temat: Wczytanie bazy IMDb do systemu SQL Server
autor: Artur Gramacki

Próba wczytania do MongoDB na prawdę **wielkiego zbioru danych** o dość złożonej strukturze [1]. Dane pobrano z serwisu filmowego *IMDb*, https://www.imdb.com/interfaces. Serwis IMDb udostępnia bazy danych w formie skompresowanych plików tekstowych, kodowanych za pomocą UTF-8. Każdy plik tekstowy w pierwszym wierszu zawiera nagłówek, a poszczególne kolumny rozdzielone zostały tabulatorem. Puste komórki danych zostały oznaczone znakami „\N". Dane przechowywane są w postaci znormalizowanej (bez redundancji danych). Serwis udostępnia siedem baz danych (https://datasets.imdbws.com). Dane są uaktualnianie codziennie.

title.akas.tsv.gz – informacje o tytułach i ich tłumaczeniach
Przykład:

```
titleId ordering title region language
types attributes isOriginalTitle
tt0253474 32 Pianista PL \N \N \N 0
```

title.basics.tsv.gz – informacje ogólne o filmie
Przykład:

```
tconst titleType primaryTitle
originalTitle isAdult startYear endYear
runtimeMinutes genres
tt0253474 movie The Pianist The Pianist 0
2002 \N 150 Biography,Drama,Music
```

title.crew.tsv.gz – informacje o reżyserze i scenarzyście
Przykład:

```
tconst directors writers
tt0253474 nm0000591 nm0367838,nm0844262
```

title.episode.tsv.gz – informacje o serialu (sezon, odcinek)
Przykład:

```
tconst parentTconst seasonNumber
episodeNumber
tt0066293 tt0058853 1 161
```

title.principals.tsv.gz – informacje o obsadzie
Przykład:

```
const ordering nconst category job
characters
tt0253474 1 nm0004778 actor \N ["Wladyslaw
Szpilman"]
```

[1] Ten punkt opracowano na podstawie pracy dyplomowej pana Piotra Busia. Tytuł pracy: *Porównanie efektywności przetwarzania danych przechowywanych w formacie JSON w relacyjnych i nierelacyjnych bazach danych*, data obrony: 2019-02-25, promotor: dr hab. inż. Artur Gramacki.

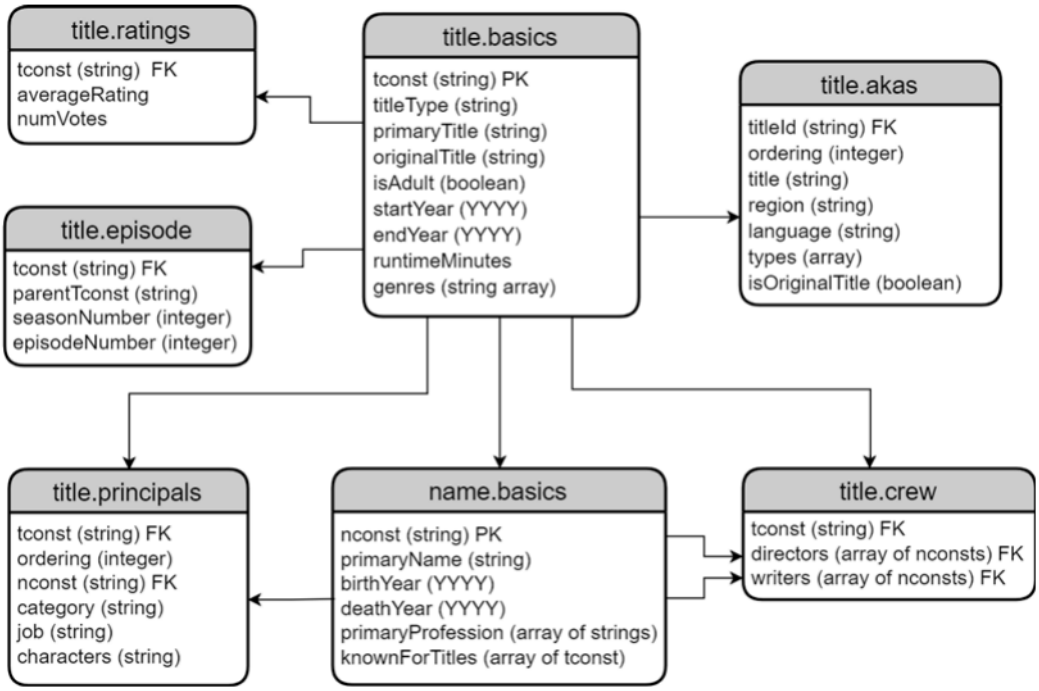
title.ratings.tsv.gz – informacje o ocenach filmu
Przykład:

tconst averageRating numVotes
tt0253474 8.5 624249

name.basics.tsv.gz – informacje na temat osób
Przykład:

nconst primaryName birthYear deathYear primaryProfession knownForTitles
nm0000591 Roman Polanski 1933 \N actor,director,writer tt0074811,tt1692486,tt0071315,tt1139328

Na poniższym rysunku pokazano model danych serwisu IMDb.



Po wczytaniu do SQL Servera danych należy zaproponować około 10 sensownych zapytań SQL. Szczegóły należy ustalić z prowadzącym zajęcia.