

Group member:

Changze Ma 36781102

Yexin Wang 86869545

David Nathan Wong 41543325

Chong Wen X1089037

1. How many unique pages did you find?

There are total 24107 unique urls under those domains.

2. What is the longest page in terms of the number of words?

The lowest page url is <http://www.ics.uci.edu/~kay/wordlist.txt> which contains words 385018.

3. What are the 50 most common words in the entire set of pages crawled under these domains?

They are in the report.txt

4. How many subdomains did you find in the ics.uci.edu domain?

They are in the report.txt.