

Allegato 1 – Verbale Terza Adunanza

Elenco Quesiti Prova Orale

1

- Considerando che l'identificazione delle unità fondamentali di analisi, i token, è requisito preliminare nella consultazione di grandi corpora, si descriva il processo di tokenizzazione e le sue problematiche specificamente per la lingua italiana e per una lingua classica.
- Allo stato dell'arte, se dovesse predisporre DA ZERO una soluzione computazionale per etichettare un corpus per parti del discorso come procederebbe? Quali tecnologie utilizzerebbe? Applicherebbe le medesime tecnologie anche a una lingua classica?
- Si descriva, nell'opportuno livello di dettaglio, la differenza che occorre tra l'esecuzione di un calcolo su una CPU e il medesimo calcolo su una GPU.
- Il Consiglio di Amministrazione
- Cosa si intende per Open Access?

2

- Se dovesse eseguire un'indagine linguistica su un fenomeno specifico dell'italiano contemporaneo (es. le frasi relative), utilizzerebbe il Web come corpus? Argomenti la risposta.
- È necessario annotare un grande corpus (almeno 100 milioni di token) ripetto alle categorie grammaticali; quali strumenti e tecnologie metterebbe in atto per raggiungere lo scopo? Si considerino anche aspetti relativi alle lingue classiche.
- Le GPU hanno consentito sviluppi impensabili anche solo 10 anni fa nell'ambito dell'analisi dei dati e delle reti neurali profonde. Si descrivano le operazioni, in questi ambiti, che possono beneficiare di simili strumenti.
- Il Senato accademico
- Cosa si intende per SVC-Spazi Virtuali di Collaborazione?

3

- L'estrazione automatica dei collocati di una parola in un corpus è uno strumento fondamentale nell'analisi linguistica; dopo aver definito la nozione di collocato si discutano le procedure informatiche per la loro estrazione e le problematiche connesse.
- I Word-Space Model a livello lessicale basati su modelli neurali sono diventati strumenti standard nell'analisi linguistica con strumenti computazionali. Dopo una breve descrizione generale di questi modelli, illustri come costruirebbe uno di tali modelli per la lingua italiana e per una lingua classica. Questi modelli possono essere utili anche per lo studio delle lingue classiche?

GR
GG
FR
P
B

- In un linguaggio di programmazione a sua scelta, descriva come implementerebbe un algoritmo che scorra tutte le linee di un file di testo.
- Gli organi di Dipartimento
- Cos'è la INTRANET?

4

- La consultazione di grandi corpora richiede procedure di text retrieval complesse. Si descrivano nel dettaglio tali procedure e si suggeriscano pacchetti software open source in grado di supportare queste procedure.
- Si discutano le metodologie connesse alla risoluzione di task nell'ambito del Natural Language Processing attraverso il Fine-tuning di pre-trained Language Model sia dal punto di vista modellistico che delle infrastrutture tecnologiche necessarie. E' possibile applicare tali metodologie anche alle lingue classiche?
- Per collegare un elaboratore alla rete Internet è necessario specificare quattro parametri, manualmente o ottenuti automaticamente dal sistema. Si descrivano i quattro parametri e la loro funzione.
- Funzioni del Direttore Generale
- Cos'è la firma elettronica?

5

- Uno studio linguistico di matrice empirista, quindi basato su corpora, come dovrebbe procedere nell'analisi di un generico fenomeno linguistico X in una varietà linguistica Y?
- Il Word-Sense Disambiguation è uno dei task fondamentali nell'ambito della linguistica computazionale. Si descrivano le risorse e le metodologie che utilizzerebbe per risolvere il task per una lingua a scelta e per una lingua classica.
- I "siti Web" sono da anni diventati uno strumento fondamentale per mostrare i risultati di ricerche scientifiche in tutte le discipline. Si descrivano gli strumenti tecnologici che proporrebbe di utilizzare per la creazione DA ZERO di un sito Web, predisposto su un server LOCALE, in ambito umanistico.
- I corsi di studio di primo e secondo ciclo
- Cosa si intende per SPID?

6

- Nell'ambito della costruzione di corpora, la codifica dei caratteri di un testo elettronico presenta numerosi problemi. Si presenti lo stato dell'arte in proposito e si discutano tali problemi anche in relazione alle lingue classiche. Quali soluzioni tecnologiche utilizzerebbe per affrontare i problemi descritti?
- Lingue con una ricca morfologia flessiva pongono problemi specifici all'analisi automatica. Si descrivano questi problemi, i task rivolti alla gestione di tali fenomeni e gli strumenti informatici necessari a una loro corretta gestione anche in relazione alle lingue classiche.

Handwritten notes on the right margin: "sp", "gg", "22", "3".

- Nell'ambito dei sistemi operativi per elaboratore, si descriva la funzione della memoria virtuale e, a grandi linee, come viene implementata.
- Funzioni del Rettore
- Differenza tra dato e informazione

7

- L'inserimento delle annotazioni linguistiche all'interno dei testi che compongono un corpus è una procedura delicata che richiede numerose accortezze. Si descrivano i problemi che comporta e le metodologie che si ritengono più appropriate considerando le diverse tipologie di annotazione (a livello di parola, a livello di struttura, a livello di relazioni, ecc.).
- Nel Natural Language Processing, una specifica architettura di reti neurali profonde è divenuta il riferimento per tutte le soluzioni applicative atte alla risoluzione dei vari task. Si identifichi tale architettura e se ne descrivano, nell'opportuno dettaglio, i tratti salienti. Quali problemi potrebbe presentare l'uso di tale architettura per l'elaborazione delle lingue classiche?
- In un linguaggio di programmazione a sua scelta, descriva come implementerebbe un algoritmo per la ricerca di una sottostringa in una stringa data, senza utilizzare funzioni già implementate per questo scopo.
- Le fonti di finanziamento dell'Università/Gli organi ausiliari di Ateneo
- Cos'è la PEC?

W
F
G
G
G