Search System
Formula 1

Diogo Nunes | up201808546   Jéssica Nascimento | up201806723   João Vítor Fernandes | up201806724

Faculdade de Engenharia da Universidade do Porto | Processamento e Recuperação de Informação | Group 23

Introduction

This project consists in a search system about Formula 1. For that we found a dataset from Kaggle that has all the information since the start of the sport.

Dataset

Formula 1 World Championship (1950 - 2021)
F1 race data from 1950 to 2021

574

Vopani • updated 4 months ago (Version 13)

Data Tasks (1) Code (26) Discussion (15) Activity Metadata

Download (19 MB) New Notebook

Usability 10.0   License CC0: Public Domain   Tags sports, auto racing

Description

Context

Formula 1 (a.k.a. F1 or Formula One) is the highest class of single-seater auto racing sanctioned by the Fédération Internationale de l'Automobile (FIA) and owned by the Formula One Group. The FIA Formula One World Championship has been one of the premier forms of racing around the world since its inaugural season in 1950. The word "formula" in the name refers to the set of rules to which all participants' cars must conform. A Formula One season consists of a series of races, known as Grands Prix, which take place worldwide on purpose-built circuits and on public roads.

Content

The dataset consists of all information on the Formula 1 races, drivers, constructors, qualifying, circuits, lap times, pit stops, championships from 1950 till the latest 2021 season.

Acknowledgements

The data is compiled from http://ergast.com/mrd/

Inspiration

"Races are won at the track. Championships are won at the factory." - Mercedes (2019)

Search Scenarios

🔍 Michael Schumacher	(Search on Drivers)

- Personal data information
- Which constructor team he raced / is racing
- Important results
- Crucial moments in which he was included (incidents, overtakes, …)
- Career description

Search Scenarios

🔍 Ferrari

(Search on Constructors)

- History of the team in Formula 1
- Memorable results
- Most successful drivers for the team
- Actual drivers on the team

Search Scenarios

🔍 2020   (Search on Seasons - year)

- Summary of the full season
- Crucial highlights
- Champions (driver and constructors)

Search Scenarios

🔍 Monaco	(Search on Circuits)

- Location
- How many times that city/country hosted a race
- Most successful drivers on that track/country
- Track description (how many turns, DRS zones, …)

Search Scenarios

(Search on specific moments)

🔍 Verstappen overtook Hamilton

🔍 Norris crash

🔍 Leclerc pole position

🔍 Soft tyres

🔍 Who started on grid 5 on 2020 Portugal GP?
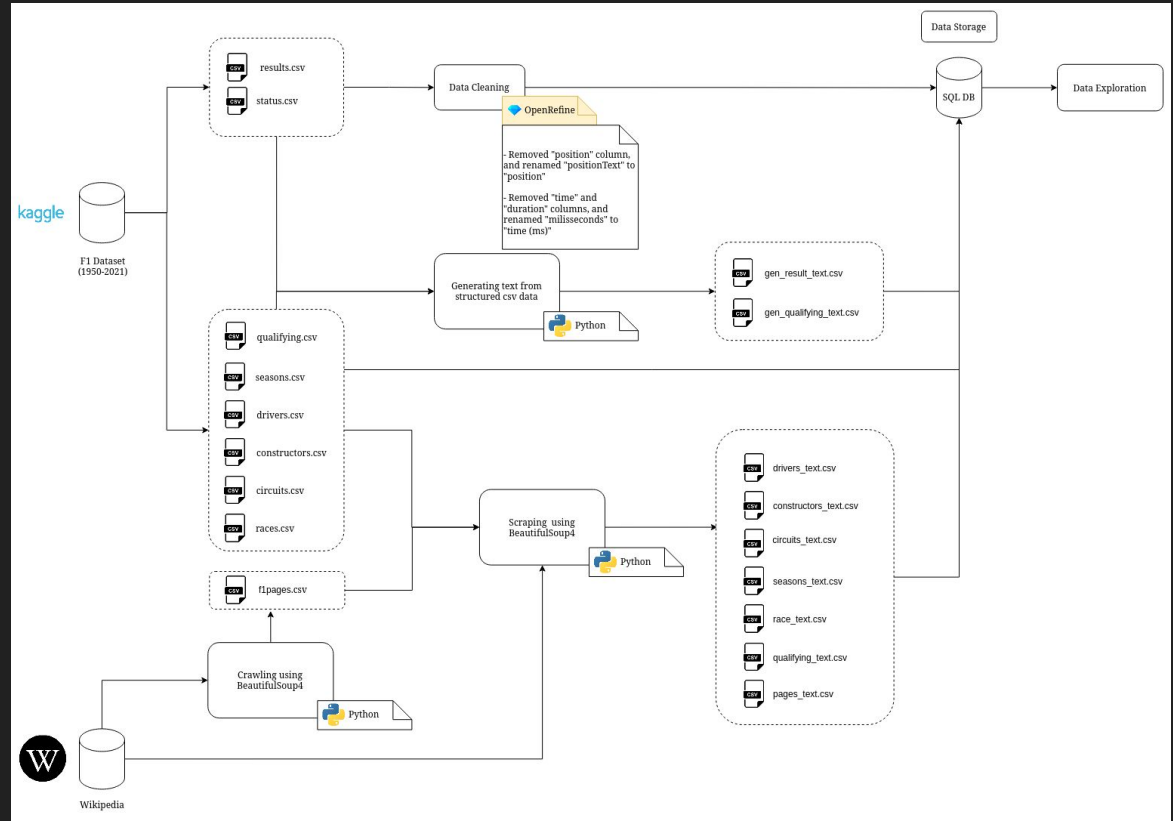
🔍 2018 Belgium Qualifying

- List of events, explanations, answer search, ...

Data Collection

- Datasets
  - Formula 1 World Championship (1950 - 2021)
  - Seasons, Drivers, Constructors, Circuits, Races, Qualifying, Results

- Wikipedia
  - Many CSV files have Wikipedia URL links to complement information given
  - We used these URLs to generate CSV files with text, so we can enrich our amount of data
  - 1 additional CSV added with text of F1 related pages: regulations, awards, licenses, ...

Data Preparation

1. Cleaning data on qualifying, results and status csvs with OpenRefine

2. Generating plain text from dataset's csvs

3. Crawling wikipedia pages using BeautifulSoup

4. Scraping 7 files using BeautifulSoup, creating new files with paragraphs from Wikipedia

Thank you!

Group 23

Diogo Nunes | up201808546   Jéssica Nascimento | up201806723   João Vítor Fernandes | up201806724