Data 612 Final Project Proposal

Anthony Pagan

7/05/2020

Project Proposal

Introduction

The ability to select best resolution to issues from a list of **FAQ** can sometimes be a time consuming task. Trial and error can cause prolonged downtime on critical systems. A recommender system that includes the experience of other similar users working on the same scope of issues would help reduce time to resolution.

Objective

The object of this project is to find a dataset that includes historical resolutions to problems or answers to questions. The rank of resolution or answers, can be given higher weights to answers that provide accurate resolutions or have high number of similar responses.

Methodology

The methodology would be to review the dataset for similar resolutions , rank and weigh best resolution to frequently asked questions. We would use nearest neighbor , frequently used phrases and accuracy scores to determine which answers are the best fit.

Data Description

The search for a proper data set is ongoing but we will begin our analysis with "question answer" dataset in Kaggle. The dataset includes 169 files with the following columns:

- ArticleTitle - the name of the Wikipedia article from which questions and answers initially came.
- Question - is the question.
- Answer - is the answer.
- DifficultyFromQuestioner - prescribed difficulty rating for the question
- DifficultyFromAnswerer - is a difficulty rating assigned by the individual who evaluated and answered the question, which may differ from the difficulty in field 4.
- ArticleFile - is the name of the file with the relevant article

Reference

https://www.kaggle.com/rtatman/questionanswer-dataset?select=S09_question_answer_pairs.txt