CS121 Project 3: Milestone 3

Team Members:

- Vincent Vu(27117314/vmvu1)
- Vinita Santhosh(51795233/ santhosv)
- Kinjal Reetoo (36923637/kreetoo)

# Test Queries

**Queries that performed well**

1. ACM
2. Cristina Lopes
3. Machine Learning
4. Year 2020
5. UCI Alumni association
6. 2019 syllabus
7. AI
8. Apocalypse 2020
9. Computer Science Master program
10. UCI netid

**Queries that performed poorly:**

1. Master of Software Engineering
2. The AI project
3. The Department of Computer Science
4. of
5. a professor
6. the schedule of classes
7. a meeting time
8. the table of periodic elements
9. cat of peter
10. Software and computer programs

The following queries did poorly as it contained stop words which caused it to be inefficient with regards to time. Evaluating all documents associated with the stop word took a significant amount of time. For example, when we searched "Master of Software Engineering" the program took approximately 0.5 seconds to run, but when we searched just "Master software engineering", the program only took 0.2 seconds.

To combat this problem, we used the method of High tf-idf query terms only, we selected a set of top A postings, k<A<N , to ensure that the cosine score is only calculated for pages with high tf-idf scores.

In terms of effectiveness, we implemented cosine similarity to replace our existing code containing Boolean Search, could not find results if one or more terms in the query did not exist in our index.