

# Manuale d'uso

WebMiningProject

Versione: 11072023\_1.0

Autori: Van Sang Ho, Marco Toffoletto

## Sommario

---

<b>1. Finalità del software.....</b>	<b>2</b>
<b>2. Funzionalità .....</b>	<b>2</b>
<b>3. Librerie utilizzate .....</b>	<b>3</b>
<b>4. Requisiti di sistema.....</b>	<b>3</b>
<b>5. Installazione .....</b>	<b>4</b>
5.1 Operazioni preliminari .....	4
5.2 Installazione del software.....	6

# 1. Finalità del software

L'applicazione offre la possibilità di scaricare e archiviare comodamente, in maniera locale, il numero di articoli che si desidera. Ciò include articoli provenienti da file CSV e articoli ottenuti tramite richieste all'API del The Guardian. Per effettuare queste richieste, è necessario utilizzare una chiave API personale e il programma leggerà e scriverà i dati in un file JSON.

Il programma è, inoltre, capace di estrarre i 50 termini più "importanti" per ordine di peso nell'insieme degli articoli scaricati e scrivere i risultati in un file di testo.

## 2. Funzionalità

- **Download di articoli dall'API The Guardian**

Il programma consente di ottenere e deserializzare articoli provenienti dall'archivio del The Guardian in base a una o più parole chiave, che vengono restituiti per ordine di rilevanza, tramite l'utilizzo dell'API (Application Programming Interface) messa a disposizione dal quotidiano. L'accesso all'API necessita di una chiave API; ne viene fornita una di default nel file "[directory programma]/Sources/TheGuardianAPIKey.txt".

- **Lettura di articoli da file CSV**

Il programma supporta la lettura e deserializzazione di articoli salvati su file in formato CSV (Comma Separated Value), opportunamente predisposti e salvati in "[directory programma]/Sources/CSV".

- **Serializzazione degli articoli letti**

Il programma effettua una serializzazione degli articoli scaricati dalle diverse sorgenti in un formato comune comprendente un ID alfanumerico in formato UUID generato casualmente e titolo e testo completi dell'articolo. Gli articoli vengono poi salvati in formato JSON (JavaScript Object Notation) nel file "[directory programma]/Storage/articles.json".

- **Estrazione del peso dei termini che compaiono negli articoli**

Il programma consente di analizzare il contenuto di tutti gli articoli salvati ed estrarre il peso (numero totale di istanze) di ogni termine che compare al loro interno. Viene utilizzata una stop list per "filtrare" le parole di poco interesse (es. congiunzioni, avverbi, articoli, etc.).

- **Salvataggio su file dei termini più rilevanti**

Il programma effettua il salvataggio nel file "[directory programma]/Storage/terms.txt" dei 50 termini con peso maggiore, ordinati alfabeticamente in caso di pesi uguali.

- **Esecuzione da riga di comando**

Il programma è eseguibile tramite terminale a riga di comando (CLI) e supporta diverse opzioni di esecuzione:

- -d "[query]": download e serializzazione degli articoli contenuti nei file CSV e ottenuti dall'API The Guardian utilizzando la query indicata come termine/i di ricerca;
- -e: estrazione e salvataggio su file dei 50 termini con peso maggiore;
- -h: visualizzazione di un menù di aiuto contenente descrizioni dei comandi disponibili.

I comandi -d e -e possono essere effettuati entrambi all'interno di una singola esecuzione, ma il comando -e non è utilizzabile se nessuna serializzazione degli articoli è stata effettuata o se il file contenente gli articoli è mancante o vuoto.

### 3. Librerie utilizzate

- **Apache Commons CLI (1.5.0)**

Libreria per il parsing degli argomenti da riga di comando.

- **Gson (2.10)**

Libreria per la serializzazione e deserializzazione di dati in (e da) file JSON.

- **Apache Commons CSV (1.10.0)**

Libreria per la lettura di dati da file CSV.

- **Unirest (1.4.9), Jackson Core (2.12.7.1)**

Librerie utilizzate dal client GuardianSourceAPI per gestire le richieste HTTP e le risposte in formato JSON.

### 4. Requisiti di sistema

- **Dipendenze software**

Dev'essere installato il pacchetto Java SE 8 oppure Java JDK 8 sul dispositivo utilizzatore.

- **Memoria**

Il sistema richiede una quantità di memoria minima per gestire il caricamento e l'elaborazione dei file degli articoli e dei termini estratti.

- **Spazio di archiviazione**

Il sistema richiede uno spazio di archiviazione minimo per memorizzare i file degli articoli scaricati, i file dei termini estratti e altri dati correlati.

- **Risorse di rete**

Il sistema richiede una connessione ad Internet attiva per il download degli articoli tramite l'API The Guardian e per eventuali altre fonti di articoli online.

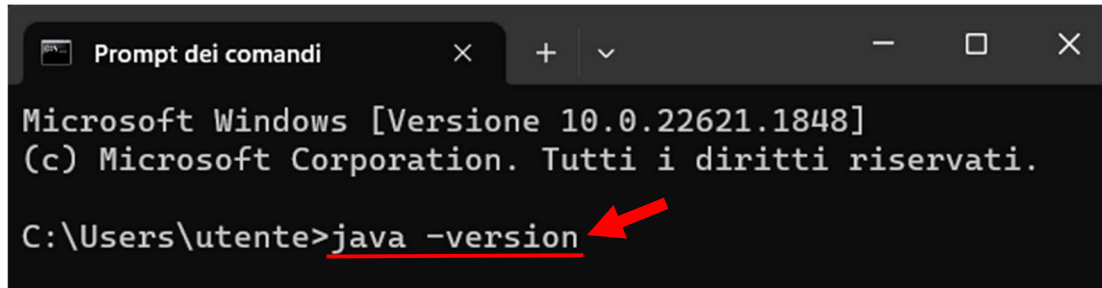
## 5. Installazione

### 5.1 Operazioni preliminari

- **Verifica della presenza di Java nel proprio dispositivo**

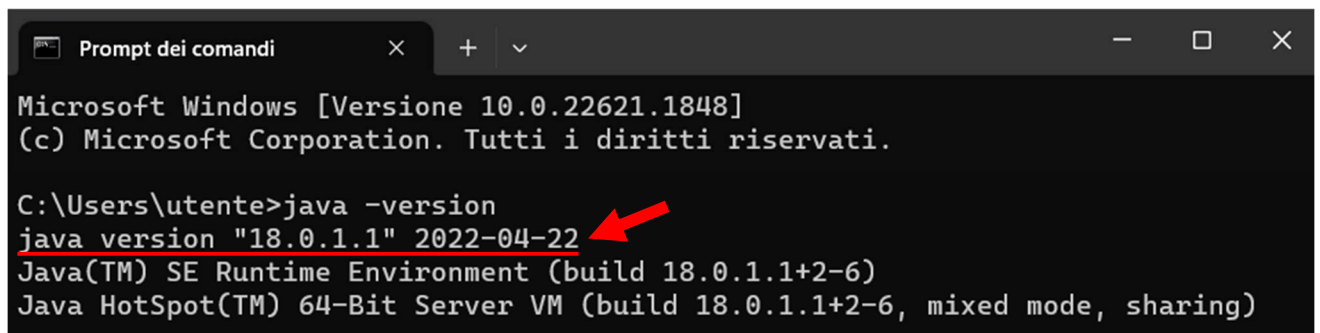
Prima di procedere all'installazione di Java, è bene verificarne la presenza nel proprio dispositivo tramite pochi e brevi passaggi:

1. Aprire il terminale del proprio sistema operativo e digitare "java -version".



```
Prompt dei comandi
Microsoft Windows [Versione 10.0.22621.1848]
(c) Microsoft Corporation. Tutti i diritti riservati.
C:\Users\utente>java -version
```

2. Se il dispositivo dispone già di Java, verrà visualizzata una schermata nella quale vengono riportate la versione installata e l'ambiente di runtime (Java SE Runtime Environment). L'importante è che la versione sia pari o superiore alla "1.8.x\_xxx".

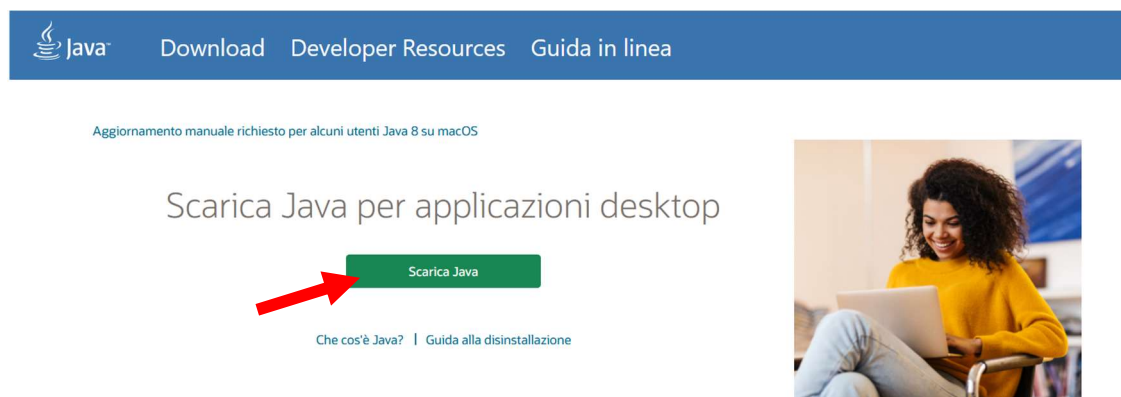


```
Prompt dei comandi
Microsoft Windows [Versione 10.0.22621.1848]
(c) Microsoft Corporation. Tutti i diritti riservati.
C:\Users\utente>java -version
java version "18.0.1.1" 2022-04-22
Java(TM) SE Runtime Environment (build 18.0.1.1+2-6)
Java HotSpot(TM) 64-Bit Server VM (build 18.0.1.1+2-6, mixed mode, sharing)
```

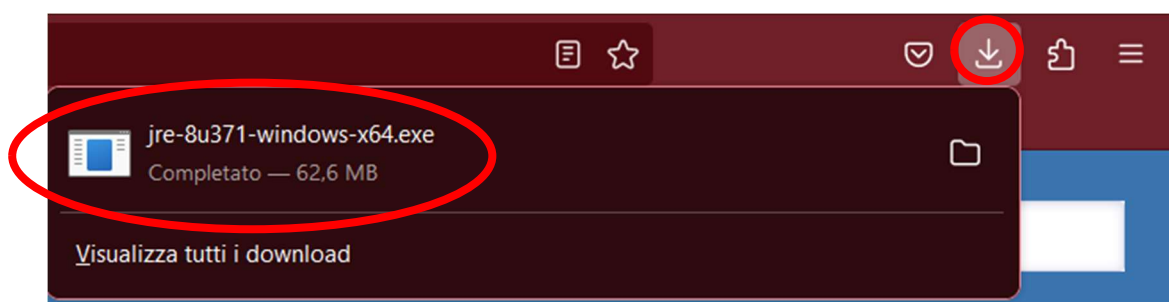
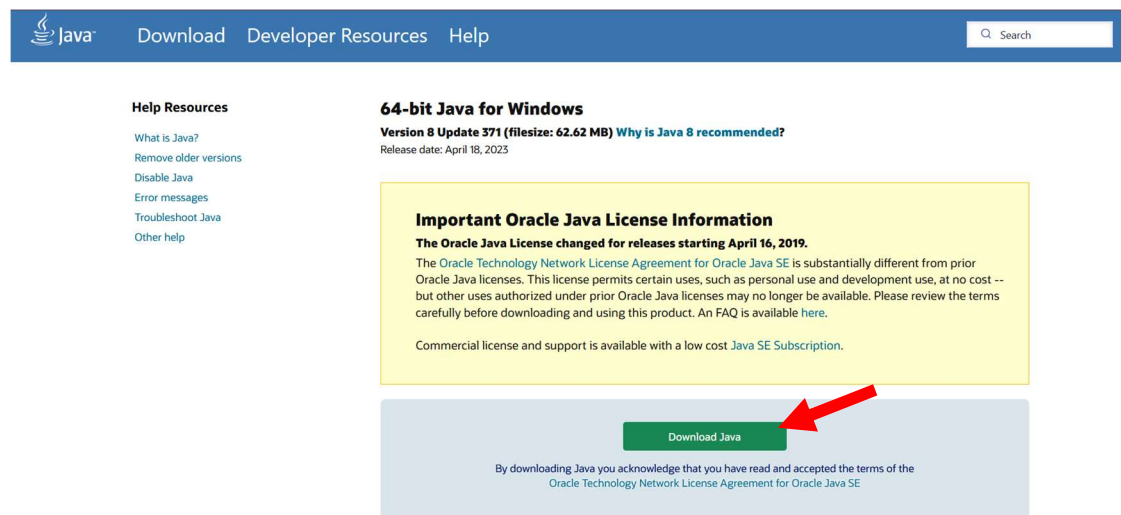
3. In caso contrario, è possibile passare all'installazione di Java tramite il sito dedicato.

- **Installazione di Java**

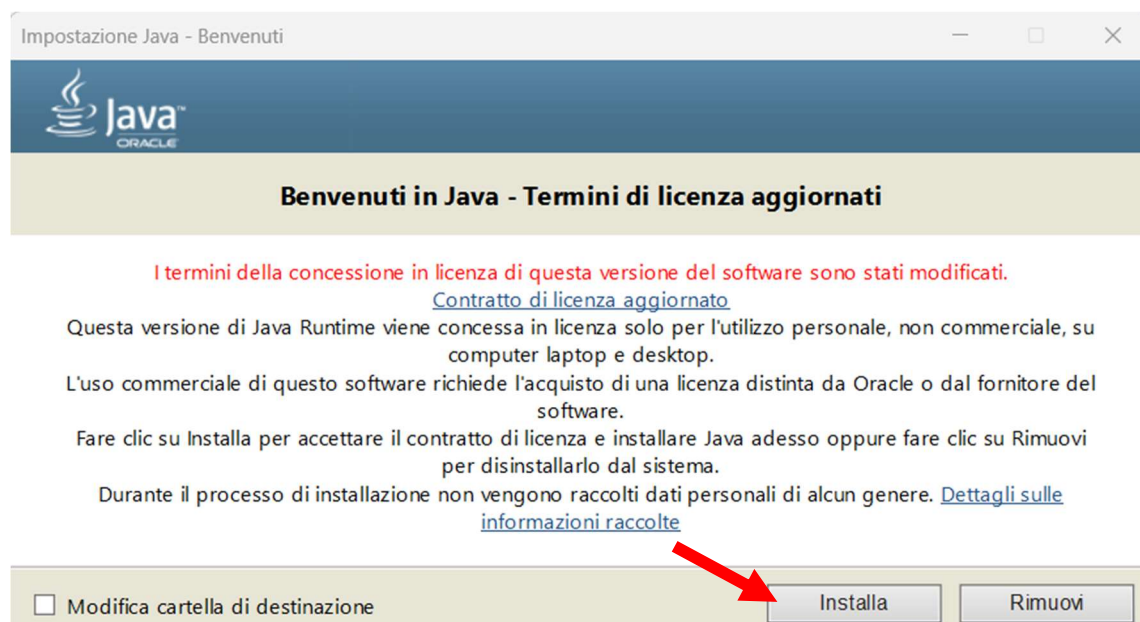
1. Aprire un qualsiasi browser e digitare nella barra degli indirizzi "<https://www.java.com/it/>", oppure cliccare il link qui presente se la lettura del manuale avviene dal dispositivo in cui si vuole installare Java.
2. Cliccare su "Scarica Java".



3. Il sito determinerà in maniera autonoma il vostro sistema operativo: cliccare quindi su "Download Java" e attendere il completamento del download.



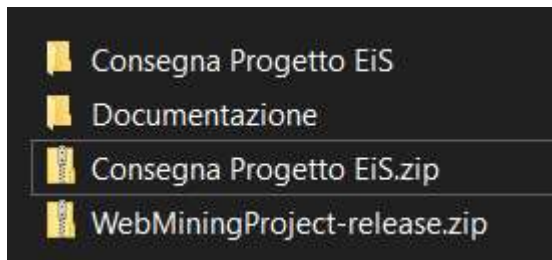
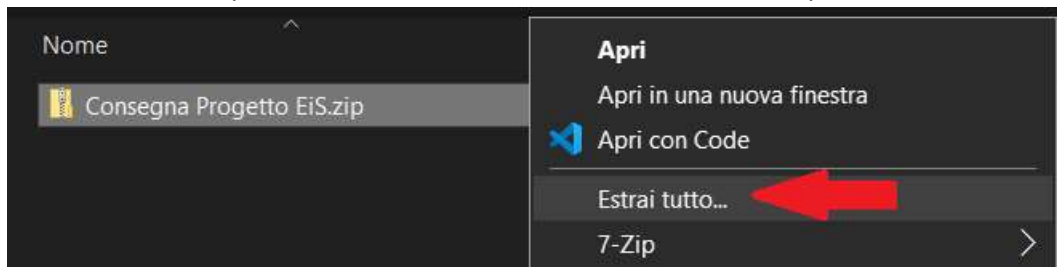
4. Una volta terminato il download, cliccare sul file di setup appena scaricato, eseguirlo e seguire le istruzioni di installazione fornite dal software.



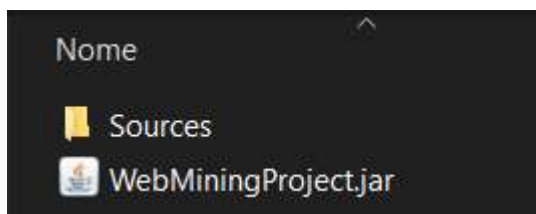
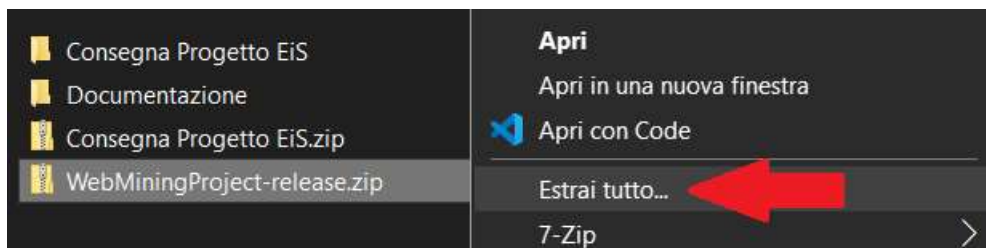
5. Terminata l'installazione, premere il tasto "Chiudi".
6. Verificare la corretta installazione di Java nel vostro dispositivo come spiegato nel punto "Verifica della presenza di Java nel proprio dispositivo".

## 5.2 Installazione del software

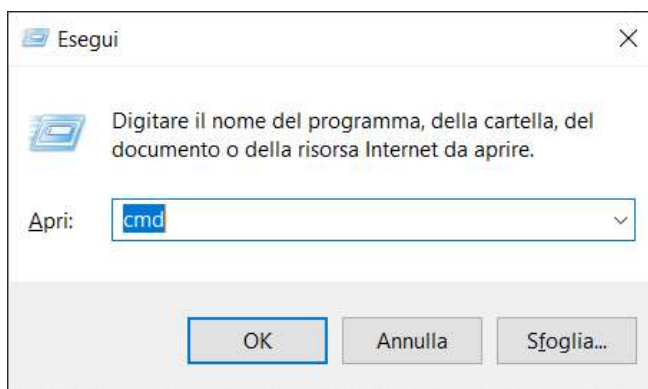
1. Scaricare l'archivio ZIP nel quale si trova il programma.
2. Estrarre tutti i file dell'archivio nella directory che si preferisce con l'utilizzo di un qualsiasi software di estrazione file (i.e. comando di estrazione Windows, WinRar).



3. Estrarre l'archivio "WebMiningProject-release.zip" contenente il file eseguibile.



4. Aprire il prompt dei comandi premendo i tasti Win + R, digitando "cmd" e premendo "OK" (oppure cercando "cmd" nella barra di ricerca di Windows).



5. Posizionarsi nella directory dove sono stati estratti i file con il comando "cd [percorso completo cartella]".

```

C:\> Prompt dei comandi
Microsoft Windows [Versione 10.0.19045.3155]
(c) Microsoft Corporation. Tutti i diritti sono riservati.

C:\Users\[redacted]>cd C:\Users\[redacted]\Documents\WebMiningProject

```

6. Eseguire il programma tramite il comando "java -jar WebMiningProject.jar [argomenti]".

```

C:\> Prompt dei comandi
Microsoft Windows [Versione 10.0.19045.3155]
(c) Microsoft Corporation. Tutti i diritti sono riservati.

C:\Users\[redacted]>cd C:\Users\[redacted]\Documents\WebMiningProject

C:\Users\[redacted]\Documents\WebMiningProject>java -jar WebMiningProject.jar -d "nuclear power" -e

```