

Trattamento Automatico delle Lingue

R. Basili

a.a. 2004-2005

Obbiettivi del Corso

- Introduzione al trattamento dei fenomeni linguistici al calcolatore
- Applicazioni del TAL
- Metodi e tecnologie principali del TAL
 - Progettazione grammatiche
 - Uso di XML
- Laboratorio di TAL
 - Analisi morfosintattica al calcolatore
 - Analisi semantica
 - Annotazioni di corpora

Propedeuticità'

- Nozioni elementari di algebra e logica
- Corsi:
 - Abilità Informatiche
 - Modelli e Linguaggi dei Dati e della Conoscenza

Orario delle lezioni

- LUNEDI'
 - h. 13:00 – 15:00 AULA **T12A**
- MERCOLEDI'
 - h. 10:00 - 12:00 AULA **P2**
- GIOVEDI'
 - h. 15:00 - 17:00 AULA **T34**
- VENERDI'
 - h. 15:00 - 17:00 AULA **T19**

Lezioni, Esercitazioni e Tutoraggio

- Lezioni su aspetti metodologici del corso
- Esercitazioni su argomenti trattati durante le lezioni e di preparazione all'esame ed al progetto finale
- Ricevimento: Ogni Mercoledì' al termine della lezione

Testi di Riferimento

- J. Lyons, *"Introduzione alla Linguistica Teorica Sez II e III"*, Universale Laterza, 1978.
- D. Jurafsky, J. H. Martin, *"Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition"*, Prentice-Hall, 2000
- G. Chierchia, S. McConnell-Ginet-, *Significato e Grammatica: una introduzione alla semantica*, Il Mulino, 1997.
- Articoli scientifici di approfondimento

Organizzazione dell'Esame finale

- Discussione orale su tutto il programma
- Progetto Finale (alla fine del corso) su un o delle linee
 - Analisi sintattica (analisi e annotazione dei dati)
 - Analisi semantica (analisi e annotazione dei dati)
 - Approfondimento a scelta (bibliografia estesa)

Uno Sguardo al Programma

- **Introduzione al TAL**
 - Cenni di Storia dell'Elaborazione Automatica del Linguaggio Naturale (NLP). Applicazioni del TAL. Ricerca Automatica di Informazione. Traduzione Automatica. Comunicazione di Agenti Intelligenti.
 - Laboratorio: Le grammatiche formali; Annotazioni dei *corpora*; Il formalismo XML. TEI: una introduzione. Uso del linguaggio Prolog.

Uno Sguardo al Programma (2)

■ Richiami di Linguistica Generale

- Linguistica Computazionale, AI e TAL. Morfologia, Sintassi e Semantica. Lessico e Sintassi.
- Semantica: scopi della semantica. L'analisi semantica e le teorie logico-deduttive. Semantica lessicale. Lessici computazionali.
- La semantica basata sui *frames*. “*The Case for case*” (C. Fillmore). Il progetto Framenet.

Uno Sguardo al Programma (3)

■ Trattamento dei fenomeni della Sintassi

- **L'analisi morfologica e la sintassi.** Il processo di etichettamento della morfosintassi (*Part-of-Speech tagging*). Approcci statistici ed approcci basati su regole al POS tagging.
- **Grammatiche formali e riconoscimento grammaticale.** La gerarchia di Chomsky. Grammatiche libere dal contesto (*Context-free grammars*). Esempi di uso.
- **Laboratorio.** Il *Tree Tagger*. Le grammatiche *context-free* in Prolog. Il progetto Si-TAL. Il parser CHAOS.

Uno Sguardo al Programma (4)

■ Trattamento dei fenomeni semantici

- **Dizionari semantici:** Wordnet; Longman Dictionary of Contemporary English. Il problema del *Word Sense disambiguation*.
- **Interpretazione semantica.** Modelli di rappresentazione semantica. Modelli cognitivi della semantica.
- **Frames.** La *Frame Semantics*. Il progetto *FrameNet*.
- **Laboratorio.** Uso di Wordnet. Il progetto Multiwordnet. Etichettatura semantica. Analisi della risorsa Framenet. Progetto: verso un Framenet per l'italiano.

Uno Sguardo al Programma (5)

■ Applicazioni del TAL

- *Information retrieval* ed *Information Extraction*
- *Question Answering*
- Sistemi di Dialogo

Elaborazione di Linguaggio Scritto e Parlato al Calcolatore

- Cos'e'?
- Realizzare programmi in grado di completare processi linguistici come:
 - Abilitare una comunicazione uomo-macchina
 - Migliorare la comunicazione tra persone (ad es. MT)
 - Elaborare oggetti linguistici (ad es. Pagine Web, documenti o chiamate telefoniche)
- Esempi:
 - Question Answering
 - Traduzione Automatica
 - Agenti che dialogano

Quali conoscenze servono?

- HAL 9000, da *"2001: A Space Odyssey"*
- Dave: *Open the pod bay doors, Hal.*
- HAL: *I'm sorry Dave, I'm afraid I can't do that.*

Qual'e' la conoscenza di HAL?

- Riconoscimento e sintesi del linguaggio parlato
 - Dizionari (pronuncia delle parole)
 - Fonetica (come riconoscere/produrre ciascun suono dell'Inglese)
- Comprensione della lingua
 - Conoscenza delle parole dell'Inglese
 - Cosa significano
 - Come si combinano (cos'e' un *`pod bay door`*?)
 - Conoscenza della struttura sintagmatica
 - *I'm I do, Sorry that afraid Dave I'm can't*

Qual'e' la conoscenza di HAL? (2)

- Dialogo e pragmatica
 - “*open the door*” e' una richiesta (non una affermazione o una ricerca di informazioni)
 - Rispondere e' un atto gentile anche se si e' pianificato di uccidere.
 - E' bene mostrarsi cooperativi (*I'm afraid, I can't...*)
 - Cosa significa *`that`* in *`I can't do that`*?
- Presino un sistema telefonico di prenotazione aerea richiede approssimativamente la stessa conoscenza

Question Answering

- Cosa significa “*porta*”?
- In quale anno e' nato Mozart?
- Quante erano le provincie italiane sino al 1995?
- C'era uno sconto sull'acquisto dei libri di linguistica da Amazon ieri?
- Cosa pensano gli scienziati riguardo alla legalizzazione della clonazione?

Modelli e Algoritmi

- **Modelli:** formalismi per esprimere diversi tipi di strutture linguistiche.
 - Automi (FSA, traduttori, modelli di markov)
 - Sistemi formali di regole (grammatiche *context-free*, sistemi di *feature*)
 - Logica (calcolo dei predicati ed inferenza)
 - Modelli Probabilistici di quelli di sopra
- **Algoritmi** usati per manipolare le rappresentazioni e creare altre strutture.
 - Alg. Di Ricerca (A*, programmazione dinamica)
 - Apprendimento Induttivo etc etc

Linguaggio, Pensiero e Comprensione

- Il Test di Turing
- La domanda “Può una macchina pensare?” può essere resa operativa.
- Versione operativa:
 - 2 persone ed un computer
 - Una persona interroga via rete l'altra E il computer
 - La macchina deve convincerla di essere umano
 - La seconda persona deve fare la stessa cosa!

Eliza: Weizenbaum (1966)

- User: You are like my father in some ways.
- ELIZA: What resemblance do you see
- User: You are not very aggressive but I think you don't want me to notice that.
- ELIZA: What makes you think I am not aggressive
- User: You don't argue with me
- ELIZA: Why do you think I don't argue with you
- User: You are afraid of me
- ELIZA: Does it please you to believe I am afraid of you

Esempio di trattamento di testo

- Trova tutte le occorrenze della parola “il” in un testo.
 - `/i\l/`
 - Perde molti esempi in maiuscolo
 - `/[iI]l/`
 - Restituisce “mille” o “Ildebrando”
 - `/\b[tT]he\b/`

Errori nel riconoscimento

- Abbiamo eliminato due tipi di errori
 - Il reperimento di stringhe che non dovevano essere trovate (Iliade, mille)
 - Falsi Positivi
 - Il mancato riconoscimento di stringhe corrette (ad es, “Il poliziotto ..”)
 - Falsi negativi

Errori (2)

- Cio' e' vero in generale nel TAL.
- Ridurre l'errore significa cercare due miglioramenti antagonisti:
 - Migliorare la accuratezza (meno falsi positivi)
 - Migliorare la copertura (meno falsi negativi).