

Elton Xue 52611936
 Chanun Sumphanphaisal 25051704
 Project Group #159

Project 3 – M1 Report

Total # of Documents	Total # Unique Tokens	Total Index Size (kb)
37497	1802295	572600

Search Terms	URLs Found	Top 10 URLs
Informatics	6090	<p>Eltons-MBP:Assignment4 eltonxue\$ python search_index.py Informatics</p> <p>1 -- www.ics.uci.edu/faculty</p> <p>2 -- www.ics.uci.edu/faculty/index.php</p> <p>3 -- www.ics.uci.edu/faculty/index.php?department=Informatics</p> <p>4 -- www.ics.uci.edu/~kay/courses/previous.html</p> <p>5 -- www.ics.uci.edu/community/news/notes/notes_2008.php</p> <p>6 -- www.ics.uci.edu/ugrad/degrees/degree_in4matx.php</p> <p>7 -- www.ics.uci.edu/~andre/teaching.html</p> <p>8 -- www.ics.uci.edu/community/news/notes/notes_2014.php</p> <p>9 -- www.ics.uci.edu/about/annualreport/2006-07/events.php</p> <p>10 -- www.ics.uci.edu/community/news/spotlight/spotlight_purpura.php</p> <p>URLs Found: 6090</p> <p>Total Unique Tokens: 1802295</p> <p>Total Documents in Corpus: 37497</p>
Mondego	10	<p>dhcp-v017-218:Assignment4 eltonxue\$ time python search_index.py Mondego</p> <p>1 -- mondego.ics.uci.edu</p> <p>2 -- www.ics.uci.edu/~lopes</p> <p>3 -- www.ics.uci.edu/community/news/notes/notes_2013.php</p> <p>4 -- www.ics.uci.edu/~lopes/datasets/sourcerer-maven-aug12.html</p> <p>5 -- www.ics.uci.edu/~lopes/datasets</p> <p>6 -- www.ics.uci.edu/~aftabh</p> <p>7 -- www.ics.uci.edu/~lopes/datasets/Koders-log-2007.html</p> <p>8 -- www.ics.uci.edu/community/news/notes/notes_2014.php</p> <p>9 -- sourcerer.ics.uci.edu</p> <p>10 -- www.ics.uci.edu/~lopes/datasets/SDS_source-repo-18k.html</p> <p>URLs Found: 10</p> <p>Total Unique Tokens: 1802295</p> <p>Total Documents in Corpus: 37497</p>
Irvine	12108	<p>dhcp-v017-218:Assignment4 eltonxue\$ time python search_index.py Irvine</p> <p>1 -- www.ics.uci.edu/~eppstein/bibs/eppstein.html</p> <p>2 -- www.ics.uci.edu/~eppstein/bibs/eppstein.bib</p> <p>3 -- cml.ics.uci.edu/category/aiml</p> <p>4 -- www.ics.uci.edu/facebook/search/str/%25E6%2595%2599%25E8%2582%2582/keywords_pages/events/246966085750971/UCIBrenICS/photos/a.10150104180706909.283411.106291486908/10154351203306909/browse/fanned_pages?id=106291486908&showauxiliary=1&fanorigin=timeline_like_chaining&ref=page_internal</p> <p>5 -- www.ics.uci.edu/facebook/pages/create/dialog/UCIBrenICS/videos/10154385262331909/search/112204368792315/p1aces-in/199405806739848/places/intersect/events/711137452397667?context=%7B%22source%22%3A%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%22source%38source_newsfeed_story_type%38regular&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%22%2C%22has_source%38source%38ref=PAGES_TIMELINE</p> <p>6 -- www.ics.uci.edu/facebook/browse/likes/login/UCIBrenICS/photos/a.10150104180706909.283411.106291486908/10154402706541909?type=3</p> <p>7 -- www.ics.uci.edu/facebook/events/340846149648374?context=%7B%22ref%22%3A%2298%22%2C%22action_history%22%3A%22null%22%7D</p> <p>8 -- www.ics.uci.edu/facebook/login/browse/fanned_pages/search/str/%25E6%2595%2599%25E8%2582%2582/keywords_pages/UCIBrenICS/photos/a.10150104180706909.283411.106291486908/10154351203306909?type=3</p> <p>9 -- www.ics.uci.edu/facebook/pages/create/dialog/UCIBrenICS/photos/a.10150104180706909.283411.106291486908/10154351203306909?type=3</p> <p>10 -- www.ics.uci.edu/facebook/login/browse/fanned_pages/search/str/%25E6%2595%2599%25E8%2582%2582/keywords_pages?ref=page_about_category</p> <p>URLs Found: 12108</p> <p>Total Unique Tokens: 1802295</p> <p>Total Documents in Corpus: 37497</p>