

KSA Ministry of Justice NLP

Abstract:

The goal of this project is to be able to reach the nearest five cases similar to the same entered case In terms of the type of trial and judgment

Design:

Data:

<https://sjp.moj.gov.sa/Filter?isFilterButtonClicked=True>

The data was extracted from the Saudi Ministry of Justice website via(web scraping)
It consists of more than 2000 rows and 13 columns.

Algorithms:

The data was extracted from the site

Data cleaning

Preprocessing (stop word, punctuation, spelling correction, delete numbers, English word)

Steamer

TF-IDF

Cosine similarity

Tools:

Python , NumPy, Pandas,Matplotlib,seaborn

Communication:

```
cosine_similarity(dt_tfidf.iloc[0:16, :],dt_tfidf)
array([[1.          , 0.04527673, 0.05592527, ..., 0.0265889 , 0.05043805,
        0.06672821],
       [0.04527673, 1.          , 0.04428619, ..., 0.03761851, 0.07428042,
        0.07205619],
       [0.05592527, 0.04428619, 1.          , ..., 0.0346366 , 0.05308088,
        0.08630334],
       ...,
       [0.0522376 , 0.05495503, 0.07531514, ..., 0.03760567, 0.08589664,
        0.11185845],
       [0.06400227, 0.07055812, 0.05787263, ..., 0.0509626 , 0.09750601,
        0.11260752],
       [0.05904896, 0.05485579, 0.07102775, ..., 0.04965179, 0.09426072,
        0.06766407]])
```