

tesi di laurea

# Realizzazione di un processo a supporto dell'interoperabilità semantica nel Web

Anno Accademico 2008/2009

**relatori**

Ch.mo prof. Angelo Chianese

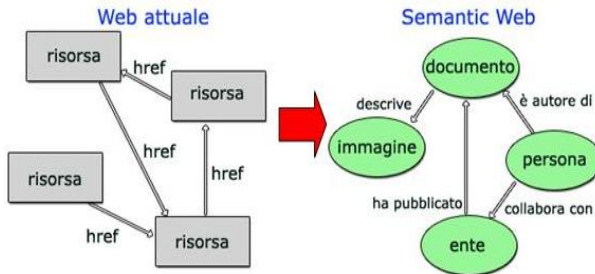
Ch.mo prof. Vincenzo Moscato

**candidato**

Chiara Chirichella

Matr. 534/3038

## Contesto e Contributo



Il significato dei dati può essere compreso senza ambiguità sia da persone che da agenti software

### Contesto

#### Evoluzione del Web: dal Web Sintattico al Web Semantico

##### ❖ Problematiche del Web Sintattico:

- assenza di significato dei collegamenti
- Web nascosto, falsi positivi, falsi negativi
- i motori di ricerca si basano su semplici tecniche di word-matching
- problemi di interoperabilità ed integrazione

##### ❖ IL Web Semantico risponde a queste problematiche, introducendo:

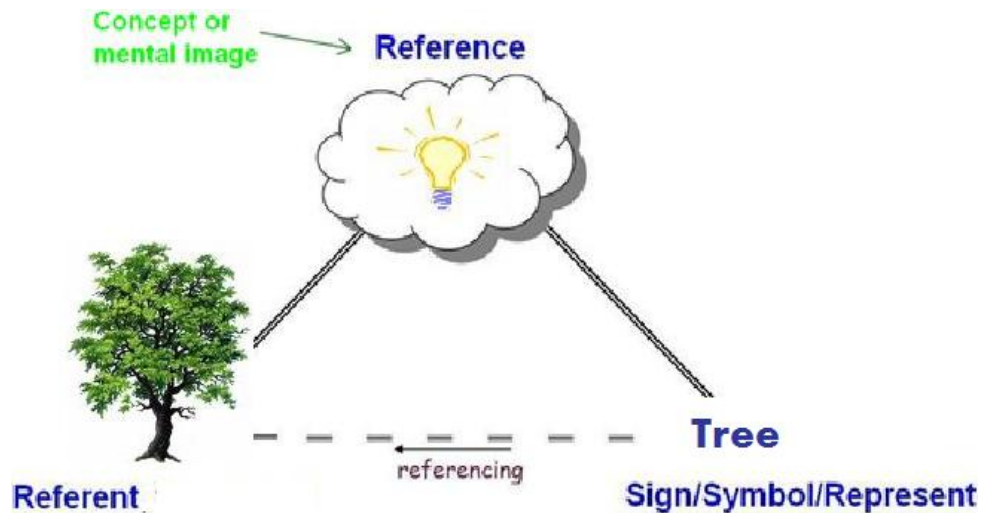
- annotazioni semantiche per le risorse (Metadati)
- relazioni tra le risorse web (Ontologie)
- appositi linguaggi e tecnologie (XML, RDF, OWL, etc..)

### Contributo

#### Realizzazione di un processo a supporto dell'Interoperabilità Semantica nel Web

- Basato sul modello ibrido a triangolo semantico (Ogden-Richards)
- Ottimizzazioni: motivazioni e tecniche di realizzazione
- Valutazione del processo ottimizzato

## Il modello di comunicazione del Triangolo Semantico



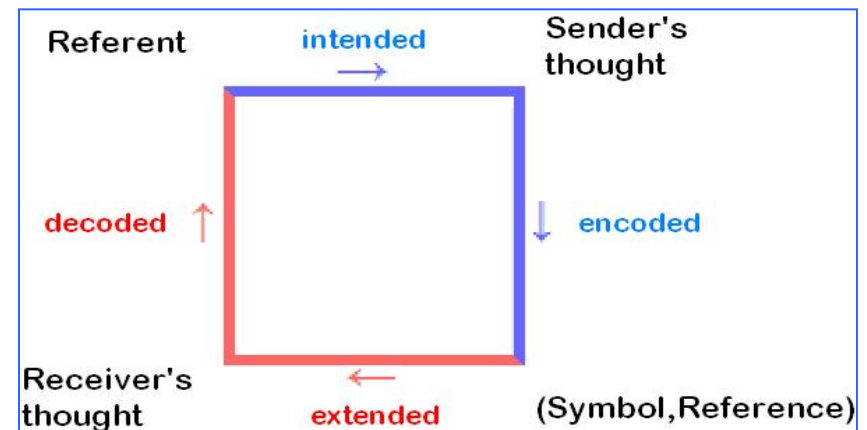
### Triangolo di Ogden-Richards

- *simbolo*: la rappresentazione
- *referenza*: concetto o immagine mentale
- *referente*: l'oggetto reale

**INTEROPERABILITA' SEMANTICA:**  
stesso insieme di referenti nella  
comunicazione

Il processo di comunicazione proposto si basa su questo modello:

- *Sender agent* - codifica del messaggio.
- *Receiver agent* - decodifica del messaggio.
- *Messaggio* - un insieme di simboli riferiti a referenti tramite le referenze.



## Processi di Codifica e Decodifica di Partenza

### Codifica

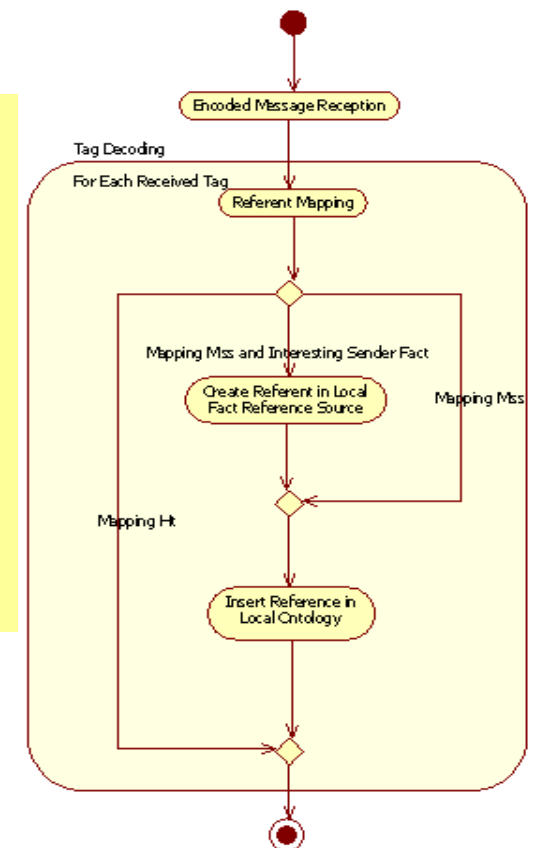


Risultato del processo  
di codifica:

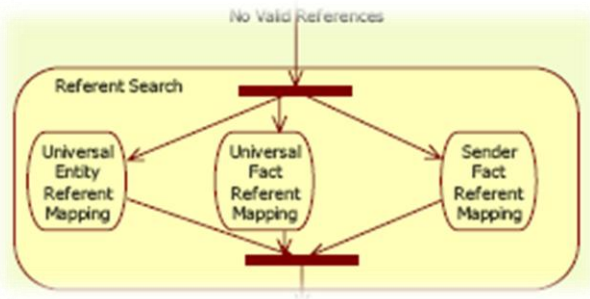
**Information Object**  
documento XML-RDF  
contenente le coppie  
<Semantic Tag,Reference>



### Decodifica



## Problemi del processo di codifica originario e soluzioni realizzate



### Problema:

Nel **processo di partenza** l'interrogazione delle fonti di conoscenza nel processo di codifica è, in realtà, un'operazione

**-limitata** nel numero di fonti consultabili

**-rigida** (ordine statico)

**-sequenziale** nel tempo



### Soluzione:

- ❖ **Estensione del numero di fonti Universali:**
  - *WordNet: Universal Entity Referent Source*
  - *Wikipedia: Universal Fact Referent Source*
  - *IMDB: Universal Fact Referent Source*
- ❖ **Algoritmo adattativo di consultazione delle fonti**
- ❖ **Parallelismo attività di validazione e di ricerca nella fonte successiva**



## Ordine di consultazione delle fonti

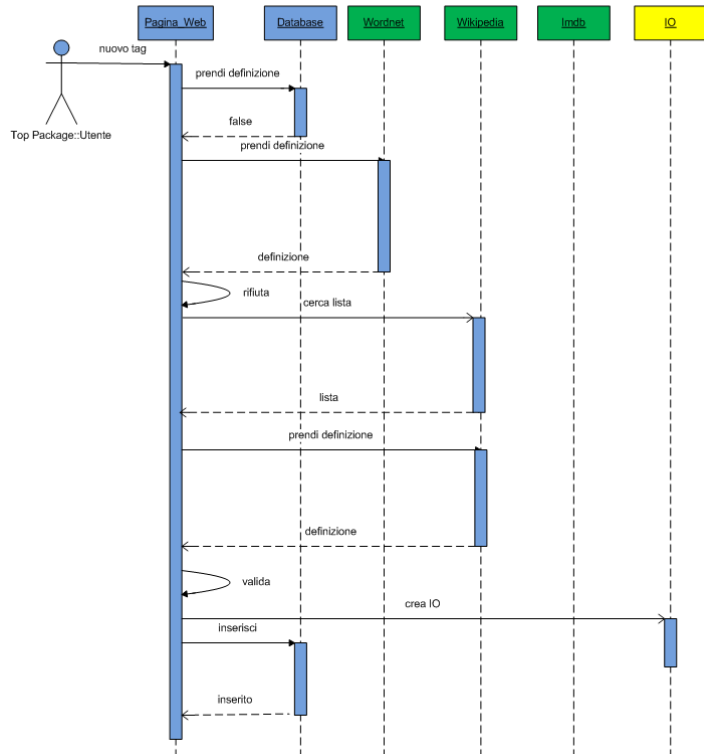


**Ordine ottimo:** ordine di consultazione delle fonti che minimizza il numero di significati da validare per ogni tag;

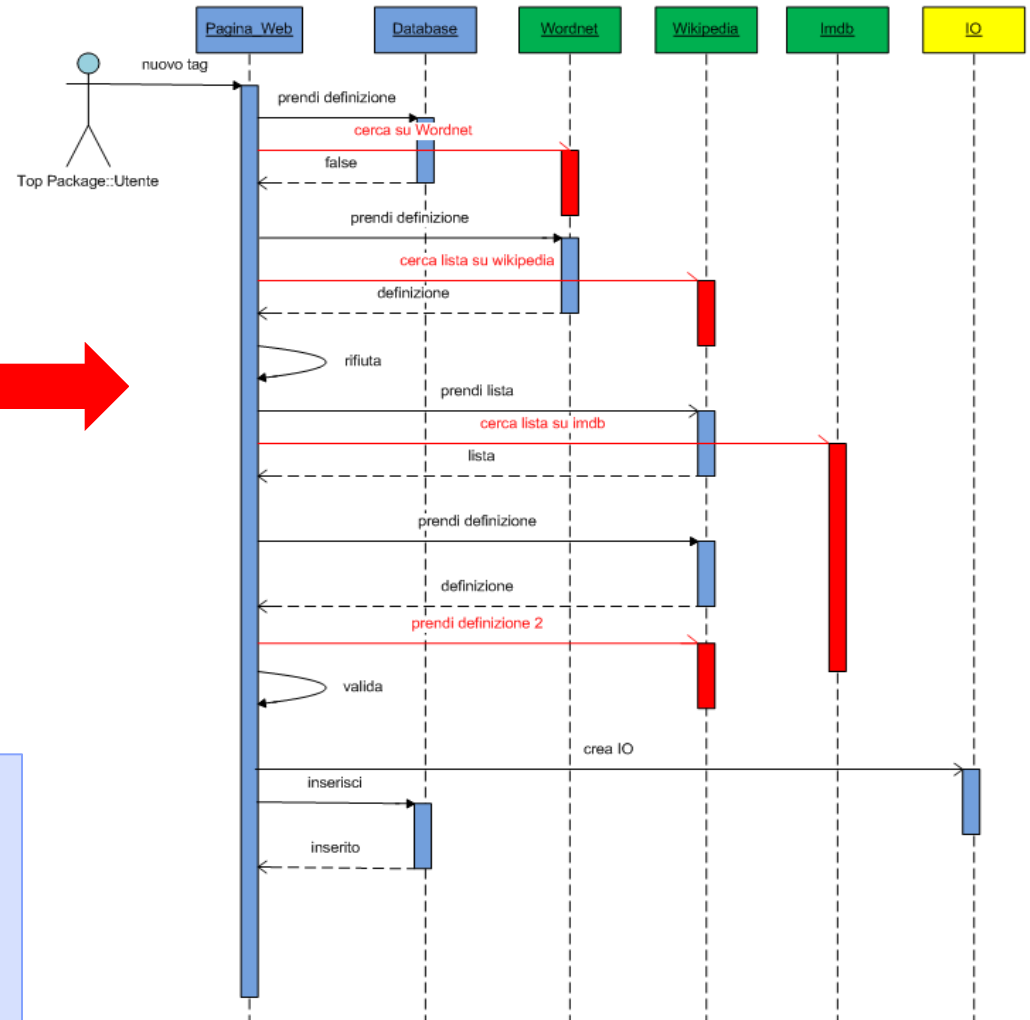
- Principio di **“località”**;
- Flessibile e adattabile automaticamente a variazioni dell'argomento dell'input;

## Attività parallele

### Processo di partenza

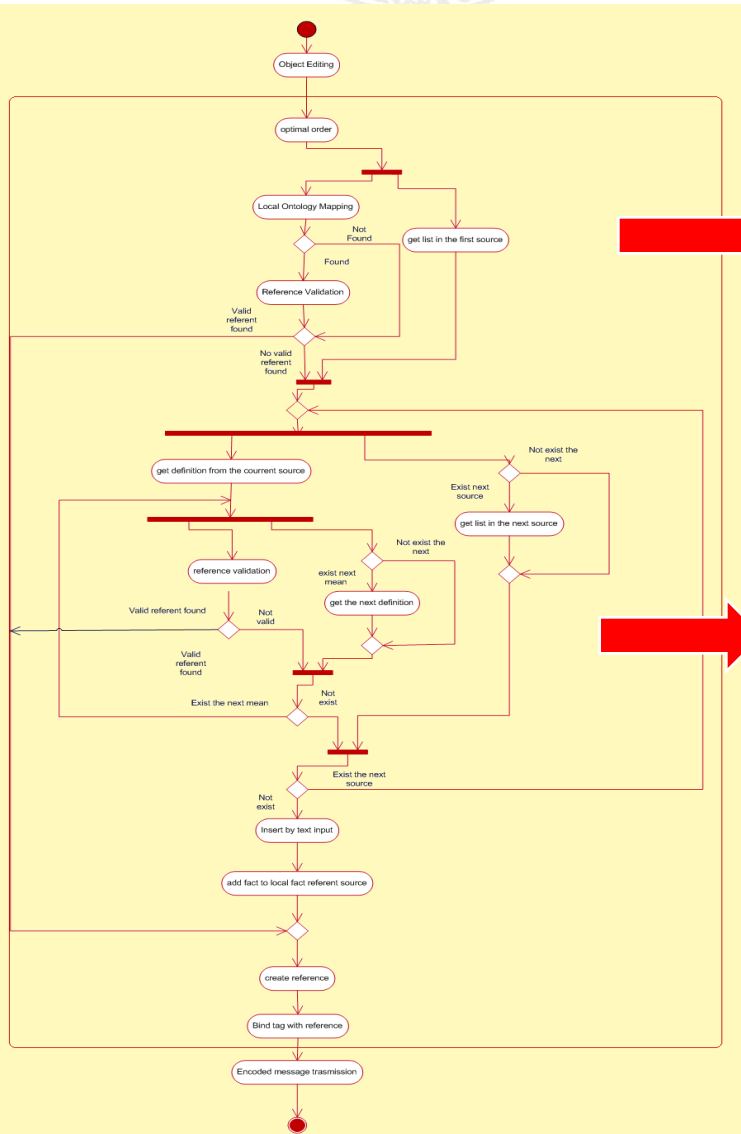


### Processo ottimizzato

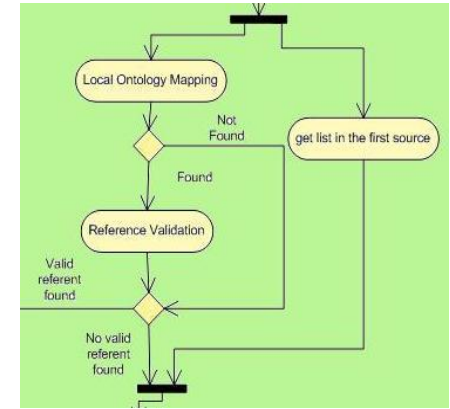


- Le operazioni sincrone di interrogazione nel processo di partenza sono scomposte in 2 operazioni, una sincrona e l'altra asincrona, nel processo ottimizzato.
- Il processo ottimizzato è attivo anche durante le fasi di validazione.

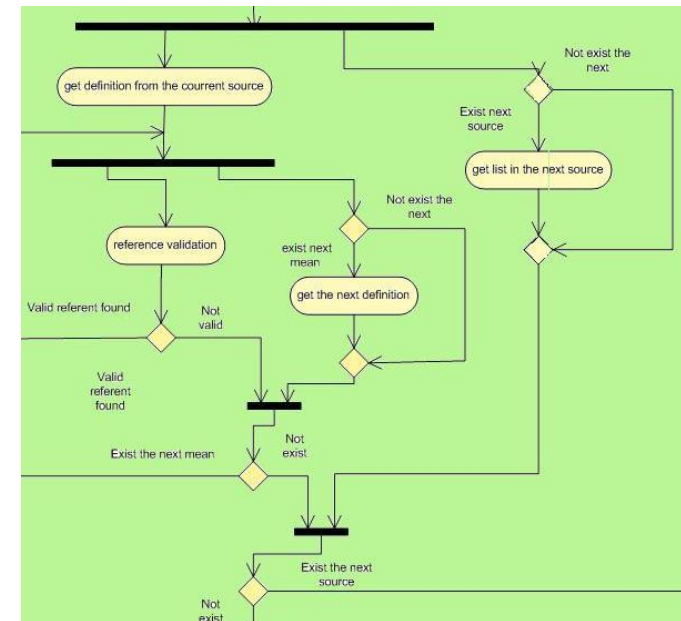
## Processo di codifica ottimizzato



-La prima fonte ad essere interrogata è sempre la **Local Ontology**, concorrentemente alla prima fonte esterna;  
-**Priorità** ai significati offerti dall'ontologia locale



-**Fork di primo livello**, tra attività di validazione e ricerca;  
-**Fork di secondo livello**, per ogni significato offerto da una fonte;  
-Non c'è necessità di sincronizzazione quando si accetta un significato;





## Sperimentazione: obiettivi e metriche

### Obiettivi

- ❖ comparazione tra processo originario e processo ottimizzato
- ❖ valutazione dello sforzo di codifica nel processo originario e nel processo ottimizzato

### Metriche

- ❖ **Indice VC (Validation Complexity):**  
 $VC(\tau, \Omega) = \text{num significati validati per tag}$   
dove  $\tau$  è il tag,  
 $\Omega$  è l'insieme ordinato delle fonti;
- ❖ **Indice AVC (Average Validation Complexity):**  
 $AVC(D, \Omega) = \sum_{\tau \in D} VC(\tau, \Omega) / |D|$   
dove  $D$  è il documento;

### Protocollo Sperimentale

- ✓ Fonti: pagine riguardanti il dominio cinematografico e il dominio sportivo
- ✓ Raccolta 800 documenti con estrazione di 1500 tag
- ✓ A partire da database dell'ontologia locale vuoto
- ✓ Scenari: dominio cinematografico e dominio sportivo
- ✓ Confronto processi originario - ottimizzato su dominio cinematografico e misto



#### IO Parameters

Source:  Sfoglia...  
(htm/doc)

Tag List:  Sfoglia...

#### Log-In

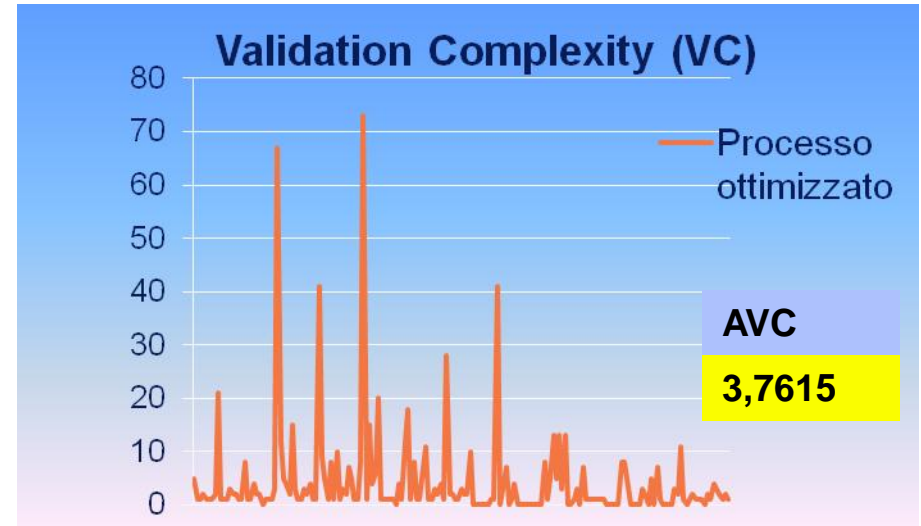
## Sperimentazione: risultati

**..confronto su dominio misto  
tra processo di partenza...**

Combinazione ordine fonti	AVC
[H1] Local, Imdb, Wordnet, Wikipedia	<b>3,962</b>
[H2] Local, Imdb, Wikipedia, Wordnet	4,306
[H3] Local, Wordnet, Imdb, Wikipedia	4,372
[H4] Local, Wordnet, Wikipedia, Imdb	5,358
[H5] Local, Wikipedia, Imdb, Wordnet	5,292
[H6] Local, Wikipedia, Wordnet, Imdb	5,702



**..e processo ottimizzato**



**..confronto tra processo di partenza e processo  
ottimizzato su dominio cinematografico**

	AVC
Processo di partenza (ordine Local, Wordnet, Wikipedia, Imdb)	10,99
Processo di partenza (ordine Local, Imdb, Wordnet, Wikipedia)	4,01
Processo ottimizzato (ordine adattativo)	<b>2,73</b>

-Il valore VC diminuisce  
all'aumentare del numero di  
tag nel processo ottimizzato

-L'algoritmo adattativo  
fornisce prestazioni migliori  
anche rispetto ad un ordine  
statico scelto in base al  
dominio dei documenti in  
ingresso...

## Conclusioni e sviluppi futuri

Lo scopo di questo elaborato di tesi è stato presentare le ottimizzazioni, e le motivazioni a loro sostegno, realizzate su un modello di interoperabilità semantica per agenti software nel Web già esistente.

### Possibili sviluppi futuri:

- Automatizzazione della fase di Tag Extraction;
- Introduzione nuove fonti di dominio specifico;