

**CONSIGLIO REGIONALE DELLA PUGLIA**  
*BIBLIOTECA MULTIMEDIALE & CENTRO DI DOCUMENTAZIONE*  
**“Teca del Mediterraneo”**

**11° WORKSHOP**

L'organizzazione della conoscenza tra identità e multiculturalità

Seminario

*“L'era digitale in biblioteca: le applicazioni innovative”*

*Terminologie a faccette nel recupero dell'informazione*  
*L'importanza del controllo del vocabolario nei sistemi di informazione*  
*contemporanei*

**Vanda Broughton**  
University College di Londra

*Bari, 23/24 giugno 2008*

Negli ultimi anni l'importanza dei vocabolari controllati come strumenti per conseguire un recupero efficace dell'informazione è stata messa più volte in discussione.

I linguaggi controllati sono spesso visti come strumenti molto complessi da costruire e usare. A causa del fatto che sono nati come strumenti biblioteconomici "tradizionali" vengono considerati inappropriati per un ambiente informatizzato e per di più non necessari, a fronte della potenza nel processare le informazioni dei moderni software di ricerca.

La nostra posizione dissente da ciò. Qui si intende invece mostrare che i **vantaggi** dei vocabolari controllati sono evidenti e dimostrabili: possono migliorare l'efficienza della ricerca; quando vengono usati insieme a tecniche automatiche di indicizzazione o classificazione producono risultati migliori; possono servire a rendere più precisi e a filtrare i risultati delle ricerche; l'utilizzo di vocabolari ben strutturati in un ambiente di ricerca può aiutare nel browsing e nella navigazione; possono essere utilizzati per strutturare e modificare automaticamente le interrogazioni dell'utente.

Al fine di consentire queste funzioni in un ambito digitale è essenziale che i vocabolari controllati siano strutturati in modo formale. Perché una macchina possa capire la struttura del vocabolario le relazioni tra i termini (o concetti) devono essere chiare e ben espresse. Quando ci riferiamo a queste relazioni intendiamo parlare di quelle di equivalenza, gerarchiche e associative come definite per i thesauri o per le classificazioni.

Un buon vocabolario dovrebbe consentire anche di esprimere **argomenti complessi** e quindi non può essere limitato alla gerarchia dei termini ma deve comprendere anche un sistema sintattico. Questo sistema sintattico dovrebbe essere comprensibile da una macchina e il modo in cui concetti o termini sono combinati dovrebbe essere logico, coerente e prevedibile.

In questo contributo il focus sarà volto a dimostrare come **l'analisi a faccette** sia fondamentale nella costruzione di vocabolari controllati e nella strutturazione di domini semantici.

L'analisi a faccette è una metodologia molto potente nell'organizzare concetti, identificare relazioni e strutturare uno spazio informativo. Consente di rendere esplicite le relazioni tra i concetti, fornisce un sistema di sintassi standard che supporta la combinazione dei termini in un modo coerente, può generare vocabolari complessi che sono abbastanza bene interpretabili da una macchina.

Se ben strutturati, i vocabolari a faccette possono essere espressi in differenti modalità:

- classificazione o strutture sistematiche
- thesauri
- liste di soggetti
- ontologie

A mio parere le potenzialità dell'analisi a faccette non sono ancora state completamente esplorate.

Per fornire un primo inquadramento della metodologia, si consideri che l'analisi a faccette in senso classico implica organizzare concetti in un dato settore raggruppando in base a caratteristiche comuni. L'analisi a faccette utilizza delle categorie standard come "contenitori" per i concetti.

Queste categorie standard nell'analisi a faccette classica sono individuate in base a considerazioni linguistiche e funzionali.

Il primo insieme di **categorie** attestato è quello di Kaiser (1911). Le categorie erano usate per produrre liste alfabetiche nell'indicizzazione per soggetti per generare intestazioni di soggetto pre-coordinate.

Un secondo set di categorie è quello elaborato da Ranganathan, ossia personalità, materia, energia, spazio e tempo (PMEST); famoso è poi il set sviluppato dall'inglese Classification Research Group, un'espansione del PMEST.

#### *Categorie CRG*

*Cose*

*Tipi*

*Parti*

*Proprietà*

*Materiali*

*Processi*

*Attività*

*Pazienti (oggetti dell'azione)*

*Prodotti*

*Prodotti intermedi*

*Agenti/Strumenti*

*Spazio*

*Tempo*

Queste ultime categorie sono state impiegate in tutte le classi della seconda edizione della Bliss Bibliographic Classification (BC2). Questa applicazione mostra che si possono applicare a tutti gli ambiti disciplinari, con particolare efficacia in quelli scientifici e tecnologici, che non in tutti i settori è necessario applicare l'intero insieme delle categorie, come pure che in certe discipline (es. arte) sarebbe richiesto aggiungere specifiche categorie (es. forma e genere).

*Esempio di applicazione delle categorie dell'analisi a faccette all'orticoltura.*

*Cose=piante*

*Tipi=alberi, piante floreali, funghi*

*Parti=petali, radici, frutti, foglie*

*Proprietà=verde, velenosità*

*Materiali=cellulosa, pigmenti*

*Processi=crescita, germinazione*

*Operazioni=coltivazione, raccolta*

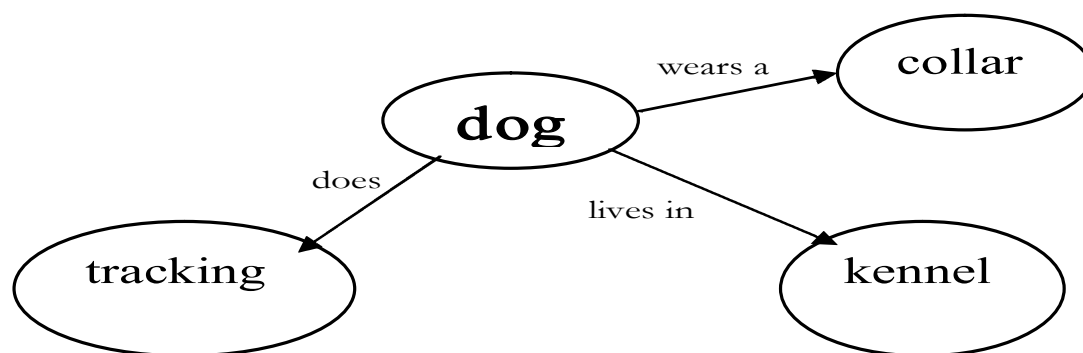
*Prodotti=fibre, droghe*

*Agenti/strumenti=giardinieri, trattori, fertilizzanti*

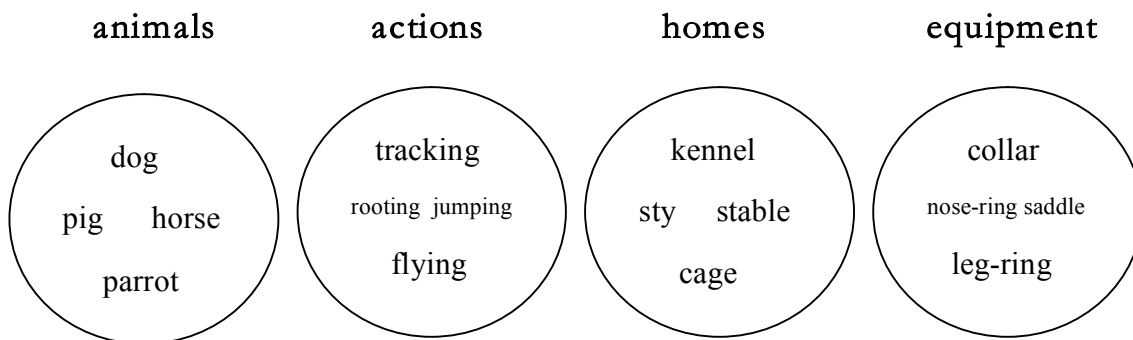
*Spazio=Brasile, tropici, aree secche*

*Tempo=Weekend, 19. secolo*

*Esempio di ontologia come definita in ambito informatico e rappresentabile in base al modello RDF*



*Esempio di organizzazione a faccette di termini / concetti di un determinato ambito (animali)*



### Concepts grouped in categories

Analizzando ora le **relazioni**, la prima relazione da prendere in considerazione è quella gerarchica. Le vere relazioni di questo tipo sono quelle che si verificano quando la classe subordinata è completamente contenuta in quella sovraordinata.

Le relazioni gerarchiche sono di tre tipi:

- 1) cosa-tipo, genere-specie;
- 2) parte-tutto;
- 3) classi di uno o esemplificative

Le altre relazioni sono di tipo non-gerarchico, e vengono impiegate per collegare termini che hanno relazioni tra di loro di vario tipo, quali causa-effetto, strumento-oggetto, azione-oggetto dell'azione...

Per esempio: raccolta (azione) con grano (oggetto), raccolta (azione) con mietitrebbia (strumento/agente), grano (oggetto) con mietitrebbia (strumento/agente)

E' importante ricordare che le relazioni possono essere paradigmatiche o sintagmatiche.

**L'ordine di citazione** è l'ordine delle categorie con cui abbiamo familiarità (cose – tipi – parti...), è il cosiddetto ordine standard di citazione.

In molti sistemi a faccette moderni è l'ordine di citazione a determinare il modo di combinare i termini.

L'appartenenza a una categoria determina la combinazione. Questo risulta implicito nella notazione. Questa appartenenza determina nella maggioranza dei casi quale sia il modo migliore per la combinazione sintattica, anche se l'utilizzo di quella prevista come default per un determinato ambito non è sempre la più adatta e quindi può essere modificata.

*Esempio di citazione standard in orticoltura:*

Winter pruning of roses = Potatura invernale delle rose

rose (entità) – pruning (operazione) + winter (tempo)

(Si noti che in italiano l'aggettivazione di potatura potrebbe portare a soluzioni diverse rispetto alla fattorizzazione dei termini N.d.T.)

Raccolta meccanizzata dei cavolini di Bruxelles

Cavolini di Bruxelles (entità) - raccolta (operazione) – machine (agente)

(Si noti che in italiano l'aggettivazione di raccolta potrebbe portare a soluzioni diverse rispetto alla fattorizzazione dei termini N.d.T.)

Una prima importante applicazione dell'analisi a faccette si ha per strutturare le classificazioni.

In particolare la “**classificazione a faccette**” usa concetti semplici, organizzati in una struttura logica, un set standard di categorie per analizzare i concetti. L'analisi a faccette, può essere usata per creare notazioni di classi o può fornire descrittori per i thesauri. Può anche essere usata come proposta in una interfaccia di ricerca

*Esempio di struttura base di una classificazione*

## [Foods]

(By physical state)

HKH PO	Essences
HKH PP	Extracts
HKH PS	Pastes

HKH PY (By operation/process used)

(By utility, etc.)

HKH QD	Convenience foods
HKH QE	Partly prepared foods
HKH QF	Instant foods
HKH QK	Artificial foods, synthetic foods

(By purpose)

(By physiological function)

HKH QS	Roughage
--------	----------

Un importante sviluppo può essere rappresentato dalla possibilità di **conversione di una classificazione a faccette direttamente in thesauro** ben strutturato secondo le norme standard.

L'ipotesi di partenza è che, essendo le relazioni in una struttura classificata a faccette sono ben definite e gestite, da questo tipo di strutture si possono generare altri tipi di linguaggi di indicizzazione.

Dalla gerarchia della classificazione si possono determinare le relazioni BT/NT o relazioni intra-faceted (paradigmatic) e alcune di tipo RT.

Altre relazioni RT possono essere identificate partendo dalle relazioni inter-facet (sintagmatiche)

Le relazioni di equivalenza sono deducibile dalle indicazioni di “synonym”.

Per costruire dei thesauri a faccette possiamo definire una metodologia standard consistente nella derivazione del thesauro da una struttura classificatoria a faccette. Questa è la metodologia che

viene esemplificata nel volume “Essential thesaurus construction”, tradotto in italiano in “Costruire thesauri”, Editrice Bibliografica, 2008 .

Una conversione automatica da classificazione a thesauro è prevista tra nello sviluppo del software che gestisce e genera la BC2. Non avendo però avuto in mente nel momento della creazione della BC2 il fatto che un thesauro dovesse esservi derivato ci sono alcuni problemi.

All'interno del **progetto CORDL** nell'ambito del 7. Programma quadro dell'Unione Europea stiamo studiando le applicazioni dei sistemi a faccette alle ontologie. Una sperimentazione di questo tipo non è ancora stata fatta, ma ci aspettiamo che possa funzionare visto, che la metodologia applicata è solida. La ricerca viene svolta insieme al Documentation Reserach & Training Center di Bangalore

Il sistema a faccette sarà la CDU (Classificazione decimale universale).

Lo scopo sarà di costruire un agente intelligente per cercare in dSpace.

Record bibliografici di varie biblioteche saranno usati per testare il sistema.

Una **conclusione** a queste considerazioni:

le terminologie a faccette costruite con il modello classico rispondono ai seguenti scopi:

- fornire lo stato funzionale dei concetti
- fornire le relazioni paradigmatiche e sintagmatiche
- fornire un ordinamento;
- dare regole per la combinazione dei termini

A mio parere, tutte gli elementi strutturali di un thesauro a faccette sono impliciti in una classificazione a faccette.

Molti elementi e relazioni possono essere inferite automaticamente

L'impiego di classificazioni e thesauri a faccette può supportare anche l'automazione dell'attività di classificazione, come la generazione automatica di strutture terminologiche.