

CS 410 Fall 2023 Final Project Proposal

Topic: Jeopardy! Topic Analyzer (Free Topic)

Group Name: BFlag

Team Members: Ella Lingafelter (Captain: ellanl2@illinois.edu), Aashish Anand (abanand2@illinois.edu), John Biggan (jbiggan@illinois.edu), John Fox (johnfox3@illinois.edu), Kush Gupta (kushg2@illinois.edu)

Project GitHub Link: https://github.com/kushg2/CS410_BFlag_FinalProject

Introduction

BFlag has decided to create a Jeopardy! topic analyzer. This falls under the category of Free Topic. The final product will consist of a web application in which the user can type a Jeopardy! style question and have it analyzed to determine the most likely category that it would fall into based on previous Jeopardy! questions. Additionally, a predictive model will be used to determine the most likely dollar value that such a question would be associated with, again based on training on previous Jeopardy! questions. This fills a niche that is currently empty. Current Jeopardy! web applications either just present old questions or allow a user to design a Jeopardy! style quiz, but without feedback on the specific questions that the user enters. With our application, the user can receive guidance as they design their questions.

Plan

Our current plan is to create a web application that allows the users to interact with the results of the topic analysis, which will be done in Python. This app will be built using the Shiny framework, which is written in R. The dataset that will be used for the topic analysis portion of the project is a JSON file containing 216,930 past Jeopardy questions:

JEOPARDY_QUESTIONS1.json, which was previously scraped from J! Archive. This site contains every single Jeopardy question that was aired on the show as well as its topic.

Due to an enormous amount of question data in the database, we plan to train the application with a data subset. When the user enters a question, they can expect the app to display its associated topic and dollar value based on the data. There will also be feedback on how to further improve the styling of the user-generated question.

Milestones

Research: Weeks 11-12 [20 hours]

- Confirm most appropriate algorithms, toolsets, and/or packages to conduct topic mining on the Jeopardy dataset.
- Confirm most appropriate framework to ultimately display the completed program to the end user.
- Confirm analysis plan.

Data Exploration: Weeks 11-12 [20 hours]

- Sample representative subset of questions from the existing database to better understand the structure of the data.
- Become familiar with the topics, dollar values, etc... using descriptive (e.g. frequencies, means) and inferential statistics (e.g. Chi-Squared, ANOVAs, and post hoc t-tests).

Data Cleaning: Weeks 11-12 [20 hours]

- Ensure data quality of the Jeopardy question database.
 - Remove null/NA values.
 - Remove duplicate entries.
 - Determine data types and create codebook.
 - Make any other changes necessary to provide clean data for next steps.

Topic Mining: Weeks 12-14 [30 hours]

- Mine Jeopardy question dataset to determine a representative topic model that can be used to label new questions.
- Train supervised predictive model to determine most likely dollar value for the submitted question.
- Validate topic model and predictive models on untrained test sets based on overall and topic specific accuracy.

Web App: Weeks 12-15 [20 hours]

- Utilize finished models in interactive web app.
- Optimize web app for speed and responsiveness.

Presentation: Week 16 [10 hours]

- A script/outline will be produced that describes the application in appropriate detail.
- Film/screen capture the product video.
- Edit video into final product.