Python: HW02

Druhý domácí úkol je zaměřený na preprocessing dat. Dostanete TSV (Tabulator Separated Values) soubor a vaším cílem bude si jej připravit pro další zpracování. Konkrétně si vyberete jen některé sloupečky, převedete si je do použitelnější podoby a uložíte si vyčištěná data jako JSON.

Zadání

Napište skript, který přečte obsah souboru `netflix_titles.tsv` obsahující filmy a převede je do seznamu slovníků, který uloží do JSON souboru `hw02_output.json`. Z každého řádku vás budou zajímat pouze následující údaje:

- `PRIMARYTITLE` (název)
- `DIRECTOR` (seznam režisérů)
- `CAST` (seznam herců)
- `GENRES` (seznam žánrů)
- `STARTYEAR` (rok vydání)

Tyto údaje převedete do slovníku s následujícími klíči a hodnotami:

Klíč	Hodnota	Příklad
`"title"`	název filmu	`"title": "V for Vendetta"`
`"directors"`	seznam všech režisérů nebo prázdný seznam, pokud není režisér uveden	`"directors": []`
`"cast"`	seznam všech herců nebo prázdný seznam, pokud není žádný herec uveden	`"cast": ["Natalie Portman", "Hugo Weaving"]`
`"genres"`	seznam všech žánrů, do kterých byl film zařazen	`"genres": ["Action", "Drama", "Sci-Fi"]`
`"decade"`	dekáda, ve které film vznikl	`"decade": 2000`

Protože formát TSV neumožňuje reprezentovat seznam, jsou herci, režiséři a žánry zadáni jako jeden řetězec a jednotlivé hodnoty jsou oddělené čárkami. Ve formátu JSON použijte pro větší přehlednost seznam, aby bylo například vidět, kolik herců nebo režisérů v seznamu je.

Může se stát, že film neobsahuje údaj o režisérech nebo hercích, ostatní jsou vždy uvedené. Pokud není uveden žádný režisér nebo herec, daná položka musí být prázdný seznam `[]`, nikoli seznam s řetězcem o nulové délce `[""]`.

Dekáda je vždy první rok desetiletí, např. rok 1987 patří do dekády 1980 a rok 2017 do dekády 2010.

Příklad

Pro lepší pochopení je zde příklad kratšího vstupu a očekávaného výstupu.

Obsah vstupního souboru:

```
tt0164334   movie   Along Came a Spider Along Came a Spider 0   2001
104 Drama,Thriller  6.4 72942   Movie   60002273    Movie   Along Came a
Spider  Lee Tamahori    Morgan Freeman, Monica Potter, Michael Wincott,
Dylan Baker, Mika Boorem, Anton Yelchin, Kim Hawthorne, Jay O. Sanders,
Billy Burke, Michael Moriarty, Penelope Ann Miller  United States, Germany,
Canada  October 1, 2019 2001    R   103 min Thrillers   When a girl is
kidnapped from a prestigious prep school, a homicide detective takes the
case, teaming up with young security agent.
tt0120484   movie   The Waterboy    The Waterboy    0   1998        90
Comedy,Sport    6.1 143770  Movie   17687959    Movie   The Waterboy
Frank Coraci    Adam Sandler, Kathy Bates, Henry Winkler, Fairuza Balk,
Jerry Reed, Lawrence Gilliard Jr., Blake Clark, Peter Dante, Jonathan
Loughran, Al Whiting    United States   March 8, 2017   1998    PG-13   90
min Comedies, Sports Movies A water boy for a college football team has a
rage that makes him a tackling machine whose bone-crushing power might
vault his team into the playoffs.
```

Očekávaný obsah výstupního souboru:

```
[
    {
        "title": "Along Came a Spider",
        "directors": [
            "Lee Tamahori"
        ],
        "cast": [
            "Morgan Freeman",
            "Monica Potter",
            "Michael Wincott",
            "Dylan Baker",
            "Mika Boorem",
            "Anton Yelchin",
            "Kim Hawthorne",
            "Jay O. Sanders",
            "Billy Burke",
            "Michael Moriarty",
            "Penelope Ann Miller"
        ],
        "genres": [
            "Drama",
            "Thriller"
        ],
        "decade": 2000
```

```
    },
    {
        "title": "The Waterboy",
        "directors": [
            "Frank Coraci"
        ],
        "cast": [
            "Adam Sandler",
            "Kathy Bates",
            "Henry Winkler",
            "Fairuza Balk",
            "Jerry Reed",
            "Lawrence Gilliard Jr.",
            "Blake Clark",
            "Peter Dante",
            "Jonathan Loughran",
            "Al Whiting"
        ],
        "genres": [
            "Comedy",
            "Sport"
        ],
        "decade": 1990
    }
]
```

Odevzdání

Svůj skript odevzdejte jako jeden soubor pojmenovaný `prijmeni_jmeno_hw02.py`. Tento soubor vložte na nějaké sdílené úložiště (Google Drive, nebo ideálně Git), vytvořte veřejně přístupný odkaz a ten odevzdejte přes moje.czechitas.cz.

K řešení (kromě řazení klíčů ve slovníku) si vystačíte s podklady z kodim.cz. Pokud se vám nebude dařit úkol vyřešit přesně dle zadání, zkuste odevzdat alespoň to, o co jste se pokoušely.

Tipy

Viz HW01.