

TEMA A

- Quesito 1
Il candidato discuta i principali approcci all'elaborazione del linguaggio naturale (rule-based e statistico), fornendo esempi di applicazione ed evidenziandone vantaggi e svantaggi.
- Quesito 2
Il candidato discuta nel dettaglio i principali strumenti informatici (pacchetti software e/o linguaggi di programmazione) che utilizzerebbe per la raccolta e organizzazione di dati testuali finalizzata alla costruzione di un corpus monolingue annotato.

TEMA B

- Quesito 1

Il candidato discuta le problematiche legate alla codifica di caratteri nei testi in formato digitale, evidenziando le problematiche legate all'utilizzo dei diversi standard di codifica.

- Quesito 2

Il candidato discuta dettagliatamente le problematiche, tecniche e legali, legate alla raccolta, utilizzo ed eventuale redistribuzione di dati linguistici dal Web.

TEMA C

- Quesito 1
Il candidato discuta vantaggi e svantaggi dell'utilizzo di database relazionali in contrapposizione a formati XML nel trattamento dei dati linguistici.
- Quesito 2
Il candidato discuta le principali applicazioni linguistiche di un corpus monolingue, esponendo nel dettaglio anche i principali strumenti informatici a disposizione (sia online che offline) per l'analisi dei corpora.