

Team: Group 5
jiliangn, 81224742
ycao16, 17878238
sibow2, 50617929

Part 1

1. The number of documents of the corpus is 37497.
2. The number of unique tokens present in the index is 577040.
3. The total size of our index on disk is 106,488 KB, which is equal to 103MB.

Part 2

The first 10 URLs retrieved for each of the search queries (“Informatics”, “Mondego”, “Irvine”):

1. Informatics

- [illegible]

4.
mondego.ics.uci.edu/projects/clonedetection/tool?C=S;O=A
5.
mondego.ics.uci.edu/projects/clonedetection/tool/latest/features?C=N;O=A
6.
mondego.ics.uci.edu/datasets/wikipedia-events?C=D;O=A
7.
mondego.ics.uci.edu/projects/clonedetection/tool/demo?C=S;O=A
8.
mondego.ics.uci.edu/projects/clonedetection/tool/latest?C=N;O=D
9.
mondego.ics.uci.edu/datasets/wikipedia-events/files
10.
mondego.ics.uci.edu/projects/clonedetection/tool?C=D;O=A

3. Irvine

1.
www.ics.uci.edu/community/events/openhouse/demos.html
2.
www.ics.uci.edu/~jesmaeln
3.
www.ics.uci.edu/~eppstein/pubs/a-smitchell.html
4.
fano.ics.uci.edu/cites/Document/Fully-dynamic-biconnectivity-and-transitive-closure.html
5.
www.ics.uci.edu/about/bren/index.php/bren_vision.php/about_contact.php/bren_vision.php
6.
cml.ics.uci.edu/aiml/page/12

7.

sli.ics.uci.edu/Site/UserCreation

8.

fano.ics.uci.edu/ca/gliders/g1434.html

9.

www.ics.uci.edu/about/bren/index.php/bren_vision.php/index.php/./search/index.php/./../grad/index.php/ICS/ics/about/about_facts

10.

www.ics.uci.edu/~ejw/authoring/collection/dt/CollSpec022.txt