M3 Report

Group Members: Xu Chen 67237167, Haolin Luo 40126533, Jason Huang 29044684

The 20 Queries:

1. kobeisgoat require ai (This is a name of a person's Othello project for ICS 46. This query proved the effectiveness of Trigram implementation of our project. As the tournament result is the First URL returned)
2. pattis (This query kept returning empty at first, I found the reason behind it is I did not do porter stemming to the word. The issue is now fixed and work fine.)
3. machine learning (This query can be used to test our bigram works fine, we implemented check sum to eliminate duplicate pages. I remember at the beginning it has two very similar pages, but as of now the similar pages are gone {cml.ics.uci.edu/category.aiml/page/2}(with and without the word #content)

   20 top results for **machine learning** (0.07648897171020508 sec)

   # 1: https://cml.ics.uci.edu/category/aiml/page/2/#content
   https://cml.ics.uci.edu/category/aiml/page/2/#content

   # 2: https://cml.ics.uci.edu/aiml/page/20/
   https://cml.ics.uci.edu/aiml/page/20/

4. a (This is a bad query. It takes a very long time even for google search engine. According to my research, google takes 0.43 seconds to do it, which is above 300ms threshold. I have to find a better way to split my index. It worked and the running time for this query

   20 top results for **a** (0.10028791427612305 sec)

   is about 100ms)
5. ACM (This is one of the query we used in M2. It continued to return empty query until I found out I never translate the query to lower case. It is now solved and worked fine)

   20 top results for **acm** (0.04904794692993164 sec)

6. Eppstein (This query is used to show the effectiveness of the search engine. This query returns all links that contains professor Eppstein.)
7. Found here (This query is also used to test the effectiveness of our bigram index, all the url found contains this bigram.)
8. Bryant (This query shows the effectiveness of the search engine. It returns the bio of Bryant Hornick as the top url.)
9. Master of Software Engineering (This query shows the completeness of cosine score calculation of the search engine. It returns master of software engineering website for the informatics department as the top url.)

   Master of Software Engineering
   (MSWE)

10. Kobeisgoat (This is a similar query of the first query. It returns different result. It only returns the tournament url but nothing else. It shows the effectiveness of our search engine. This is not a regular word but it is a project name so it should only appear in the url returned but nothing else.)

11. Lakers Clippers (This query shows the effectiveness of our bigram search. The top result is the Web API explanation Professor Thornton used in ICS 32. He use lakers clippers as

– say, lakers clippers, if you like basketball –

an example)

12. Frank (This is my name. I enter the query just for the best of it. I found a graduate students names AJ.Frank, and his bio is the top url returned. This proves the effectiveness of our search engine)

13. David Kay (This query showed the effectiveness of the bigram. The top urls are all about professor Kay)

14. Webreg (This query showed the effectiveness of the search engine. The top url returned

ICS 23 Spring 2012
Enrollment Information

is a website of enrollment information.)
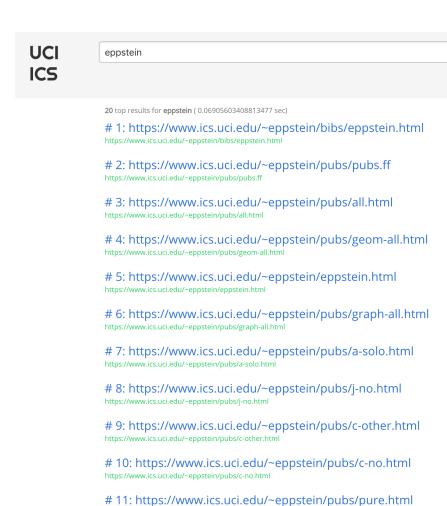
15. Cristina lopes (Another query showed up in M2. All the url returned contains relevant information of professor lopes)

16. I (A bad query similar to 'a'. It took a long time to search before, but it is now under the threshold after the optimization of indexes.)

17. This is a bad query (This is a query used to show the effectiveness of cosine score calculation of our search engine.)

18. this is a very very very very long query (This shows the search engine can handle very long query. The top query returned contained a huge number of word 'very'. It is successfully computed under 300 ms)

20 top results for **this is a very very very very long query** (0.24449491500854492 sec)

19. acggggggggggggggggggccccccccccggggggggggggggggggggggg (I found a very interesting token in my index file. I found it is DNA strings used in one of the professor kay classes in the end. It is found with in 0.04 second)

20. Alex Thornton (It proved the effectiveness of our search engine. The top urls are all professor Thornton related)

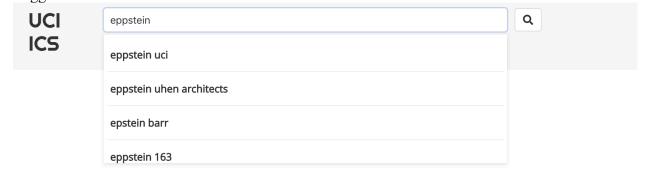Our WEB GUI (Implemented using Django Web Framework):
Front Page

UCI ICS search engine

search engine

Result Page

UCI ICS

eppstein [🔍]

20 top results for **eppstein** (0.06905603408813477 sec)

 # 1: https://www.ics.uci.edu/~eppstein/bibs/eppstein.html
https://www.ics.uci.edu/~eppstein/bibs/eppstein.html

 # 2: https://www.ics.uci.edu/~eppstein/pubs/pubs.ff
https://www.ics.uci.edu/~eppstein/pubs/pubs.ff

 # 3: https://www.ics.uci.edu/~eppstein/pubs/all.html
https://www.ics.uci.edu/~eppstein/pubs/all.html

 # 4: https://www.ics.uci.edu/~eppstein/pubs/geom-all.html
https://www.ics.uci.edu/~eppstein/pubs/geom-all.html

 # 5: https://www.ics.uci.edu/~eppstein/eppstein.html
https://www.ics.uci.edu/~eppstein/eppstein.html

 # 6: https://www.ics.uci.edu/~eppstein/pubs/graph-all.html
https://www.ics.uci.edu/~eppstein/pubs/graph-all.html

 # 7: https://www.ics.uci.edu/~eppstein/pubs/a-solo.html
https://www.ics.uci.edu/~eppstein/pubs/a-solo.html

 # 8: https://www.ics.uci.edu/~eppstein/pubs/j-no.html
https://www.ics.uci.edu/~eppstein/pubs/j-no.html

 # 9: https://www.ics.uci.edu/~eppstein/pubs/c-other.html
https://www.ics.uci.edu/~eppstein/pubs/c-other.html

 # 10: https://www.ics.uci.edu/~eppstein/pubs/c-no.html
https://www.ics.uci.edu/~eppstein/pubs/c-no.html

 # 11: https://www.ics.uci.edu/~eppstein/pubs/pure.html

Suggested Search Term:

UCI ICS

eppstein [🔍]

eppstein uci

eppstein uhen architects

epstein barr

eppstein 163