

## Testo del progetto

Obiettivo del progetto è la realizzazione di un prototipo di motore di ricerca (semplificato) che, una volta acquisiti i dati relativi ad una collezione di testi, consenta di effettuare la ricerca di una o più parole all'interno dei testi considerati. Il progetto dovrà essere realizzato su piattaforma windows ma a "linea di comando" cioè con una interfaccia testuale, come specificata nel seguito.

### Specifiche del programma da realizzare

Il sistema deve presentare all'utente una interfaccia principale dalla quale è possibile attivare le varie funzionalità offerte dal sistema:

1. Acquisizione e memorizzazione in forma opportuna dei dati da elaborare (i file di testo su cui effettuare le interrogazioni).
2. Interrogazione - singolo termine - risultati non ordinati.
3. Interrogazione - singolo termine - risultati ordinati.
4. Interrogazione - termini multipli - risultati non ordinati.
5. Interrogazione - termini multipli - risultati ordinati.

Nessuna interrogazione può essere eseguita se non si è preventivamente effettuata almeno una operazione di acquisizione dei dati. Chiaramente, l'interfaccia principale deve anche consentire l'uscita dal sistema.

#### *Acquisizione dati*

All'attivazione di tale funzionalità, l'utente dovrà specificare il percorso completo della cartella in cui si trovano i file di testo da acquisire. Se la cartella viene correttamente rilevata all'interno del file system, allora tutti i dati dei file contenuti in tale cartella dovranno essere opportunamente acquisiti e memorizzati, altrimenti si dovrà stampare un messaggio di errore e si dovrà dare la possibilità o di specificare un nuovo percorso oppure di ritornare all'interfaccia principale.

#### *Interrogazioni varie*

Le varie funzionalità di interrogazione consistono nella ricerca di una o più parole all'interno dei file acquisiti dal sistema tramite l'ultima operazione di acquisizione dati. Per "parola" o "termine" si intende una qualunque stringa alfanumerica (escludendo quindi segni di punteggiatura e simboli speciali).

In particolare, nel caso delle funzionalità 2. e 3. l'utente dovrà specificare la parola da cercare e il sistema dovrà restituire in output l'elenco dei nomi di file contenenti almeno una occorrenza della parola cercata. Se viene attivata la funzionalità 2. l'ordine dei file elencati può essere qualsiasi, mentre se si attiva la funzionalità 3. l'elenco con i nomi dei file dovrà essere ordinato in base al numero di occorrenze della parola presenti nei file (dal file con più occorrenze a quello che ne contiene di meno).

Le funzionalità 4. e 5. differiscono rispettivamente dalle funzionalità 2. e 3. solo in quanto è possibile specificare più di una parola. Per realizzare l'ordinamento dei file (funzionalità 5) si deve considerare la somma totale delle occorrenze per ogni parola cercata. In questo caso un file apparterrà al risultato solo se la somma delle occorrenze dei singoli termini è maggiore di zero (i termini sono posti in OR tra loro).

Chiaramente il risultato sarà un elenco vuoto in caso di ricerca con esito negativo. Dopo aver mostrato il risultato della ricerca, l'utente potrà eseguire una nuova ricerca (dello stesso tipo dell'ultima effettuata) oppure ritornare all'interfaccia principale del sistema.

## **Organizzazione e regole:**

La valutazione complessiva dell'esame consiste:

- in una valutazione delle prestazioni del progetto (massimo 10 punti, esclusi eventuali bonus)
- una discussione della relazione (massimo 5 punti)
- una valutazione orale (massimo 15 punti).

Il progetto è obbligatorio per tutti gli studenti.

Il progetto è individuale. Non sono ammessi gruppi.

Non è messo alcun limite al linguaggio di programmazione da usare ed alle librerie e strutture dati. L'unico vincolo (per una valutazione uniforme dei vari progetti) è che il programma deve poter essere eseguito su piattaforma windows.

Il progetto va consegnato al massimo entro la data che sarà indicata in ogni seduta d'esame; è necessario consegnare **personalmente o via e-mail** (alla Dott.ssa Cozza):

- Il file (.exe o .jar) del programma avente come nome il numero di matricola, il cognome ed il nome dello studente (es. 10025RossiMario.exe)
- I file sorgenti del programma
- Una relazione contenente almeno tre sezioni distinte in cui si inserisce:
  1. una descrizione dettagliata dell'approccio e delle strutture dati adottate,
  2. l'analisi di complessità asintotica nel caso peggiore dell'algoritmo sviluppato per ciascuna funzionalità,
  3. considerazioni sulle caratteristiche dell'input per le quali sono state fatte le scelte progettuali (ad es. struttura dati scelta) e le motivazioni per cui si ritengono tali scelte vincenti

Si noti che se si sceglie di inviare il progetto via e-mail è necessario comprimerlo in un file .zip in quanto molti programmi per la posta elettronica cancellano automaticamente i .exe dalla posta in uscita.

## **Valutazione**

- Se il programma non è corretto, cioè dà errori di esecuzione, non riesce ad elaborare correttamente l'input dai file, o non calcola correttamente quello che dovrebbe, lo studente deve riconsegnare il progetto. In questo caso lo studente non è ammesso alla prova orale.

Puramente ai fini della valutazione positiva del progetto ed all'accesso all'orale ed alla discussione della relazione è sufficiente che il sistema implementi correttamente le funzionalità 1. e 2.

La corretta realizzazione di ognuna delle funzionalità porterà ad una assegnazione di un punteggio pari a:

- 2 per i punti 1. e 2.
- 2 per il punto 3.
- 3 per il punto 4.
- 3 per il punto 5.

Per un totale di 10 punti complessivi.

In ciascun appello inoltre, sarà stilata una graduatoria in base alla velocità di esecuzione del punto 5. di ciascun progetto, su un insieme di file in input di riferimento (che non saranno resi pubblici prima della valutazione!):

- Il sistema più veloce in assoluto (sia rispetto all'appello in corso, sia rispetto a sistemi consegnati in eventuali appelli precedenti) riceverà un bonus di 5 punti che verranno conteggiati solo se con tutte le altre prove si raggiunge almeno il 18.

- I primi 5 classificati (i 5 sistemi più veloci sull'input di riferimento) riceveranno un bonus in punti, legato al voto dell'orale secondo le modalità specificate di seguito; tale bonus verrà sommato al voto dell'orale solo se esso è superiore a 9.

I due tipi di bonus suesposti non sono cumulabili.

Il bonus assegnato ai primi cinque classificati avrà il seguente valore:

- Voto orale tra 9 e 10 : Bonus = 1 punto
- Voto orale tra 11 e 12 : Bonus = 2 punti
- Voto orale tra 13 e 15 : Bonus = 3 punti

N.B. Il tempo di esecuzione registrato per il sistema consegnato (e quindi tutto il meccanismo dei bonus) verrà considerato solo per l'appello in cui esso è stato consegnato. Nell'eventualità che lo studente non superi l'orale, ma il progetto dia comunque esito positivo, verrà conservato solo l'esito del progetto per i successivi due appelli. L'eventuale ripartecipazione alla competizione per il bonus potrà essere ottenuta dallo studente solo riconsegnando una nuova versione del progetto (un nuovo eseguibile).