Team APS
Members: Andrew Sciotti (sciotti2)

1. **What are the names and NetIDs of all your team members? Who is the captain? The captain will have more administrative duties than team members.**

Member/Captain: Andrew Sciotti, sciotti2

2. **What system have you chosen? Which subtopic(s) under the system?**

I have chosen to improve the EducationalWeb System, specifically by providing more context and improved reading comprehension to the explanations in the form of summarization. For instance, the explanation of "PLSA" is (roughly) 63 sentences long!

3. **Briefly describe the datasets, algorithms or techniques you plan to use**

The goal is to implement extractive summarization on the retrieved relevant explanations provided by EducationalWeb System. The dataset will be provided via the textbook, "Text data management and analysis: a practical introduction to information retrieval and text mining". The fundamentals of the extractive summarization are based on PageRank, but for text, coined (not so creatively), TextRank (reference: https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf). This is an unsupervised algorithm.

4. **If you are adding a function, how will you demonstrate that it works as expected? If you are improving a function, how will you show your implementation actually works better?**

The most straightforward method of evaluation is to manually obtain a few examples of explanations that are lengthy and evaluate the effective conciseness provided by the function. Quantitative measures are unlikely to be evaluated, so the results will be qualitatively evaluated. Ideally, if there were sufficient users, something like A/B testing would be implemented, but that is entirely out of scope.

5. **How will your code communicate with or utilize the system? It is also fine to build your own systems, just please state your plan clearly**

The plan is to identify where the explanation is retrieved and forwarded to the webpage, and intercept that function call to be routed through the extractive summarizer.

6. **Which programming language do you plan to use?**

Python

7. **Please justify that the workload of your topic is at least 20*N hours, N being the total number of students in your team. You may list the main tasks to be completed, and the estimated time cost for each task.**

Project Plan:

- Setup environment, **2 hours**
- Familiarization with EducationalWeb System, **5 hours**
- Progress Report, **1 hour**
- Deep dive into the retrieval/explanation system, **3 hours**
- Implement extractive summarization, **3 hours**
- Debug & evaluated extractive summarization, **3 hours**
- Document code & prepare tutorial, **3 hours**