

Documento dei requisiti di specifica Software

Indice dei contenuti

- 1. Prefazione**
- 2. Introduzione**
 - 2.1 Obiettivo
 - 2.2 Esplicazione dell'obiettivo
- 3. Presentazione Software**
 - 3.1 Funzionalità software
 - 3.2 Vincoli software
 - 3.3 Requisiti funzionali
 - 3.4 Interfaccia utente
- 4. Requisiti non funzionali**
- 5. Diagrammi**
 - 5.1 Diagramma dei casi d'uso
 - 5.2 Diagramma di classe
 - 5.3 Diagrammi di sequenza
 - 5.4 Diagramma di attività

1.Prefazione

Il software è stato realizzato da un gruppo di 4 studenti universitari ed è stato implementato puramente a scopo didattico, pertanto, non è indirizzato ad un particolare tipo di utenza, se non alla sola lettura del docente universitario.

2.Introduzione

Il seguente documento è stato creato per specificare i requisiti del progetto software “Top 50 words”, realizzato ai fini dell’esame del corso Elementi di Ingegneria del Software presso la facoltà di Ingegneria Informatica dell’Università di Padova.

2.1 Obiettivo

Lo scopo del progetto è quello di implementare un software che sia in grado di fare il download di articoli da testate giornalistiche online resi disponibili da diverse sorgenti e di estrarre e visualizzare i termini più “importanti” nell’insieme degli articoli scaricati.

2.2 Esplicazione dell’obiettivo

Nel software, le sorgenti disponibili saranno i file .CSV¹ e The Guardian API². Da quest’ultima verrà eseguito il download di 1000 articoli della testata giornalistica The Guardian. Mentre nella prima sono disponibili 1000 articoli di NYTimes all’interno di un unico file.

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”.

Verranno, infine, estratti i 50 termini con maggior peso e memorizzati in un file di testo (txt).

¹ .CSV: Comma Separated Values

² API: Application Programming Interface

3. Presentazione Software

3.1 Funzionalità Software

Il software deve:

- poter supportare nuove sorgenti
- effettuare, dopo la fase di download, la persistenza su file degli articoli usando lo stesso formato per tutti gli articoli di tutte le sorgenti
- poter supportare nuove modalità di memorizzazione ed accesso agli articoli
- poter supportare nuove strutture per memorizzare ed avere accesso ai termini più importanti

L'utente deve poter specificare se:

- eseguire solo il download
- eseguire solo l'estrazione dei termini a partire dai file in cui sono stati memorizzati gli articoli
- eseguire entrambe le azioni in sequenza

L'utente si interfaccia con il programma tramite *CLI*³

3.2 Vincoli Software

Il software necessita che sia installata una versione di Java pari o superiore alla 8.

3.3 Requisiti funzionali

3.3.1 Download articoli da API

Consente all'utente di scaricare 1000 articoli dal sito The Guardian e salvarli nel formato Json.

Selezionare l'opzione nel menù interattivo "Download dei file json del TheGuardian", digitando da tastiera il numero 1. Dopo aver verificato che la API key sia valida, il sistema interroga l'API Endpoint ed effettua il download degli articoli.

Gli articoli vengono elaborati e salvati in un file di formato Json.

Successivamente c'è la possibilità di serializzare questi articoli in un file .txt digitando da tastiera il numero 2 nel menù interattivo, e poi premendo 1.

3.3.2 Lettura articoli da CSV

Consente all'utente di esaminare 1000 articoli da file .CSV, situato nel percorso `./Files/NY Times/nytimes articles v2.csv`. Selezionare l'opzione nel menù interattivo "Serializza i file sorgente in file txt", digitando da tastiera il numero 2.

Successivamente, digitare il numero 2 per effettuare la serializzazione degli articoli del New York Times. Gli articoli, dopo la serializzazione, vengono salvati in un file di formato .txt.

³ CLI: Command Line Interface

3.3.3 Estrazione dei termini

Consente all'utente di visualizzare i 50 termini più frequenti all'interno degli articoli New York Times/The Guardian serializzati.

Il programma procederà a deserializzare gli articoli in oggetti di tipo Article, successivamente verranno analizzati i seguenti oggetti e verranno estratte le 50 parole più utilizzate. Nel file "output.txt" verranno salvati i termini ed il corrispettivo numero di occorrenze.

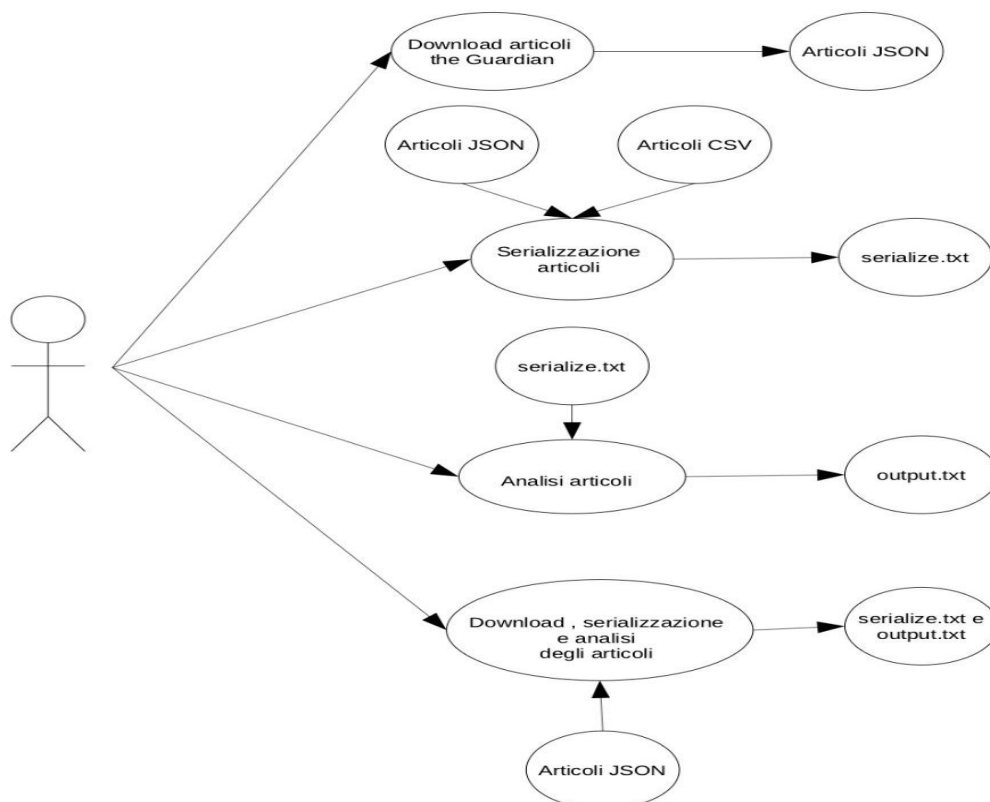
4. Requisiti non funzionali

In caso di acquisizione dei dati tramite l'API, il sistema dipende dall'utilizzo di risorse disponibili su Internet. In assenza di una connessione Internet, risulterà dunque impossibile recuperare gli articoli.

Inoltre è indispensabile possedere una chiave API valida per effettuare il download degli articoli.

5. Diagramma

Diagramma dei casi d'uso



Il programma permette all'utente di scegliere tra 4 opzioni:

- Download articoli the Guardian: scarica gli articoli tramite API salvandoli in Json
- Serializzazione articoli: serializza i file sorgente in file txt.
- Analisi articoli: analizza gli articoli selezionati, estraendo i termini con maggior peso salvandoli in un file txt.
- Download, serializzazione ed analisi articoli The Guardian: esegue tutti i passaggi precedenti.

Diagramma di classe

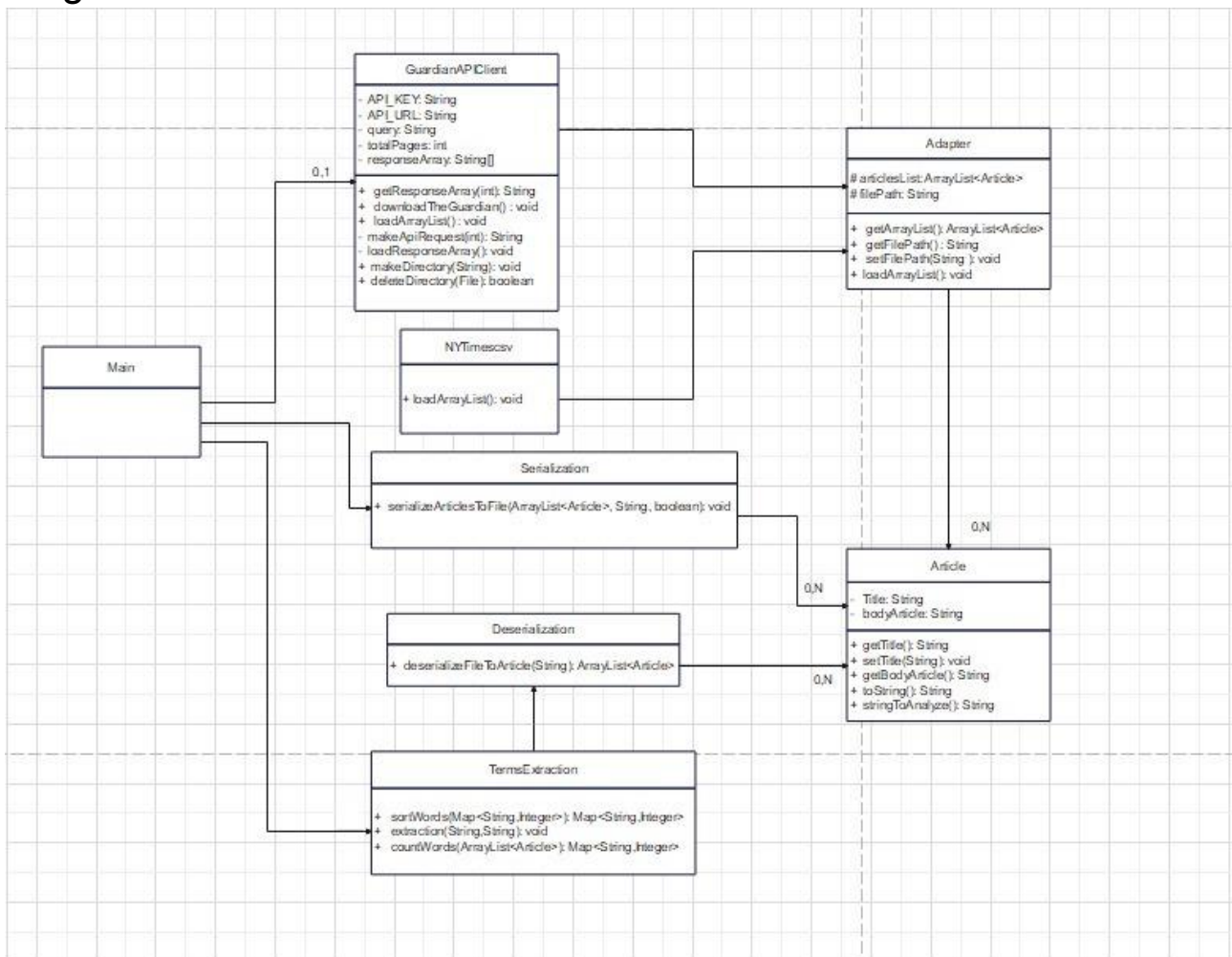


Diagramma di Sequenza

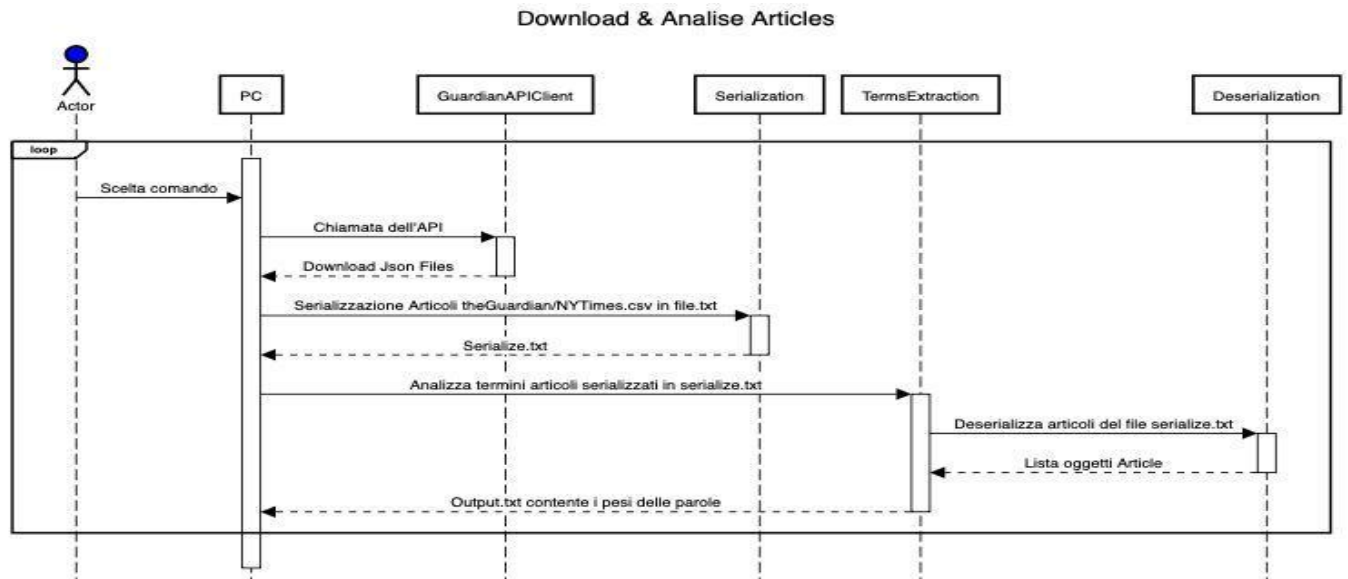


Diagramma di attività

