

MovieLens

Zadání úlohy

- vytvořte doporučovací systém pro filmy nad datasetem MovieLens pomocí metody collaborative filtering

Dataset

- <https://grouplens.org/datasets/movielens/>
- použijeme jen malou podmnožinu dat:
 - <https://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

Úkoly

- vyzkoušejte doporučování pomocí content filtering
 - ke každému filmu načtěte štítky (tags) a z nich vytvořte text, který film charakterizuje
 - text převedte pomocí Bag of Words a TF-IDF na vektor
 - porovnejte vektory pomocí kosinové vzdálenosti a na základě tohoto údaje udělejte doporučení
- vyzkoušejte vizualizovat filmy (skrz jejich hodnocení od uživatelů) metodou t-SNE

Nápověda

- vytvoření textu ze štítků

```
mixed = pd.merge(movies, tags, on='movieId', how='left')
mixed.fillna("", inplace=True)
mixed = pd.DataFrame(mixed.groupby('movieId')['tag'].apply(
    lambda x: "%s" % ' '.join(x)))
Final = pd.merge(movies, mixed, on='movieId', how='left')
Final['metadata'] = Final[['tag', 'genres']].apply(
    lambda x: ' '.join(x), axis = 1)
Final[['movieId', 'title', 'metadata']].head(3)
```