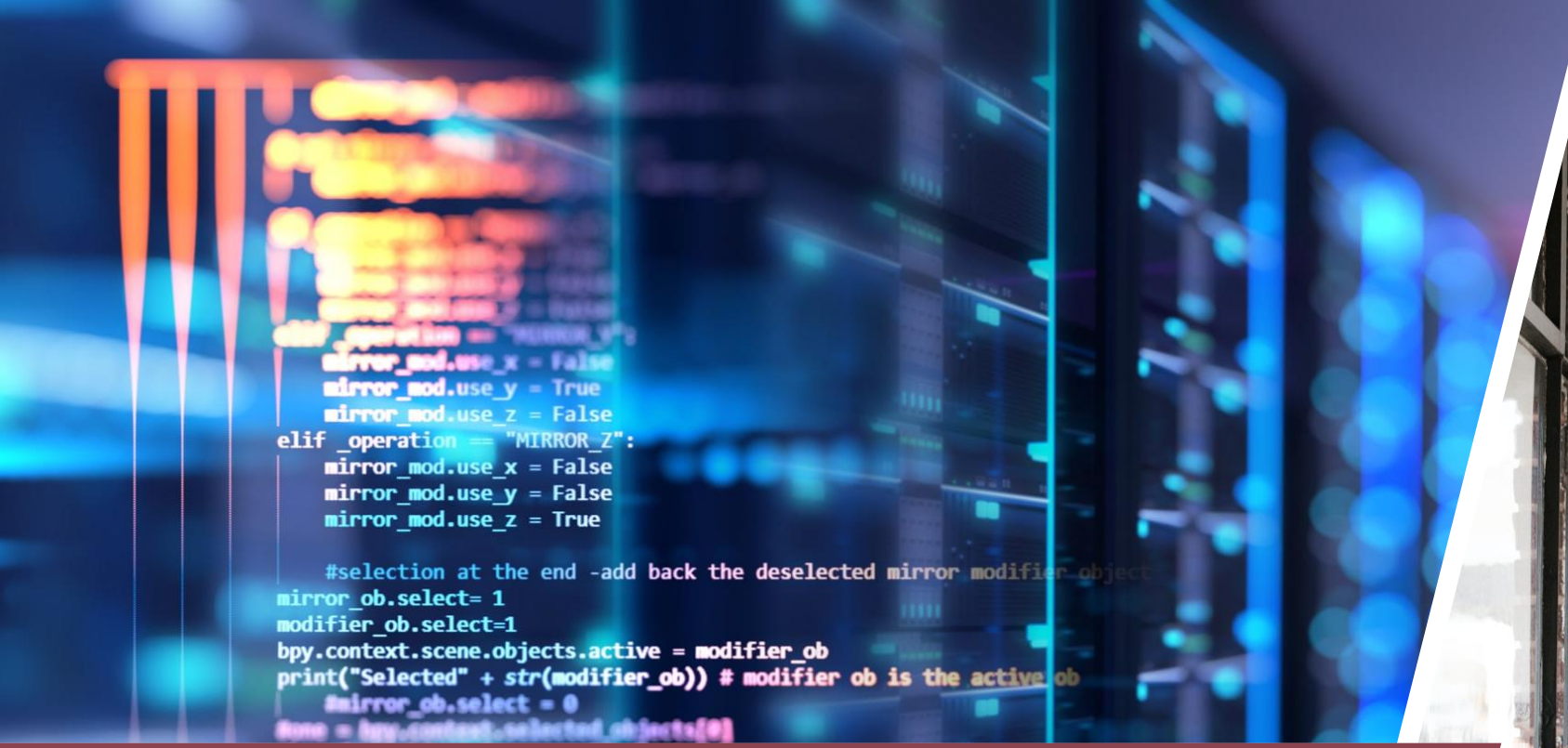
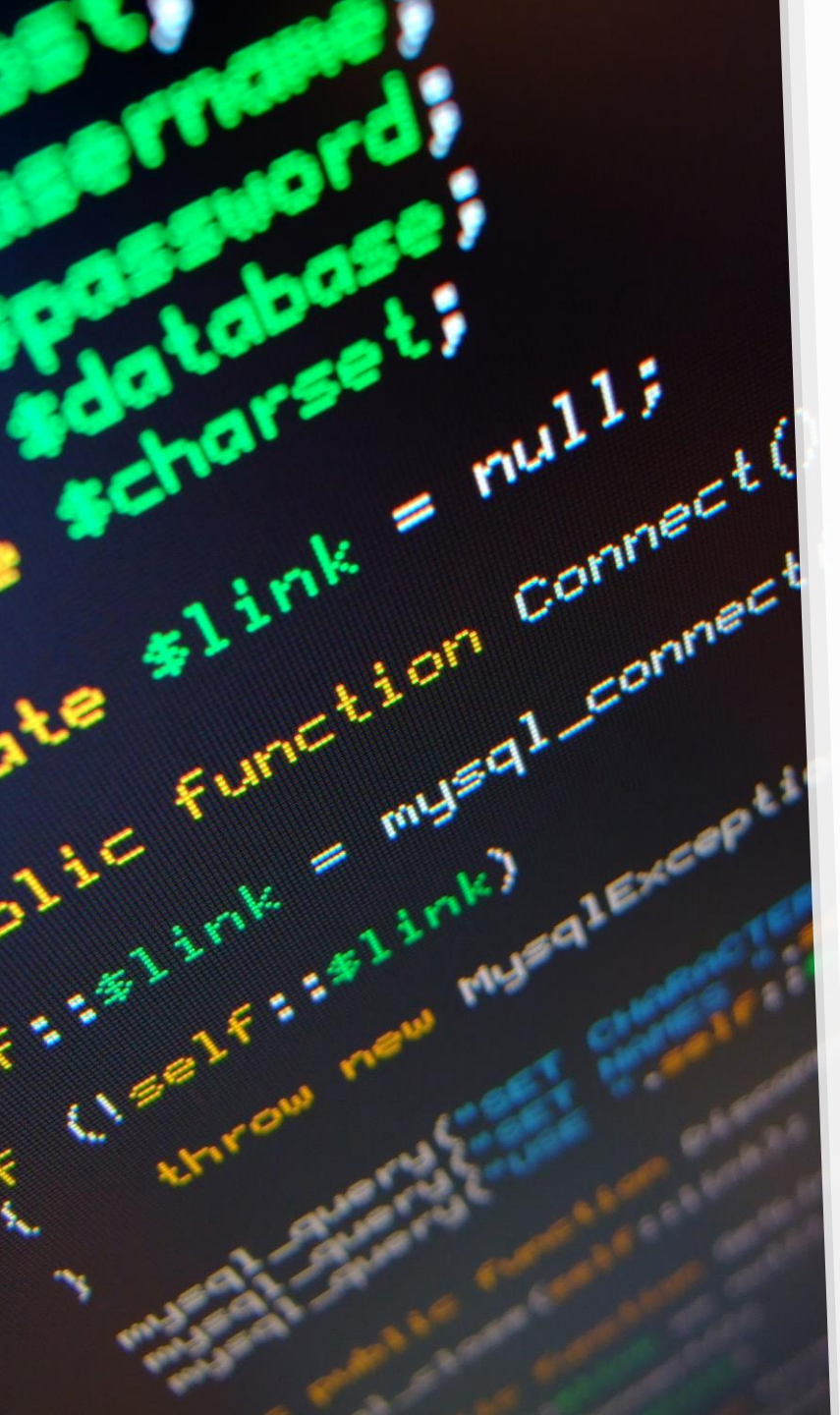



# LUCENE

PRÁCTICA 1
Gestión de Información en la Web

Álvaro de la Flor Bonilla

ETSIIT
Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación



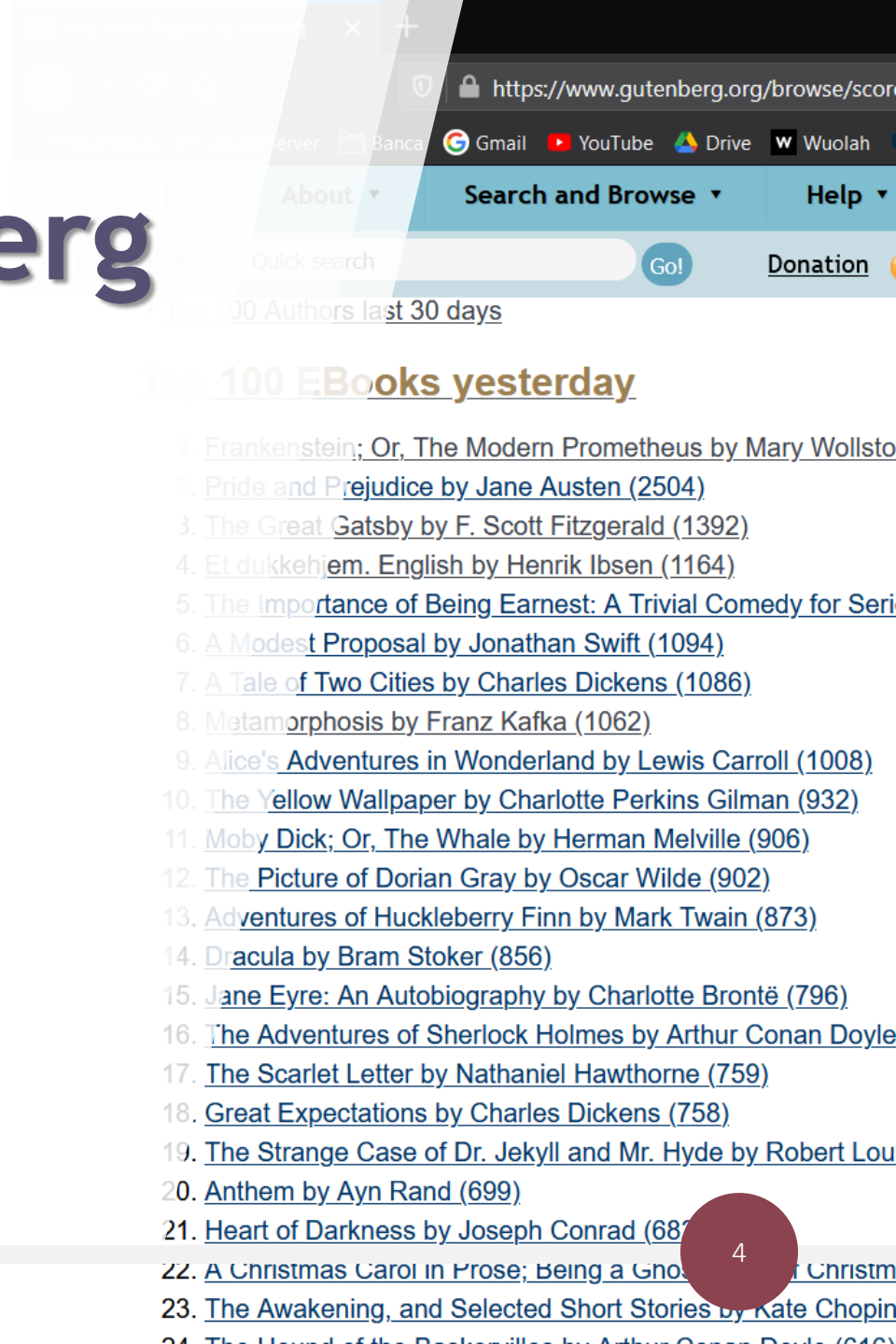

# CONFIGURACIÓN

LENGUAJES

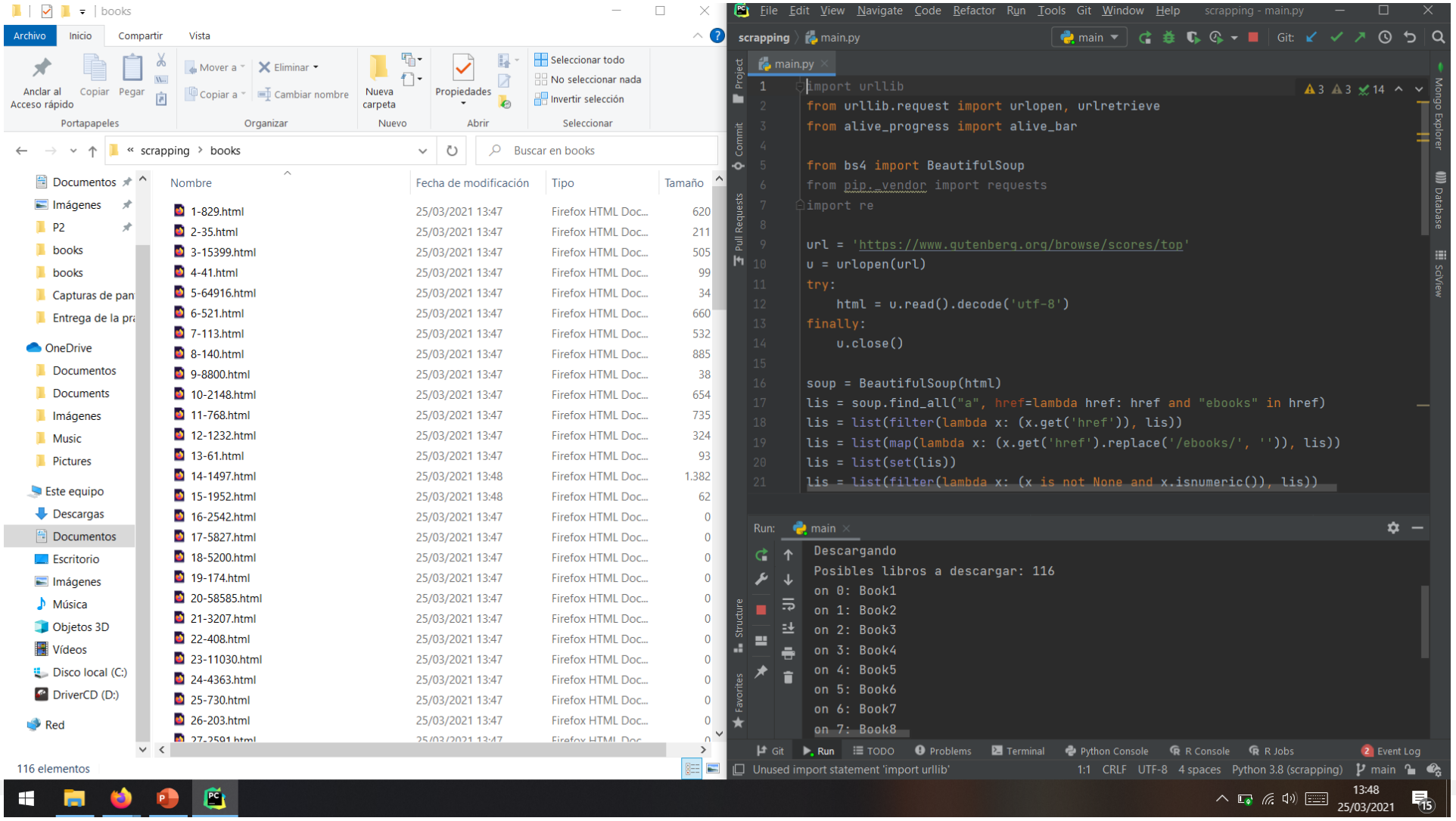
BIBLIOTECA

# Project Gutenberg

## *SCRAPING*



# TOP 100 LIBROS



# Desarrollo

## *INDEX, SEARCH*

# Index

```
public Indexer(String indexDirectoryPath) throws IOException {
    // Add stop words
    List<String> words = Execute.readWords();
    CharArraySet stopSet = StopFilter.makeStopSet(words);
    // Builds an analyzer with stop words
    StandardAnalyzer analyzer = new StandardAnalyzer(stopSet);
    // A Directory provides an abstraction layer for storing a list of files. A directory contains only files (no sub-folder hierarchy)
    Directory indexDirectory =  FSDirectory.open(Paths.get(indexDirectoryPath));
    // Holds all the configuration that is used to create an IndexWriter
    IndexWriterConfig iwriter = new IndexWriterConfig(analyzer);
    // An IndexWriter creates and maintains an index
    writer = new IndexWriter(indexDirectory, iwriter);
}
```

# Search

```
// Using umbral
if (Constants.useUmbral) {
    scoreDocs = scoreDocs.stream().filter(x -> calculateIndividualScore(x.score, searchQuery, maxScore, minScore)).collect(Collectors.toList());
}
```

```
public static Boolean calculateIndividualScore(Float score, String searchQuery, Float maxScore, Float minScore) {
    Boolean res = false;
    Integer words = Arrays.stream(searchQuery.split( regex: " ")).filter(x -> !x.equals(" ")).collect(Collectors.toList()).size();

    Float scoreNormalize = ((score + words * 0.47f) * 10) / maxScore;

    if ((1 - normalize(scoreNormalize, minScore)) < Constants.umbral) {
        res = true;
    }

    return res;
}

public static Float normalize(Float val, Float min) { return (val - min) / (10 - min); }
```

# Desarrollo

*DEMO*

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help p1 - a9496001.txt

p1_code > src > index > docs1 > a9496001.txt

Main Git:

Execute.java | a9496002.txt | a9496001.txt | ExecuteSearch.java

Run: Main

```
"C:\Program Files\Java\jdk-14.0.1\bin\java.exe" ...
En primer lugar, que desea hacer:
   (1)  Realizar búsqueda
   (2)  Volver a indexar
   (3)  Configurar el sistema
   (0)  Salir
Elija entre [0, 1, 2, 3]
```

9: Git | 4: Run | 6: Problems | Build | TODO | Terminal | Hadoop | Spark monitoring | Zeppelin | Event Log

Build completed successfully in 1 s 83 ms (moments ago)

LF UTF-8 4 spaces main

14:05 25/03/2021


# ¡GRACIAS!

## ¿ALGUNA DUDA?