Compte rendu aymane hinane

Jsoup es tune librairie java qui permet d'analyser les pages html , on analysant leur code html

Le but de Jsoup est de extraire les information d'un code HTML a partir des balises

On installe jsup via le gestionnaire se package maven

```
<dependency> <groupId>org.jsoup</groupId>
<artifactId>jsoup</artifactId> <version>1.15.3</version>
</dependency>
```

On integre jsoup via la library

```
import org.jsoup.*;
```

on se connect au site web qu'on veut cibler

```
Document doc = Jsoup.connect("https://quotes.toscrape.com/").get();
```

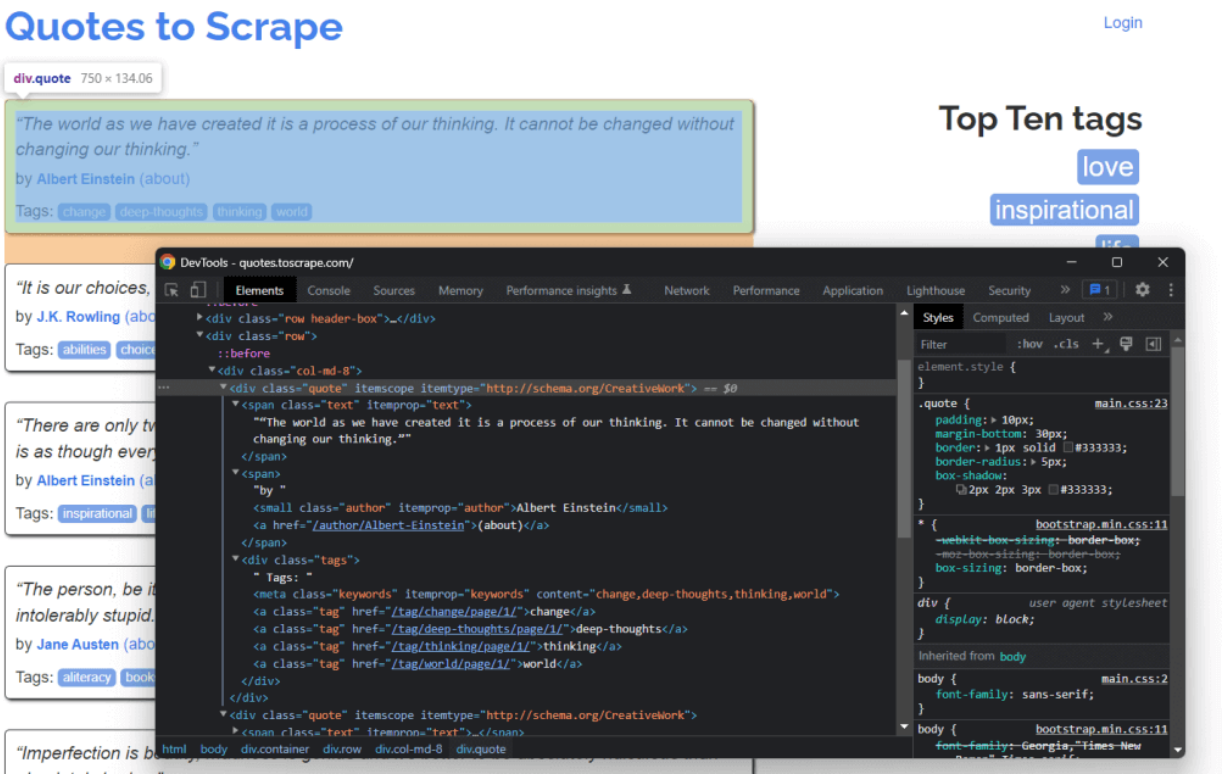
```
Document doc =
Jsoup.connect("https://quotes.toscrape.com/").get();
```

On cas d'echec de connexion une exception de type `IOException`. Serat declecher

1-premier etape

Il faut commencer par analyser le code de la page html qu'on veut cibler
On utilisant le developer tools de n'importe qu'elle navigateur

Et aussi de connaitre l'element html qui contienne la donne que on veut cibler



2-selectionner les elements avec jsoup

Selectionner toute les balises div

```
doc.getElementsByTag("div");
```

selectionner tous les elements html qui ont la class quote

```
Elements quotes = doc.getElementsByClass(".quote");
```

Pour cibler les balise enfant d'un parent

```
for (Element quoteElement: quoteElements) { Element text = quoteElement.select(".text").first(); Element author = quoteElement.select(".author").first(); Elements tags = quoteElement.select(".tag"); }
```

```
//dans cette class nous allons stocker les donner recuperer
List<Quote> quotes = new ArrayList<>();
```

selectionner tous les elements html qui ont la class quote

```
Elements quoteElements = doc.select(".quote");

// iterer sur l'element parent

(Element quoteElement : quoteElements) {

// créer l'objet quote qui servirat a stocker les donnees
Quote quote = new Quote();

 // on vat netoyer la donner stocker dans la balise text on
 // eliminent les quotes

String text = quoteElement.select(".text").first().text()
      .replace("“", "")
      .replace("”", "");


String author = quoteElement.select(".author").first().text();

List<String> tags = new ArrayList<>();

for (Element tag : quoteElement.select(".tag")) {
tags.add(tag.text());
}

// storing the scraped data in the Quote object
quote.setText(text);
quote.setAuthor(author);
quote.setTags(String.join(", ", tags));
quotes.add(quote);

}
```