

ELEMENTI DI INGEGNERIA DEL SOFTWARE

Progetto per Esame - v0.2

Emanuele Di Buccio

19/04/2023

Laurea in Ingegneria Informatica, A.A. 2022/2023

REGOLE GENERALI

- Il progetto dovrà essere realizzato in **gruppo composto da 4 persone**.
- Utilizzare il **linguaggio Java** (unit testing utilizzando JUnit).
- Il codice dovrà essere opportunamente commentato e indentato.
Commentare il codice in modo da poter utilizzare Javadoc.
- La documentazione dovrà essere prodotta come gdoc o, in alternativa, come pagina Web usando il Maven site plugin.
- Per i diagrammi UML si consiglia di utilizzare un tool; si consiglia draw.io, ma potete utilizzare anche altri strumenti, e.g., PlantUML, Lucidchart, miro.
- Si consiglia di utilizzare un IDE, e.g., IntelliJ o Eclipse.
- Per le dipendenze potete usare Maven.

GRUPPI PER PROGETTO

- Inviare una email con la composizione del gruppo riportando per ciascun membro del gruppo
 - Nome e cognome
 - Numero di matricola
 - Email istituzionale

Inviare una sola email e mettere in CC gli altri membri del gruppo.

- Se non avete trovato un gruppo inserite la vostra email qui.
È necessaria l'autenticazione tramite account di ateneo.
Potete usare le email per contattarvi e creare i gruppi.

Se non riceverò i gruppi entro il giorno 01/05/2023, i gruppi verranno estratti a random tra i nominativi presenti nel file.

- Il codice dovrà essere consegnato **mediante un archivio** (zip, gz, bz2, tar.gz, tar.bz2) **caricato su compito moodle** relativo all'appello in cui si desidera fare la presentazione. Il compito moodle verrà reso disponibile nelle prossime settimane in una sezione della pagina del corso.
- L'archivio dovrà includere codice e documentazione
- Date di consegna:
 - 14/06/2023 per appello del 21/06/2023
 - 11/07/2023 per appello del 18/07/2023
- **Nota:** **la consegna è di gruppo** ed è quindi sufficiente che un solo membro del gruppo la effettui tramite il compito moodle. Per poter effettuare la consegna è necessario aver comunicato la composizione del gruppo al docente in modo da creare un gruppo corrispondente su moodle.

1. **Documento requisiti.** Gli use case rappresentati con diagrammi UML e con descrizione in linguaggio naturale strutturato.
2. **Domain model.** In UML, se necessario con descrizione testuale.
3. **Design Model.** In UML, con descrizione testuale dove necessario. Deve includere modelli strutturali e dinamici.
4. **Codice** (java + eventuali file di compilazione/jar)
5. **Test** (documento + report generato dai test)
6. **Manuale.** Deve contenere:
 - una panoramica concisa e ad alto livello del progetto (come sono state implementate le funzionalità richieste)
 - istruzioni su come installare ed eseguire il software
 - indicazione di quali funzioni sono state riutilizzate da librerie esistenti, specificando la versione delle librerie

Verifiche finali prima della consegna

- Rileggere attentamente le specifiche di progetto
- E' stato consegnato tutto quello che era richiesto nelle specifiche di progetto?
- I diagramma sono completi? Nel diagramma delle classi tutte le associazioni hanno un nome e cardinalità? Le classi hanno gli attributi necessari?
- Il codice rispecchia la documentazione prodotta?
- Le istruzioni nel manuale sono sufficienti ad eseguire il codice?

La valutazione del progetto sarà basata sui seguenti aspetti:

- la realizzazione delle specifiche funzionali
- l'organizzazione e la leggibilità del codice
- l'adeguatezza della documentazione allegata al progetto
- la corretta definizione dei test
- la discussione del progetto

Progettare ed implementare un sistema software in grado di scaricare (download) articoli da testate giornalistiche online resi disponibili da diverse sorgenti e di estrarre e visualizzare i termini più “importanti” nell’insieme degli articoli scaricati.

Sorgenti (sources)

- file CSV che verrà fornito su moodle
 - la prima riga sarà l'header, con la lista di attributi/campi
- The Guardian API (<https://open-platform.theguardian.com/>)
 - È necessaria una API Key; registrarsi [qui](#) per un accesso da Developer.
 - Documentazione disponibile [qui](#); vi servirà lo API Endpoint per accedere al contenuto (search); la documentazione per tale endpoint è disponibile [qui](#)
 - È possibile usare il client Java disponibile su github:
<https://github.com/matarrese/content-api-the-guardian>
 - Sarà messo a disposizione su moodle un esempio di risposta data dallo API Endpoint; la risposta è in formato [JSON](#)

PROGETTO - ESTRAZIONE TERMINI 1/-

- Con **termine** si intende una parola che compare nel testo dell'articolo (titolo + corpo)
 - *Esempio:* nella stringa "this is a test" i termini sono: "this", "is", "a", "test"
- Estrarre i 50 termini con maggior peso e memorizzarli in un file di testo (txt), dove ciascuna riga deve essere nel formato: "<termine> <peso>"
 - *Esempio:*
test 32
nuclear 16
fusion 7
power 7

Se ci sono pareggi in termini di peso, utilizzare l'ordine alfabetico come criterio di ordinamento aggiuntivo.

- Per estrarre i termini potete usare StringTokenizer oppure le funzionalità messe a disposizione da una libreria come CoreNLP (vedere le Pipeline ed in particolare il “passo” Tokenization)

PROGETTO - RICHIESTE

1. Il sistema deve poter supportare nuove sorgenti
2. Dopo la fase di download, deve essere effettuata la persistenza su file degli articoli usando lo stesso formato per tutti gli articoli di tutte le sorgenti; dovete scrivere voi il serializzatore ed il deserializzatore. Il deserializzatore serve per ricaricare gli articoli a partire dai file.
3. Il sistema deve poter supportare nuove modalità di memorizzazione ed accesso agli articoli.
4. Per estrarre i termini ed il loro peso (numero di documenti in cui appare), è necessario partire dai file in cui gli articoli sono memorizzati (punto 2).
5. Il sistema deve poter supportare nuove strutture per memorizzare ed avere accesso ai termini più importanti
6. L'utente deve poter specificare se eseguire solo il download, solo l'estrazione dei termini a partire dai file in cui sono stati memorizzati gli articoli, o entrambe le azioni in sequenza.

- La versione di Java consigliata è la 8; se è necessario usare una versione differente, motivare perché e specificare la versione esatta nel manuale