

## CERL *thesaurus file*

Claudia Fabian  
Bayerische Staatsbibliothek

### Che cos'è?

Il *thesaurus* del Consorzio delle biblioteche europee di ricerca (CERL) esiste dal 1999 ed è liberamente accessibile a tutti sul Web. Viene gestito dal Data Conversion Group a Göttingen: il formato si ispira all'UNIMARC. Comprende tre categorie: nomi di luogo (luoghi di stampa, di pubblicazione), nomi di persone dedite ad attività collegate alla stampa (stampatori, librai) e nomi di persona in generale (autori, curatori, traduttori). Essendo il *thesaurus* ancora in corso di costruzione, sono in discussione ulteriori aggiunte principalmente per enti, nonché per precedenti proprietari e provenienze.

Il *thesaurus* del CERL si occupa esclusivamente – come rivela la scelta delle suddette entità – di libri antichi, che, nel contesto delle attività del CERL, sono libri stampati con pressa a mano, a partire dagli inizi (1450) fino intorno al 1830/40. Il *thesaurus* è pertanto strettamente collegato al database HPB del Consorzio che raccoglie record bibliografici di libri antichi, provenienti da diversi progetti europei di catalogazione, in un singolo e *one-stop shopping* database gestito dal *provider* californiano RGL. La costruzione di questo database, che contiene attualmente 1,5 milioni di record, va avanti dal 1991 con l'intenzione di sviluppare un database comune per il patrimonio culturale europeo di libri antichi che metta insieme virtualmente beni e ricchezze dispersi per l'Europa. Al di là della sua importanza sotto il profilo della catalogazione e del recupero, il database è un mezzo di ricerca altamente specializzato che permette una visione globale del passato della stampa in Europa. Il *thesaurus file* rispetta moltissimo lo spirito e i metodi di questo comune database europeo e pertanto ci può fornire importanti indicazioni metodologiche sull'*authority work*.

Mi occuperò ora del *thesaurus* del CERL sotto questo punto di vista.

### Creazione e logica del *thesaurus* del CERL

1. Il nome “thesaurus” non è arbitrario, ma indica un importante dato di fatto nella catalogazione e creazione di *authority file* in Europa. Le nostre differenti regole nazionali di catalogazione, ma forse anche le nostre differenti tradizioni culturali e quindi le diverse esigenze e aspettative dei nostri utenti, che parlano lingue diverse, fanno sì che l'*authority control* a livello nazionale (oppure regionale o locale) porti a differenti definizioni di voci di autorità, cioè di forme standard.

Un luogo “universale” come Roma, la città santa, è noto sotto almeno tre nomi: Roma (italiano e latino), Rome (inglese e francese – per caso), Rom (tedesco e svedese). E questo è un esempio facile, come noi tutti sappiamo. Immaginate però quante varianti si possono trovare nei libri per questo facile nome di luogo (Romae, ecc.) Non c'è nessuna uniformità, non nella storia, non nella lingua, non nella catalogazione e non nelle aspettative degli utenti. Quello che fa l'*authority control* è definire (di nuovo in modo più o meno arbitrario) una di queste forme come standard e quindi elencare tutte le altre come varianti della prima. Grazie a questo moderno strumento di catalogazione sarebbe teoricamente risolto il problema di definire quale forma è quella standard. Una corretta implementazione di un *authority file* in

un sistema di catalogazione o meglio di ricerca dovrebbe sempre consentire di cercare il nome contenuto nell'*authority record* in *qualsiasi* forma. La forma standard rimane quella rilevante, ad esempio per la presentazione dell'*authority record* o per la presentazione della forma del nome nel record bibliografico. Noi compilatori di cataloghi sembriamo non essere in grado di tradire le nostre aspettative nazionali e di abbandonare le nostre "forme preferite". Il *thesaurus* è una brillante reazione a questo stato di fatto. Il *thesaurus* del CERL ammette tutti gli *authority record* creati da ciascuna agenzia catalografica e mantiene la forma standard proposta, indicando che questa è la forma preferita dall'una o dall'altra agenzia. Lo scopo è di non porre regole nella scelta della forma standard. Per questo motivo si deve dire che il formato del *thesaurus* CERL è solo "ispirato" all'UNIMARC, in quanto quest'ultimo permette solo *una* forma standard, non diverse forme parallele. Ma anche la decisione di adottare una tale strategia, così intelligente dal punto di vista internazionale, non permette di evitare tutti i problemi. Quale forma mostrare per prima all'utente del *thesaurus*? Si è deciso di prendere la prima in ordine alfabetico – almeno questo è qualcosa su cui i compilatori e gli utenti di cataloghi non possono non mettersi d'accordo. I risultati tuttavia non sono sempre convincenti:

"Rom" per "Roma", "Parigi" per "Paris", ma "Firenze" per "Florence". Dovremmo indicare la forma nazionale corretta? Questo potrebbe però in Europa creare nuovi problemi per i luoghi, così come per la maggior parte delle altre entità che nel corso del tempo hanno cambiato nome o confini nazionali. Non c'è nessuna risposta soddisfacente a questo problema multilinguistico e multistorico. I computer si possono accontentare molto più facilmente: basta un numero che può essere usato come standard, ad esempio il codice ISADN. Personalmente sono convinta che i codici numerici nell'ambito dell'*authority control* diventeranno inevitabili. In via di principio, il CERL potrebbe essere l'organo adatto ad attribuire codici numerici a luoghi, stampatori, librai e nomi di persone relativi ai libri antichi.

2.2. Il *thesaurus* del CERL mette insieme *authority record* di provenienza diversa. Questa è l'intera filosofia del CERL: combinare in un unico file distinti file provenienti da progetti diversi e indipendenti. Questo è quanto ha fatto il CERL creando il file HPB. Questa metodologia risulta interessante per il lavoro di *authority* a livello internazionale, in quanto, mentre altrove si fanno esperimenti con motori di ricerca per la ricerca incrociata, qui è stata adottata la procedura della fusione fisica di file. Non sarete sorpresi se nutro qualche dubbio sulla razionalità e sull'efficacia di questa impresa. Tuttavia penso che sia fattibile nel campo, tutto sommato limitato, dei libri antichi. Le difficoltà consistono non tanto nella fusione dei file originari oggi possibile usando tutte le risorse computazionali disponibili, quanto piuttosto nell'aggiornamento automatico delle informazioni necessarie nel caso in cui siano stati apportati cambiamenti nei file originari. Secondo lo stato attuale del *thesaurus* del CERL, questa combinazione di dati è possibile per entità quali i nomi di luogo e siamo fiduciosi che lo sia anche per i nomi di persona. Tuttavia occorre ancora un'accurata analisi del file originario e un'accurata conversione del formato nonché un controllo per evitare duplicazioni con il file preesistente.

Per la parte del *thesaurus* relativa ai nomi di luogo, il file è stato inizializzato usando un documento creato originariamente dalla Biblioteca nazionale bavarese e pubblicato come libro. Successivamente è stato integrato con le forme standardizzate provenienti da Stoccolma; si prevede di ampliarlo ulteriormente con le forme di San Pietroburgo. In entrambi i casi, non si tratta di *authority file* ma di liste di nomi che forniscono la forma standardizzata e le varianti. L'integrazione automatica dei file provenienti dal Catalogo delle English Cathedral Libraries e dei nomi di luogo contenuti in EDIT16 è stata meno soddisfacente ma, siccome la quantità di questi dati è relativamente piccola, l'integrazione verrà fatta manualmente. In effetti, gli aggiornamenti manuali generalmente riescono meglio.

La fusione automatica di file deve essere integrata da una laboriosa attività manuale di revisione e non siamo ancora in grado di implementare funzioni di aggiornamento. Per quanto riguarda i nomi di stampatori e librai l'*authority file* dell'Aia, di Parigi e di Zagabria fanno già parte del *thesaurus* del CERL. EDIT16 e le Cathedral Libraries verranno integrati nel prossimo futuro. Qui sorgono grossi problemi perché le catalogazioni, o meglio le definizioni delle forme standardizzate, differiscono di molto. Gli stampatori sono talvolta considerati come nomi di persona, talvolta come enti e ciò si riflette in una diversa struttura. L'uniformità è però essenziale per permettere la fusione automatica.

Per quanto riguarda i nomi di autori, ecc., il primo *input* nel *thesaurus* del CERL viene anche dalla Libreria Reale dell'Aia. Disponibile è quindi l'*authority file* ESTC e il passo successivo sarà quello di includere la parte PND (PersonenNamenDatei) dell'*authority file* "Nomi tedeschi", che è altamente specializzato in nomi presenti nei libri antichi, contenendo tutti i nomi di autori classici e medioevali e tutti i nomi provenienti dalla conversione dei cataloghi prima del 1850.

La mia impressione è che il *thesaurus* del CERL diventerà molto rapidamente uno strumento indipendente per l'*authority work* sui libri antichi. Tutti i file esistenti possono essere utili al suo sviluppo, ma poi dovrà vivere di vita propria, la qual cosa ha senso nell'area particolarmente specializzata della catalogazione di libri antichi, ma non è una soluzione per il lavoro di *authority* in generale.

### **La funzione del *thesaurus* del CERL rispetto all'HPB**

Perché il CERL ha incominciato a costruire questo *thesaurus file*? Ci sono due ragioni complementari: la prima è che sono arrivati, per essere inclusi nell'HPB, molti file che, nel loro originario contesto, erano basati su *authority file*, principalmente per nomi di persona (e di enti) nonché di stampatori e librai. Per quanto riguarda i nomi di luogo, c'era spesso una standardizzazione implicita. Il record bibliografico contiene il nome come si trova nel libro e la sua forma standardizzata.

L'originario link a un *authority file* è mantenuto conservando il numero per la connessione, ma non è di aiuto alla ricerca in quanto questo numero perde il valore di *linking*. Allo stesso tempo tutte le informazioni contenute negli *authority record* relativi andavano perse per il HPB. Solo la forma standardizzata, che è solo *una* forma, veniva mantenuta.

La seconda ragione è il HPB stesso. Come già detto, questo è uno strumento di informazione altamente specializzato; le informazioni contenute nei record vengono accuratamente conservate, trasformate e sono utilizzabili in maniera ottimale nella ricerca sui libri antichi, molto meglio di quanto si possa fare con i nostri OPAC. Tuttavia questo sofisticato sistema di ricerca non può da solo venire a capo delle differenti tradizioni di catalogazione importate nei record. Il *thesaurus* è così stato concepito per aiutare direttamente l'utilizzatore dell'HPB a muoversi nella molteplicità delle diverse catalogazioni e a fornirgli il più completo insieme di risultati, senza pretendere da lui l'intero lavoro intellettuale (come per esempio conoscere tutte le varianti di un nome di luogo).

Per il momento si è deciso di *non* implementare alcuna nuova struttura di *linking* tra il record del *thesaurus* del CERL e i record bibliografici nell'HPB. Il *thesaurus* del CERL rimane un database completamente distinto che può essere gestito in qualsiasi luogo e su qualsiasi sistema. La cosiddetta "ricerca assistita" offerta dal HPB porta l'utilizzatore nel *thesaurus* del CERL, dove può identificare il record che gli interessa. Tutte le forme rilevanti del nome sono quindi trasmesse al HPB per effettuare una ricerca più completa.

A prima vista questa procedura è soddisfacente e permette risultati di ricerca migliori. Tuttavia non sono convinta che questa sia la risposta definitiva ai problemi di ricerca nel HPB. Alcuni di questi non possono essere risolti senza una struttura fissa di *linking*.

Questo risulta da tre esempi tratti dai nomi di luogo:

Nomi di luogo fittizi non possono far parte di questa ricerca assistita perché un nome fittizio come “Eleutheropolis” può essere usato per diversi luoghi reali e perché denominazioni del tipo “Rom i.e. Schwabach” darebbero risultati sbagliati.

Omonimi creano grossi problemi e falsano i risultati di ricerca.  
Frankfurt – Frankfurt am Main, Frankfurt an der Oder

Le varianti omonimi creano la stessa confusione.  
Lugdunum: Lyon, Leyden

#### **4. Funzione del *thesaurus* del CERL quale mezzo di informazione in sé**

Anche se la ricerca assistita non funziona perfettamente, è tuttavia di aiuto. Ma il vero valore del *thesaurus* del CERL consiste nella sua funzione quale mezzo di informazione in sé – come uno strumento che ci dà la possibilità di mettere insieme e redigere informazioni di *authority* per i libri antichi in un unico contesto rilevante per l'*authority* e forse di interesse scientifico. Tutto ciò offre eccezionali possibilità di valorizzare le informazioni contenute negli *authority record* combinando diversi strumenti di informazione.

Alcuni primi esempi sono già disponibili. La parte del *thesaurus* del CERL relativa ai nomi di luogo è stata costruita sulla base di una pubblicazione della Biblioteca nazionale bavarese del 1991 che doveva essere aggiornata. L'aggiornamento viene fatto all'interno del *thesaurus* del CERL e noi trarremo profitto dai miglioramenti offerti da altri file. Stiamo ora cercando di sviluppare sistematicamente questo file, come già per gli autori classici e medioevali. Anche in questo caso il numero di entità da prendere in considerazione è limitato e si può fare il lavoro redazionale estraendo nomi da fonti di riferimento, spesso del diciannovesimo secolo (un'epoca in cui si è fatto anche molto lavoro specializzato di *authority*). Fino ad ora sono stati aggiunti al file circa 900 nuovi *authority record* di nomi di luogo. Aggiornare significa anche trovare forme alternative di nomi, indicare fonti di riferimento nel record e fornire altre informazioni importanti sul luogo. Una particolare attenzione è rivolta ai nomi di luogo fittizi. Così il *thesaurus* file dei nomi di luogo può diventare uno strumento prezioso di informazioni sui luoghi dei libri antichi.

Uno degli scopi principali è ampliare questo file aggiornato con le coordinate geografiche contenute nel *Getty place name thesaurus file*. Questo si potrà fare automaticamente una volta che il nostro file sia abbastanza completo. Con queste coordinate si può creare una carta elettronica che potrebbe liberarci dal problema terribile, dal punto di vista storico, dei codici nazionali e forse potrebbe fornirci risultati migliori in ricerche del tipo “la stampa del XVI secolo in Francia”.

Informazioni storico-geografiche sono anche fornite da altri progetti non limitati alla storia della stampa. La Baviera, per esempio, sta creando uno strumento elettronico per la sua storia regionale. Perché non collegare queste informazioni ai record del CT cosicché si possa passare da uno strumento all'altro?

Uno di questi elementi di valorizzazione è già diventato realtà all'interno del CT. Avendo cercato un nome di luogo è possibile vedere i nomi degli stampatori e dei librai che lavoravano in questo luogo. Lo strumento necessario è a disposizione: è stato realizzato collegando internamente il file dei nomi di luogo con quello degli stampatori e dei librai. Questo è possibile se il nome di luogo è contenuto nel record degli stampatori. Non tutti i file originali avranno questa caratteristica: in questo caso sarà necessario un lavoro manuale di redazione.

I record degli stampatori e dei librai permettono anche ulteriori collegamenti. Qui abbiamo cominciato a prendere in considerazione la possibilità di collegarci ai ricchi progetti italiani di digitalizzazione delle marche. Anche se le immagini non portassero direttamente all'edizione in cui la marca si trova (questa informazione è disponibile all'interno dei progetti originari), potrebbero fornire ulteriori informazioni utili.

*Last but not least*, questo potrebbe essere un luogo appropriato in cui accogliere o collegare informazioni importanti sui caratteri di stampa, che al momento sono contenuti in volumi tradizionali di grande valore, che potrebbero venire digitalizzati e soddisfare, così, vecchie esigenze in una nuova forma.

Per quanto riguarda i nomi di persona, la situazione potrebbe essere diversa. In questo caso, le agenzie nazionali potrebbero essere molto più vicine al mantenimento dell'informazione, ma tuttavia, perché non immaginare uno dei miei progetti preferiti, quello di collegare i record dei nomi degli autori classici e medioevali contenuti in PAN e PMA (che in realtà appartengono all'intera Europa, se non addirittura oltre) a copie digitalizzate di quelle fonti citate nei record? Ciò sarebbe di grande aiuto all'identificazione del corretto *authority record* per ogni autore. In questo campo l'attribuzione di un numero standard sarebbe anche preziosa, perché la discussione su qual è la forma standard migliore non finirà mai e l'imposizione di regole non risolve il problema.

Concludendo: Avrete notato che con queste ultime considerazioni ho cercato di rendere giustizia all'invito rivoltomi, che era di parlare dell'*authority file* "Nomi di persona nel medioevo" (PMA), a cui il mio nome è legato – per lo meno nella mente di Mauro Guerrini, che colgo l'occasione di ringraziare per l'invito. Parlare del *thesaurus* del CERL mi ha offerto l'opportunità di riferire su un'attività in corso di svolgimento in cui sono tuttora coinvolta e di parlare di riutilizzazione, riorganizzazione e valorizzazione del lavoro attuale in tema di *authority control*. Credo fermamente che l'*authority work* sia il modo migliore per schiudere le ricchezze dell'universo bibliografico a coloro che hanno bisogno dei nostri servizi. Dobbiamo fare tutto questo in modo razionale, unendo le nostre forze, sfruttando al meglio quanto già esiste e imparando dalle esperienze altrui.

Mi rallegro che questa conferenza ci abbia offerto una occasione ideale per raggiungere questo scopo.