



Presentazione Tesi di Laurea

Valutazione di strumenti semiautomatici per la creazione di ontologie da risorse testuali



Relatore: Prof. Ing. Fulvio Corno

Candidata: **Sara Valente**

Relatore esterno: Ing. Alberto Ciaramella



Il problema

Gli utenti umani si orientano nel Web grazie alla loro esperienza di navigazione e alla capacità di evocazione che possono avere parole chiave.

Queste caratteristiche non appartengono invece a nessuna applicazione, che in definitiva non è in grado di interpretare il contenuto delle pagine.

Problema: Trovare il più velocemente possibile l'informazione necessaria.



La soluzione

Web semantico:

un web che, integrato con gli opportuni linguaggi, sia in grado di esplicitare e quindi comprendere il significato del testo contenuto all'interno delle pagine.

➡ Ricerca di un sistema comune e condiviso per la comunicazione e comprensione dell'informazione.



L'ontologia

L'interazione tra persone e sistemi software porta alla ricerca di un sistema comune e condiviso per la comunicazione e comprensione dell'informazione.

L'idea è quella di usare un vocabolario condiviso per descrivere il contenuto delle risorse, la cui semantica è descritta in un formato non ambiguo e processabile da una macchina.

Definiamo l'**ontologia** come *il tentativo di formulare uno schema concettuale esaustivo e rigoroso nell'ambito di un dato dominio.*

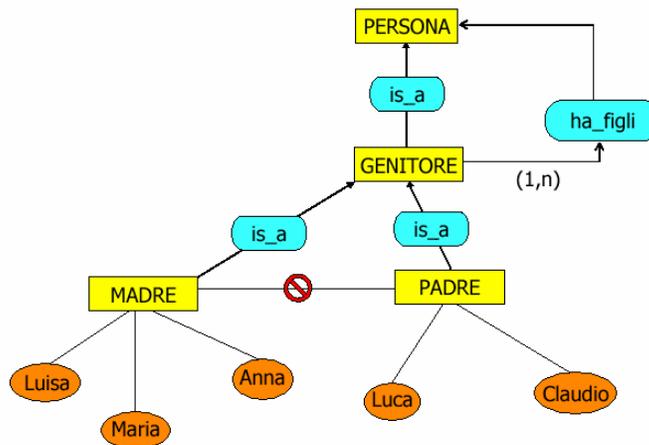
L'ontologia

Gli elementi di un'ontologia:

- I concetti
- Le relazioni
- Le funzioni
- Gli assiomi
- Le istanze

Costruire un'ontologia da una risorsa testuale vuol dire quindi andare ad estrarre questi elementi.

Esempio di ontologia





Costruire un'ontologia

Un'ontologia è costruita seguendo questi step:

- Acquisire la conoscenza del dominio
- Organizzare l'ontologia, prevede le fasi:
 - Flat glossary
 - Structured glossary
 - Identificare tutte le relazioni
- Popolare l'ontologia
- Controllare il proprio lavoro
- Considerare il riuso di risorse esistenti
- Consegnare l'ontologia



Obiettivo del lavoro

Creare le ontologie non è semplice: occorre una buona conoscenza del dominio e spesso la vastità dei dati di partenza rende il lavoro alquanto complicato.

Avere un software in grado di agevolare il compito è molto importante.



Ricerca di tools per la creazione semiautomatica di ontologie da risorse testuali.

Tools for ontology learning

Sono stati analizzati i seguenti tools per l'apprendimento dell'ontologia dal testo:

ONTOLOGY LEARNING LAYERS								
SYSTEM	Terms	Synonyms	Concept Formation	Concept Hierarchy	Relations	Relation Hierarchy	Axiom Schemata	General Axioms
JATKE	?				?			
TEXT2ONTO	X	X	X		X			
ONTOLT	?				?			
ONTOLEARN	?	?						
ONTOGEN 2.0	X		X	X	X			
RELEXT					?			

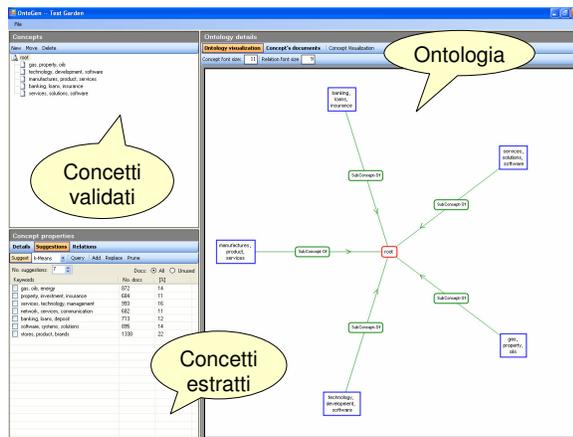
Text2Onto

The screenshot shows the Text2Onto application window. It features a left sidebar with a tree view of project files, a central area displaying a list of extracted terms and their relationships, and a bottom pane showing a log of operations. Callout boxes highlight specific parts: 'algoritmi' points to the sidebar, 'corpus' points to the file list, 'output del processo di ontology learning' points to the main list, and 'Operazioni svolte ed errori' points to the log.

Principali caratteristiche:

- Algoritmi di estrazione
- Probabilistic Ontology Model (POM)
- Validazione dell'utente
- Traccia del processo

OntoGen 2.0



Principali Caratteristiche:

- Estrazione tramite Centroid Vector e Support Vector Machine
- Visualizzazione grafica
- Dati in ingresso

Confronto tra i due tools

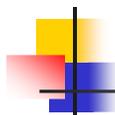
	<i>ONTOGEN 2.0</i>	<i>TEXT2ONTO</i>
FILE IN INGRESSO	HTML o txt (Folders), Bag-of-word	HTML, PDF
FILE IN USCITA	Proton Topic Ontology, RDF Schema, OWL	RDFS, OWL, F-Logic
CONCETTI	Suggerimento del concetto rispetto alle occorrenze	Estrazione di tutti i termini di cui calcola la rilevanza
FORMAZIONE DEI CONCETTI	Validazione dei suggerimenti o creazione di un nuovo concetto	Validazione o cancellazione dei termini estratti
GERARCHIA DEI CONCETTI	I concetti possono essere strutturati automaticamente o manualmente	Non è possibile dare una gerarchia ai concetti
RELAZIONI	Sub-concept-of, similar	Calcola la rilevanza per ogni coppia di concetti
VISUALIZZAZIONE GRAFICA	Mostra l'ontologia come un grafo che si sviluppa da un concetto radice	Mostra il POM come una lista di concetti e relazioni



Confronto tra i due tools

Sono adoperati come dati di input dei brevetti, reperiti con Intellipatent, appartenenti alla categoria G03B3/00 (Physics – Photography - Apparatus or arrangements)

Text2Onto	OntoGen 2.0
Inutilizzabile per grandi quantità di dati.	Lavora molto bene con una grande mole di documenti.
Non necessita di alcuna struttura per i dati in ingresso.	I dati in ingresso devono avere una struttura organizzata.
Text2Onto è molto più preciso, problema di troppi dati da vagliare.	È meno preciso, problema di perdere dati importanti.



Conclusioni finali

- Non esiste ancora un tool in grado di realizzare semiautomaticamente delle ontologie di elevata qualità;
- Tra quelli visti Text2Onto e OntoGen 2.0 sembrano i migliori, e possono essere un punto di partenza del processo (magari congiuntamente);
- In ogni caso meritano attenzione e sarebbe necessario approfondirne le performance;
- Nei laboratori di ricerca e nelle università si è al lavoro.