

Semestrálny projekt z predmetu ISJ (Skriptovací jazyky) dokumentácia

Varianta: Automatické sťahovanie a srovnávanie tituliek k filmům

Projekt bol implementovaný v jazyku Python verzia 3.2.3

Pri spúšťaní skriptu program prijíma jediný parameter, URL adresu českých tituliek z databázy <http://www.opensubtitles.com>.

Príklady spustenia:

```
./projekt.py http://www.opensubtitles.org/cs/subtitles/3258272/the-mist-cs  
python projekt.py http://www.opensubtitles.org/cs/subtitles/3258272/the-mist-cs  
./projekt.py http://www.opensubtitles.org/cs/subtitles/4454648/the-shawshank-  
redemption-cs
```

Detaily implementácie:

Skript na začiatku vyparsuje číslo z URL adresy, ktoré je potrebné k stiahnutiu zadáných českých tituliek. Na samotné sťahovanie som použil modul `urllib.request`.

Nasledovne skript zistí IMDB identifikačné číslo filmu ktoré vyparsuje z xml súboru tituliek. Potom sa použije API¹ databázy opensubtitles, používam funkciu `SearchSubtitles`, ktorá prijíma parametre IMDB ID filmu a požadovaný jazyk (v našom prípade angličtina). Funkcia vracia ID čísla nájdených tituliek. Nasledne ich skript všetky stiahne a porovná podľa veľkosti súboru s českými titulkami. Podľa žiadosti užívateľa skript môže ostatné anglické titulky (okrem tých, ktoré sú najvhodnejšie) vymazať.

Skript potom extrahuje anglické titulky do adresára „en“ a české do adresára „cz“.

Pri sťahovaní titulkov môže skript naraziť na captcha overovanie, ktoré bohužiaľ nieje možné obísť.

Bohužiaľ, skript nevyhovuje naplno špecifikácii, keďže nerieši zrovnávanie stiahnutých titulkov. Nevedel som si s tým poradiť.

¹ <http://trac.opensubtitles.org/projects/opensubtitles/wiki/XMLRPC>