

# Task-scrapper

Autor: Boris Španić

# Uvod

## Ideja

U programiranju postoji alat koji se zove evaluator. Taj alat provjerava točnost rješenja poslanog zadatka. Kako ne postoji evaluator za sve HONI i Infokup zadatke, napravljen je CTK evaluator. Kada se zadatak stavlja na evaluator najbitniji dio, pored samih testnih primjera, je tekst zadatka. Kako raznovrsni autori zadataka imaju kreativnu slobodnu, struktura teksta je različita svake godine. Zato sam napravio task-scrapper, konzolnu aplikaciju u pythonu koja parsira zadatke s domaćih natjecanja.

## O autoru

Autor ovog projekta je moja malenkost, Boris Španić, učenik 2. razreda Gimnazije Andrije Mohorovičića Rijeka. Boris Španić (u daljnjem tekstu *ja*) je vrsni poznavatelj C++-a i natjecatelj na spomenutim domaćim natjecanjima. No za potrebe projekta ja sam odlučio kucati u python-u.

# Aplikacija

Projekt je konzolna aplikacija. Kako bi ste mogli koristiti aplikaciju morate prvo instalirati sljedeće python datoteke:

- pip install pdfminer3
- pip install tkinter

Zatim, u direktoriju s kodom, u konzolu ukucate python task-scrapper.py.

Zatim morate izabrati želite li otvoriti .pdf ili .txt datoteku (ovdje morate upisati slovo p za pdf ili slovo t za txt). Zatim će vam se otvoriti nativni file browser u kojem trebate izabrati datoteku koju želite parsirati. Nakon toga slijediti će niz pitanja ovisno o izabranoj vrsti ekstenzije datoteke.

```
C:\Windows\system32\cmd.exe

C:\Users\Boris\Documents\task-scrapper> python task-scrapper.py
Parsiranje iz .pdf ili .txt filea? [p/t]
t
HONI ili Infokup? [h/i]
h
Koja godina? [Ako je HONI i npr. sezona 2019/2020 stavi "2020"]
2013
Ime zadatka? [Onako kako je napisano npr. "MANIPULATOR" i "Skandi"]
Hiperacijevi
```

*pitanja za .txt ekstenziju*

```
C:\Windows\system32\cmd.exe

C:\Users\Boris\Documents\task-scrapper> python task-scrapper.py
Parsiranje iz .pdf ili .txt filea? [p/t]
p
HONI ili Infokup? [h/i]
h
Koja godina? [Ako je HONI i npr. sezona 2019/2020 stavi "2020"]
2020
Ime zadatka? [Onako kako je napisano npr. "MANIPULATOR" i "Skandi"]
Ulica
Od koje do koje stranice? [npr. "6 9" i "1 1"]
3 3
```

*pitanja za .pdf ekstenziju*

Na kraju aplikacija otvori vaš nativni alat za tekst s parsiranim zadatkom.

## Korištenje

Za svaku godinu možete parsirati .txt file. Problem se javlja kod parsiranja .pdf fileova. Kao što sam naveo prije, kreativna sloboda autora i pisanje u čudnim alatima otežava parsiranje pdf-ova te pdfminer koji sam koristio u ovom projektu ne radi na starim HONI-jima i nekim infokupima, pa preporučam korištenje pdf parsiranja samo za 2020., <sup>1</sup>2017. i 2016. godinu (zato jer su zadaci pisani u LaTeX-u).

---

<sup>1</sup> 2016. i 2017. samo za infokup

# Način rada aplikacije

Sam kod koji se koristi za izvršavanje parsiranja je *task-scrapper.py*, no postoji i kod *parsing\_parameters.py*. Taj kod je zapravo riječnik koji sadrži parametre za parsiranje teksta.

Samo parsiranje izvodi se na način da se učitaju .pdf ili .txt datoteka te prebaci u listu. Nakon toga, ovisno o natjecanju, iz *parsing\_parameters.py* kod se parsira pomoću provjeravanja elementa riječnika i početka linije. Parsirani tekst na kraju bi trebao ovako izgledati:

```
Tekst zadatka
##Ulazni podaci
Opis ulaznih podataka
##Izlazni podaci
Opis izlaznih podataka
##Bodovanje
Opis bodovanja
##Probni primjeri
Probni primjeri iz zadatka
###ulaz
...

ulaz probnog primjera
...

###izlaz
...

izlaz probnog primjera
...
###Pojašnjenje trećeg probnog primjera:
pojašnjenje, samo ako zadatak sadrži pojašnjenje
```