

Notazione e interfaccia di ricerca per una classificazione a livelli

CLAUDIO GNOLI - GABRIELE MERLI

Negli anni Sessanta il Classification research group studiò la possibilità di uno schema di classificazione generale "a faccette libere", basato sulla teoria dei livelli di integrazione. Uno schema di questo tipo è stato da noi sperimentato per indicizzare una bibliografia di cultura locale. Sono presentate in particolare le caratteristiche della notazione, progettate per facilitare il recupero di descrizioni bibliografiche in ambiente digitale. Il sistema è stato sperimentato realizzando un'interfaccia di ricerca via Web scritta in linguaggio PHP, che estrae i dati registrati in una base-dati MySQL. La struttura a livelli permette di esprimere le relazioni fra concetti sia per semplice giustapposizione dei corrispondenti simboli, sia in modi più sofisticati, compresa la relazione di dipendenza che sussiste tra fenomeni di livelli di integrazione diversi, quali "piante" e "boschi" o "montagne" e "alpinismo".

Parole chiave: Classification research group – Classificazione a faccette libere – Interfacce di ricerca – Livelli di integrazione – Notazione – PHP – Relazioni tesaurali

Classificazione a livelli

Il Classification research group (CRG) si è costituito a Londra nel 1952, e nei decenni successivi ha condotto importanti ricerche sui fondamenti della classificazione bibliografica. Assumendo come punto di partenza l'analisi a faccette introdotta da Ranganathan, il CRG da un lato l'ha applicata alla realizzazione di schemi speciali e di uno schema generale (la seconda edizione della Bliss classification), dall'altro ne ha esplorato possibili evoluzioni, introducendo nozioni tratte dalla teoria dei livelli di integrazione, dalla teoria generale dei sistemi e dalla grammatica generativa trasformazionale (Austin, 1976). Il prodotto più tangibile delle ricerche innovative è il sistema di indicizzazione per soggetto PRECIS; ma gli studi hanno riguardato anche la creazione di un nuovo schema di classificazione generale fondato, invece che sulle tradizionali divisioni disciplinari, su un ordinamento dei fenomeni secondo una sequenza di livelli di integrazione.

La teoria dei livelli di integrazione (*integrative levels*) ha numerosi precursori sia tra i pensatori anglosassoni (Blitz, 1992) che tra quelli di area germanofona (Poli, 2001), ma fu formulata più esplicitamente dal filosofo statunitense James Feibleman (1954). Essa afferma che la realtà è strutturata in una successione di livelli di cre-

sciente complessità: atomi, molecole, organismi, società, tecnologie, culture... Ciascun livello fonda la propria esistenza sui livelli precedenti, ma se ne differenzia per il possesso di nuove proprietà emergenti, non riducibili ai livelli precedenti: ad esempio, gli organismi sono costituiti di molecole, ma l'essere vivo o morto è una proprietà degli organismi, che non è applicabile alle molecole.

La teoria dei livelli fu presentata da Douglas J. Foscett nei lavori del CRG, e la sua applicazione venne esplorata nel corso degli anni Sessanta grazie a un finanziamento della NATO (Austin, 1969; CRG, 1969; Foscett, 1970; Foscett, 1978; Gnoli - Poli, 2004; Spiteri, 1995). Si arrivò a produrre una bozza di schema così strutturata:

- a-b termini relativi: es. molto, superiore
- d-f termini posizionali: tempo, spazio, persona
- g-q proprietà: forme, aspetti, suoni, sapori, odori, stati, strutture...
- r-x attività: es. resistenza, cessazione, velocità, equilibrio, movimento, assemblaggio...
- A sistemi generali
- B energia e fenomeni
 - B2 forza
 - B3 gravità
 - B4 suono
 - B5 calore
 - B6 energia elettrica
 - B8 particelle elementari
- C materia
 - C2 sistemi di particelle integrate [atomi, ...]
 - C3 stati molecolari
 - C6 elementi
 - C7 composti
 - C8 elementi individuali, loro stati, composti ecc.
- D sistemi minerali
- E sistemi alla base della vita
- G universo astronomico
- H-Z sistemi geo-centrati
- H terra come ambiente
- J atmosfera
- K idrosfera
- L geosfera
- M sistemi viventi geo-centrati
- N virus

Q piante
 R animali
 S uomo
 ...

L'idea chiave è che il concetto di un dato fenomeno della realtà, es. "rame", conservi sempre la stessa notazione indipendentemente dal contesto disciplinare in cui è impiegato, es. chimica, industria mineraria, tecnologia, arte... Tale notazione deve essere definita al livello di integrazione più basso al quale il fenomeno è riconoscibile, ossia a quello che Jason Farradane chiamò *place of unique definition*: nel caso del rame, quindi, il livello degli elementi chimici. In questo modo, ciascun concetto costituisce un isolato liberamente combinabile con qualsiasi altro. La combinazione avviene con tecniche simili alle faccette in una classificazione disciplinare, ma si estende liberamente attraverso l'intero schema: Derek Austin (1976) parla perciò di *freely faceted classification*, e in un senso simile J. C. Gardin (1965) di *free classification*. La combinazione può avvenire con un sistema di operatori sintattici scritti tra parentesi tonde (da cui deriveranno in seguito gli operatori del PRECIS):

C88	rame
U66C88	tappi di rame
V67(2)U66	tappi per bottiglie
V67(5)z96T3	lavaggio automatico di bottiglie

Secondo Austin questo tipo di notazione massimamente espressiva è particolarmente adatto alla ricerca in ambiente digitale (cfr. anche Gardin, 1965; Gnoli, 2003), mentre risulterebbe meno adeguato agli scopi tradizionali dell'ordinamento di libri a scaffale. Comunque dopo il 1970 la ricerca non venne ulteriormente sviluppata, sia per l'esaurimento dei fondi che per i nuovi impegni di Austin con la *British national bibliography*. Inoltre nel CRG, descritto dallo stesso Austin come «un'eterogenea raccolta di individualisti», inevitabilmente coesistevano idee e approcci diversi, tutti stimolanti ma non sempre convergenti. Le ricerche del Gruppo costituiscono perciò un patrimonio non ancora interamente studiato e sfruttato: appare quindi interessante sperimentarne possibili applicazioni e sviluppi, come ci si propone di fare nel presente lavoro.

Dopo qualche prova embrionale, condotta con Lorena Zuccolo sulle collezioni del Laboratorio regionale di Educazione ambientale del Friuli Venezia Giulia, uno schema a livelli è stato utilizzato per indicizzare una bibliografia sulla cultura locale di un'area dell'Appennino settentrionale, e permetterne l'interrogazione attraverso

un'interfaccia web (Merli *et al.*, 2004). Le limitate dimensioni della bibliografia (attualmente circa 400 documenti fra monografie, articoli e filmati) e la varietà dei soggetti in essa contenuti si prestano alla sperimentazione di una struttura generale e del suo possibile sfruttamento con strumenti di recupero digitali.

Struttura della notazione

Le classi principali di una classificazione a livelli, dunque, non sono discipline di studio, bensì fenomeni di vari livelli di integrazione. I livelli sono disposti in una successione crescente di complessità, cosicché ogni livello è identificato da una lettera successiva nell'ordine alfabetico alle lettere dei livelli inferiori da cui dipende. Conservando la sostanza dello schema abbozzato dal CRG, le classi principali sono state riformulate come segue:

- B energia, particelle elementari
- C tempo
- D spazio
- E atomi
- F molecole
- G masse, oggetti
- H corpi celesti
- I strutture cosmiche: sistemi planetari, galassie...
- J atmosfera
- K litosfera: rocce...
- L strutture geomorfologiche: monti, fiumi...
- M organismi unicellulari: batteri...
- N organismi pluricellulari: funghi, piante, animali...
- O società
- P ecosistemi
- Q mente
- R linguaggio
- U artefatti tecnologici
- V lavoro, economia
- W organizzazioni, istituzioni
- X arte, attività del tempo libero
- Y cultura

Le sottoclassi sono espresse da ulteriori lettere successive alla prima:

Nf funghi pluricellulari
Nl alghe pluricellulari
Nm muschi
Np piante vascolari
 Npb felci
 Npc gimnosperme: conifere...
 Npf piante con fiore
Nq animali

Si è preferito utilizzare sempre lettere, invece di cifre come nella bozza del CRG, per esprimere il fatto che la natura delle sottoclassi non è nettamente differenziata da quella delle classi principali: anzi, un'analisi più accurata potrebbe portare a decidere, supponiamo, che gli animali sono tanto differenziati per complessità dagli altri organismi pluricellulari da meritare lo *status* di classe principale, e quindi spostarli da Nq a O, senza con questo infrangere la logica generale del sistema. Anche a livello della seconda e della terza lettera, come si vede, i fenomeni sono disposti tendenzialmente in ordine crescente di complessità: la dimensione verticale dello schema esprime quindi complessità crescente, mentre la dimensione orizzontale esprime (come nelle classificazioni tradizionali) specificità crescente.

L'attribuzione dei simboli è per il momento intuitiva, al solo scopo di sperimentare il funzionamento di questo tipo di sistema. In un approccio più preciso, l'albero di classificazione sarebbe rappresentabile matematicamente come un grafo orientato, nel quale non mancano casi di *assorbimento* (un livello si origina in effetti dall'interazione di più livelli inferiori: ad esempio gli ecosistemi comprendono costituenti sia inorganici che organici); la teoria matematica dei codici offrirebbe allora metodi per l'attribuzione della notazione appropriata a ciascun nodo del grafo (Gnoli - Doldi, 2005).

Le classi così definite possono essere combinate, nel modo concepito dal CRG, giustapponendo la notazione dei diversi fenomeni in relazione fra loro. Ad esempio, un documento sulla fauna (Nq) di montagna (Lm) può essere indicizzato come Nq Lm. Per facilitare il recupero automatico in ambiente digitale, i simboli di ciascun fenomeno componente sono stati scritti separati da uno spazio. Le iniziali maiuscole sono una scelta di comodo per riconoscere le parti componenti, sia visivamente che in fase di recupero automatico in modalità *case-sensitive*. Tecnicamente, spazi e maiuscole svolgono la stessa funzione in modo ridondante: non sarebbe difficile realizzare applicazioni che funzionino utilizzando soltanto la prima o soltanto la seconda di queste caratteristiche.

L'abbandono delle discipline come classi principali non significa che le discipline non siano trattabili da un sistema di questo tipo. Quando fosse necessario indicizzare documenti che trattino, ad esempio, della botanica in quanto scienza dal punto di vista storico, sociale ecc., si potrà utilizzare una notazione definita al livello della cultura, ottenuta per combinazione di Y_S "scienza" con la notazione dell'oggetto di studio tipico della disciplina, come $Y_S N_P$. In questo modo, lo schema dei fenomeni e quello delle discipline possono essere collegati in un rapporto *a fisarmonica* (Gnoli, 2005).

I limiti dell'analisi

In questo sistema, i soggetti dei documenti sono dunque rappresentati mediante la combinazione di concetti che li compongono, appartenenti a diversi livelli di integrazione. Alcuni concetti possono essere formati per composizione di concetti più semplici: essi possono essere rappresentati unendo le notazioni dei concetti semplici con la lettera riservata *a*: così per il concetto "toponimi" è stato usato il simbolo $Ro a D$, composto di Ro "parole" e D "spazio".

La scomposizione dei concetti in elementi più semplici fu teorizzata già da Lullo e poi da Leibniz: rappresentando ogni concetto come una combinazione di simboli degli elementi componenti, quei filosofi vagheggiavano una vera e propria algebra del pensiero, *l'ars combinatoria*. A questo riguardo si pone però il problema di un criterio per identificare gli elementi semplici. Il concetto di semplice dipende infatti dallo stato delle conoscenze presenti: per esempio, la parola *idrogeno* significava originariamente "produttore di acqua", ed era quindi definita nei termini di una sostanza tradizionalmente ritenuta semplice (addirittura, nell'antichità, un costituente fondamentale del mondo); ma successivamente lo sviluppo della chimica ha portato a rendersi conto che al contrario è l'acqua ad essere un composto di idrogeno e ossigeno, due sostanze più fondamentali, ossia appartenenti a un livello di integrazione inferiore, sebbene meno direttamente percepibili nell'esperienza quotidiana. Questa conoscenza si riflette nella formula chimica dell'acqua, scritta come una combinazione di idrogeno e ossigeno.

Un moderno lulliano, a questo punto, potrebbe ingenuamente proporre di rappresentare ogni cosa come una combinazione di elementi chimici. Ma chiaramente anche questa non sarebbe che una soluzione provvisoria: in realtà anche gli atomi degli elementi sono costituiti da particelle di livello inferiore, quali protoni ed elettroni, i protoni a loro volta sono costituiti da quark... È evidente quindi che un linguaggio esclusivamente analitico offrirebbe rappresentazioni troppo dipendenti dalle conoscenze del tempo, oltre che poco convenienti (come osservò Bertrand

Russell, esse risulterebbero terribilmente prolisse). Non solo: nella prospettiva della teoria dei livelli di integrazione, esse peccherebbero di riduzionismo, riducendo l'essenza di un fenomeno alla somma delle sue parti componenti e trascurando invece le proprietà emergenti caratteristiche del nuovo livello di integrazione: l'acqua è caratterizzata dall'essere liquida tra 0 e 100 gradi Celsius, una proprietà che non si ritrova nei singoli atomi di cui è costituita.

Lo stesso Foskett intuì questo aspetto degli schemi analitici: egli osserva che la situazione è simile a quella dei dialetti creoli, nei quali gli oggetti nuovi come il pianoforte sono inizialmente descritti con locuzioni composte da serie di molte parole più comuni, come *scatola-tu-manovrare-lei-piangere*, ma col tempo i nuovi concetti tendono a stabilizzarsi ed essere denotati da un termine apposito (Foskett, 1961). È quindi opportuno che i fenomeni di un livello superiore, caratterizzati da proprietà emergenti, non siano rappresentati con combinazioni di simboli dei livelli componenti inferiori, ma con nuovi simboli. L'acqua, nel nostro schema, avrà un proprio simbolo al livello dei composti molecolari. Solo le combinazioni di concetti che non formino nuovi livelli di integrazione saranno rappresentabili con composti, come nel caso di "toponimi", o con semplici giustapposizioni di classi separate da uno spazio, nel caso di relazioni accidentali come "fauna di montagna".

Classe preferita

In una bibliografia specializzata, è naturale che ricorrano con particolare frequenza alcune classi che in uno schema generale corrisponderebbero a simboli molto specifici. Nel caso della bibliografia da noi indicizzata, la specificità risiede in un'area geografica, della quale sono coperti tutti gli aspetti naturali, sociali, storici e culturali. La bibliografia infatti fa parte di un sito web dedicato al territorio in questione, un'area montuosa a cavallo tra Lombardia, Emilia, Piemonte e Liguria collettivamente indicata come "le Quattro Province".

La notazione completa per il territorio in questione e le sue sottoclassi sarebbe costituita da molte lettere e cifre, appartenendo a una catena gerarchica come "Territorio geografico. Terre emerse. Entità individuali. Europa. Penisola italiana. Regione appenninica. Tratto pavese-genovese". Conviene allora riassumere questo valore specifico definendolo come la *classe preferita* (Ranganathan, 1967, capitolo DG), rappresentata nel nostro schema dalla cifra 9. Le sue sottoclassi saranno per definizione sottoinsiemi del territorio considerato:

9 Quattro Province

9b bacino dello Scrivia

9 bk bacino del Borbera
 9d bacino del Curone
 9f bacino dello Stàffora
 9r bacino del Tidone
 9t bacino del Trebbia
 9tn bacino dell'veto

Attributi

Nella bozza del CRG, le classi di *entità* rappresentate da lettere maiuscole sono precedute da altre classi con lettere minuscole. Queste contengono *termini relativi*, *termini posizionali*, *proprietà*, *attività*, ossia tutti i concetti che non corrispondono ad entità stabili ma a loro *attributi*. Austin riconosce che gli attributi possono essere disposti in una sequenza di specificità crescente (Austin, 1969; CRG, 1969): ad esempio l'attributo "rosso" è applicabile a una maggior varietà di entità che l'attributo "ligneo", e pertanto nelle tavole deve essere elencato prima.

Tuttavia questi aspetti dello schema vengono da Austin solo schizzati, dichiaratamente per mancanza di tempo. In effetti, nel quadro dell'approccio a livelli appare coerente che anche gli attributi siano definiti a determinati livelli di integrazione: il concetto di "femmina", per esempio, poggia la sua essenza sul livello degli organismi viventi, e non ha senso applicarlo a livelli inferiori; analogamente, una relazione come "stipulare un contratto" ha senso al livello economico. Ciò suggerisce che gli attributi, anziché essere elencati separatamente, vengano integrati nella sequenza generale dei livelli, identificandoli al loro interno con un apposito indicatore. Come indicatori sono state scelte delle cifre, finora libere da altri usi nella presente sperimentazione. Ad esempio, poiché come si è detto il sesso è una *proprietà* del livello degli organismi, si avrà:

N organismi
 N3 proprietà degli organismi
 N3s sesso
 N3sf femmine
 N3sm maschi

Le *parti* sono fenomeni che non esistono indipendentemente, ma solo in quanto appartenenti a un sistema più complesso, come i manubri rispetto alle biciclette. Austin nota che perciò non devono essere trattate come livelli autonomi, ma come

"sottolivelli" (CRG, 1969; Gnoli - Poli, 2004). Anche le parti dunque sono definibili in corrispondenza di un determinato livello:

N organismi
 N4 parti degli organismi
 N4r sistema respiratorio
 N4s sistema circolatorio
 N4sd arterie
 N4sh cuore

La categoria delle *attività*, utilizzata dal CRG in modo analogo all'energia di Ranganathan, sembra essenzialmente associata alle idee di cambiamento e movimento: queste a loro volta sono connesse al tempo, che, per quanto costituisca un livello assai fondamentale, non è preesistente a ogni altro fenomeno, come evidenziato dagli sviluppi recenti della fisica. Perciò è sembrato preferibile utilizzare, invece che attività, una più generale categoria di *relazioni*, ossia concetti che stabiliscono un nesso tra altri concetti. Le relazioni sono infatti definite a un livello estremamente fondamentale, in logica, immediatamente dopo le stesse classi o insiemi. Anche in questo caso, comunque, esistono relazioni particolari definite a specifici livelli:

N organismi
 N2 relazioni degli organismi
 N2g generare, relazione genitore-figlio

Le relazioni reggono implicitamente degli argomenti, formando così delle costruzioni sintattiche. Queste potranno essere esplicitate, quando sia opportuno, al posto delle associazioni più generiche di concetti espresse dal semplice spazio bianco:

Nqvr N3sm rettili : maschi
 (Nqvr ; N3sm)N2g generazione di figli maschi da parte di rettili

Le cifre che rappresentano gli attributi sono state scelte con l'intento di produrre una successione utile tra le diverse categorie, secondo il buon principio classificatorio che il generale preceda lo specifico. La successione relazioni-proprietà-parti è apparsa al momento come quella più conveniente; le proprietà precedono le parti anche nell'ordine di citazione standard del CRG, introdotto da Brian Vickery e applicato a molti schemi. Inoltre gli attributi caratteristici di un certo livello prece-

deranno nell'ordinamento alfabetico le sottoclassi di quel livello, poiché in ambiente digitale le cifre vengono automaticamente ordinate prima delle lettere. Ancora precedenti risulteranno i segni di punteggiatura e altri simboli, di cui si potrebbe valutare l'impiego per altri scopi, sebbene sia auspicabile che la notazione non risulti eccessivamente complessa.

Ordine di citazione

Per gli scopi dell'indicizzazione è opportuno che siano definite regole per l'ordine di citazione, ossia per la successione delle classi che combinandosi formano il soggetto di ciascun documento. In questa sede continua ad apparire valida la logica della classificazione a faccette: in particolare, il *principio di inversione* prescrive che per ciascun documento le classi siano disposte in ordine inverso rispetto a quello delle tavole. Questo significa che in prima posizione (corrispondente all'entità o alla personalità degli schemi a faccette) verrà a trovarsi il fenomeno componente di più alto livello, seguito da quelli dei livelli man mano inferiori. I composti risultanti sono formati da un insieme di classi liberamente combinate: a differenza degli schemi a faccette classici, però, le classi non sono rigidamente costrette entro categorie standard, bensì qualsiasi classe può funzionare come una faccetta.

Per il campione di bibliografia utilizzato, la semplice giustapposizione delle classi componenti in ordine invertito è risultata nella maggior parte dei casi abbastanza chiara, producendo successioni di soggetti coerenti e significative:

Xiino Qy	pifferi : personalità
Xiino Qy 9f	pifferi : personalità : valle Staffora
Xiino Um Qy 9x	pifferi : artigianato : personalità : val Fontanabuona
Xiino Xic Qy 9f	pifferi : canto : personalità : valle Staffora
...	
Xn Xiino	ballo : pifferi
...	
Yt 9f	eventi tradizionali : valle Staffora
Yt Ch 9f 9r	eventi tradizionali : storia : valle Staffora, val Tidone
Ytl Ule 9f	carnevale : paesi : valle Staffora
Ytl Xn 9f	carnevale : ballo : valle Staffora
Ytm 9r 9t	maggio : val Tidone, val Trebbia

Solo in qualche caso, la semplice combinazione delle classi componenti è risultata oscura e ambigua. Il volume di Luigi Heilmann *La parlata di Portàlbera e la*

terminologia vinicola nell'Oltrepò pavese risulterebbe così indicizzato:

Ufr Ry9e Ro Npfvv 9f-p agricoltura:dialetti lombardi:parole:viti:Oltrepò pavese

In questi casi, la successione più significativa ("agricoltura" dovrebbe trovarsi vicino a "viti") non corrisponde al semplice ordine invertito, e andrebbe quindi espressa con relazioni che regolino i propri argomenti secondo un ordine definito (relazioni asimmetriche, quale "generare" nell'esempio visto sopra) o con classi composte.

Altri casi in cui l'ordine di citazione invertito appare inadeguato si presentano quando due concetti, invece che genericamente associati, siano coordinati da una congiunzione, come in "geologia e botanica della val Trebbia". Qui (come in alcuni casi del campione sopra riportato) appare opportuno che essi vengano citati secondo lo stesso ordine delle tavole, come se i fenomeni coordinati venissero elencati nell'ambito di uno stesso ruolo sintattico:

K Np 9t rocce, piante : val Trebbia

Realizzazione dell'interfaccia di ricerca

Per la gestione e l'interrogazione della bibliografia è stata creata una base-dati utilizzando il DBMS MySQL 3.2. Questa scelta è stata motivata dalla sua leggerezza e velocità, e dal fatto che si presta bene ad attività orientate al Web; esso inoltre è rilasciato sotto licenza *open source*.

Sono state create due tavole, una per i documenti e una per lo schema di classificazione. La tavola dei documenti comprende campi corrispondenti ai principali elementi della descrizione bibliografica (titolo, autore, editore, anno, ecc.) nonché un campo per i simboli di classe associati a ciascun documento; quest'ultimo campo contiene uno o più simboli, elencati in ordine invertito e separati da spazi, come negli esempi visti sopra. La tavola dello schema di classificazione comprendeva inizialmente solo due campi, uno per la notazione e uno per gli equivalenti verbali; successivamente è stata ampliata con altri campi per le funzioni di classauro descritte nel séguito. Non sono stati generati *authority file* per elementi diversi dalla classe, in quanto la forma delle intestazioni descrittive è secondaria ai fini della presente sperimentazione. Nella creazione delle tavole non è stato necessario attenersi scrupolosamente al rigore in fase di progettazione che sarebbe imposto da altri DBMS come Oracle; questo ci ha permesso una grande flessibilità, indispensabile

in questa fase di sviluppo e sperimentazione, non precludendo comunque la possibilità di un maggior rigore in future fasi.

Per creare l'interfaccia alla base-dati è stato usato il linguaggio di *scripting* PHP [Hypertext preprocessor], il più usato nella creazione di pagine web dinamiche, da noi scelto per le sue doti di semplicità e immediatezza, nonché per le sue potenti funzionalità di elaborazione di stringhe (*parsing*). A questi due strumenti si aggiungono l'uso della piattaforma Linux e del *server web* Apache, andando così a formare quell'infrastruttura denominata LAMP, dalle iniziali dei quattro strumenti, che ha permesso in tutto il mondo lo sviluppo di *software open source* e ha stimolato la creatività di numerosissimi programmatori.

L'interfaccia sviluppata si articola in una decina di pagine PHP. La pagina principale <index.php> offre diverse possibilità di ricerca: per titolo, per autore, per classe o per equivalente verbale; le ultime due sfruttano lo schema qui presentato. Nella ricerca per classi l'utente seleziona una classe dalla lista di quelle disponibili, mentre nella ricerca per equivalente verbale egli inserisce un termine corrispondente al soggetto sul quale desidera trovare documenti.

La ricerca per classi utilizza delle semplici *query* che hanno il compito di rilevare un determinato *pattern* corrispondente alla classe componente ricercata, es. *Np*, all'interno del campo dei simboli di classe nella tavola dei documenti. La *query* non si limita a ricercare una corrispondenza esatta, ma effettua un *pattern matching* dell'espressione '%Np%', includendo così tra i risultati tutte le combinazioni sia di *Np* che delle sue sottoclassi con qualsiasi altro simbolo, ad esempio *Y Npf*.

Il risultato ottenuto è una nuova pagina <libri.php> che presenta tutti i record restituiti dalla *query*, ciascuno accompagnato dai corrispondenti equivalenti verbali recuperati dalla tabella dello schema di classificazione. Questi ultimi vengono richiamati grazie a operazioni di *parsing* e lettura della stringa classe, che permettono di isolare gli elementi componenti in essa contenuti (*Npf*) e di andare così a recuperarne la descrizione ("piante con fiore").

La ricerca per equivalente verbale estrae dalla tavola dello schema di classificazione le classi che contengono il termine richiesto nel campo equivalente verbale. Dopo aver preso visione delle classi corrispondenti, è possibile sceglierne una e proseguire la ricerca secondo le modalità descritte nel caso precedente.

Classauro

La ricerca per equivalenti verbali, naturalmente, è efficace soltanto nel caso che l'utente richieda termini esattamente coincidenti con quelli utilizzati nella base-dati. Per aumentare le possibilità di recupero, lo schema di classificazione può essere

arricchito da relazioni di rinvio simili a quelle di un tesoro. Si produce così un tipo di strumento che offre la massima ricchezza informativa unendo le funzioni di classificazione e tesoro, e che perciò Bhattacharyya (1982) ha definito *classauro*.

Nella tavola dello schema di classificazione, ai campi destinati alla notazione e agli equivalenti verbali ne sono stati aggiunti altri per i termini affini con cui gli utenti potrebbero cercare gli stessi concetti. Le varianti più ovvie sono quelle morfologiche, come le forme singolari anziché plurali, e quelle lessicali, come sinonimi e quasi-sinonimi. Quanto alle relazioni di associazione dei tesori, qui esse sono state distinte in tre tipi: associazione generica, discipline e dipendenza. Le ultime due conseguono dalla struttura di uno schema a livelli di integrazione.

Poiché le classi sono definite da fenomeni e non da discipline, l'equivalente verbale della classe $N_{\alpha\nu\sigma}$ è "uccelli" e non "ornitologia". Il termine "ornitologia" è allora registrato nel campo per le *relazioni di disciplina*, e utilizzato per convogliare alla classe $N_{\alpha\nu\sigma}$ le ricerche a partire da quel termine. Nel caso che lo schema sia sviluppato a fisarmonica ed esista la classe dell'ornitologia in quanto disciplina ($Y_{\sigma\alpha n\alpha\nu\sigma}$), il termine "ornitologia" si troverà stavolta nel campo dell'equivalente verbale, e una ricerca con "ornitologia" otterrà tra i risultati anche questa classe insieme alla $N_{\alpha\nu\sigma}$.

Infine, le *relazioni di dipendenza* sono quelle tra un fenomeno e i fenomeni di livelli superiori da esso derivati, quando il rapporto non sia già espresso nella notazione. Ad esempio, poiché la notazione per "acqua" non è ottenuta per composizione di "idrogeno" e "ossigeno", tra "idrogeno" e "acqua" potrebbe essere registrata una relazione di dipendenza. Si osserva che l'opportunità di questa relazione dipende dal contesto: in ambito chimico essa può essere rilevante, mentre nello schema presente sarebbe probabilmente poco utile. Casi in cui la relazione di dipendenza è invece utile sono stati osservati anche nella nostra sperimentazione. Il concetto "boschi" è definito al livello degli ecosistemi ($P_{\mathcal{F}}$), in quanto un bosco è costituito da molte componenti organiche e inorganiche; tuttavia le piante ne costituiscono un componente particolarmente importante; perciò nella registrazione di $N_{\mathcal{P}}$ "piante" il campo per le relazioni di dipendenza è stato compilato con la notazione $P_{\mathcal{F}}$. Analogamente si è fatto con $X_{\mathcal{G}\sigma}$ "alpinismo" e $L_{\mathcal{M}}$ "monti". In tal modo, l'interfaccia ha potuto essere programmata affinché l'utente che sceglie la classe $N_{\mathcal{P}}$ "piante" ottenga in risposta sia la classe $N_{\mathcal{P}}$ stessa con tutte le sue sottoclassi, sia la classe $P_{\mathcal{F}}$ "boschi": la *query* di selezione dati cioè ricerca applicando il criterio di *pattern matching* non solo all'interno del campo principale della notazione, ma anche all'interno di quello del campo ausiliario per le relazioni di dipendenza.

Quest'ultima funzione di uno schema a livelli è prefigurata di sfuggita dallo stesso Foskett: «A new feature has to be consciously incorporated in the alphabetical

index, namely, a system of references from certain combinations of terms to at least the level above». A suo avviso, è opportuno che questo tipo di relazione sia registrato in una sola direzione, ossia dal livello inferiore verso quello superiore: «I surmise that only the upward references would be necessary» (Foskett, 1961).

Sviluppi

L'applicazione qui descritta ha soprattutto lo scopo di mostrare le caratteristiche essenziali di funzionamento di uno schema a livelli in ambiente digitale. Essa rientra peraltro in un più ampio progetto di ricerca, illustrato sul sito web del capitolo italiano dell'International Society for Knowledge Organization (ISKO Italia, 2004), comprendente aspetti filosofici, matematici, tecnici e informatici, che potranno essere gradualmente approfonditi in studi distinti.

L'attribuzione della notazione, come si è accennato, è provvisoria e probabilmente migliorabile in vari dettagli. In particolare sarebbe necessario individuare criteri più precisi per stabilire la successione dei livelli e l'attribuzione di un fenomeno a una classe più o meno specifica rispetto ai fenomeni vicini. Inoltre la notazione potrebbe essere arricchita e precisata con altri simboli e funzioni (quantità, relazioni simmetriche e transitive, deittici, differenze specifiche...) una volta chiarita la loro relazione logica con i simboli esistenti e la loro effettiva funzionalità.

Occorrerebbe poi verificare l'efficacia di uno schema a livelli quando sia applicato in maggiore dettaglio a uno specifico settore disciplinare. È diffusa l'opinione secondo cui l'approccio a livelli sarebbe adatto a indicizzare letteratura scientifica (i livelli più bassi) ma non letteratura umanistica: secondo Austin (1976), in quest'ultima assume maggiore importanza la sequenza di combinazione tra le classi, il che farebbe propendere in favore di faccette predefinite. D'altra parte l'approccio a livelli è cresciuto all'interno del CRG come una sorta di sviluppo logico di quello a faccette, ed è immaginabile che i due possano essere conciliati all'interno di un quadro più generale; questo chiamerebbe in causa importanti questioni notazionali. Acquisire esperienza attraverso nuove sperimentazioni sembra essere l'unica via per esplorare questa strada.

Ringraziamenti

A Viviana Doldi per i dialoghi sulla modellizzazione matematica dei livelli; a Lorena Zuccolo per i commenti emersi dalle prime sperimentazioni; a Fulvio Bisi per la creazione del database; a Emanuela Casson e Aida Slavić per le discus-

sioni "semantiche" e la scoperta dei classauri; alla Biblioteca Servizi bibliografici della regione Toscana per l'eccellente servizio di fornitura di documenti.

Bibliografia citata

- Derek Austin, *Prospects for a new general classification*. "Journal of librarianship", 1 (1969), n. 3, p. 149-169
- Derek Austin, *The CRG research into a freely faceted scheme*, in *Classification in the 1970s: a second look*, Arthur Maltby ed. Revised ed. London : Bingley ; Hamden : Linnet books, 1976, p. 159-194
- G. Bhattacharyya, *Classaurus: its fundamentals, design, and use*, in *Universal classification. 1: Subject analysis and ordering systems. Proceedings of the 4th International study conference on classification research, 6th annual conference of Gesellschaft für Klassifikation. Augsburg, June 28-July 2, 1982*, Ingetraut Dahlberg ed. Frankfurt : Indeks, 1982, p. 139-148
- David Blitz, *Emergent evolution: qualitative novelty and the levels of reality*. Dordrecht [etc.] : Kluwer, 1992
- Classification research group, *Classification and information control. Papers representing the work of the Classification research group during 1960-1968*. London : Library association, 1969
- James K. Feibleman, *The integrative levels in nature*. "British journal for the philosophy of science", 5 (1954), n. 17, p. 59-66. Ripubbl. in *Focus on information*, Barbara Kyle ed. London : Aslib, 1965, p. 27-41
- Douglas J. Foskett, *Classification and integrative levels*, in *The Sayers memorial volume*, D. J. Foskett and B. I. Palmer eds. London : Library association, 1961, p. 136-150. Ripubbl. in *Theory of subject analysis: a sourcebook*, Lois Mai Chan, Phyllis A. Richmond, Elaine Svenonius eds. Littleton (Colorado) : Libraries unlimited, 1985, p. 210-220
- Douglas J. Foskett, *Classification for a general index language: a review of recent research by the Classification research group*. London : Library association, 1970
- Douglas J. Foskett, *The theory of integrative levels and its relevance to the design of information systems*. "Aslib proceedings", 30 (1978), n. 6, p. 202-208
- J. C. Gardin, *Free classifications and faceted classifications: their exploitation with computers*, in *Classification research. Proceedings International conference on classification research. Elsinore (Denmark) 1964*, Pauline Atherton ed. Copenhagen : Munksgaard, 1965, p. 161-176

- Claudio Gnoli, *Coordinazione, ordine di citazione e livelli integrativi in ambiente digitale*. "Bibliotime", n. s., 6 (2003), n. 1, <<http://www.spbo.unibo.it/bibliotime/num-vi-1/gnoli.htm>> [consultato in data 2005-01-20]
- Claudio Gnoli, *BC2 classes for phenomena: an opportunity to apply the theory of integrative levels*. "Bliss classification bulletin", 47 (2005)
- Claudio Gnoli - Viviana Doldi, *Grafì e classificazione. Intervento presentato alla Tavola rotonda sul Web semantico: sintesi e proposte. Prato, 26 gennaio 2005*. "E-LIS", 2005 <<http://eprints.rclis.org/archive/00003076>> [consultato in data 2005-02-01]
- Claudio Gnoli - Roberto Poli, *Levels of reality and levels of representation*. "Knowledge organization", 31 (2004), n. 3, p. 151-160
- ISKO Italia, *Integrative level classification: research project*. ISKO Italia, <<http://www.iskoi.org/ilc>>, 2004 [consultato in data 2005-01-20]
- Gabriele Merli - Fulvio Bisi - Claudio Gnoli, *Libri e articoli, in Dove comincia l'Appennino: note culturali e naturalistiche sul territorio delle Quattro Province*, Paolo Ferrari Magà *et al.* <<http://www.iskoi.org/ilc/4p>> [consultato in data 2005-03-29]
- Roberto Poli, *The basic problem of the theory of levels of reality*. "Axiomathes", 12 (2001), n. 3-4, p. 261-283
- Shiyali R. Ranganathan, *Prolegomena to library classification*, with assistance of M. A. Gopinath. Bangalore : Sarada Ranganathan endowment for library science, 1967
- Louise M. Spiteri, *The Classification research group and the theory of integrative levels*. "The Katharine Sharp review", 1 (1995) <<http://www.lis.uiuc.edu/review/summer1995>> [consultato in data 2005-01-20].