# Milestone 3 Report

## Team Member:

1. Vini Patel
2. Oscar Mark
3. Alexander Butarita
4. Naveen Sastri

## Queries

1. Master of Software Engineering
- This query is long so it took around 600ms for the first time to search. But took only 10 ms if searched again because it was already stored in cache. After splitting the inverted index into multiple csv files according to the first letter of token. This query now takes about 190 ms to search for the first time.

2. aburtsev143a2018falllectureslecture01introfig
- This query shows the effectiveness of search engine. There is only 1 url that has an exact match of this query. There are many more queries that do good and where it is only found in 1 url.

3. a
- This is a very bad query and initially took a long time to show the result, the reason being there are tons of urls that match this query. This query also shows the effectiveness of ranking.

4. Cristina Lopes

5. This is a very very very very long query
- This query shows the effectiveness of having long query and still performing well under 300ms. Takes around 60ms.

6. Alex Thornton

7. ACM

8. Machine learning

9. Eppstein

10. cs121

11. UCI

12. Cyber Security

13. Graduate Courses

14. Graduation Application

15. Computer engineer

16. Prerequisite courses for Computer Science

17. UC Transfer Courses

18. UCI Summer Course Schedule

19. Donald Bren Hall

20. ICS Course Schedule for 2024

## **Fix:**

Our team fixed the issue of going over 300ms by splitting the inverted index into different files. By splitting the index into 27 different files, 1 for each first letter, we were able to optimize our search engine as this is only searching the files according to the first letter of the query tokens. On top of that we were using json files initially and switched to csv as it is a lot faster to access the content in the csv files. We also implemented cache, so recently search queries will take even less time than that.