

Fakulta Informatiky a Informačných Technológií
Slovenská Technická Univerzita v Bratislave

Marcel Valent

Údaje o slovenskej futbalovej lige, hráčoch

Vyhľadávanie informácií

Študijný program: Inteligentné Softvérové Systémy

Cvičiaci: Ing. Igor Stupavský

December 2023

Téma

Údaje o slovenskej futbalovej lige, hráčoch

Stránka

Nikeliga.sk

GitHub

<https://github.com/MarcelValent/VINF>

Zámer projektu

V tomto projekte mám za cieľ získať údaje o zápasoch a hráčoch zo slovenskej Nike ligy. Budem extrahovať informácie o góloch, žltých a červených kartách.

Na začiatku budem crawlovať údaje z HTML stránok slovenskej Nike ligy a systematicky ich uložiť. Ďalej plánujem spracovať a doplniť naše údaje o štatistikách gólov, žltých a červených kariet z relevantných zápasov.

Pre demografické informácie týkajúce sa národností hráčov v slovenskej Nike lige využijem distribuované spracovanie v Apache Spark a príslušné demografické údaje z Wikipédie.

Po spracovaní údajov vytvorím index pomocou Lucene a implementujem vyhľadávanie, aby som mohol získať štatistiky o hráčoch na základe zadaných dopytov o góly, žlté a červené karty.

Na záver plánujem vytvoriť aj jednotkové testy na overenie správnosti vyhľadávacích dotazov a získaných štatistík o hráčoch v slovenskej Nike lige.

Podobné riešenia

Podobné riešenia ako je toto moje som nedokázal na internete nájsť. Pravdepodobne je to z dôvodu, že podobné riešenia majú kluby interne, ale nie sú dostupné verejnosti. Je však taktiež možné, že takéto riešenia vôbec neexistujú aj vzhľadom na malú veľkosť, známosť a atraktivnosť našej slovenskej futbalovej ligy v doma či v zahraničí.

Inštalácia a spustenie projektu

Inštalácia a spustenie projektu je jednoduchá. Stačí si tento projekt stiahnuť do zložky, kde sa dá následne spúšťať jednotlivé funkcie. Na spustenie indexer.py a search.py je potrebné tieto funkcie spúšťať cez docker.

Pseudokód

Na wiki stránke.

Dáta

Ukážka dát bude dostupná na Github/wiki.

Miera úspešnosti po spracovaní menšieho wiki dumpu:

Po spracovaní menšieho wiki dumpu (cca 800MB) som obohatil cca 16.5% dát. Na väčšom dume som netestoval z dôvodu nedostatočnej výkonnosti internetu doma.

- **Blank percentage: 83.46%**

Výsledky unit testov:

```
Name= Milan Ristovski
```

```
goals=28
```

```
Output: goals=28
```

```
Expected output: goals=28
```

```
Result: Successful
```

```
Name= Aleksandar Čavrič
```

```
goals=45
```

```
Output: goals=45
```

```
Expected output: goals=45
```

```
Result: Successful
```

```
Name= Matúš Marcin
```

```
yellow_cards=12
```

```
Output: yellow_cards=12
```

```
Expected output: yellow_cards=12
```

```
Result: Successful
```

```
Name= Boris Godál
```

```
yellow_cards=39
```

```
Output: yellow_cards=39
```

```
Expected output: yellow_cards=39
```

```
Result: Successful
```

```
Name= Boris Godál
```

```
red_cards=3
```

```
Output: red_cards=3
```

```
Expected output: red_cards=3
```

```
Result: Successful
```

Ako môžeme vidieť, všetky naše unit testy boli úspešné, keďže vrátili nám očakávaný výstup.

Konzultácie

Na wiki stránke.