



***The Thesaurus. Review,  
Renaissance, and Revision***

Sandra K. Roe and Alan R. Thomas (eds.), London, The Haworth Information Press, 2004

Presentare la storia, le modalità di realizzazione e le potenzialità dei thesauri nel mondo digitale e degli ambienti di rete, questo l'obiettivo del volume, curato dalla bibliotecaria Sandra K. Roe e dal professore di Library and Information Science Alan R. Thomas, che riunisce una serie di contributi di esperti e docenti internazionali.

Gli standard<sup>1</sup> definiscono un thesaurus come un vocabolario controllato e strutturato, in cui i termini sono

organizzati secondo tre tipi di relazioni semantiche: relazioni di equivalenza, gerarchiche e associative. Lo scopo dei thesauri è quello di promuovere la consistenza nell'indicizzazione dei documenti e facilitarne la ricerca: sono in pratica delle collezioni di terminologie, strutturate in modo tale da mostrare le relazioni tra termini e concetti. Il processo che conduce alla creazione di un vocabolario controllato permette sia all'indicizzatore che all'utente di accedere ai medesimi concetti attraverso termini autorizzati, noti come "descrittori". Negli ultimi anni i thesauri hanno progressivamente attirato l'attenzione di vari ambiti professionali, commerciali e scientifici, poiché risultano uno strumento prezioso per descrivere, organizzare e fornire accesso all'enorme quantità di documenti disponibili online. Questi contenuti hanno infatti un reale valore solo se accompagnati da adeguati sistemi di rappresentazione di documenti digitali e da sistemi di recupero efficaci ed efficienti (che forniscono cioè accesso solamente ai documenti di una collezione rilevanti per l'utente). Per fornire una soluzione all'information overload, le strategie di information retrieval seguono due diversi approcci metodologici (che tuttavia dovrebbero essere sinergicamente integrati): il *retrieval term-based* (indicizzazione dei dati tramite termini estratti dal linguaggio naturale dei documenti, oppure vocabolari controllati come thesauri, classificazioni e soggettazioni) e il *retrieval content-based*<sup>2</sup> (sistemi di *visual retrieval*, *video* e *audio retrieval* che indicizzano e permettono di ricercare i documenti digita-

li attraverso i loro attributi audiovisivi). A differenza degli altri tipi di vocabolari controllati utilizzati nel recupero *term-based*, i thesauri consentono maggiore flessibilità, consistenza e specificità. E per questo motivo risultano utili nella fase di data entry, per la descrizione e catalogazione: come *online search assistant* (permettendo di usare sinonimi, varianti lessicali, inversioni, traduzioni, nomi storici, forme alternative e omografi) e come *knowledgebase*, mostrando le relazioni tra termini organizzati in una sorta di mappa che guida l'utente all'informazione.

La loro creazione può risultare onerosa in termini di tempo e competenze necessarie. Ma ciò che, almeno per il momento, fa la differenza tra i sistemi di recupero automatici e un indicizzatore umano sono sostanzialmente due aspetti: la valutazione del pubblico (che permette di dedurre quale linguaggio deve essere utilizzato per rendere ricercabili collezioni di documenti) e l'avere in mente una mappa mentale (una "struttura sindetica") che consente di creare un indice con riferimenti incrociati. Dal punto di vista tecnico, l'utilizzo di un thesaurus potrebbe costituire una barriera per l'utente se ad esempio non fosse integrato nel sistema di ricerca, oppure se il suo browser fosse disponibile nell'interfaccia solo come lista alfabetica. È stato nondimeno dimostrato che i thesauri migliorano il recupero dei dati,<sup>3</sup> e come ha scritto Patricia Harpring del Getty Vocabulary Program, "it is clear that vocabularies are the key to navigating and retrieving meaningful results from the massive amount of largely inchoate

information now potentially available in digital form".<sup>4</sup> Sui thesauri e le tematiche ad essi correlate è disponibile varia documentazione, sia a stampa che online. Tuttavia, nel panorama editoriale mancava una pubblicazione esauriente e aggiornata sull'argomento. Il pregio del volume curato da Sandra K. Roe e Alan R. Thomas è quello di costituire uno strumento di riferimento approfondito e ad ampio spettro su un cinquantennio di esperienze nel settore. I temi trattati spaziano infatti dall'evoluzione delle definizioni, delle caratteristiche, delle funzioni e delle applicazioni dei thesauri, all'importanza degli standard per la loro creazione; dalla valutazione, selezione e utilizzo dei software al ruolo dei consulenti nella loro progettazione e manutenzione; dalle problematiche dei thesauri multilingue alla formazione di esperti in questo settore. Il tutto alla luce delle possibilità offerte da HTML e XML, mappe topiche e Web semantico.

Questa raccolta di contributi, come indicano i curatori, intende soddisfare tre scopi. Il primo è quello di rendere familiari o ricordare alla comunità della Library and Information Science la storia dei thesauri e gli standard per la loro realizzazione. In apertura del volume, Jean Aitchinson<sup>5</sup> e Stella Destre Clarke<sup>6</sup> raccontano la storia dell'utilizzo del termine "thesaurus", dalla prima volta in cui è apparso nel contesto dell'information retrieval fino alle sfide odierne dell'interoperabilità e dell'usabilità da parte dell'utente finale.

Il secondo scopo del volume è quello di fornire un'ampia bibliografia ragionata e dei

tutorial pratici per la costruzione, l'utilizzo e la valutazione dei thesauri. In *Teach yourself thesauri* Alan R. Thomas<sup>7</sup> indica le competenze necessarie e gli strumenti per la creazione di un thesaurus, segnalando vari riferimenti bibliografici utili. James R. Shearer<sup>8</sup> suggerisce una procedura a nove fasi per costruire un thesaurus in modo sistematico, con esercizi per passare dalla teoria alla pratica. Un altro approccio viene offerto da Marianne Lykke Wielsen,<sup>9</sup> che analizza i punti chiave per la creazione dei thesauri e le metodologie per acquisire conoscenze sull'ambiente informativo tramite l'analisi del comportamento degli utenti. Leonard Will<sup>10</sup> ci presenta invece il ruolo e la funzione del consulente nelle varie fasi di progettazione e creazione di un thesaurus, mentre Leslie A. Owens<sup>11</sup> e Pauline A. Cochrane<sup>12</sup> descrivono i metodi di valutazione applicabili nella creazione *ex novo* o nella revisione di un thesaurus già esistente.

Infine, il terzo scopo del volume è quello di affrontare tematiche relative ai thesauri nel mondo digitale e degli ambienti di rete. Jane Greenberg<sup>13</sup> presenta i risultati di uno studio sulla comprensione dell'utente e sull'utilizzo di un thesaurus nella banca dati ABI/INFO. Eric H. Johnson<sup>14</sup> propone un network di servizi di accesso e navigazione a un thesaurus in cui applicare l'XML, mentre in *Tools of the trade: vocabulary management software* Melissa A. Riesland<sup>15</sup> presenta i criteri per valutare gli applicativi per la gestione dei vocabolari controllati. In *Multilingual subject access* Patrice Landry<sup>16</sup> descrive il progetto MACS, in cui vengono collegati i termini equivalenti nei sog-

gettari inglese, francese e tedesco, in modo da permettere all'utente di accedere a un catalogo nella lingua che preferisce. Infine, in un'intervista con i due curatori del volume Amy Warner<sup>17</sup> descrive le attuali iniziative della National Information Standard Organization<sup>18</sup> per la revisione delle ANSI/NISO Z39.19 per i thesauri monolingue.

Nel loro insieme, questi contributi informano il lettore in modo puntuale e aggiornato sulle modalità di costruzione dei thesauri, contestualizzandone la storia e al tempo spesso speculando sul loro futuro nel mondo digitale. Probabilmente le competenze coinvolte nella creazione dei thesauri non smetteranno di essere utili, ma non è detto che essi continueranno a rappresentare lo strumento terminologico preferito negli ambienti di rete.

Sarebbe perciò interessante se, in una prossima pubblicazione approfondita sui thesauri, fossero inclusi anche argomenti quali tassonomie, ontologie e *mind map* (o qualche altro strumento che nascerà nei prossimi anni), tematiche relative alla visualizzazione semantica nell'interfaccia utente dei thesauri e casi di studio in cui l'utilizzo dei thesauri sia efficacemente integrato con altri sistemi di *information retrieval content-based*.

Perla Innocenti

Sistema bibliotecario di ateneo  
Politecnico di Milano  
perla.innocenti@polimi.it

### Note

<sup>1</sup> ANSI/NISO Z39.19-2003: *guidelines for the construction, format, and management of monolingual thesauri*; UNI ISO 2788: *Linee guida per la costruzione e sviluppo di thesauri monolingue*.

<sup>2</sup> Cfr. il volume *MultiMedia information retrieval. Metodologie ed esperienze internazionali di content-based retrieval per l'informazione e la documentazione*, a cura di R. Raieli e P. Innocenti, AIDA, Roma 2004.

<sup>3</sup> M. HEARST, *User interfaces and visualization*, in *Modern information retrieval*, R. Baeza-Yates and B. Ribeiro-Neto (eds.), Addison-Wesley, MA 1999, p. 257-325.

<sup>4</sup> P. HARPRING, *How forcible are right words! Overview of applications and interfaces incorporating the Getty Vocabularies*, in *Museum and the Web conference*, 1999, <<http://www.archimuse.com/mw99/papers/harpring/harpring.html>>.

<sup>5</sup> Creatrice di vari thesauri, autrice del pionieristico *Thesaurus facet* (1969) e successivamente di varie pubblicazioni tra cui il noto manuale *Thesaurus construction and use*.

<sup>6</sup> Consulente per la creazione di thesauri e presidente del Working group of the British Standard Institution.

<sup>7</sup> Visiting associate professor del Pratt Institute di New York.

<sup>8</sup> Senior lecturer della University of Westminster, Londra, e consulente.

<sup>9</sup> Department of Information Studies della Royal School of Library and Information Science di Aalborg.

<sup>10</sup> Consulente nell'ambito dell'information management, specializzato in thesauri e schemi di classificazione, creatore della Willpower Information.

<sup>11</sup> Consulente nell'ambito del knowledge management, costruzione di thesauri e indicizzazione.

<sup>12</sup> Professore emerito della University of Illinois di Urbana Champaign e della Syracuse University, ha fornito consulenza sui thesauri per molte organizzazioni e condotto ricerche per decenni sull'information retrieval.

<sup>13</sup> Assistant professor alla School of Information and Library Science, University of North Carolina at Chapel Hill. Le sue attività didattiche e di ricerca sono focalizzate su problematiche relative ai metadati e alla classificazione.

<sup>14</sup> Knowledge architect, specializzato in Library and Information Science, sociologia e Computer Science.

<sup>15</sup> Consulente per la S&T Consult-

ing di Seattle, specializzata in tassonomie. Tesoriere di una delle sezioni della American Society for Information Science and Technology (ASIS&T), ha sviluppato tassonomie per Singingfish.

<sup>16</sup> Direttore della sezione Subject Indexing della Swiss National Library.

<sup>17</sup> Capoprogetto del Thesaurus Development Team della NISO.

<sup>18</sup> NATIONAL INFORMATION STANDARDS ORGANIZATION, *Developing the next generation of standards for controlled vocabularies and thesauri*, NISO, Bethesda 2003, <<http://www.niso.org/committees/MT-info.html>>.