CS410 Team Project Proposal

Team Info

Name	NetID
Xiao Luo (captain)	xiaoluo6
Shi Zhang	shiz4
Wenyan Zhang	wenyanz4
Arindam Chakraborty	arindam4

Topic Overview

Topic

Friends scripts sentiment analysis

Description

We aim to explore the sentiments expressed in the Friends TV show scripts. Our focus involves processing the screenplay script text data and conducting an in-depth analysis using text retrieval and sentiment analysis.

Task

We've divided the project into the following sub-tasks, assigned to different team members:
1. Data pre-processing: Clean and organize the raw screenplay script text data.
2. Text retrieval: Retrieve a ranked list of lines based on relevance for a given topic of interest.
3. Sentiment analysis: Analyze sentiments based on the retrieved lines.
4. Data visualization: Develop a Tableau dashboard to consolidate and visualize sentiments related to selected topics.

Why are we interested in the topic

As avid Friends TV series fans, we're eager to analyze the numerous dialogues from a sentiment perspective. Additionally, the recent loss of Matthew Perry, who portrayed Chandler, motivates us to revisit the show's amusing and heartwarming moments.

Planned approach

Our approach involves downloading relevant script text data for all seasons and episodes, utilizing packages such as NLTK for text processing, Gensim for text retrieval, and Hugging Face for sentiment analysis. We'll complement this with the development of a Tableau dashboard to showcase our insights.

Tools and datasets

Tools & systems:
- Pytorch: deep learning training
- NLTK/spaCy: text processing
- Gensim: topic modeling, document indexing, retrieval by similarity, etc.
- Hugging Face: sentiment analysis

Dataset:
- Friends screenplay script text data

Expected outcome

We anticipate delivering a Tableau dashboard allowing users to explore our findings on various sentiments within the TV series through visualizations.

Evaluation of Work

Our evaluation approaches include identifying a small ground truth dataset with labeled sentiments to measure sentiment analysis performance metrics (e.g., precision, recall, F1). Additionally, we'll use a list of test queries to validate if the expected results are triggered.

Programming language

We plan to use Python for text retrieval and sentiment analysis, and create a dashboard using Tableau to consolidate and visualize the results

Workload justification

The project's sub-tasks, including data pre-processing, text retrieval, sentiment analysis, and dashboard building, amount to an estimated workload of approximately 80 hours for the team.