
ANNO ACCADEMICO 2024/2025

Tecnologie del Linguaggio Naturale

Teoria - Radicioni

Altair's Notes



UNIVERSITÀ
DI TORINO



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1	SEMANTICA LESSICALE	PAGINA 5
1.1	Introduzione	5
	Programma della Seconda Parte del Corso — 5 • Che cos'è la semantica lessicale? — 5 • Alcune Definizioni di Ontologia — 6 • Design di Ontologie — 8	
1.2	Knowledge Bases	8
CAPITOLO 2	KNOWLEDGE RAPRESENTATION	PAGINA 11
CAPITOLO 3	WORDNET E FRAMESET	PAGINA 13
CAPITOLO 4	NATURAL LANGUAGE TOOLKIT (NLTK)	PAGINA 15
CAPITOLO 5	NGRAMS	PAGINA 17
CAPITOLO 6	UN'INTRODUZIONE AI TRANSFORMERS	PAGINA 19
CAPITOLO 7	SEMINARI	PAGINA 21
7.1	Eye Tracking	21
7.2	Dementia Detection	21

Premessa

Licenza

Questi appunti sono rilasciati sotto licenza Creative Commons Attribuzione 4.0 Internazionale (per maggiori informazioni consultare il link: <https://creativecommons.org/version4/>).



Formato utilizzato

Box di "Concetto sbagliato":

Concetto sbagliato 0.1: Testo del concetto sbagliato

Testo contenente il concetto giusto.

Box di "Corollario":

Corollario 0.0.1 Nome del corollario

Testo del corollario. Per corollario si intende una definizione minore, legata a un'altra definizione.

Box di "Definizione":

Definizione 0.0.1: Nome delle definizioni

Testo della definizione.

Box di "Domanda":

Domanda 0.1

Testo della domanda. Le domande sono spesso utilizzate per far riflettere sulle definizioni o sui concetti.

Box di "Esempio":

Esempio 0.0.1 (Nome dell'esempio)

Testo dell'esempio. Gli esempi sono tratti dalle slides del corso.

Box di "Note":

Note:-

Testo della nota. Le note sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive.

Box di "Osservazioni":

Osservazioni 0.0.1

Testo delle osservazioni. Le osservazioni sono spesso utilizzate per chiarire concetti o per dare informazioni aggiuntive. A differenza delle note le osservazioni sono più specifiche.

1

Semantica Lessicale

1.1 Introduzione

1.1.1 Programma della Seconda Parte del Corso

Per quanto riguarda questa parte del corso ci si concentrerà sulla semantica lessicale, in particolare sui seguenti punti:

- Problema della Knowledge Representation.
- WordNet e BabelNet.
- FrameNet.
- N-grams.
- Word/Sense embeddings.
- Encoder/Decoder.

1.1.2 Che cos'è la semantica lessicale?

Definizione 1.1.1: Semantica Lessicale

La semantica lessicale è lo studio del significato delle parole e delle loro relazioni. Lo studio di cosa i singoli oggetti lessicali significano, cosa fanno, come possono essere rappresentati e combinati.

Sono scelte le ontologie per:

- *Parte metodologica*: adottare un piano molto interdisciplinare a un alto livello di generalità.
- *Parte architetturale*: centralità del ruolo che un'ontologia può giocare in un sistema informativo.

Le parole sono flessibili:

- Capacità di ripetere le parole nei *sensi* già noti, combinarle in enunciati anche nuovi, capacità di estendere una parola o una frase per esprimere nuovi sensi:
 - Queste capacità sono ereditate dagli esseri umani.
 - Non si sviluppano in ogni circostanza.
- La cooperazione tra le capacità del punto precedente è importante per sviluppare la *metalinguistica*:
 - Che cosa significa questa parola? Che cosa vuol dire? Si può dire così? Si scrive gocce o gocce?
 - Le parole hanno funzione riflessiva: si usa la *lingua per riflettere sulla lingua*.

Rigidità vs. Deformabilità:

- La chimica e la matematica parlano di determinati aspetti dell'esperienza (sono *rigide*).
- Le lingue invece sono *deformabili*: se ne può alterare o dilatare il significato.

1.1.3 Alcune Definizioni di Ontologia

Domanda 1.1

Cosa significa il termine ontologia?

Definizione 1.1.2: Ontologia Filosofica

In filosofia il termine ontologia indica lo studio dell'essere e delle sue categorie fondamentali. Un'ontologia definisce un insieme di primitive rappresentazionali con il quale modellare un dominio di conoscenza o di discorso. È un sistema organizzato in categorie e relazioni. Le loro definizioni includono informazioni sul loro significato e vincoli su come applicarle in maniera consistente.

Note:-

Per esempio una categoria rappresenta tutti i tipi di entità che possono fungere da soggetto in un predicato.

Problema: questa definizione ammette che quasi tutto può essere considerato un'ontologia.

Definizione 1.1.3: Ontologia come Concettualizzazione

Un'ontologia è una specifica esplicità di una concettualizzazione. Una struttura formale di un pezzo della realtà come percepito e organizzato da un agente indipendentemente da:

- Vocabolario utilizzato.
- L'occorrenza in una situazione specifica.

Note:-

Situazioni diverse coinvolgenti lo stesso oggetto descritte da un vocabolario diverso possono avere la stessa concettualizzazione.

Ontologia e semantica:

- Un'ontologia riguarda ciò che c'è, la semantica si riferisce a ciò che c'è.
- Differenti aspetti del linguaggio hanno differenti ruoli nell'ontologia.

Ontologie e Knowledge Bases:

- Componente *terminologica* (ontologia):
 - Indipendente da un particolare stato.
 - Pensata per supportare servizi terminologici.
- Componente *asserzionale*:
 - Riflette specifici stati.
 - Pensata per problem solving.

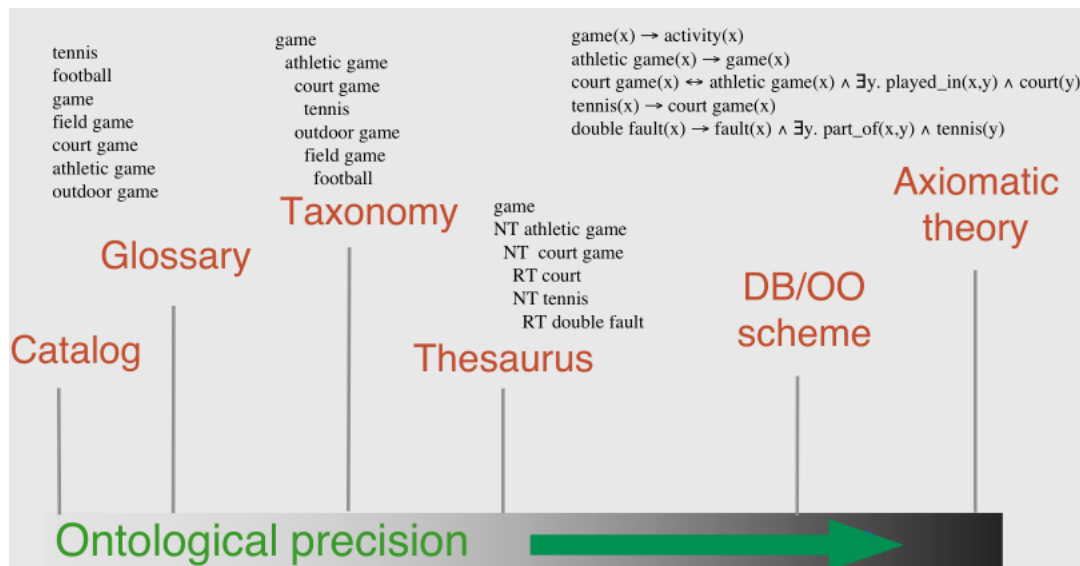


Figure 1.1: Livelli di precisione delle ontologie.

Ontologia e lessico:

- Livello lessicale: il *lessico* è un elenco di parole in una lingua (vocabolario + conoscenza sull'utilizzo delle parole).
- Relazioni lessicali: sinonimia (termini con lo stesso significato), iponimia (relazione di sottoclasse), iperonimia (relazione di superclasse), meronimia (costituente di qualcosa), olonimia (intero), antonimia (termini di significato opposto).

Caratteristiche del lessico:

- *Overlapping Word Senses:*
 - Nelle ontologie le sottocategorie di una categoria generale sono *mutualmente esclusive*.
 - Nel lessico esistono *sovrapposizioni di significato* (e.g. quasi sinonimi).
- *Buchi nel lessico:*
 - Il lessico di un linguaggio omette riferimenti a *categorie ontologiche* che non sono lessicalizzate in quel linguaggio.
 - Sono categorie che richiedono perifrasi (giri di parole) per essere definite.
- *Lessici tecnici:* nei contesti tecnici il linguaggio è molto vicino all'ontologia del dominio.

Domanda 1.2

Perché costruire ontologie?

- *Condivisione* della comprensione delle entità di un certo dominio.
- *Riutilizzo* dei dati e dell'informazione.
- Creare comunità di ricercatori.

1.1.4 Design di Ontologie

Entità ed Eventi:

- *Entità*: oggetti che continuano per un periodo di tempo mantenendo la propria identità.
- *Eventi*: oggetti che accadono, si svolgono o si sviluppano nel tempo.

Definizione 1.1.4: Ontologie Fondazionali

Un'ontologia fondazionale cattura un insieme di distinzioni base valide in vari domini.

Note:-

Alcune celebri sono DOLCE, SUMO e CYC.

Definizione 1.1.5: DOLCE

DOLCE è un'ontologia fondazionale con lo scopo di rappresentare le strutture concettuali di base che emergono dal linguaggio naturale e dalla cognizione umana.

Scelte alla base di DOLCE:

- *Endurant*: entità che sono completamente presenti in ogni momento della loro esistenza.
- *Perdurant*: eventi che si estendono nel tempo e sono parzialmente presenti in ogni istante.
- *Quality*: proprietà specifiche di un'entità, dipendenti da essa.
- *Approccio moltiplicativo*: differenti oggetti ed eventi possono essere co-localizzati nello spazio-tempo.

Definizione 1.1.6: Criteri di Identità

I criteri di identità sono utilizzati per valutare se due entità sono in relazione di sottoclasse e sono proprietà necessarie delle entità confrontate.

Note:-

Per esempio "intervallo di tempo" non può essere sottoclasse di "durata di tempo" perché utilizzano criteri diversi.

Endurants vs. Perdurants in DOLCE:

- Gli endurants possono cambiare nel tempo: lo stesso endurant può avere proprietà incompatibili in tempi diversi.
- I perdurants non cambiano, hanno una locazione spaziale ben definita dagli endurants che vi partecipano.
- La relazione tra endurants e perdurants è la *partecipazione*: un endurant vive partecipando in qualche perdurant.

Qualities e Quality regions in DOLCE:

- Le qualities possono essere viste come entità base che possono essere percepite e misurate (e.g. forma, colore, suono, etc.).
- Le qualities sono caratteristiche per specifici individui.
- Si distingue tra una qualità e il suo valore (e.g. il colore di una specifica rosa e quale sia la particolare sfumatura di rosso).

1.2 Knowledge Bases

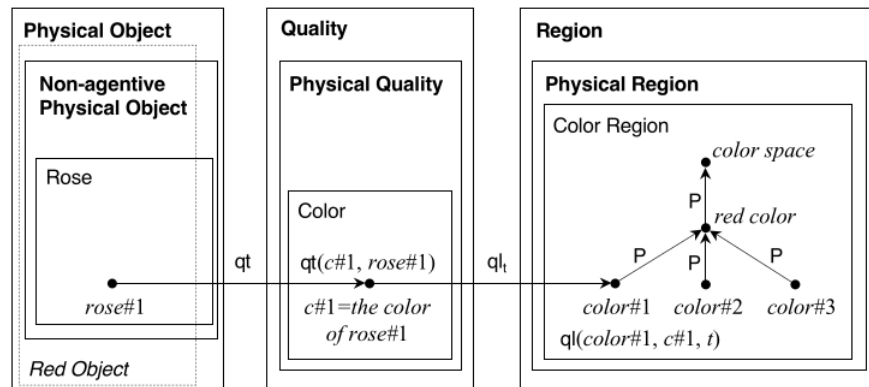


Figure 1.2: Rappresentazione del colore di una rosa.

2

Knowledge Representation

3

WordNet e FrameNet

4

Natural Language ToolKit (NLTK)

5

Ngrams

6

Un'Introduzione ai Transformers

7

Seminari

7.1 Eye Tracking

7.2 Dementia Detection

