

Project 3 Milestone #1 Report

Dayue Bai
dayueb@uci.edu
dayueb
CS 121

Part 1:

Number of the documents of the corpus: 37497 documents

Number of [unique] tokens present in the index: 669957 tokens

The total size (in KB) of my index on disk: 42839.6KB

Part 2: URL retrieved for each of the queries below

-----Informatics-----
wics.ics.uci.edu/category/news/info-session?afg12_page_id=2&afg44_page_id=4&afg34_page_id=4&afg15_page_id=2&afg42_page_id=2&afg22_page_id=2

[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/faculty/centers/about/search/search_graduate_all.php/grad/policies/index/about/about_contact.php/about/about_deanmsg.php/grad/funding/index](http://www.ics.uci.edu/faculty/profiles/view_faculty.php/grad/degrees/index/faculty/centers/ugrad/sao/centers/ugrad/index/involved/about/ugrad/policies/Withdrawal%20Readmission/faculty/centers/about/search/search_graduate_all.php/grad/policies/index/about/about_contact.php/about/about_deanmsg.php/grad/funding/index)

www.ics.uci.edu/alumni/mentor/hall_of_fame/hall_of_fame/stayconnected/hall_of_fame/stayconnected/stayconnected/stayconnected/hall_of_fame/stayconnected/hall_of_fame/hall_of_fame/stayconnected/stayconnected/hall_of_fame/stayconnected/hall_of_fame/stayconnected/hall_of_fame/inductees.aspx.php

[illegible]

wics.ics.uci.edu/category/news/info-session?afg12_page_id=2page%2F2%2Fpage/2/&afg42_page_id=2&afg44_page_id=4&afg43_page_id=3&afg34_page_id=4

www.ics.uci.edu/ugrad/degrees/Double_Majors.php

www.ics.uci.edu/about/visit/../../bren/index.php/bren_advance.php/../../brenhall/index.php/bren_vision.php/../../ugrad/index.php/bren_vision.php/bren_press.php/../../about_mission.php/../../about_safety.php/index.php/bren_advance.php/../../search/index.php/bren_press.php/bren_vision.php/../../brenhall/index.php/../../visit/index.php/../../ugrad/index.php/../../about_safety.php/bren_vision.php/../../visit/index.php/../../ugrad/index.php/../../about_contact.php/bren_about.php/../../brenhall/index.php/bren_vision.php/bren_press.php/index.php/bren_press.php/../../search/index.php/../../search/index.php/ICS/ics/about/about_facts/../../ugrad/index.php/../../ugrad/index.php/../../ugrad/index.php/ICS/ics/about/about_facts/bren_press.php/../../grad/index.php/../../about_deanmsg.php/../../about_deanmsg.php/../../about_mission.php/../../visit/index.php/../../about_contact.php/bren_about.php

www.ics.uci.edu/faculty/profiles/view_faculty.php/faculty/centers/faculty/about/about_contact.php/about/about_contact.php/ugrad/QA_Petitions/community
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?type=3

www.ics.uci.edu/alumni/mentor/stayconnected/stayconnected/stayconnected/hall_of_fame/stayconnected/stayconnected/stayconnected/hall_of_fame/stayconnected/hall_of_fame/stayconnected/stayconnected/index.php

-----Mondego-----

mondego.ics.uci.edu/projects/clonedetection/tool/source?C=S;O=A

www.ics.uci.edu/community/news/notes/notes_2013.php

mondego.ics.uci.edu/projects/clonedetection/tool/demo

mondego.ics.uci.edu/projects/clonedetection/tool/latest/features?C=D;O=A

mondego.ics.uci.edu/datasets/wikipedia-events/files?C=M;O=A

mondego.ics.uci.edu/projects/clonedetection/tool/demo?C=S;O=A

mondego.ics.uci.edu/projects/clonedetection/tool/latest/plugins

flamingo.ics.uci.edu/releases/1.0/src/filtertree?C=S;O=A