

# Lezione

---

## Ontology Learning e Open Information Extraction

Luigi Di Caro

# Ontology Learning e Open Information Extraction

---

- Panoramica
  - Cosa significa Ontology Learning, varianti e definizioni
  - Livelli di profondità
  - Task e metodi
- Lab

**Riferimento:**

**Libro “Ontology Learning and Population from Text”**

Philipp Cimiano

# Comprensione del linguaggio naturale

---

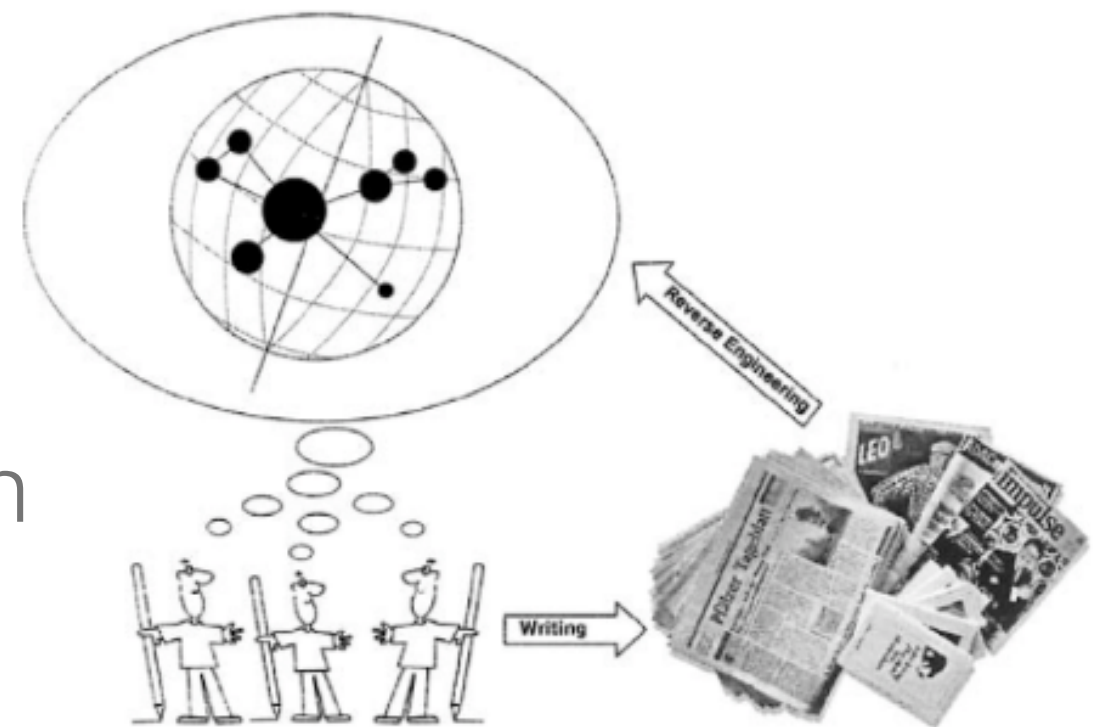
- Ontology Learning

## Problemi principali

- Definizione (“*reverse engineering*”) World knowledge is not usually encoded  
Domain knowledge is not used completely

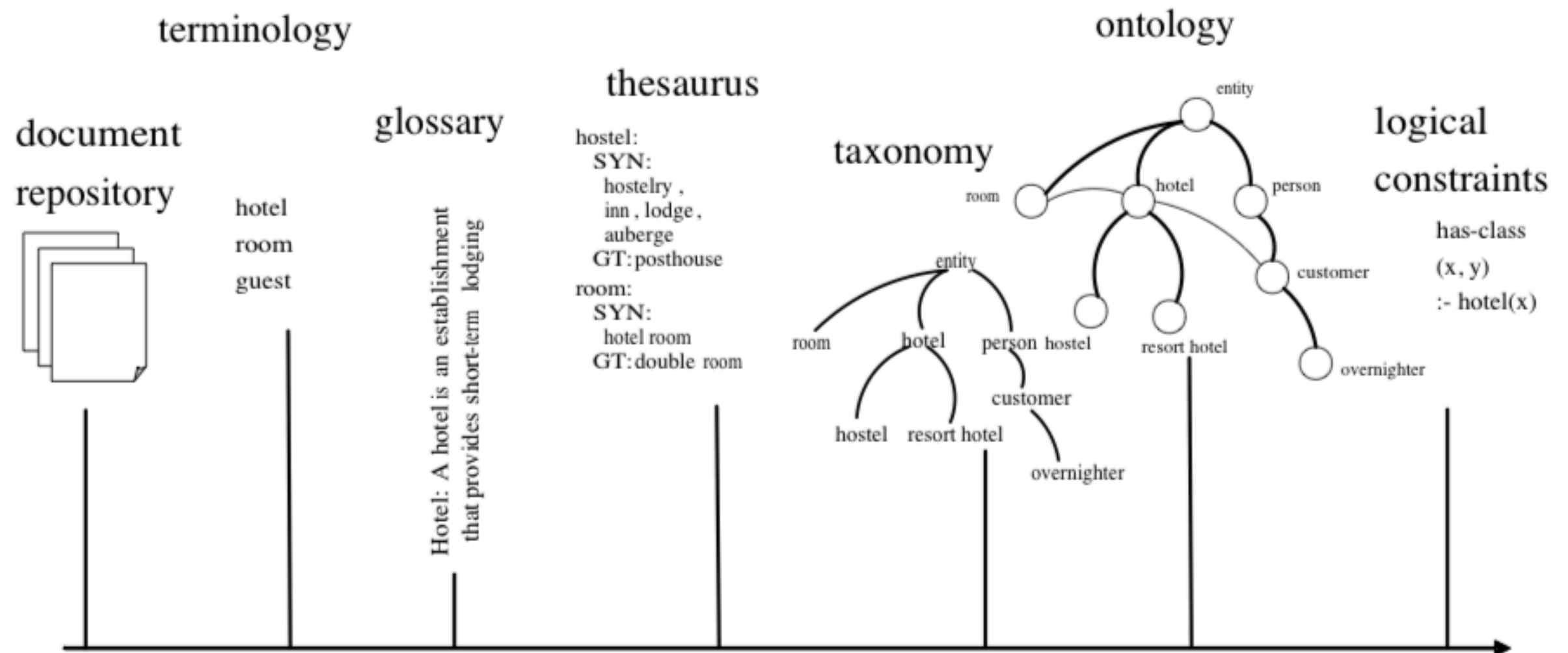
- Differenze con (e tra)

- Ontology Population
- Ontology-based Annotation
- Ontology Enrichment



# Comprensione del linguaggio naturale

- Ontologie: livelli di profondità



[http://wwwusers.di.uniroma1.it/~navigli/pubs/Navigli\\_Oxford\\_chapter\\_2015.pdf](http://wwwusers.di.uniroma1.it/~navigli/pubs/Navigli_Oxford_chapter_2015.pdf)

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Tasks
    - Term Extraction
      - trovare nomi per concetti e relazioni
    - Synonym Extraction
    - Concept Extraction
      - intension (gloss learning)
      - extension
    - Concept Hierarchies Induction
      - dati i concetti
    - Relation Extraction
    - Population (NER, IE)

# Comprensione del linguaggio naturale

---

- Ontology Learning

- Tasks

- Term Extraction
      - trovare nomi per concetti e relazioni

- Synonym Extraction

- Concept Extraction

- intension

- extension

- **Concept Hierarchies Induction**

- dati i concetti

- Relation Extraction

- Population (NER, IE)

*Paradigmi*

Patterns lessico-sintattici

Distributional Hypothesis

Nozione di Sussunzione

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
    - Natural Language Processing
    - Formal Concept Analysis
    - Machine Learning

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
    - Verbi
      - azioni, eventi, etc.
    - Aggettivi
      - attributi
    - Nomi propri
      - “individuals”
    - Nomi
      - classi
    - Avverbi
      - modi
    - Preposizioni
      - vincoli spazio-temporali
- Natural Language Processing
  - **Ruolo delle part-of-speech**
- Formal Concept Analysis
- Machine Learning



# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
- Natural Language Processing
  - Ruolo delle part-of-speech
  - **Preprocessing**
- Formal Concept Analysis
- Machine Learning
- Tokenization
- Normalization (abbrev., etc.)
- Stemming / Lemmatizzazione
- POS tagging
- Named Entity Recognition
- Risoluzione dei pronomi

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
- Natural Language Processing
  - Ruolo delle part-of-speech
  - Preprocessing
  - **Analisi sintattica**
- Formal Concept Analysis
- Machine Learning
- Regole su parse tree

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
- Natural Language Processing
  - Ruolo delle part-of-speech
  - Preprocessing
  - Analisi sintattica
  - **Similarità**
- Formal Concept Analysis
- Machine Learning
- Vector Space Models (VSMs)
  - Distribuzioni di probabilità
  - Term relevance (tf, tfidf, etc)

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
    - Natural Language Processing
      - Ruolo delle part-of-speech
      - Preprocessing
      - Analisi sintattica
      - Similarità
      - **Risorse linguistiche**
    - Formal Concept Analysis
    - Machine Learning
  - WordNet
  - BabelNet
  - ConceptNet
  - VerbNet
  - FrameNet
  - PropBank
  - ...

# Comprensione del linguaggio naturale

---

- Ontology Learning

- Methods

- Natural Language Processing
    - Formal Concept Analysis (FCA)
      - **Formal Context**
    - Machine Learning

- Oggetti
      - Attributi
      - Incidenza

Nozioni di formal context  
e operatori

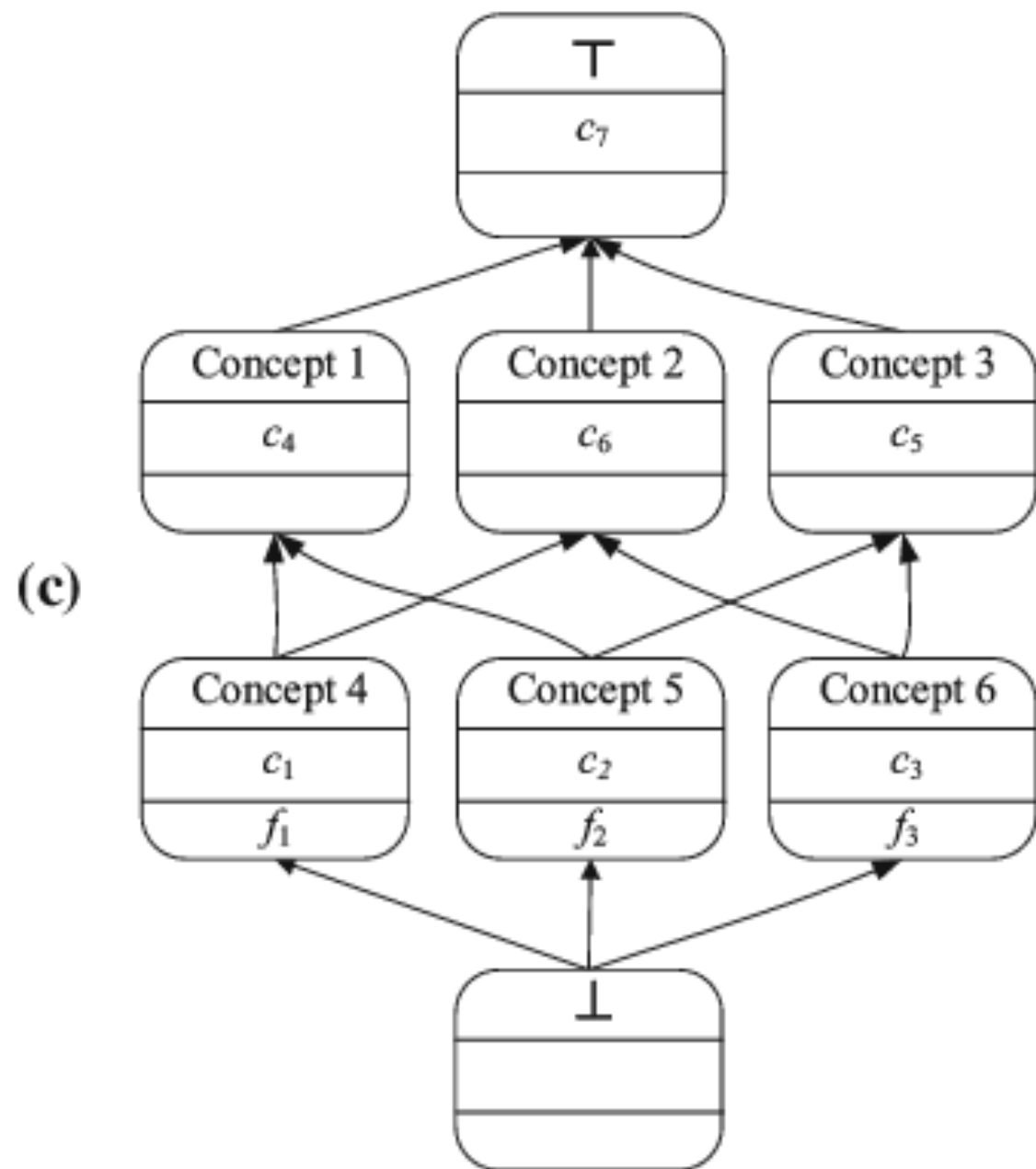
# Comprensione del linguaggio naturale

	$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$
$f_1$	x			x		x	x
$f_2$		x		x	x		x
$f_3$			x		x	x	x

(a)

$\top$	$\{f_1, f_2, f_3\}, \{c_7\}$
Concept 1	$\{f_1, f_2\}, \{c_4, c_7\}$
Concept 2	$\{f_1, f_3\}, \{c_6, c_7\}$
Concept 3	$\{f_2, f_3\}, \{c_5, c_7\}$
Concept 4	$\{f_1\}, \{c_1, c_4, c_6, c_7\}$
Concept 5	$\{f_2\}, \{c_2, c_4, c_5, c_7\}$
Concept 6	$\{f_3\}, \{c_3, c_5, c_6, c_7\}$
$\perp$	$\{\emptyset\}, \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$

(b)



# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
    - Bayesian classifiers
    - Decision trees
    - Support Vector Machines
    - Neural Networks
- Natural Language Processing
- Formal Concept Analysis
- Machine Learning
  - **Supervised methods**

# Comprensione del linguaggio naturale

---

- Ontology Learning
  - Methods
    - Natural Language Processing
    - Formal Concept Analysis
    - Machine Learning
      - Supervised methods
      - **Unsupervised methods**
  - Clustering
  - Hierarchical clustering
    - divisivo
    - agglomerativo



# Open Information Extraction

---

- Nato dalla necessità di estrarre grandi quantità di informazioni da grandi corpora
- Legato all'OL, ma differente in natura
  - Estrazione di triplette (nella maggior parte dei casi) che costituiscono *relational phrases*
    - Solitamente *arg1 - verbal phrase - arg2*
    - *esempio:*
      - “*Faust made a deal with the devil*” —> (*Faust, made a deal with, the devil*).

# Open Information Extraction

---

- OIE nasce e si sviluppa usando pattern semplici e tecniche NLP
- Esempi di pattern (in stile OL-Hearst)

Authors	Patterns
Yamada and Baldwin	N BE worthy of V[+nom]; V[+ing] Noun; N BE Adverb-V[ed]; N (deserves merits) V[+nom]
Cimiano and Wenderoth	NP comprises NP <sub>C</sub> ; NP are made up of NP <sub>C</sub> ; NP <sub>F</sub> ,? such as NP
Caselli and Russo	used to (make put) V; made of N; of ADJ color; produced by N; a kind of N

# Open Information Extraction

---

- OIE nasce e si sviluppa usando pattern semplici e tecniche NLP
- Si possono imporre vincoli sugli argomenti, o sulle frasi verbali (usando risorse semantiche, o domain needs)

# Open Information Extraction

---

- OIE nasce e si sviluppa usando pattern semplici e tecniche NLP
- Solitamente, la maggior parte dei sistemi usa due steps
  - POS tagging (e/o dependency parsing)
  - Extraction step
    - Può usare WSD, check dipendenze sintattiche subj/obj tra verbal phrase e argomenti, check su ontologie esistenti, etc.

# Open Information Extraction

---

- OIE nasce e si sviluppa usando pattern semplici e tecniche NLP
- A volte, si usa un step di *filtering*
  - In base alle triple estratte, si cerca di rimuovere triplette rare, oppure effettuare clustering e revisione delle estrazioni

# Open Information Extraction

---

- OIE
  - Problematiche
    - Molti sistemi, difficili da valutare
    - Molte estrazioni, difficili da comparare
    - Triplette non di facile utilizzo in applicazioni reali

# Open Information Extraction

---

- OIE
  - Esempi di sistemi esistenti
    - ReVerb (uno dei primi, basato su vincoli sintattici)
    - KrankeN (uso di WSD)
    - ClausIE (estrae relazioni N-ary)
    - DefIE (combina Parsing e WSD, creando un grafo)
    - NELL (Never Ending Language Learning)
    - ...

# Lab - Esercizio “FCA”

---

- Implementare un sistema di Ontology Learning facendo uso della FCA
  - Libreria Python per FCA
    - <https://pypi.org/project/concepts/>
  - Dati
    - A vostra scelta (sia i “concetti” che le features)
    - Es. di features: proprietà semantiche, related words, synsets, occorrenza in documenti, ecc.
      - *Un esempio (usabile come no):*
        - Date 2 lingue in input (es. ita e eng)
        - *Concepts*: termini (di entrambe le lingue)
        - *Features*: synset di appartenenza