

formazione, per stanare quei dati che si nascondono nelle pieghe di una produzione documentaria da sempre e per sempre ampia e complicata.

L'indicizzazione è il sogno nascosto dentro a ogni sistema documentario, forse in senso assoluto un'utopia, ma Roberto Guarasci e Mauro Guerrini si adoperano per riportare il tema alla sua indispensabile concretezza, che risponde soprattutto ai bisogni di una società perseguitata dalla propria bulimia informativa.

Spiegare cos'è l'indicizzazione, e farlo con un approccio che stia sul confine tra un'ineludibile dimensione tecnica e il desiderio di tradurla in prassi accessibili ai più, non è certo compito da poco. Significa mettere in gioco le tradizioni e gli umori di discipline inevitabilmente sospese tra la loro ingombrante dimensione storica, l'anelito a governare l'informazione che le sostanzia e i crescenti appetiti di una lunga congiuntura digitale. In questo senso la cifra scelta da Guarasci e Guerrini risulta decisamente efficace e consente loro di muoversi con larghezza di vedute nel perimetro di un tema solo apparentemente circoscritto ad algebrici tecnicismi. L'intero volume si mantiene in equilibrio tra due discipline che, sia pure intese qui in senso ampio, alla fine trovano la similitudine più solida proprio nella reciproca consapevolezza della loro diversità relativa. Allo stesso modo ci si muove con la dovuta accortezza tra approcci di dominio in lunga e costante evoluzione, che si sono tradotti nel tempo in buone pratiche e standard frutto di suggestioni culturali di diverso ordine e grado. Si va costantemente in cerca di una sintesi possibile tra queste variabili o, quanto meno, della costruzione di un consapevole sguardo di insieme.

Il vero collante di questo libro è la volontà di difendere lo spessore

culturale delle pratiche di indicizzazione, inseguendo un'efficienza che non sia solo meccanica. Si tenta di rispondere al bisogno di interpretare il dato, ancora prima di classificarlo, per comprenderne le logiche e farne oggetto di un'euristica possibile. Gli autori, pur senza rinunciare a dar conto di prassi concrete, ricorrendo spesso ad esempi tangibili, valutano il tema alla luce di una matura consapevolezza contestuale. Come avremo modo di vedere meglio, non ci troviamo di fronte a una indiscriminata caccia al dato ma, soprattutto, al bisogno di controllare la qualità e l'affidabilità dell'informazione di cui si va in cerca.

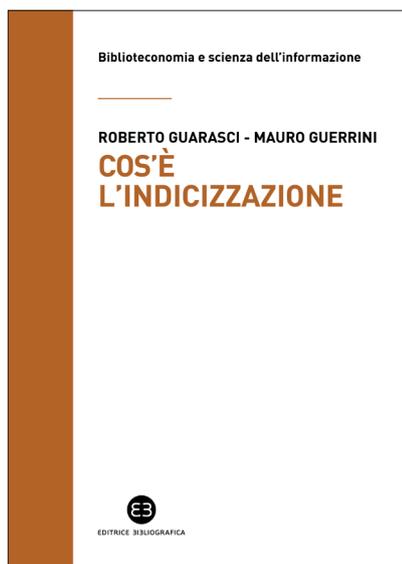
La posizione di Guarasci e Guerrini al riguardo sembra chiara fin dalle prime pagine, quando scrivono che "il termine indicizzazione indica un'operazione quanto mai complessa: è l'attività di valutazione del messaggio, cioè del contenuto concettuale di un testo" (p. 13). Indicizzare, quindi, significa innanzitutto capire, cioè riuscire a individuare i concetti chiave di una risorsa prima ancora di renderla cercabile. "La mia coscienza ha visto solo una macchia di inchiostro su un foglio di carta", scriveva Fernando Pessoa, e probabilmente lo sforzo dell'indicizzazione sta tutto qui, nel tentativo di andare oltre a quella macchia (o a quel pugno di bit), verso la comprensione rotonda di ciò che si trova.

Ogni libro, e questo non fa eccezione, ha poi gli occhi e la voce di chi lo ha pensato. Bisogna intanto notare, allora, come i due autori e i loro percorsi scientifici rappresentino nei rispettivi ambiti disciplinari delle eccezioni sempre disponibili a un confronto tra due dominî diversi e uguali. Forse è proprio questo il tratto distintivo di un lavoro di questo tipo: mettere il concetto di informazione al centro, prima ancora di preoccuparsi di come le

Roberto Guarasci, Mauro Guerrini
COS'È L'INDICIZZAZIONE

Milano, Editrice Bibliografica, 2022, 120 p.

DOI: 10.3302/0392-8586-202203-067-1



Il titolo di questo libro, a portarlo alle estreme conseguenze, si assume la responsabilità di aiutarci a svelare i segreti più intimi dell'in-

diverse discipline lo abbiamo declinato, descritto e cercato. Le scienze del libro e quelle degli archivi hanno dovuto sempre confrontarsi con le loro urgenze euristiche e proprio intorno al bisogno di catturare l'informazione sono cresciute e si sono consolidate. I diversi standard e le diverse pratiche mirano in fondo soprattutto a una concettualizzazione che vivifichi una tassonomia ineludibile per chi studia le sedimentazioni documentarie.

Non si tratta però di meri esercizi retorici o di speculazioni sulla insondabile vastità della conoscenza: "Lo scopo dell'indicizzazione è quello di stabilire l'accesso ai contenuti significativi trattati dal documento e ne consegue che *qualsiasi documento o informazione è utile solo se è accessibile e ricercabile* (p. 71, citando Paola Costanzo Capitani).

L'indicizzazione, letta così, esce dalle brume della speculazione teorica, cui pure deve qualcosa, e diventa un bisogno pubblico, un'esigenza sociale di un mondo che senza l'andirivieni delle informazioni si estinguerebbe.

Detto tutto questo occorre precisare, come già anticipato, che la struttura del volume si articola in due sezioni, la prima delle quali si concentra su aspetti biblioteconomici, lasciando alla seconda il compito di riflettere sulla dimensione archivistica, sia pure interpretata con la giusta dose di relativismo metodologico.

Le differenze disciplinari ci sono ma vengono affrontate con la consapevolezza di un docuverso che esiste e va tenuto presente. Segnalano ricchezza, non contrapposizione, e attivano un circolo virtuoso che dagli archivi va alle biblioteche e viceversa, arricchendosi di opportunità e di idee.

Al riguardo molto centrata risulta la premessa terminologica che da un lato punta a disambiguare e dall'altro cerca una sintesi tra i di-

versi *oggetti* affrontati dallo studio. Ci permettiamo di notare al riguardo come la questione "del nome delle cose" sia assolutamente rilevante. Nelle scienze documentarie e nella logica di chi cerca, intanto, i nomi vengono prima delle cose, l'impronta prima del piede, perché ogni rappresentazione è un'interpretazione e dalla qualità dell'interpretazione deriva la tracciabilità reale dell'oggetto. Nella lunga transizione digitale, poi, è inevitabile porsi il problema terminologico, anche in ragione di un ineludibile bisogno di riconcettualizzazione indotto da nuove modalità di concezione, produzione e uso delle risorse documentarie. Le parole totemiche, archivio e biblioteca, insieme a tutti i loro derivati, legati all'esercizio delle discipline che ne scaturiscono, non bastano più a se stesse e metterle in discussione significa accettare lo sforzo necessario a ridefinirne perimetri, suoni e contenuti.

Il tempo trasforma interi universi e chi tratta con senso di responsabilità l'informazione qualificata sa che deve fare i conti con la storia e l'evoluzione delle società che si rispecchiano puntualmente nelle risposte cangianti delle filiere informative. È quindi opportuno notare come l'elemento diacronico, in termini di valutazione storica e comparativa delle diverse tecniche di indicizzazione, dalle più rudimentali a quelle tecnologicamente più sofisticate, sia un altro tratto distintivo di questo lavoro. E risulta tanto più apprezzabile in quanto dimostra che le discipline documentarie, come la natura, non fanno salti. Il loro è un lungo e mai concluso percorso evolutivo, inscindibile dall'evoluzione dei mezzi di comunicazione e dei bisogni informativi.

Se entriamo a questo punto nel merito del testo constatiamo che il primo dei due capitoli si apre alludendo immediatamente alla

complessità del tema: "In ambito biblioteconomico per *indicizzazione* si intende l'intero processo di descrizione di una risorsa e attribuzione di punti d'accesso, nominali e semantici, al record o dataset ottenuto; in quest'accezione, indicizzazione coincide con la catalogazione *tout court*, ovvero, in termini contemporanei, con la metadattazione" (p. 15). In questo passaggio iniziale l'allusione al processo di metadattazione è molto più di una sfumatura. Nella documentazione contemporanea essa ha infatti un ruolo centrale e finisce con l'assumere la fisionomia di una catalogazione aumentata, capace di tenere conto di tutta la meta informazione necessaria non solo a descrivere e a cercare gli oggetti digitali ma anche a gestirli e conservarli correttamente.

Dal punto di vista teorico emerge poi una distinzione importante tra indicizzazione semiotica e indicizzazione semantica. "La prima mira a costruire degli indici in base ai segni che una risorsa presenta su se stessa, ovvero in base agli attributi, alle formulazioni, alle dichiarazioni che una risorsa comunemente presenta in luoghi precisi" (p. 14). Da questa rappresentazione fisica si passa alla seconda tipologia di indicizzazione, quella semantica, che pone al centro "l'azione di descrivere o identificare un documento nei termini del suo contenuto concettuale" (p. 15).

Proprio mentre si fa riferimento ad approcci prevalentemente tecnici si entra così nell'aspetto più qualificante di un'indicizzazione consapevole, quello che deve confrontarsi con l'eterno femminile della mediazione. Si legge infatti che "colui che indicizza *media* tra la risorsa e il lettore, ovvero corrobora il concetto di catalogo come *medium* tra la collezione e l'utente. L'autore offre il soggetto, l'utente desidera ottenere l'informazione e l'indicizzatore funge da collegamento tra i due,

con le difficoltà connaturate alle due azioni” (p. 15). La mediazione nelle discipline documentarie è un tratto genetico da cui non si sfugge, pena l’impaludamento in una sterile autoreferenzialità. In maniera molto opportuna la trattazione prosegue perciò confrontandosi con questo costante lavoro, ricostruendolo nella sua dimensione storica e tecnica e fornendo indicazioni puntuali e contestualizzate in merito alle diverse procedure.

Si riflette intorno all’idea, agli strumenti e alle prassi della catalogazione e della catalogazione per soggetto, entrando con puntualità nel merito di passaggi tecnici e metodologici come l’analisi concettuale o il controllo terminologico, e valutando i relativi di strumenti, quali soggettari e *thesauri*. Un passaggio importante di questa parte del volume sembra quello in cui si enunciano i principi di fondo dell’indicizzazione per soggetto, individuati a livello internazionale e implementati poi a livello nazionale dal GRIS, per arrivare poi alla descrizione del nuovo soggettario.

Da segnalare, infine, le pagine dedicate alla classificazione, braccio armato delle procedure di catalogazione e terreno di confine tra le due discipline in gioco. Se infatti “classificare, teoricamente, significa raggruppare idee e concetti che presentano caratteristiche simili tra loro, separandoli da altri” (p. 59), indipendentemente dalla peculiare fisionomia della classificazione archivistica o bibliografica, dietro questa attività si nasconde il soffio divino dell’ordine, l’idea di un’organizzazione dell’informazione che vada oltre la sua triste radice tassonomica.

Il secondo capitolo di *Cos’è l’indicizzazione* si concentra invece quelli che vengono definiti documenti testuali, “intendendo con questa definizione i documenti e gli atti prodotti da una qualunque persona

fisica o giuridica nell’esercizio della sua attività, (che) non nascono in base a una logica di volontarietà ma in base a un principio di necessità che è quello del raggiungimento delle finalità proprie del soggetto produttore” (p. 71). Ad essere precisi, anzi, con qualche personale perplessità di lungo periodo sul carattere esclusivo di quel *giuridicamente rilevanti*, “le attività di indicizzazione automatica si riferiscono in questo testo ai documenti informatici contenuti nei sistemi di gestione documentale che il Codice dell’amministrazione digitale così definisce: “Documento elettronico che contiene la rappresentazione informatica di atti, fatti o dati giuridicamente rilevanti” (p. 88).

In questa parte del volume si guarda quindi a una generica dimensione archivistica, ma, coerentemente agli intenti degli autori, ci si concentra a ragione sui singoli veicoli dell’informazione, a scapito di più ampi e distesi sistemi di relazioni logiche e fisiche tra le diverse entità informative e strutturali. Per così dire, si guarda all’archivio dal basso, alla ricerca della sua massima funzionalità per l’utente, consapevoli del fatto che l’attività di indicizzazione “è particolarmente rilevante in relazione alla necessità di dare tempestiva risposta agli utenti” (p. 71).

Il profilo archivistico emerge comunque nel qualificare come oggetti *involontari* e *necessari* i documenti considerati e si ripropone nei riferimenti alle procedure di trattamento tipiche della funzione archivistica, a partire dal protocollo e dalla classificazione. A queste prassi, che potremmo definire di pre o proto-indicizzazione, viene riconosciuto il giusto valore, ponendo in particolare l’accento sulla “necessità di indicizzare per classificare” (p. 72). Bisogna insomma fare in modo che le cose, ma direi anche le azioni, ricevano i loro nomi per poter

le poi collocare dentro agli schemi logici e concettuali propedeutici all’ordine e quindi funzionali a una più affidabile reperibilità del dato. Anche in questo capitolo la profondità cronologica della ricostruzione dei metodi e delle azioni di indicizzazione ha il suo significato, soprattutto quando marca la distinzione fondamentale (e non gerarchica) tra l’indicizzazione manuale e quella automatica. Sembra molto opportuna in questa valutazione di insieme la constatazione del ritardo accumulato dalle scienze del documento nei confronti di quelle del testo: “La distanza concettuale tra le scienze del testo e le scienze del documento nei decenni passati non ha agevolato l’avvio di una riflessione in ambito archivistico delle problematiche legate all’indicizzazione delle unità documentali complice anche l’assioma del rigido rispetto delle volontà del soggetto produttore e quindi anche delle sue esplicazioni testuali” (p. 75). Almeno su questo terreno, insomma, gli archivi per certi versi inseguono le biblioteche. Esistono diverse spiegazioni per questo stato di cose, a partire dalle peculiarità dei rispettivi sistemi di oggetti informativi. Si può però concordare soprattutto con i limiti che derivano da un approccio ideologico e fortemente storicizzato alla descrizione archivistica. Da qui scaturiscono rigidità metodologiche e un forte particolarismo descrittivo che, per la sua natura “antropologica”, nemmeno i numerosi standard sono riusciti davvero a debellare. Si ha il dubbio che dietro a questi processi l’archivistica intraveda soprattutto il demone sinistro dell’ordinamento per materie, percepito esclusivamente come oscena violazione *ex post* del principio di provenienza, piuttosto che come possibile contributo in progress alla gestione e all’organizzazione dell’informazione. Si aggiunga a questo che la deriva

storico culturale che ha trascinato gli archivi per buona parte del Novecento (e che neppure il nuovo millennio sembra saper contenere) ha fatto prevalere di necessità i contesti sui contenuti, malgrado risulti sempre più evidente che c'è una forte esigenza di ripensare gli archivi anche in quanto risorse strategiche di informazione qualificata, utili più alle urgenze del presente che ai tempi lunghi della retroflessione del pensiero storico. La realtà quotidiana ci impone ormai di pensare ad archivi dinamici, interoperabili e ben piantati nell'operatività quotidiana di cui sono strumenti essenziali. Ammettere questo significa aprire prospettive nuove alla disciplina e ai suoi adepti e comporta una riflessione approfondita sul modo di esercitare un ruolo e una professione. Come ci ricordano gli autori siamo infatti spinti anche verso il superamento dell'indicizzazione handmade: "La quantità di dati e documenti digitali circolanti è la motivazione alla base della necessità dell'indicizzazione automatica" (p. 87). Siamo cioè di fronte a un passaggio delicato e per certi versi inquietante, al tempo del *machine learning* e/o del *deep learning* che fanno intravedere la possibilità di una selezione robotica dell'informazione.

L'apprezzabile equilibrio dell'approccio di Guarasci e Guerrini però ci tranquillizza, almeno per quanto si possa stare tranquilli in tempi in cui le tecnologie corrono effettivamente all'impazzata e il controllo umano rischia davvero di trovarsi con il fiato corto. Partendo dall'assunto che "un documento può essere recuperato e quindi utilizzato solo se indicizzato correttamente" (p. 87) e ribadendo quindi il bisogno di un'indispensabile metadattazione "meccanica", si sottolineano ancora gli elementi quantitativi, dal momento che "l'indicizzazione automatica o, per meglio dire, semi-au-

tomatica è una attività necessitata dalle quantità dei documenti contenuti in un qualunque sistema di gestione documentale" (p. 97). Ma molto saggiamente si precisa anche che tale attività "richiede sempre e comunque una supervisione attenta dell'indicizzatore e dell'esperto di dominio per evitare un erroneo o improprio recupero dell'informazione" (p. 97). "Sempre e comunque" è una presa di posizione molto netta che mi sembra frutto soprattutto di una lucida consapevolezza in merito all'impatto reale e potenziale delle risorse di calcolo disponibili. A questo punto entrano però in gioco anche le possibili contraddizioni che si agitano sotto la superficie di una lettura meccanica di questi fenomeni. Credo si possa dire, nello specifico, che indicizzare correttamente significa confrontarsi con i limiti sia delle macchine che degli uomini, facendo i conti con le conseguenze potenzialmente disumane delle performance sempre più spregiudicate delle prime e con la inesorabile condanna alla soggettività dei secondi. L'efficacia dell'indicizzazione risiede anche e soprattutto nell'affidabilità almeno relativa dei suoi risultati e l'attenzione a un fattore umano "compensato" è sicuramente una garanzia, perché, come recita uno slogan di qualche tempo fa, la potenza è niente senza controllo. Quello che è sicuro è che le procedure di indicizzazione automatica impongono una cooperazione partecipativa tra diverse discipline e diverse figure professionali, dal momento che "sono per loro natura disciplinarmente trasversali intersecandosi in misura e quantità variabili con le metodiche e le tecniche della Linguistica Computazionale, della Statistica Testuale, del Data Analysis e del Text mining" (p. 90).

Un ultimo problema, tra le molte suggestioni che il libro solleva per

quanto sia nel complesso piuttosto asciutto, è quello della strutturazione, o non strutturazione, dei documenti, ai confini di una diplomazia applicativa. Come ci ricordano gli autori, il problema della struttura o della sua assenza va letto anche in ragione degli utenti designati e del linguaggio delle risorse e dei documenti. Da quegli stessi documenti, visti alla luce di una lunga serialità, scaturiscono o non scaturiscono corpora efficaci e altrettanto efficaci risultati di restituzione. In questo senso il documento strutturato risponde meglio a certe sollecitazioni, mentre oggetti più fluidi comportano un'elaborazione più complessa. Serve un'azione che in qualche modo scarnifichi il testo e ne ricomponga gli elementi linguistici essenziali in termini di occorrenze, significati e relazioni quantitative che possano poi trovare posto nella logica del grafo in quanto strumento di organizzazione e disseminazione relazionale.

In questo modo, partendo dalle cose in cerca di un nome, si arriva ai ponti di Eulero (p. 96) e ai grafi capaci di attraversare i fiumi informativi. *Cos'è l'indicizzazione* sembra quindi tenere fede al suo titolo, mettendo in lettore in condizione di apprezzare uno stato dell'arte storicizzato di questa pratica preziosa e consentendogli di prendere confidenza critica con le teorie, le prassi e gli strumenti in ambiti anche diversi tra loro.

Se ne ricava l'impressione di una miscela equilibrata, utile soprattutto alla comprensione del senso e del bisogno di indicizzare, e di indicizzare seguendo procedure di controllo e verifica che non mirano solo alla quantità o alla rapidità dell'azione euristica, componenti pure non trascurabili, ma alla qualità dell'informazione che si otterrà. Sullo sfondo si muovono questioni metodologiche di lungo periodo e si colgono gli usi e i costumi di due

comunità disciplinari che da sempre si confrontano con il demone della conoscenza nascosta e con il bisogno di sconfiggere le resistenze che le diverse risorse informative oppongono agli utenti. In questo senso si può parlare di un libro animato da un dotto spirito di servizio che contribuisce a mettere a fuoco temi centrali e che dimostra come certe barriere di dominio siano frutto di una rigidità di metodo e

di approccio che è la prima nemica della dinamica fluidità richiesta ormai dal confronto con qualsiasi risorsa informativa.

Alla fine di questa lettura sembra in definitiva un po' più chiaro il percorso che porta dall'archivistica e dalla biblioteconomia a nuove e nemmeno troppo futuribili discipline documentarie, capaci di sopportare la complessità di una modernità che si vorrebbe invece inaridita o

semplificata dalle tecniche e dalle macchine. Parlare di indicizzazione, insomma, significa, almeno in questo libro, cercare di confrontarsi serenamente con la complessità, in cerca di quella conoscenza che serve ad ogni nostra azione quotidiana.

FEDERICO VALACCHI

Università di Macerata
federico.valacchi@unimc.it