettiva (il 'che cosa' o il tema) e quella soggettiva (il 'chi' o il punto di vista e il 'come' o il metodo). La dimensione soggettiva comprende le teorie, i punti di vista, le scuole e sistemi di pensiero, i metodi, gli strumenti, ecc. ('contesti epistemici'), che sono anche gli elementi di discriminazione delle opere che trattano di un medesimo tema. A questi contesti occorre aggiungere quello linguistico: la lingua nella dimensione sincronica e diacronica.

Clare Beghtol individua quattro tipi fondamentali di *warrant*: *literary warrant*, *scientific/philosophical warrant*, *educational warrant* e *cultural warrant*, che interagiscono a costituire la base semantica di un sistema di classificazione/indicizzazione. Al di là di queste differenziazioni, qui interessa sottolineare la caratteristica essenziale implicita nel concetto di *warrant*: «the close correlation between meaning and function or use»⁷¹. In altre parole, la validità dei concetti/classi, dei termini e delle loro relazioni in un sistema di classificazione/indicizzazione ha come fonte autorevole il loro uso all'interno delle opere e nell'ambito scientifico, culturale, linguistico, ecc.

Se questa (descritta qui in estrema sintesi) è la natura dei dati primari coinvolti nel processo di *information retrieval*, allora i dati utilizzati dall'IA per realizzare i suoi compiti (per esempio, di aiuto nell'analisi e nell'indicizzazione delle opere) non possono che essere 'dati storici', di qualità, accuratamente selezionati, altamente affidabili e continuamente aggiornati. Questi dati comprendono, da un lato, i documenti, le opere e, dall'altro, tutte quelle fonti di conoscenza, linguistiche ed extralinguistiche (storiche, scientifiche, bibliografiche, ecc.), necessarie per analizzare le opere e generare metadati, compreso il linguaggio di indicizzazione con le sue regole e strumenti. In questo ambito, sarà sempre indispensabile il gatto reale ('dati storici' + 'regole vincolanti'); mentre non appare perseguibile il «modello del gatto senza il gatto», che rappresenta l'esempio paradigmatico dell'IA ('dati sintetici' + 'regole costitutive')⁷².

70 Karen Sparck Jones, *Information retrieval and artificial intelligence*, «Artificial Intelligence», 114 (1999), n. 1-2, p. 257-281: p. 260.

71 Clare Beghtol, *Semantic validity: concepts of warrant in bibliographic classification systems*, «Library resources & technical services», 30 (1986), n. 2, p. 109-125: p. 111.

72 Riprendo la metafora del gatto e la distinzione tra le due coppie di nozioni *dati storici/regole vincolanti* e *dati sintetici/regole costitutive* dal cap. 3 (*Futuro: lo sviluppo prevedibile dell'IA*) del libro di L. Floridi *Etica dell'intelligenza artificiale* cit., p. 65-88. La distinzione tra dati storici e dati sintetici

Perciò, analizzare la natura dei dati e dei processi che caratterizzano l'*information retrieval* è un buon punto di partenza per riflettere sui compiti da affidare a un sistema di IA, sui vincoli e le restrizioni di cui tenere conto nel disegnarne le applicazioni. È una questione di ingegno, che non può essere separata dalla saggezza:

Osservando non solo indietro, ma anche in avanti, si evidenzia un problema che archivisti, bibliotecari e studiosi affrontano in continuazione: la scelta di cosa scartare e cosa invece conservare, considerata l'immensa mole di materiale proveniente non solo dal passato, ma che si accumula nel presente. La saggezza consiste nel collegare il passato con il futuro, suggerendo cosa fare nel presente. Si tratta di rendere la conoscenza recuperabile al fine di rispondere a domande che non sono ancora state poste. [...] Se il futuro ha bisogno di saggezza è necessario allora praticare un *ethos* che consiste nel trovare modi di attingere alle risorse che il passato e il presente ci offrono per guidare i comportamenti, ampliare gli orizzonti e aiutare a progettare le nuove istituzioni di cui avremo bisogno in futuro, capaci di fornire risposte per i problemi di domani. La saggezza risiederà allora nei meccanismi capaci di colmare il divario tra gli individui e la comunità, tra gli individui e ciò che avviene al livello in cui operano i sistemi complessi. La saggezza è necessaria per rispondere a ciò che non è ancora prevedibile. Non riguarda le 'soluzioni', anche se potrebbe aiutare a farle sorgere. È il contrario di una soluzione rapida, che sia tecnologica o di altro tipo⁷³.

Di fronte al 'nuovo' dell'intelligenza artificiale, tornano in mente le parole di Luigi Crocetti a proposito di un altro 'nuovo', pronunciate a Reggio Emilia in un memorabile pomeriggio di novembre di quarant'anni fa. Anche Crocetti parlò della funzione di filtro, di selezione della massa crescente di documenti come di una 'nuova' funzione professionale delle biblioteche. Mi limito a citare la conclusione del suo intervento, sebbene tutto il testo meriti di essere riletto.

Nel mondo della comunicazione selvaggia la funzione di garanti della competenza e della pertinenza può essere assunta dalle biblioteche: di qui l'esaltazione del momento della guida, della consulenza, degli scambi coi profani e con gli specialisti. Se questo avverrà, ricordiamo che sarà potuto accadere per l'incontro della biblioteca con le nuove tecnologie. Dalla metamorfosi nascerà una nuova vitalità della biblioteca, se i bibliotecari avranno il coraggio di viverla, questa biblioteca, anche se forse non è la biblioteca sognata; di vivere questo mondo, anche se non è il migliore dei mondi possibili⁷⁴.

riguarda la provenienza dei dati: i dati storici sono dati 'reali', provenienti da fonti esterne (per esempio, un certo numero di scansioni di immagini 3D di parti del corpo umano o di cartelle cliniche di pazienti utilizzate per addestrare un sistema di IA a compiti diagnostici); mentre i dati sintetici sono dati 'artificiali', ossia generati interamente dall'IA facendo affidamento esclusivamente su regole costitutive. Floridi aggiunge anche la nozione di *dati ibridi*, che sono un prodotto di dati storici e sintetici. Per addestrare un sistema di IA a generare metadati è indispensabile utilizzare dati storici (i documenti, le opere, le fonti autorevoli in campo linguistico, scientifico e culturale) e le regole del linguaggio di indicizzazione, il cui carattere regolativo e pragmatico è discusso in Alberto Cheti, *Seguire una regola. Regole e prassi nell'indicizzazione per soggetto*, «JLIS.it», 13 (2022), n. 3, p. 100-109.

⁷³ H. Nowotny, *Le macchine di Dio* cit., p. 129, 131.

⁷⁴ L. Crocetti, *Il nuovo in biblioteca* cit., p. 28.

Conclusioni

Nonostante le aspettative e le proposte di miglioramento suscitate dall'avvento dei cataloghi online, pochi progressi sono stati fatti nel campo dell'accesso per soggetto alle risorse bibliografiche. Le nuove tecnologie hanno influito più sul rinnovamento delle norme e degli strumenti di indicizzazione che sulle procedure di ricerca mediante il catalogo.

Questo esito è imputabile presumibilmente a un insufficiente allineamento, talvolta a un disallineamento, tra linguaggio e tecnologia all'interno della catena mezzi-fini. L'espressione di Karen Coyle che «tutto è tecnologia tranne l'utente» vale solo se in entrambi i lati, quello della tecnologia (come mezzo) e quello dell'utente (come fine), includiamo anche il sistema di indicizzazione con il suo linguaggio, le sue norme e procedure.

Un sistema rinnovato, ristrutturato, adeguato alle possibilità tecnologiche, come quello auspicato nella transizione dal catalogo a schede al catalogo online, avrebbe dovuto costituire la base per arricchire e facilitare l'accesso per soggetto, mediante la varietà dei soggetti, la molteplicità dei punti di accesso, la disponibilità di una complessa ma coerente rete semantica, alle quali far corrispondere un'interfaccia semplice, amichevole, interattiva. Con un giudizio forse eccessivo, sicuramente sommario, possiamo affermare che, al contrario, è stata seguita prevalentemente la tendenza a sterilizzare i dati, ridotti a parole chiave, privandoli dei loro contesti e depotenziandone la capacità di migliorare il successo della ricerca.

Oggi, di fronte a una nuova transizione, è utile riflettere sugli esiti di quella precedente, porsi le domande di Karen Coyle, prendere spunto dai suggerimenti di Marcia Bates. Soprattutto, è indispensabile analizzare la natura dei dati e dei processi che caratterizzano il recupero dell'informazione per individuare i compiti da affidare a un sistema di IA, stabilire i vincoli e le restrizioni di cui tenere conto nel progettare le applicazioni, scegliere con cura il dataset di qualità con cui addestrarlo, compreso il linguaggio di indicizzazione con le sue norme e strumenti, acquisire competenze nel formulare le domande (*prompting*) per ottenere risposte accurate e pertinenti... Poiché niente di tutto questo sarà facile, è meglio iniziare subito, a casa, a scuola, al lavoro, e nelle nostre società⁷⁵, con la consapevolezza che affrontare la crisi non da soli è la condizione per uscirne⁷⁶.

⁷⁵ Cfr. Luciano Floridi; Massimo Chiriatti, *GPT-3: its nature, scope, limits, and consequences*, «Minds & machines», 30 (2020), p. 681-694; p. 692-693, <<https://doi.org/10.1007/s11023-020-09548-1>>.

⁷⁶ L. Crocetti, *Il nuovo in biblioteca* cit., p. 28.

Articolo proposto il 23 gennaio 2025 e accettato il 4 marzo 2025.

ABSTRACT

AIB studi, vol. 64 n. 3 (settembre/dicembre 2024), p. 321-352. DOI 10.2426/aibstudi-14127
ISSN: 2280-9112, E-ISSN: 2239-6152 - Copyright © 2024 Alberto Cheti

ALBERTO CHETI, e-mail: chetialberto@infinito.it.

L'accesso per soggetto: dal catalogo a schede al catalogo online, con uno sguardo all'Intelligenza Artificiale

Prendendo spunto da alcune osservazioni di Karen Coyle sul ruolo dei metadati semantici nei cataloghi online e, in particolare, sulla perdita della loro funzione organizzativa, il saggio descrive il fermento di idee che ha caratterizzato la transizione dal catalogo a schede al catalogo online nell'ambito dell'accesso per soggetto: le sfide, le aspettative e i rischi connessi all'utilizzo delle nuove tecnologie, le loro implicazioni nel rinnovamento degli strumenti di indicizzazione e per un nuovo design del catalogo. Una situazione, quella di quarant'anni fa, analoga alla situazione attuale, in cui si sta profilando l'opportunità di applicare alla produzione di metadati i nuovi modelli di intelligenza artificiale. Il saggio si conclude proprio con uno sguardo all'intelligenza artificiale, chiedendosi quali sono le condizioni e i confini entro i quali essa può operare con successo in questo ambito, quali i vincoli e le restrizioni di cui tenere conto nel progettare le applicazioni. Tra lo spunto iniziale e lo sguardo finale, il saggio ripercorre alcune tappe significative dello sviluppo dei cataloghi e del loro design riguardo all'accesso per soggetto: le lezioni di Pauline Cochrane del 1984 su *Modern subject access in the online age*, incluse le numerose voci di esperti da lei chiamati a confrontarsi su questo tema; il fermento della transizione in Italia, quale si riflette in alcuni interventi di Luigi Crocetti e nel Convegno di Trieste del 1985 su *Indicizzazione per soggetto e automazione*; i principi di design suggeriti da Marcia Bates in quegli anni. Non si tratta tanto di un excursus storico, quanto di una riflessione critica e una discussione su temi e problemi significativi per un nuovo design del catalogo: la complessità dell'accesso per soggetto, l'efficacia in questo ambito di nozioni come 'indeterminazione', 'varietà', 'ridondanza' e la loro connessione con quelle tradizionali di 'pre-coordinazione', 'specificità' e 'sommarrizzazione', il rapporto tra vocabolario e tecnologia.

Subject access: from card catalogs to online catalogs, with a look at Artificial Intelligence

Building on Karen Coyle's observations about the role of semantic metadata in online catalogs and the loss of their organizing function, the paper examines the intellectual ferment that characterized the transition from card catalogs to online catalogs in the field of subject access. It explores the challenges, expectations, and risks associated with the adoption of new technologies, their implications for modernizing indexing tools, and their impact on catalog design.

The historical context from four decades ago bears striking resemblance to contemporary developments in the field, as institutions now explore the potential application of artificial intelligence technologies to enhance metadata generation processes.

The paper concludes with an examination of artificial intelligence, considering the conditions and boundaries within which it can operate effectively, and the constraints that must be addressed when designing AI applications.

Between the initial reference to Coyle and the final examination of AI, the paper traces significant developments in catalog evolution and design regarding subject access. It discusses Pauline Cochrane's 1984 lectures on *Modern subject access in the online age*, including expert proposals she solicited on this issue; the excitement surrounding this transition in Italy, as reflected in Luigi Crocetti's presentations and the 1985 Trieste Conference on *Subject indexing and automation*; and the design principles proposed by Marcia Bates during this period.

Rather than presenting a mere historical overview, the paper offers a critical reflection on key issues relevant to modern catalog design: the complexity of subject access; the effectiveness of concepts such as 'uncertainty', 'variety', and 'redundancy'; their relationship to traditional notions of 'pre-coordination', 'specificity', and 'summarization'; and the interaction between vocabulary and technology.