

Laurea Magistrale in Editoria e Giornalismo
Sistemi per il recupero delle informazioni
Pre-appello del 19 dicembre 2014

Cognome e nome: _____ **Matricola:** _____

Domanda 1	Domanda 2	Domanda 3	Domanda 4	Domanda 5	Domanda 6	Totale
/6	/6	/6	/6	/5	/4	/33

Istruzioni:

- ◆ È vietato portare all'esame libri, eserciziari, appunti e dispense. Chiunque venga trovato in possesso di documentazione relativa al corso, in formato analogico e/o digitale, – anche se non strettamente attinente alle domande proposte – vedrà annullata la propria prova.
- ◆ Scrivere solo sui fogli distribuiti, cancellando le parti di brutta con un tratto di penna. Non separare questi fogli. Non utilizzare la penna rossa. Scrivere nome e cognome su tutti i fogli utilizzati.
- ◆ Tempo a disposizione: 1 ora e 30 minuti.

- 1) Rappresentare graficamente il processo di indicizzazione di documenti testuali, e spiegare brevemente cosa è la lemmatizzazione.
- 2) Descrivere le misure di “precisione” e “richiamo”. Quale rapporto vi è tra esse?
- 3) Rappresentare graficamente e descrivere l'architettura centralizzata crawler-indexer.
- 4) Cosa è TF-IDF? Per cosa viene usato e su quali idee si basa?
- 5) Descrivere la legge di Zipf. In cosa può essere di aiuto nell'ambito dell'indicizzazione dei documenti testuali?
- 6) Cosa è una sitemap e quali scopi ha?