Project Group 83

Damian Ly / dsly #82239081

Kelly Tran / trankc1 #81718615

Tam Hoang / tchoang1 #32758731

(INF/CS 141/121) Project 3 – Search Engine: Index Report

Number of Files	37497
Number of Documents	3,256,093
Number of Unique Words	71,478
Total Document Size	1.0 GB
Average Document Size	339.2 B
Total Size of Index (KB)	422,662 KB
Size of Each Index (KB)	1. (_id_) : 32,890 KB 2. (term,folderID,docID,tf,title,h1,h2,h3,h4,h5,h6,strong,body): 78,680 KB 3. (url) : 311,091 KB
Query: "Informatics" **(10 results of 1643)**	www.ics.uci.edu/community/news/view_news?id=1031 www.ics.uci.edu/~andre/informatics121f2015.html 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 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/~kay/courses/i42/hw wics.ics.uci.edu/category/news/info-session?afg41_page_id=2page%2F2%2F&afg44_page_id=5 wics.ics.uci.edu/category/news/info-session?afg12_page_id=2page%2F2%2Fpage%2F2%2F&afg34_page_id=2&afg15_page_id=1&afg43_page_id=1 www.ics.uci.edu/prospective/en/degrees/data-science wics.ics.uci.edu/category/news/info-session?afg41_page_id=2page/2/&afg44_page_id=3&afg15_page_id=3&afg42_page_id=3&afg12_page_id=2 www.ics.uci.edu/faculty/index.php?letter=Z
Query: "Mondego" **(3 results)**	mondego.ics.uci.edu www.ics.uci.edu/~lopes www.ics.uci.edu/~aftabh
Query: "Irvine" **(10 results of 5862)**	www.ics.uci.edu/facebook/login/browse/fanned_pages/events/252272701892321?acontext=%7B%22source%22%3A3%2C%22source_newsfeed_story_type%22%3A%22regular%22%2C%22action_history%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%22%3Atrue%7D&source=3&source_newsfeed_story_type=regular&action_history=%5B%7B%22surface%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 www.ics.uci.edu/about/bren/index.php/bren_vision.php/about_contact.php/bren_vision.php/ICS/index.php/bren_advance.php/../about_mission.php/bren_press.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/~mlearn/datasets/Abalonedatasets/datasets/datasets/datasets/datasets/datasets/datasets/Poker+Hand
	www.ics.uci.edu/~mlearn/datasets/Abalonedatasets/datase
	www.ics.uci.edu/about/bren/index.php/bren_vision.php/grad/index.php/search/index.php/bren_press.php/ICS/ics/grad/index.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/../search/index.php/index.php/../brenhall/index.php/bren_advance.php/ICS/ics/about/about_facts/ICS/ics/about/about_facts/bren_vision.php/../../ugrad/index.php/../../grad/index.php/ICS/ics/about/about_facts/ICS/ics/about/about_facts/bren_vision.php
	www.ics.uci.edu/~jesmaeln/publication.html
	www.ics.uci.edu/~mlearn/datasets/data

	www.ics.uci.edu/~mlearn/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/datasets/Car+Evaluation
Query: "Artificial" (10 of 313)	www.ics.uci.edu/~dvk/pub/SIGMODR04_dvk.html sli.ics.uci.edu/Classes/2016S-274b www.ics.uci.edu/~pazzani/Research.html www.ics.uci.edu/~mlearn/datasets/Abalonedatasets/datase cml.ics.uci.edu/2017/01/winter-2017 cml.ics.uci.edu/category/news/page/4?page=events archive.ics.uci.edu/ml/datasets/Concrete+Compressive+Strength archive.ics.uci.edu/ml/datasets/Auto+MPG fano.ics.uci.edu/cites/Document/Neural-networks-adaptive-optimization-and-RNA-secondary-structure-prediction.html www.ics.uci.edu/community/news/notes/notes_2013.php
Query: "computer" (10 of 9109)	www.ics.uci.edu/alumni/stayconnected/stayconnected/hall_of_fame/hall_of_fame/stayconnected/stayconnected/hall_of_fame/hall_of_fame/hall_of_fame/stayconnected/hall_of_fame/hall_of_fame/hall_of_fame/stayconnected/index.php www.ics.uci.edu/facebook/login/browse/fanned_pages/events/252272701892321?acontext=%7B%22source%22%3A3%2C%22source_newsfeed_story_type%22%3A%22regular%22%2C%22action_history%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%22%3Atrue%7D&source=3&source_newsfeed_story_type=regular&action_history=%5B%7B%22surface%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 www.ics.uci.edu/about/bren/index.php/bren_vision.php/about_contact.php/bren_vision.php/ICS/index.php/bren_advance.php/../about_mission.php/bren_press.php fano.ics.uci.edu/cites/Author/Murray-Sherk.html 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/about/bren/index.php/bren_vision.php/grad/index.php/search/index.php/bren_press.php/ICS/ics/grad/index.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/../../ugra

	d/index.php/../about_safety.php/bren_vision.php/../visit/index.php/../../ugrad/index.php/../about_contact.php/bren_about.php/../search/index.php/index.php/../brenhall/index.php/bren_advance.php/ICS/ics/about/about_facts/ICS/ics/about/about_facts/bren_vision.php/../../ugrad/index.php/../../grad/index.php/ICS/ics/about/about_facts/ICS/ics/about/about_facts/bren_vision.php www.ics.uci.edu/community/news/view_news?id=1031 fano.ics.uci.edu/cites/Document/Optimization-over-zonotopes-and-training-support-vector-machines.html www.ics.uci.edu/prospective/en/contact/student-affairs/contact/contact/contact/contact/contact/contact/contact/student-affairs/contact/student-affairs/contact/contact/student-affairs/contact/student-affairs

[Ignore] For Reference Only:

MongoDB Shell Commands
> db.dictFile.count()
> db.dictFile.distinct('term').length
> db.dictFile.stats()
> db.dictFile.find({term:"informatics"})

MongoDB Compass:

searchEngine.dictFile

Documents   Aggregations   Schema

FILTER   {term:"mondego"}

PROJECT   {url:1}

SORT

COLLATION

VIEW   ≔ LIST   ⊞ TABLE

🏠 dictFile

	_id ObjectId	url String
1	5bf5081d54d48f2b8ca06510	"mondego.ics.uci.edu"
2	5bf508ff54d48f07f8c2fed8	"www.ics.uci.edu/~lopes"
3	5bf5950154d48f30582bc215	"www.ics.uci.edu/~aftabh"