
ANNO ACCADEMICO 2024/2025

Tecnologie del Linguaggio Naturale

Teoria

Altair's Notes



UNIVERSITÀ
DI TORINO



DIPARTIMENTO DI INFORMATICA

CAPITOLO 1

INTRODUZIONE ALLE TECNOLOGIE DEL LINGUAGGIO NATURALE **PAGINA 5**

1.1 Prologo

5

La Complessità del Linguaggio Naturale — 6 • I Livelli di Conoscenza del Linguaggio — 8

CAPITOLO 2

TEST2 **PAGINA 10**

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

Introduzione alle Tecnologie del Linguaggio Naturale

1.1 Prologo

La prima parte del corso sarà incentrata sulla linguistica computazionale generale, in cui ci si soffermerà sugli aspetti più tradizionali e linguistici¹. In questa parte verrà anche trattato il parsing. Nella seconda parte si andranno a studiare la semantica lessicale e le ontologie. Infine, nella terza parte del corso si andrà a studiare NLP statistico e distribuzionale.

Parte prima: keywords

- NLP
- CL
- Lexicon
- Morphology
- Syntax
- semantics
- Conversational Interface
- Conversational agent
- Dialogue System
- Parsing
- NLG
- MT
- Grammar
- Treebank
- NL ambiguity
- BOT
- LLM

Figure 1.1: Il giorno prima dell'esame bisogna sapere cosa significano tutte queste parole :3

¹Libro di riferimento: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. La prima e la seconda edizione, perché Jurafsky non riesce a finire il draft della terza :(

Le 4 ere della linguistica computazionale:

1. 1940 - 1969: primi tentativi.
2. 1970 - 1992: formalizzazione.
3. 1993 - 2012: apprendimento automatico.
4. 2013 - 2018: deep learning.

Note:-

Tutto cambiò nel 2018, quando NLP fu il primo successo su larga scala di rete neurale autosupervisionata.

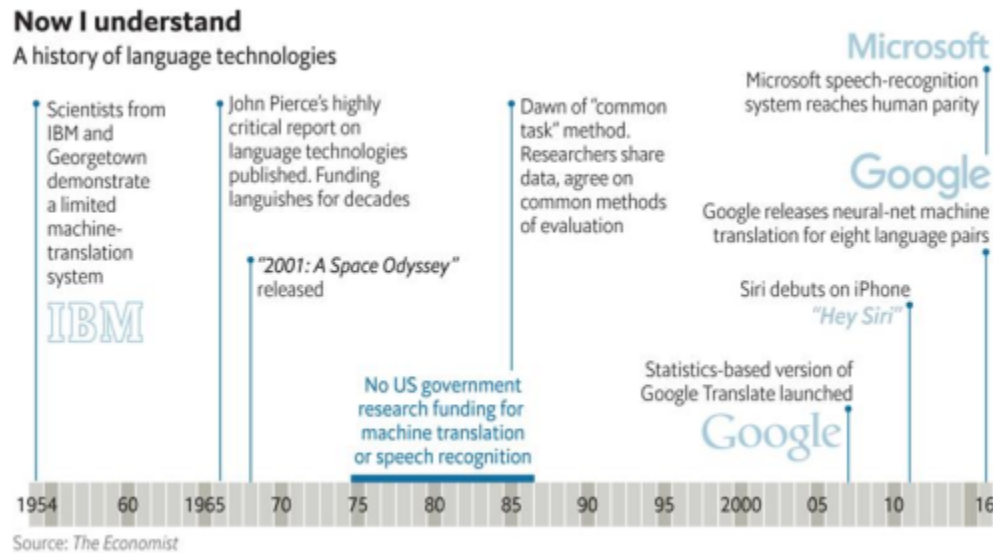


Figure 1.2: Il passato delle tecnologie del linguaggio naturale.

1.1.1 La Complessità del Linguaggio Naturale

C'è un legame tra linguaggio umano e intelligenza. Già Turing sosteneva che se si potesse parlare in un certo modo si fosse intelligenti (test di Turing). La differenza tra il linguaggio umano e un linguaggio di programmazione è l'*ambiguità*: C o Java non sono ambigui.

Il linguaggio umano:

- *Discretezza* (esistenza di elementi):
 - Api: Ritmo, orientamento, durata.
 - Esseri umani: Fonemi, morfemi, parole.
- *Ricorsività*:
 - Scimpanze: Gesti atomici.
 - Uomo: Gianni vede Pietro, Maria vuole che Gianni veda Pietro, Paolo crede che Maria voglia che Gianni veda Pietro.
- *Dipendenza dalla struttura*:
 - Non “una parola dietro l'altra” ma c'è una struttura: La ragazza parte, I ragazzi di cui mi ha parlato la ragazza partono.
- *Località*:
 - Gianni lo ha guardato.
 - Gianni ha detto che Pietro lo ha guardato.

Intelligenza e linguaggio nel il test di Turing:

- Possono le macchine pensare?
- Se riesco a parlare come un essere umano allora penso.
- Gioco dell'imitazione: un giudice deve capire se quello che ha davanti è un uomo oppure un computer.

Note:-

Ci sono una serie di obiezioni a questo test: teologia, matematica, coscienza, etc.

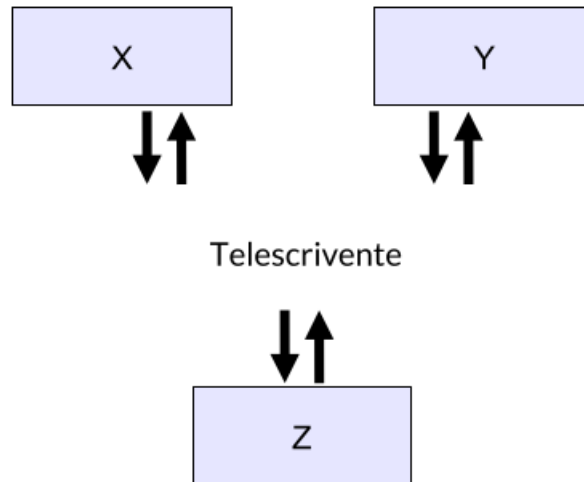


Figure 1.3: Il gioco dell'imitazione.

Nel 1966, Weizenbaum crea Eliza. Una macchina in grado di "comprendere" e ingannare gli esseri umani.

Note:-

Il punto debole del test di Turing e di Eliza è il giudice: se è coinvolto emotivamente potrebbe far passare un computer per un essere umano^a.

^aBlade runner moment

Definizione 1.1.1: Winograd Schema

Evoluzione del Turing test: un test a scelta multipla che utilizza domande con una specifica struttura. In questi test gli esseri umani sono molto bravi a rispondere, i computer no.

Note:-

Rimuove il giudizio, quindi tecnicamente più accurato.

Corollario 1.1.1 Captcha

Un test di Turing inverso per capire se l'interlocutore è umano. Non c'è linguaggio, ma riconoscimento cognitivo.

Corollario 1.1.2 Voight-Kampff Test

Test in Blade runner basato sulle emozioni, evoluzione del test di Turing.

1.1.2 I Livelli di Conoscenza del Linguaggio

HAL 9000, in "2001: Odissea nello spazio" mostra un esempio di comunicazione.

Domanda 1.1

Come fa HAL a rispondere?

- Riconoscimento vocale.
- Comprensione del linguaggio naturale.
- Generazione del linguaggio naturale.
- Sintesi vocale.
- Recupero ed estrazione di informazioni.
- Inferenza.

Livelli della conoscenza:

1. Il suono: HAL deve essere in grado di analizzare e produrre dei segnali audio che contengono le parole: foni e fonemi.
2. Le parole: HAL deve essere in grado di riconoscere le singole parole.
3. Raggruppare le parole: HAL deve essere in grado di distinguere la struttura della frase.
4. Significato: HAL deve conoscere il significato delle singole parole e deve essere in grado di comporre questi significati per trovare il significato complessivo della frase.
5. Contesto e scopi: HAL deve avere delle conoscenze del mondo che gli permettono usare il linguaggio in maniera contestuale: *I'm afraid*, *I can't* invece di *I won't*.
6. Conversazione: HAL deve essere in grado di conversare, dando delle risposte e facendo delle domande pertinenti al discorso.

A ogni livello corrisponde una parte del linguaggio:

1. Fonetica e Fonologia: lo studio del suono della lingua.
2. Morfologia: lo studio delle parti significative delle parole.
3. Sintassi: lo studio sulla struttura e sulle relazioni tra le parole.
4. Semantica: lo studio del significato.
5. Pragmatica: lo studio di come il linguaggio è usato per compiere goal. Il passivo serve per mettere in luce/enfatizzare alcune parti della frase.
6. Discorso: lo studio delle unità linguistiche rispetto alla singola dichiarazione.

Note:-

Jurafsky è un chad nerd.

2

Test2

