

# Appunti: Verso il Semantic Web - Metadati e Vocabolari

## Introduzione

- **Il Web dei dati** si evolve dal web dei documenti e dei programmi.
- **Obiettivo:** Rendere le informazioni riutilizzabili per applicazioni automatiche, non solo leggibili dagli umani.
- **Problema:** Difficoltà nella ricerca di informazioni a causa di:
  - Differenza tra termini di ricerca e termini nei documenti.
  - Molteplicità di termini per stile o abitudine.
  - Ambiguità intrinseca di alcuni termini.
- **Vantaggi della serializzazione (XML, JSON, YAML):**
  - Codifica chiara e universale (UTF-8).
  - Delimitazione netta tra dati e metadati.
  - Etichettatura gerarchica per dare senso ai dati.
- **Manca la semantica:** Il significato non è nei dati, nel markup, nel DTD, o in HTML. Risiede nell'applicazione o nella mente umana.

## Il Semantic Web

- **Soluzione:** Il Semantic Web.
- **Approccio:** Astratto, sintattico, basato su affermazioni su classi o proprietà del dominio.
- **Strumenti:**
  - RDF: Meccanismo per esprimere affermazioni.
  - OWL: Meccanismo per organizzare e strutturare le affermazioni.
- **Scopo:** Generare informazioni riutilizzabili da applicazioni automatiche, non solo leggibili.
- **Motori di ricerca:** I metadati permettono ricerche più intelligenti.

## PICS (Platform for Internet Content Selection)

- **Risposta del W3C** a preoccupazioni su contenuti web inappropriati (1995).
- **Obiettivo:** Evitare censure centralizzate, promuovere la responsabilità autoriale.
- **Meccanismo:**
  - Valutazione (rating) dei contenuti da parte di associazioni.
  - Utenti scelgono a quali associazioni affidarsi.
  - Architettura distribuita per i rating.
- **Flessibilità:** PICS non definiva categorie predefinite, ma un meccanismo generale.
- **Limitazioni:** Richiedeva di identificare categorie e valori in anticipo.
- **Evoluzione:** Sostituito da POWDER (basato su RDF), ma entrambi non hanno avuto grande successo.

## Organizzazione delle Informazioni

- **Termini chiave:**
  - **Metadati:** Dati su dati (es. titolo, autore, argomento).
  - **Vocabolario controllato:** Insieme di termini precisi, non ridondanti, non ambigui.
  - **Tassonomia:** Gerarchia di termini in un vocabolario controllato (es. relazioni di generalità/specificità).

- **Thesaurus:** Tassonomia con relazioni aggiuntive tra termini (es. sinonimi, relazioni associative).
- **Classificazione a faccette:** Descrizione di un oggetto tramite un insieme di proprietà con valori da tesauri.
- **Ontologia:** Composizione di classi con relazioni espresse tramite proprietà (valori = riferimenti ad altre classi).
- **Folksonomia:** Tassonomia generata dagli utenti (tag cloud), senza vocabolario controllato.
- **Tipi di metadati:**
  - **Esterni/Interni/Riflessivi:** Parlano di risorse terze/contenuti nel documento/parlano del documento stesso.
  - **Autoriali/Redazionali:** Forniti dall'autore/da membri della catena di produzione.
  - **Oggettivi/Soggettivi:** Non discutibili/interpretazioni personali.
  - **Dissettivi/Anti-dissettivi:** Si applicano a ogni parte del documento/solo al documento nel suo insieme.
- **Usi dei metadati:**
  - Classificare, catalogare, organizzare documenti.
  - Ricerche concettuali.
  - Analisi e statistiche.
- **Problemi dei metadati:**
  - Ambiguità dei valori e degli scopi.
  - Varietà eccessiva di termini.

## Thesauri (o Thesauri)

- **Definizione (ISO 2788-1986):** Vocabolario di un linguaggio di indicizzazione controllato, organizzato per esplicitare relazioni tra concetti.
- **Scopo:**
  - Incontro tra lessico dell'autore e dell'utente.
  - Univocità semantica (un termine per concetto, un concetto per termine).
- **Relazioni tra termini:**
  - **Gerarchica:** Subordinazione (es. matematica/geometria).
  - **Preferenziale (sinonimica):** Termine preferito tra sinonimi (es. regola/norma).
  - **Associativa:** Relazione non gerarchica né di equivalenza (es. ecologia/inquinamento).
- **Relazione preferenziale (dettagli):**
  - USE (termine non preferito) / UF (Use For, termine preferito).
  - Tipi di sinonimia: vera, varianti ortografiche, sigle, preferenze linguistiche, ecc.
  - Sinonimia convenzionale (quasi-sinonimia, upward posting, antinomia).
- **Relazione gerarchica (dettagli):**
  - NT (Narrower Term) / BT (Broader Term).
  - Tipi:
    - **is\_a (genere/specie):** Categoria e membri (es. felino/gatto).
    - **has\_a (parte/tutto):** Concetto complesso e componenti (es. sistema circolatorio/vene).

- **part\_of (gerarchia compositazionale):** Concetto complesso e parti (non strettamente gerarchica, es. automobile/motore).
  - **instance\_of (classe/istanza):** Classe e individuo (es. Pontefici/Francesco I).
- **Monogerarchie vs. poligerarchie:** Gerarchie semplici vs. multiple (una classe deriva da più classi).

## Classificazioni a Faccette

- **Definizione (Ranganathan):** Descrizione di un oggetto tramite proprietà con valori da tesauri.
- **Schema fisso:** Deve identificare (non solo descrivere) una risorsa specifica (chiave).
- **Esempio (Dublin Core):** Tipo documento, destinatari, titolo, autore, URL, formato, data.

## Ontologie

- **Definizione:** Composizione di classi con relazioni espresse tramite proprietà (valori = riferimenti ad altre classi).
- **Progressione:** Dai metadati (casino) -> vocabolario controllato -> tassonomie/thesauri -> classificazione a faccette -> ontologie.
- **Vantaggi:** Evita ambiguità, associa metadati e proprietà ai valori.
- **Esempio:** "Fabio Vitali" non è una stringa, ma un'istanza della classe "Persona".
- **Uso delle ontologie (esempio JSON-LD):**

```
{
  "@context": {
    "@vocab": "http://www.fabiovitali.it/"
  },
  "@type": "document",
  "author": {
    "@type": "person",
    "name": "Fabio Vitali",
    "affiliation": {
      "@type": "organization",
      "name": "Università di Bologna"
    }
  },
  "title": "Metadati",
  "coverage": {
    "@type": "lesson",
    "date": "3 maggio 2024",
    "context": {
      "@type": "course",
      "authority": {
        "@type": "organization",
        "name": "Università di Bologna"
      }
    },
    "title": ["Tecnologie Web"]
  }
},
```

```
"subject": ["Metadata and ontologies", "Raw metadata collections vs. ontologies",
"Using metadata"]
}
```

## Folksonomie

- **Alternativa alle ontologie:** Tassonomie generate dagli utenti (tag cloud).
- **Caratteristiche:**
  - Nessun vocabolario controllato o modello concettuale.
  - Risorsa associata a categorie tramite termini usati dagli utenti.
  - Prevalenza statistica dei termini.
  - Nessuna inferenza o deduzione sui termini (stringhe opache).
- **Vantaggi:** Gratuito, flessibile, democratico, decentralizzato, olistico, scala bene.
- **Svantaggi** Non permette ragionamenti di alto livello, non essendoci un modello concettuale forte.

## Vocabolari nel Semantic Web

- **Definizione:** Insieme di termini per descrivere un dominio concettuale.
- **Tipicamente descrivono:** Entità, classi, proprietà (semplici o relazioni).
- **Serializzazioni RDF:**
  - RDF/XML (verboso, poco usato).
  - Turtle (minimale).
  - JSON-LD (estensione di JSON).
  - RDFa (estensione di HTML5).

## Vocabolari Famosi

- **Organizzazioni classiche (non digitali):**
  - Classificazione Dewey: Struttura numerica gerarchica (problemi: rigida, eurocentrica).
  - Classificazione Library of Congress.
  - Modello PMEST (Ranganathan): Personality, Matter, Energy, Space, Time.
  - Marc 21: Catalogazione informatica di risorse librarie (stringa di 24 byte).
- **Ontologie:**
  - **Dublin Core:** Metadati per risorse di rete (classificazione a faccette, 15 categorie, qualificatori).
    - Esempi di categorie: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Relation, Source, Language, Coverage, Rights Management.
    - Esempi in JSON-LD, Turtle, XML-RDF.
  - **FRBR (Functional Requirements for Bibliographic Records):** Modello concettuale per identificare requisiti minimi delle descrizioni bibliografiche.
    - Gerarchia WEMI: Work, Expression, Manifestation, Item.
    - Relazioni primarie tra entità FRBR.
    - Integrazione con Dublin Core.
  - **FOAF (Friend Of A Friend):** Descrizione machine-readable di persone, gruppi, aziende, ecc.
    - Classi: agent, document, OnlineAccount, project.
    - Esempio in JSON-LD.
  - **SKOS (Simple Knowledge Organisation System):** Modello per esprimere schemi concettuali (tesauri, classificazioni, tassonomie, ecc.).

- Classe principale: concept.
- Proprietà: label, relazioni semantiche (broader, narrower, related).
- Esempi in JSON-LD, Turtle, RDF/XML.

## Conclusioni

- I modelli (ontologici e non) possono essere combinati rispettando le regole di mapping.
- I metadati, i vocabolari controllati, le tassonomie, i thesauri, e le ontologie sono fondamentali per il semantic web.
- Il Semantic Web è un superamento al modo classico di intendere il web, e ne estende le potenzialità.