CS 410 Project Proposal

Predictive Maintenance Scheduling Using Text Analysis (Free Topic)

Team Members

Captain: Gabe West (NetID: gawest2)
Team Member 1: Anastasia Serebryakova (NetID: as124)

Description

Our project aims to develop a predictive maintenance scheduling system by analyzing historical planned maintenance tasks using text analysis techniques. This system will group maintenance events based on their textual description and generate suggested timeframes. The input for the system will include JSON files containing the following information about past maintenance tasks: task summary, start datetime, end datetime, status.

We plan to use different topic modeling strategies including Expectation-Maximization (EM), Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) to create task clusters using the textual data.
The topic modeling techniques can uncover patterns within historical data which will enable us to generate recommendations for future maintenance schedules via event placement. Ultimately, the goal of the project is to improve maintenance efficiency and reduce downtime.

Technical Details

Programming Language: Python
Tools for text analysis: Numpy, Matplotlib, ScikitLearn, Gensim
Data: JSON-formatted data of historical maintenance schedules

Workload

The workload for this project is expected to be at least 40 hours. Here are the main tasks and their estimated time costs:

Generating a dataset from the raw data	5 hours	Gabe / Anastasia
Data Preprocessing	5 hours	Gabe / Anastasia
Topic Modeling (EM, LDA, LSA)	20 hours	Gabe / Anastasia
Evaluation and comparison between the models	10 hours	Gabe / Anastasia
Code Review	10 hours	Gabe / Anastasia
Testing	5 hours	Gabe / Anastasia
Documentation	5 hours	Gabe / Anastasia

The total estimated time cost is 60 hours, demonstrating the complexity and depth of our project.