Recuperación de Información

Práctica 3. Implementación de un Sistema de Recuperación de Información

Santiago Álvarez Valdivia

Ioannis Efthymiou

22 of october of 2023

Episode Stats

The code reads all provided csv files, and gathers some stats from them. First, it uses the CapitulosUnidos files to extract average rating and votes for all episodes. Then, it uses the Capitulos files to count the characters in each episode, and obtain the average.

```
Average rating: 7.282269
Average votes: 848.23047
Average characters: 35.5646
```

Analyzers to use by field

Capitulos (Dialogues)

episode_id	No analyzer, just using the value as is
number	No analyzer, just using the value as is
timestamp_in_ms	No analyzer, just using the value as is
raw_character_text	Simple Analyzer
raw_location_text	Keyword Analyzer
spoken_words	English Analyzer

CapitulosUnidos (Episode Summary)

episode_id	No analyzer, just using the value as is
spoken_words	English Analyzer
raw_character_text	Custom Analyzer (tokenize by commas, lowercase filter)
imdb_rating	No analyzer, just using the value as is
imdb_votes	No analyzer, just using the value as is
number_in_season	No analyzer, just using the value as is

original_air_date	No analyzer, just using the value as is
original_air_year	No analyzer, just using the value as is
season	No analyzer, just using the value as is
title	English Analyzer
us_viewers_in_millions	No analyzer, just using the value as is
views	No analyzer, just using the value as is

The custom analyzer made for raw_character_text splits the text by its commas, and transforms it to lowercase

```
Custom Analyzer demo:
Initial text: Ralph Wiggum, Miss Hoover, Chief Wiggum, Milhouse Van Houten, JANEY, Hayseed #2, Agnes
Skinner, Apu Nahasapeemapetilon, Marge Simpson, Class, Nelson Muntz, Company, Homer Simpson, Sherri
Mackleberry, Baby Bart, Lisa Simpson, Gary Chalmers, Seymour Skinner, Groundskeeper Willie, Rowdy Soldier,
Colonel, Private, Ned Flanders, Martin Prince, Flanders's Beatnik Dad, Kearney Zzyzwicz, Hayseed #1,
Leopold, Luigi, Jimbo Jones, Bart Simpson, Edna Krabappel-Flanders, Lunchlady Doris
Obtained tokens:
    ralph wiggum
    miss hoover
    chief wiggum
    milhouse van houten
    janey
    hayseed #2
    agnes skinner
    apu nahasapeemapetilon
    marge simpson
    class
    nelson muntz
    company
    homer simpson
    sherri mackleberry
    baby bart
    lisa simpson
    gary chalmers
    seymour skinner
    groundskeeper willie
    rowdy soldier
    colonel
```

Queries

Guiones (Dialogues)

Query: Halloween ghost trick treat

Description: Trying to find episodes related with Halloween, and kids dressed up as ghosts for Trick or Treat

Index: D

Relevant documents: 10

Cap 405 l 135

Cap 207 l 115

Cap 552 l 22

Cap 542 l 3

Cap 177 l 139

Cap 105 l 153

Cap 405 l 141

Cap 405 l 148

Cap 556 l 172

Cap 396 l 219

CapitulosUnidos (Episode Summary)

Query: Halloween ghost trick treat

Description: Trying to find episodes related with Halloween, and kids dressed up as ghosts for Trick or Treat

Index: C

Relevant documents: 10

Cap 405 l 1

Cap 532 l 1

Cap 207 l 1

Cap 395 l 1

Cap 42 l 1

Cap 270 l 1

Cap 290 l 1

Cap 80 l 1

Cap 333 l 1

Cap 359 l 1