

# Tassonomie dinamiche

La ricerca tradizionale è basata sul **recupero delle informazioni**.

La maggior parte dei compiti comuni però prevede un'**esplorazione**

- trovare delle relazioni
- diminuire le alternative

Si pensi ad esempio ad un utente che vuole comprare una macchina fotografica ma non sa esattamente quale; comincerà innanzitutto a valutare il parametro prezzo, fatta la scelta (es. prezzo basso) il numero di macchine risultanti sarà diminuito, a questo punto può continuare in questo modo scegliendo altri parametri di interesse (sensibilità iso, pixel ecc.) e continuando così a sfoltire le possibili scelte finché non arriverà ad un insieme di fotocamere che soddisfino la sua scelta.

All'utente è richiesto di:

- trovare tutte le possibili caratteristiche
- pesare queste caratteristiche e focalizzarsi sulla più rilevante
- esplorare e trovare tutte le caratteristiche correlate
- ripetere il processo finché il numero di elementi selezionati è sufficientemente piccolo per un'ispezione manuale

Rappresentazione

- parte intensionale: il database è rappresentato da una tassonomia disegnata da un esperto
- parte estensionale: i documenti possono essere classificati ad ogni livello di astrazione ed ogni documento è **classificato sotto due o più concetti**. (Per documenti intendiamo qualunque oggetto, non solo testuali)

Un concetto è un'etichetta che identifica un insieme di documenti classificati sotto tale concetto.

Nessuna relazione a parte la sussunzione (IS-A, PART OF) deve essere mantenuta nello schema.

La sussunzione richiede di mantenere un **vincolo di inclusione** per cui: se  $D(C)$  denota un insieme di documenti classificati sotto  $C$ , e  $C'$  è un discendente di  $C$  nella gerarchia allora  $\rightarrow$

$$D(C') \subseteq D(C)$$

Come sono correlati tra loro i concetti?

- tramite **sussunzione** (is-a, part-of)
- tramite regola di **inferenza estensionale**: due concetti  $C$  e  $C'$  sono correlati se c'è almeno un documento classificato sia sotto  $C$  che sotto  $C'$  o uno dei loro discendenti. Risultato importante  $\rightarrow$  una relazione tra due elementi non rappresentata esplicitamente nello schema viene inferita sulla base di un'evidenza empirica

Idea della tassonomia dinamica:

1. settare il focus di interesse
2. riassumere i concetti correlati ed eliminare quelli non correlati

Conseguenza importante: Le relazioni tra concetti non devono essere anticipate ma possono essere inferite dalla classificazione attuale.

Vantaggi:

- schema più semplice
- si adatta a nuove relazioni (dinamico)
- si scoprono delle relazioni inaspettate

Il termine “**tassonomie dinamiche**” è usato per indicare che la tassonomia può adattarsi a, e riassumere ogni sottoinsieme dell'universo, mentre le tradizionali tassonomie statiche possono riassumere solo l'intero universo.

**Benefici (relativi anche all'esempio di ricerca con zoom delle slide):**

- interfaccia più semplice e familiare
- l'utente è effettivamente guidato alla raggiunta del suo obiettivo. Ad ogni stadio ha una lista completa dei concetti correlati
- trasparenza: l'utente sa esattamente cosa sta succedendo
- interazione completamente simmetrica: se A e B sono correlati, l'utente troverà B facendo lo zoom su A e troverà A facendo lo zoom su B
- scoperta di relazioni inaspettate
- nessun risultato vuoto
- ogni combinazione di concetti (AND, OR, NOT) è supportata dalle corrispondenti operazioni insiemistiche sulle loro estensioni profonde
- supporto multilingua semplice: (basta tradurre le etichette)
- semplice raccogliere informazioni sull'interesse dell'utente
- schema semplice e minimale: non sono necessari argomenti composti. Es. quadri in Lettonia non deve essere un concetto perchè si può derivare dall'intersezione tra Latvia e quadri (ottenuti a runtime).
- Integrazione con altre tecniche di recupero (information retrieval, database):
  1. è possibile dare un contesto concettuale grazie alla tassonomia dinamica e poi sugli elementi rimasti applicare una normale tecnica di ricerca
  2. è possibile eseguire una normale ricerca e grazie alla tassonomia dinamica riassumere concettualmente i risultati ottenuti da tale ricerca.  
(possono essere combinati)

## Linee guida per il disegno di tassonomie

Organizzare la tassonomia come un insieme di sottotassonomie **ortogonalmente indipendenti**. Queste sottotassonomie indipendenti sono generalmente chiamate “facets”.

Avvertimento: stare attenti alle **false correlazioni**, che possono nascere da documenti composti (documenti che non sono atomici, ma che contengono altri documenti)

In ogni caso l'eliminazione di concetti correlati è una linea guida, non un obbligo. Quando dei concetti correlati sono conosciuti e largamente usati, si possono tranquillamente inserire (es. usare internet e non sintetizzare con: computer e network)

**fanout**: il numero di figli per ogni concetto non dovrebbe andare oltre 10-20

**profondità**: la tassonomia non dovrebbe essere più profonda di 3-4 livelli

Riassumendo quanto detto finora, i punti fondamentali della tassonomia dinamica sono:

- un modello minimale ancora effettivo. Gli utenti lo capiscono
- la regola di inferenza estensionale definisce un meccanismo per adattare la tassonomia al variare del focus dell'utente
- il modello è basato sui metadati. I documenti che devono essere gestiti non sono necessariamente testuali
- chiara separazione tra classificazione ed uso (esplorazione)

## Analisi della convergenza

Tassonomie **monodimensionali**: massimo raffinamento a livello terminale. Numero di terminali un ordine di grandezza minore del corpus

Tassonomie **multidimensionali**:

1. **senza composizioni** di concetti: massimo raffinamento a livello terminale. Il numero di terminali richiesti può essere più grande del corpus (peggiore delle tassonomie monodimensionali)
2. **con composizione** AND di concetti
  - con **classificazione a facet**: decisamente migliore delle altre. Es. una tassonomia con 1.000 terminali e 10 facets è in grado, con 3 operazioni di zoom di passare da 10.000.000 di documenti a 10. (non è una sensazione psicologica. È realmente più veloce!!)

Le tassonomie dinamiche hanno un range di applicazioni molto ampio, es: siti di e-commerce, cataloghi online, brokeraggio di lavoro, sistemi di diagnostica ecc.