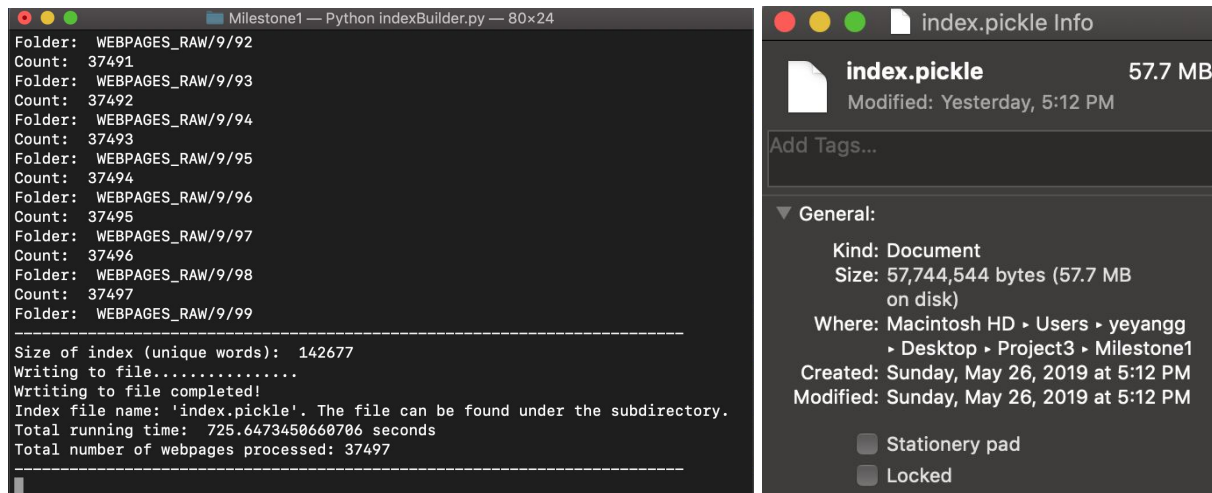Yeyang Gu

Justin Lock

Project Group 65

INF 141

May 27 2019

INF 141 Project 3 Milestone 1 Report

## Part 1: Index



| Number of Unique Words | 142,677 |
|------------------------|---------|
| Number of documents | 37,497 |
| Size of the index file | 57.7 MB |

# Part 2: Searching

### 2.1 Table

| Query Name | # of URLs retrieved |
| --- | --- |
| "Informatics" | 1838 |
| "Mondego" | 47 |
| "Irvine" | 10798 |

### 2.2 Output for searching the query "**Informatics**"

```
Please enter a search query:
Enter number '0' to exit the program
Informatics
You have entered the word: informatics
Searching results for the word: informatics ........
Only showing the first 20 URLs
[ 1 ]
wics.ics.uci.edu/category/news/info-session?afg12_page_id=2&afg44_page_id=4&afg34_pag
e_id=4&afg15_page_id=2&afg42_page_id=2&afg22_page_id=2
[ 2 ]
wics.ics.uci.edu/category/news/info-session?afg12_page_id=2page%2F2%2Fpage/2/&afg42_p
age_id=2&afg44_page_id=4&afg43_page_id=3&afg34_page_id=4
[ 3 ]
www.ics.uci.edu/facebook/login/browse/fanned_pages/events/252272701892321/search/str/
%25E9%259D%259E%25E8%2590%25A5%25E5%2588%25A9%25E7%25BB%2584%25E7%25BB%2587/keywords_
pages/UCIBrenICS/photos/a.10150104180706909.283411.106291486908/10154408253796909?typ
e=3
[ 4 ] sdcl.ics.uci.edu/2015/06/end-of-quarter-lunch-2
[ 5 ] www.ics.uci.edu/~kay/courses/i42/hw
[ 6 ]
www.ics.uci.edu/facebook/pages/create/dialog/UCIBrenICS/videos/10154385262331909/even
ts/1849114518693808/browse/fanned_pages?id=106291486908&showauxiliary=1&fanorigin=tim
eline_like_chaining&ref=page_internal
[ 7 ] sdcl.ics.uci.edu/2013/12/new-workshop-on-crowdsourcing-in-software-engineering
[ 8 ] www.ics.uci.edu/~kay/courses/4/refs.html
[ 9 ]
www.ics.uci.edu/facebook/login/browse/fanned_pages/events/252272701892321?acontext=%7
B%22source%22%3A3%2C%22source_newsfeed_story_type%22%3A%22regular%22%2C%22action_hist
ory%22%3A%22%5B%7B%5C%22surface%5C%22%3A%5C%22newsfeed%5C%22%2C%5C%22mechanism%5C%22%
3A%5C%22feed_story%5C%22%2C%5C%22extra_data%5C%22%3A%5B%5D%7D%5D%22%2C%22has_source%2
2%3Atrue%7D&source=3&source_newsfeed_story_type=regular&action_history=%5B%7B%22surfa
ce%22%3A%22newsfeed%22%2C%22mechanism%22%3A%22feed_story%22%2C%22extra_data%22%3A%5B%
5D%7D%5D&has_source=1&hc_ref=PAGES_TIMELINE
[ 10 ] cml.ics.uci.edu/category/news/page/6?page=events&subPage=aiml
[ 11 ]
wics.ics.uci.edu/category/news/info-session?afg41_page_id=2page%2F2%2F&afg44_page_id=
5
```

```
[ 12 ]
wics.ics.uci.edu/category/news/info-session?afg12_page_id=2page%2F2%2Fpage%2F2%2F&afg
34_page_id=2&afg15_page_id=1&afg43_page_id=1
[ 13 ] www.ics.uci.edu/community/news/view_news?id=1031
[ 14 ] www.ics.uci.edu/prospective/en/degrees/data-science
[ 15 ] vcp.ics.uci.edu/biblio/author/439
[ 16 ] www.ics.uci.edu/~andre/informatics121f2015.html
[ 17 ] sdcl.ics.uci.edu/2012/10/welcome-sergio
[ 18 ]
wics.ics.uci.edu/category/news/info-session?afg41_page_id=2page/2/&afg44_page_id=3&af
g15_page_id=3&afg42_page_id=3&afg12_page_id=2
[ 19 ]
wics.ics.uci.edu/category/news/info-session?afg15_page_id=2&afg41_page_id=2&afg34_pag
e_id=4&afg42_page_id=1
[ 20 ]
wics.ics.uci.edu/category/news/info-session?afg12_page_id=2page%2F2%2F&afg13_page_id=
2&afg34_page_id=4&afg42_page_id=3&afg22_page_id=2&afg44_page_id=3&afg15_page_id=3
Search Completed.
Total number of results:  1838
Time to search: 0.0012149810791015625 seconds
****************************************************
```

## 2.3 Output for searching the query "**Mondego**"

```
You have entered the word: mondego
Searching results for the word: mondego .........
Only showing the first 20 URLs
[ 1 ] mondego.ics.uci.edu/projects/clonedetection/tool/source?C=S;O=A
[ 2 ] mondego.ics.uci.edu/projects/clonedetection/tool/demo
[ 3 ] mondego.ics.uci.edu/projects/clonedetection/tool/latest/features?C=D;O=A
[ 4 ] mondego.ics.uci.edu/datasets/wikipedia-events/files?C=M;O=A
[ 5 ] mondego.ics.uci.edu/projects/clonedetection/tool/demo?C=S;O=A
[ 6 ] mondego.ics.uci.edu/projects/clonedetection/tool/latest/plugins
[ 7 ] mondego.ics.uci.edu/datasets/wikipedia-events/files?C=M;O=D
[ 8 ] mondego.ics.uci.edu
[ 9 ] mondego.ics.uci.edu/datasets?C=N;O=D
[ 10 ] www.ics.uci.edu/~lopes
[ 11 ] mondego.ics.uci.edu/projects/clonedetection/tool/latest
[ 12 ] mondego.ics.uci.edu/projects/clonedetection/tool/latest/plugins?C=M;O=A
[ 13 ] mondego.ics.uci.edu/datasets?C=S;O=A
[ 14 ] mondego.ics.uci.edu/datasets/wikipedia-events/files
[ 15 ] mondego.ics.uci.edu/datasets/wikipedia-events?C=D;O=A
[ 16 ] mondego.ics.uci.edu/projects/clonedetection/tool/latest?C=D;O=A
[ 17 ] mondego.ics.uci.edu/datasets
[ 18 ] mondego.ics.uci.edu/projects/clonedetection/tool/demo?C=D;O=A
[ 19 ] mondego.ics.uci.edu/projects/clonedetection/tool/source?C=S;O=D
[ 20 ] mondego.ics.uci.edu/projects/clonedetection/tool/source?C=D;O=A
Search Completed.
Total number of results:  47
Time to search: 0.00026988983154296875 seconds
****************************************************
```

## 2.4 Output for searching the query "**Irvine**"

```
Please enter a search query:
Enter number '0' to exit the program
Irvine
Searching results for the word: irvine ........
Only showing the first 20 URLs
[ 1 ] fano.ics.uci.edu/cites/Author/Murray-Sherk.html
[ 2 ] fano.ics.uci.edu/cites/Document/Fully-dynamic-min-cut.html
[ 3 ] fano.ics.uci.edu/cites/Author/Stefano-Leonardi.html
[ 4 ] www.ics.uci.edu/~dvk/pub/SIGMODR04_dvk.html
[ 5 ] vision.ics.uci.edu/papers/RamananBK_ICCV_2007
[ 6 ] fano.ics.uci.edu/cites/Publication/GesHofWal-TR-03.html
[ 7 ]
www.ics.uci.edu/~mlearn/datasets/datasets/datasets/datasets/datasets/datasets/dataset
s/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/da
tasets/datasets/datasets/datasets/datasets/Human+Activity+Recognition+Using+Smartphon
es
[ 8 ]
www.ics.uci.edu/faculty/profiles/view_faculty.php/grad/degrees/index/faculty/centers/
ugrad/sao/centers/ugrad/index/involved/about/ugrad/policies/Withdrawal_Readmission/fa
culty/centers/about/search/search_graduate_all.php/grad/policies/index/about/about_co
ntact.php/about/about_deanmsg.php/grad/funding/index
[ 9 ] cgvw.ics.uci.edu/the-future-of-research-in-computer-games-and-virtual-worlds
[ 10 ] fano.ics.uci.edu/cites/Document/On-reset-sequence-length.html
[ 11 ] flamingo.ics.uci.edu/releases/1.0/src/filtertree?C=S;O=A
[ 12 ] fano.ics.uci.edu/cites/Author/Bob-Hertzberger.html
[ 13 ]
fano.ics.uci.edu/cites/Organization/Univ-di-Palermo-Dip-di-Matematica-Appl-Lab-for-Co
mputer-Science.html
[ 14 ] fano.ics.uci.edu/ca/gliders/g21668.html
[ 15 ] sli.ics.uci.edu/Classes-CS271-Notes/SearchHeuristic
[ 16 ] fano.ics.uci.edu/cites/Organization/Cambridge-Press.html
[ 17 ] www.ics.uci.edu/~sjordan/research/net neutrality.html
[ 18 ] fano.ics.uci.edu/cites/Publication/Epp-TR-95-50.html
[ 19 ] fano.ics.uci.edu/cites/Publication/FraOss-TR-99.html
[ 20 ] sli.ics.uci.edu/Profiles/Ihler
Search Completed.
Total number of results:  10798
Time to search: 0.0015232563018798828 seconds
****************************************************
```