Codey Huntting 82661345
Dong Tran 10876970
Alan Ton 70738853

Search Engine Report

Table of Contents

Queries tested

1. cristina lopes
2. master of software engineering
3. acm
4. machine learning
5. computer science
6. professor
7. UCI
8. University of California Irvine
9. ICS
10. Information and computer science
11. recursion
12. class
13. coronavirus
14. pandemic
15. 2020
16. biology
17. alumni
18. pathfinding algorithm

19. object oriented programming
20. uml diagram
21. information retrieval

Old Indexer

Results

Results for query: 'cristina lopes'
Completed in 34.69049072265625 secs
Top results:
https://www.ics.uci.edu/~hsajnani/
https://www.ics.uci.edu/~sjavanma/
http://mondego.ics.uci.edu/
https://www.informatics.uci.edu/explore/faculty-profiles/cristina-lopes/#content
https://www.informatics.uci.edu/explore/faculty-profiles/cristina-lopes/

Results for query: 'master of software engineering'
Completed in 69.74587535858154 secs
Top results:
https://www.ics.uci.edu/~kay/courses/h22/hw/DVD.txt
https://www.ics.uci.edu/~kay/courses/h22/hw/DVD-random.txt
https://www.ics.uci.edu/~eppstein/bibs/kpath.bib
https://www.ics.uci.edu/~gbowker/classification/
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/news

Results for query: 'acm'
Completed in 17.46534252166748 secs
Top results:
https://www.ics.uci.edu/~eppstein/pubs/pubs.ff
https://www.ics.uci.edu/~eppstein/pubs/all.html
https://www.ics.uci.edu/~kay/courses/131/s04readings.html
https://www.ics.uci.edu/~eppstein/bibs/eppstein.html
https://www.ics.uci.edu/~eppstein/pubs/geom-all.html

Results for query: 'machine learning'
Completed in 35.025192737579346 secs
Top results:
http://mondego.ics.uci.edu/datasets/maven-contents.txt
https://cml.ics.uci.edu/category/aiml/page/2/#content
https://cml.ics.uci.edu/category/aiml/page/2/
https://www.ics.uci.edu/~pazzani/Publications/OldPublications.html#1989
https://www.ics.uci.edu/~pazzani/Publications/APubs.html

Results for query: 'computer science'
Completed in 34.93456411361694 secs
Top results:
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/bib
https://www.ics.uci.edu/~eppstein/bibs/eppstein.html
https://www.ics.uci.edu/faculty/index.php
https://www.ics.uci.edu/faculty/
https://www.ics.uci.edu/community/news/notes/notes_2007.php

Results for query: 'professor'
Completed in 17.51003360748291 secs
Top results:
https://www.ics.uci.edu/community/news/notes/notes_2007.php
https://www.cs.uci.edu/faculty/
https://www.ics.uci.edu/faculty/index.php
https://www.ics.uci.edu/faculty/
https://www.informatics.uci.edu/very-top-footer-menu-items/people/

Results for query: 'UCI'
Completed in 17.593652725219727 secs
Top results:
https://www.ics.uci.edu/~magda/Courses/ics156/Ch1_files/WS_FTP.LOG
https://www.informatics.uci.edu/very-top-footer-menu-items/people/
https://www.informatics.uci.edu/very-top-footer-menu-items/people/#Affiliated
https://grape.ics.uci.edu/wiki/asterix/raw-attachment/wiki/cs122a-2019-spring/CompatibleHomework6Dump.sql
https://grape.ics.uci.edu/wiki/asterix/raw-attachment/wiki/cs122a-2019-spring/Homework5Dump.sql

Results for query: 'University California Irvine'
Completed in 52.420467376708984 secs
Top results:
https://cml.ics.uci.edu/category/aiml/page/2/#content
https://cml.ics.uci.edu/category/aiml/page/2/
https://cml.ics.uci.edu/category/aiml/#content
https://cml.ics.uci.edu/category/aiml/
https://redmiles.ics.uci.edu/publication/

Results for query: 'ICS'
Completed in 17.3059720993042 secs
Top results:
https://www.ics.uci.edu/~magda/Courses/ics156/Ch1_files/WS_FTP.LOG
https://www.ics.uci.edu/~kay/courses/previous.html

https://www.ics.uci.edu/~magda/Courses/ics156/Ch3_files/WS_FTP.LOG
https://www.ics.uci.edu/~thornton/ics45c/ProjectGuide/Project0/
https://www.ics.uci.edu/~pattis/ICS-46/lectures/notes/template.txt

Results for query: 'Information computer science'
Completed in 52.397966146469116 secs
Top results:
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/bib
https://www.ics.uci.edu/~gbowker/converge.html
https://www.ics.uci.edu/~eppstein/bibs/eppstein.html
https://www.ics.uci.edu/community/news/notes/notes_2007.php
https://www.ics.uci.edu/~eppstein/bibs/kpath.bib

Results for query: 'recursion'
Completed in 17.729698419570923 secs
Top results:
https://www.ics.uci.edu/~kay/courses/i42/wildride/data/1000customers.txt
https://www.ics.uci.edu/~pattis/ICS-46/lectures/notes/recursion.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/recursion.txt
https://www.ics.uci.edu/~kay/courses/i42/wildride/data/customers2.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/functionalprogramming.txt

Results for query: 'class'
Completed in 17.57003951072693 secs
Top results:
https://www.ics.uci.edu/~schark/simulator/javadoc/AllNames.html
https://www.ics.uci.edu/~arcadia/Teamware/docs/AllNames.html
https://www.ics.uci.edu/~pattis/ICS-21/lectures/writingclasses/lecture.html
https://www.ics.uci.edu/~pattis/ICS-33/lectures/inheritancei.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/class.txt

Results for query: 'coronavirus'
Completed in 17.73116183280945 secs
No results found

Results for query: 'pandemic'
Completed in 18.348912715911865 secs
Top results:
https://www.ics.uci.edu/~dan/genealogy/Miller/langsam/louser.htm
https://www.ics.uci.edu/~kay/wordlist.txt
https://www.informatics.uci.edu/2017/08/
https://www.informatics.uci.edu/forbes-plagued-by-workplace-interruptions-set-some-boundaries-mark-quoted/
https://www.informatics.uci.edu/2017/08/#content

Results for query: '2020'
Completed in 17.60988140106201 secs
Top results:
http://mondego.ics.uci.edu/datasets/maven-contents.txt
https://mcs.ics.uci.edu/prospective-students/cost-and-financial-aid/
https://mswe.ics.uci.edu/program/curriculum/
https://www.ics.uci.edu/community/news/view_news?id=1465
https://www.ics.uci.edu/community/news/view_news.php?id=1465

Results for query: 'biology'
Completed in 17.400765419006348 secs
Top results:
https://www.ics.uci.edu/~emj/Mjolsness_CV_V67p.htm
https://emj.ics.uci.edu/papers/computational-biology-papers/#content
https://emj.ics.uci.edu/papers/computational-biology-papers/
http://computableplant.ics.uci.edu/papers/#essays
http://computableplant.ics.uci.edu/papers/

Results for query: 'alumni'
Completed in 17.351154088974 secs
Top results:
https://www.ics.uci.edu/about/annualreport/2006-07/alumni.php
https://www.ics.uci.edu/community/news/view_news.php?id=1136
https://www.ics.uci.edu/community/alumni/index.php
https://www.ics.uci.edu/about/annualreport/2005-06/alumni.php
https://www.ics.uci.edu/community/alumni/

Results for query: 'pathfinding algorithm'
Completed in 34.36037564277649 secs
Top results:
https://www.ics.uci.edu/~kay/wordlist.txt
https://www.ics.uci.edu/~dechter/courses/ics-271/fall-08/project/index.html

Results for query: 'object oriented programming'
Completed in 51.84824824333191 secs
Top results:
https://www.ics.uci.edu/~pattis/ICS-21/lectures/usingclasses/lecture.html
https://www.ics.uci.edu/~pattis/ICS-33/lectures/review.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/class.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/inheritancei.txt
https://www.ics.uci.edu/~pattis/quotations.html

Results for query: 'uml diagram'

Completed in 34.44801568984985 secs
Top results:
http://mondego.ics.uci.edu/datasets/maven-contents.txt
https://www.ics.uci.edu/~thornton/inf122/ProjectGuide/Project1/
https://www.ics.uci.edu/~alspaugh/cls/shr/msc.html
https://www.ics.uci.edu/~alspaugh/cls/shr/ontology.html
https://www.ics.uci.edu/~thornton/inf122/ProjectGuide/Project3/

Results for query: 'information retrieval'
Completed in 34.67146706581116 secs
Top results:
https://www.ics.uci.edu/~gbowker/converge.html
https://www.ics.uci.edu/~gbowker/classification/
https://www.ics.uci.edu/~kobsa/privacy/German.htm#nr34
https://www.ics.uci.edu/~gbowker/forget.html
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/bib

Comments

The old indexer was *extremely* slow. Even a single word query took 17 seconds to complete! This indexer's performance was unsatisfactory for a few reasons:

1. **Unmerged index:** the index used in the old indexer was not merged or alphabetized. Because of this, a single word and its postings could exist multiple times in the file. Thus the query had to search the *entire* index files for all instances of a query token
2. **Unweighted ranking:** the indexer does not add any special weight to words found in bold or in headers in the HTML. The only score used is the term frequency, which gave some unsatisfactory results. For instance, the first result for the query "cristina lopes" takes us to the profile of a grad student loosely associated with Prof. Lopes, rather than Prof. Lopes' actual profile.

New Indexer

Results

Results for query: 'cristina lopes'
Completed in 0.0029985904693603516 secs
Total results: 83
Top results:
https://www.informatics.uci.edu/explore/faculty-profiles/cristina-lopes/
https://www.informatics.uci.edu/explore/faculty-profiles/cristina-lopes/#content
https://www.informatics.uci.edu/lopes-analyzes-big-code-with-funding-from-darpa/
https://www.informatics.uci.edu/lopes-analyzes-big-code-with-funding-from-darpa/#content

https://www.informatics.uci.edu/lopes-featured-speaker-at-2016-opensimulator-community-conference/

Results for query: 'master of software engineering'
Completed in 0.19024324417114258 secs
Total results: 1631
Top results:
https://www.ics.uci.edu/~kay/courses/h22/hw/DVD-random.txt
https://www.ics.uci.edu/~kay/courses/h22/hw/DVD.txt
https://www.ics.uci.edu/~eppstein/bibs/kpath.bib
https://www.ics.uci.edu/~gbowker/classification/
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/news

Results for query: 'acm'
Completed in 0.007000446319580078 secs
Total results: 2789
Top results:
https://www.ics.uci.edu/~gmark/Home_page/Publications.html
https://www.ics.uci.edu/~eppstein/pubs/pubs.ff
https://www.ics.uci.edu/~eppstein/pubs/all.html
https://www.ics.uci.edu/~kay/courses/131/s04readings.html
https://www.ics.uci.edu/~eppstein/bibs/eppstein.html

Results for query: 'machine learning'
Completed in 0.052004337310791016 secs
Total results: 3475
Top results:
https://www.ics.uci.edu/~welling/publications/publications.html
https://cml.ics.uci.edu/category/aiml/page/2/
https://cml.ics.uci.edu/category/aiml/page/2/#content
https://cml.ics.uci.edu/category/aiml/
https://cml.ics.uci.edu/category/aiml/#content

Results for query: 'computer science'
Completed in 0.14101243019104004 secs
Total results: 11303
Top results:
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/bib
https://www.ics.uci.edu/~ics1c/hw4/33.html
https://www.ics.uci.edu/faculty/
https://www.ics.uci.edu/faculty/index.php
https://www.ics.uci.edu/community/news/notes/notes_2007.php

Results for query: 'professor'

Completed in 0.014002561569213867 secs
Total results: 4031
Top results:
https://www.ics.uci.edu/community/news/
https://www.ics.uci.edu/community/news/index.php
https://www.ics.uci.edu/faculty/
https://www.ics.uci.edu/faculty/index.php
https://www.cs.uci.edu/faculty/

Results for query: 'UCI'
Completed in 0.05600333213806152 secs
Total results: 20154
Top results:
https://www.ics.uci.edu/community/news/
https://www.ics.uci.edu/community/news/index.php
https://www.ics.uci.edu/~magda/Courses/ics156/Ch1_files/WS_FTP.LOG
https://www.ics.uci.edu/~rickl/courses/cs-171/0-ihler-2016-fq/Lectures/Lathrop/CS-171%20Fall%20Quarter%202016.htm
https://www.ics.uci.edu/~rickl/courses/cs-171/2016-fq-cs171/CS-171-FQ-2016_20160918.htm

Results for query: 'University California Irvine'
Completed in 0.11100435256958008 secs
Total results: 6176
Top results:
https://www.ics.uci.edu/~cbdaviso/
https://www.ics.uci.edu/~cbdaviso/index.html
https://cml.ics.uci.edu/category/aiml/page/2/
https://cml.ics.uci.edu/category/aiml/page/2/#content
https://cml.ics.uci.edu/category/aiml/

Results for query: 'ICS'
Completed in 0.06000161170959473 secs
Total results: 18643
Top results:
https://www.ics.uci.edu/~rickl/courses/ics-h197/2012-fq-h197/ICS-H197-FQ-2012.htm
https://www.ics.uci.edu/~rickl/courses/ics-h197/2014-fq-h197/ICS-H197-FQ-2014.htm
https://www.ics.uci.edu/~rickl/courses/ics-h197/2013-fq-h197/ICS-H197-FQ-2013.htm
https://www.ics.uci.edu/~rickl/courses/ics-h197/2011-fq-h197/ICS-H197-FQ-2011.htm
https://www.ics.uci.edu/~magda/Courses/ics156/Ch1_files/WS_FTP.LOG

Results for query: 'Information computer science'
Completed in 0.20001435279846191 secs
Total results: 10187
Top results:

https://www.ics.uci.edu/~ics1c/hw4/33.html
https://www.ics.uci.edu/~dan/class/267P/datasets/calgary/bib
https://www.ics.uci.edu/community/news/notes/notes_2007.php
https://www.ics.uci.edu/~gbowker/converge.html
https://www.ics.uci.edu/faculty/

Results for query: 'recursion'
Completed in 0.0010008811950683594 secs
Total results: 188
Top results:
https://www.ics.uci.edu/~kay/courses/i42/wildride/data/1000customers.txt
https://www.ics.uci.edu/~ejw/authoring/munich/issues/tsld010.htm
https://www.ics.uci.edu/~pattis/ICS-46/lectures/notes/recursion.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/recursion.txt
http://tutors.ics.uci.edu/index.php/79-python-resources/123-recursion-examples

Results for query: 'class'
Completed in 0.03200101852416992 secs
Total results: 11238
Top results:
https://www.ics.uci.edu/~rickl/courses/cs-171/0-ihler-2016-fq/Lectures/Lathrop/CS-171%20Fall%20Quarter%202016.htm
https://www.ics.uci.edu/~rickl/courses/cs-171/2016-fq-cs171/CS-171-FQ-2016_20160918.htm
https://www.ics.uci.edu/~rickl/courses/cs-171/2016-fq-cs171/CS-171-FQ-2016_20160919.htm
https://www.ics.uci.edu/~rickl/courses/cs-171/2016-fq-cs171/CS-171-FQ-2016_draft.htm
https://www.ics.uci.edu/~schark/simulator/javadoc/AllNames.html

Results for query: 'coronavirus'
Completed in 0.0 secs
Total results: 0
No results found

Results for query: 'pandemic'
Completed in 0.0 secs
Total results: 7
Top results:
https://www.ics.uci.edu/~dan/genealogy/Miller/langsam/louser.htm
https://www.ics.uci.edu/~kay/wordlist.txt
https://www.informatics.uci.edu/2017/08/
https://www.informatics.uci.edu/2017/08/#content
https://www.informatics.uci.edu/forbes-plagued-by-workplace-interruptions-set-some-boundaries-mark-quoted/

Results for query: '2020'

Completed in 0.0010001659393310547 secs
Total results: 348
Top results:
https://www.informatics.uci.edu/cpri-hosts-workforce-2020-panel-discussion-and-networking-reception-for-ics-90-students/
https://www.informatics.uci.edu/cpri-hosts-workforce-2020-panel-discussion-and-networking-reception-for-ics-90-students/#content
https://www.cs.uci.edu/cpri-hosts-workforce-2020-panel-discussion-and-networking-reception-for-ics-90-students/
https://www.cs.uci.edu/cpri-hosts-workforce-2020-panel-discussion-and-networking-reception-for-ics-90-students/#more-1916
https://www.ics.uci.edu/community/news/view_news.php?id=1465

Results for query: 'biology'
Completed in 0.0009999275207519531 secs
Total results: 614
Top results:
https://www.ics.uci.edu/~emj/Mjolsness_CV_V67p.htm
https://emj.ics.uci.edu/papers/computational-biology-papers/
https://emj.ics.uci.edu/papers/computational-biology-papers/#content
https://www.ics.uci.edu/~rickl/rickl-publications
https://www.ics.uci.edu/~rickl/rickl-publications.html

Results for query: 'alumni'
Completed in 0.009001493453979492 secs
Total results: 3709
Top results:
https://www.ics.uci.edu/community/news/
https://www.ics.uci.edu/community/news/index.php
https://www.ics.uci.edu/about/annualreport/2006-07/alumni.php
https://www.ics.uci.edu/about/annualreport/2005-06/alumni.php
https://mailman.ics.uci.edu/mailman/listinfo/alumni.mcs

Results for query: 'pathfinding algorithm'
Completed in 0.010001659393310547 secs
Total results: 2
Top results:
https://www.ics.uci.edu/~kay/wordlist.txt
https://www.ics.uci.edu/~dechter/courses/ics-271/fall-08/project/index.html

Results for query: 'object oriented programming'
Completed in 0.02800726890563965 secs
Total results: 231
Top results:

https://mswe.ics.uci.edu/program/curriculum/
https://www.ics.uci.edu/~pattis/ICS-21/lectures/usingclasses/lecture.html
https://www.ics.uci.edu/~pattis/ICS-33/lectures/review.txt
https://www.ics.uci.edu/~pattis/ICS-33/lectures/class.txt
https://www.ics.uci.edu/~pattis/ICS-21/lectures/interfaces/lecture.html

Results for query: 'uml diagram'
Completed in 0.003993511199951172 secs
Total results: 14
Top results:
http://mondego.ics.uci.edu/datasets/maven-contents.txt
https://www.ics.uci.edu/~thornton/inf122/ProjectGuide/Project1/
https://www.ics.uci.edu/~alspaugh/cls/shr/msc.html
https://www.ics.uci.edu/~alspaugh/cls/shr/ontology.html
https://www.ics.uci.edu/~thornton/inf122/ProjectGuide/Project3/

Results for query: 'information retrieval'
Completed in 0.07100629806518555 secs
Total results: 707
Top results:
https://www.ics.uci.edu/~gbowker/converge.html
http://www-db.ics.uci.edu/pages/research/mars/
http://www-db.ics.uci.edu/pages/research/mars/#people
http://www-db.ics.uci.edu/pages/research/mars/index.shtml
https://www.ics.uci.edu/~gbowker/classification/

Comments

We are pleased to report that the new indexer is *extremely* fast compared to the old indexer. The new indexer performs better using the following techniques:

1. **Index merging:** we implemented index merging so that each term has only one position in the index file, and all terms are in alphabetical order.
2. **Lexicon:** we built a lexicon to look up the exact position of the terms in the index, allowing us to search the index in O(1) time!
3. **Fast posting merge algorithm:** we exploited the fact that posting lists are sorted by document url to build a fast algorithm that merges the posting lists.
4. **Weighted term scoring:** the indexer adds additional weights to each term based on its presence in important html tags such as and <h1>. This produced exceptional ranking results, especially for the query “cristina lopes” which finds Prof. Lopes’ profile page as the first result!