***Milestone #3 Report***

1. ***NDCG criteria:*** When calculating NDCG, we consider urls ranked 1-2 in Google results have relevance value 3, and urls ranked 3-5 in Google have relevance value2, and urls ranked 6-10 as 1, and all other urls as 0.

***NDCG calculation results***: Our average DCG@5 is 1.3656 before improvement, and 4.1543 after improvement. As the idea DCG@5 is . As a result, our average NDCG@5 is 0.1571 without heuristic improvement, and such a number increases to 0.4779 with significantly 204% improvement.

1. ***Heuristics:*** We tried to improve NDCG@5 roughly in the below 5 approaches, mainly including the aspect of word processing, two-gram optimization, link conferral etc., more specifically:
2. Stemming, we notice that before improvement, “graduate course” and “graduate courses” would be considered as different terms, so we applied the stemming dictionary from [***http://snowball.tartarus.org/algorithms/english/stemmer.html***](http://snowball.tartarus.org/algorithms/english/stemmer.html)
3. Two-gram optimization. Instead of only indexing the single word, we indexer will also indexes two-grams and calculate their tfidf, which can take into consideration the coherence between words Hence, if the number of query words is no smaller than 2, the tfidf values for each two-grams in queries will be applied for evaluating the weight of a page.
4. Title of each URL is stored, and a score is calculated which measures the similarity between title and query
5. Anchor text of each URL is stored, and a score is calculated which measures the similarity between anchor text and query
6. PageRank is calculated for each URL, taking into account the conferral authority between links.

The formula for the overall score of a URL is indicated as below (tfidf\_score is calculated based on single words or two-grams, which depends on the number of words in query):

Explanation: In the above formula, are the tuning parameters for our search engine. We experimented with different weight values to determine optimal parameters. We tried w1 = 0.0 ~ 3.0, w2 = 0.0 ~ 1.0, w3 = 0.0 ~ 1.0. And experiment results show that w1 = 2.40, w2 = 0.1, w3 = 0.1 achieves the best NDCG@5 score.

Comparison of results before and after improvement

|  |  |  |
| --- | --- | --- |
| mondego | | |
| Results from Google( top 10 ) | Before improvement | After improvement |
| http://nile.ics.uci.edu:9000/wifi  http://mondego.ics.uci.edu/  http://mondego.ics.uci.edu/index.html  http://mailman.ics.uci.edu/mailman/listinfo/mondego  http://nile.ics.uci.edu:9000/wifi/user/account/  http://mailman.ics.uci.edu/mailman/admin/mondego  http://sdcl.ics.uci.edu/mondego-group/  http://nile.ics.uci.edu:9000/wifi/forgotpassword  http://nile.ics.uci.edu:9000/wifi/admin/estates/new  http://sdcl.ics.uci.edu/2012/05/calico-for-the-mondego-group/  http://mailman.ics.uci.edu/mailman/private/mondego/ | **https://mailman.ics.uci.edu/mailman/listinfo/mondego**  **http://mondego.ics.uci.edu/**  http://www.ics.uci.edu/~lopes/  https://www.ics.uci.edu/community/news/notes/index.php  http://www.ics.uci.edu/community/news/notes/index | **http://mondego.ics.uci.edu/**  http://www.ics.uci.edu/~lopes/  http://www.ics.uci.edu/~djp3/classes/2006\_03\_30\_ICS105/Resources/AnteaterIdol.html  http://mondego.ics.uci.edu/datasets/wikipedia-events/files/?C=D%3BO%3DA  http://mondego.ics.uci.edu/datasets/ |

|  |  |  |
| --- | --- | --- |
| machine learning | | |
| Results from Google(top 10) | Before improvement | After improvement |
| http://archive.ics.uci.edu/ml/machine-learning-databases/  http://archive.ics.uci.edu/ml/about.html  http://cml.ics.uci.edu/  http://archive.ics.uci.edu/ml/  http://mlearn.ics.uci.edu/MLRepository.html  http://sli.ics.uci.edu/Classes/2013F-273a  http://sli.ics.uci.edu/Classes/2015W-273a  http://archive.ics.uci.edu/ml/datasets.html  http://archive.ics.uci.edu/ml/datasets/Credit+Approval  http://archive.ics.uci.edu/ml/datasets/Diabetes | http://www.ics.uci.edu/~pazzani/Publications/OldPublications.html  http://www.ics.uci.edu/~pazzani/Publications/APubs.html  http://cml.ics.uci.edu/?page=events&subPage=aiml\_1213  http://cml.ics.uci.edu/?page=events&subPage=aiml  http://isg.ics.uci.edu/events.html | http://archive.ics.uci.edu/ml/datasets/ser+Knowledge+Modeling+Data+%28Students%27+Knowledge+Levels+on+DC+Electrical+Machines%29  http://www.ics.uci.edu/~pazzani/Publications/OldPublications.html  http://www.ics.uci.edu/~pazzani/Publications/APubs.html  http://cml.ics.uci.edu/?page=events&subPage=aiml\_1213  http://www.ics.uci.edu/~qliu1/MLcrowd\_ICML\_workshop/ |

|  |  |  |
| --- | --- | --- |
| software engineering | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/prospective/en/degrees/software-engineering/  http://www.ics.uci.edu/grad/degrees/degree\_se.php  http://www.ics.uci.edu/ugrad/degrees/degree\_se.php  http://www.ics.uci.edu/prospective/en/degrees/computer-science-engineering/  http://www.ics.uci.edu/~ziv/ooad/intro\_to\_se/tsld008.htm  http://www.ics.uci.edu/~djr/DebraJRichardson/SE4S.html  http://www.ics.uci.edu/faculty/area/area\_software.php  http://se.ics.uci.edu/  http://www.ics.uci.edu/~emilyo/SimSE/se\_rules.html  http://se4s.ics.uci.edu/ | http://www.ics.uci.edu/~wscacchi/Papers/Vintage/Software\_Productivity.html  http://www.ics.uci.edu/~taylor/Publications.htm  http://www.ics.uci.edu/~wscacchi/publications.html  http://www.ics.uci.edu/~andre/publications.html  http://www.ics.uci.edu/~wscacchi/ | http://sdcl.ics.uci.edu/2013/12/new-workshop-on-crowdsourcing-in-software-engineering/  **http://www.ics.uci.edu/prospective/en/degrees/software-engineering/**  **http://www.ics.uci.edu/ugrad/degrees/degree\_se.php**  http://cradl.ics.uci.edu/?p=602  http://www.ics.uci.edu/~emilyo/SimSE/se\_rules.html |

|  |  |  |
| --- | --- | --- |
| Security | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://ftp.ics.uci.edu/pub/centos0/ics-custom-build/BUILD/nagios-3.0.6/html/docs/security.html  http://sconce.ics.uci.edu/  http://www.ics.uci.edu/computing/linux/security.php  http://www.ics.uci.edu/computing/linux/file-security.php  http://www.ics.uci.edu/~gts/  http://sprout.ics.uci.edu/  http://www.ics.uci.edu/~keldefra/manet.htm  http://sprout.ics.uci.edu/past\_projects/odb/  http://sprout.ics.uci.edu/projects/privacy-dna/  http://www.ics.uci.edu/~sbruntha/ | http://www.ics.uci.edu/~gts/pubs.html  http://www.ics.uci.edu/community/news/notes/notes\_2009.php  https://www.ics.uci.edu/community/news/notes/notes\_2009.php  http://www.ics.uci.edu/community/news/notes/notes\_2010.php  http://www.ics.uci.edu/community/news/notes/notes\_2007.php | http://www.ics.uci.edu/~jie/SecureSoftwareArchitecture.html  http://sli.ics.uci.edu/pmwiki/pmwiki.php?n=PmWiki%2FSecurity  http://sli.ics.uci.edu/PmWiki/Security  http://www.ics.uci.edu/~goodrich/teach/ics8/notes/  http://www.ics.uci.edu/~goodrich/teach/ics8/hw/hw1.html |

|  |  |  |
| --- | --- | --- |
| student affairs | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/prospective/en/contact/student-affairs/  http://www.ics.uci.edu/about/search/search\_sao.php  http://www.ics.uci.edu/ugrad/  http://www.ics.uci.edu/grad/sao/  http://www.ics.uci.edu/about/visit/  http://www.ics.uci.edu/about/annualreport/2005-06/sao.php  http://www.ics.uci.edu/ugrad/sao/  http://www.ics.uci.edu/about/annualreport/2006-07/sao.php  http://www.ics.uci.edu/grad/  http://www.ics.uci.edu/about/about\_contact.php | http://www.ics.uci.edu/ugrad/qa/index.php  http://www.ics.uci.edu/ugrad/qa/  http://www.ics.uci.edu/community/news/notes/notes\_2007.php  **http://www.ics.uci.edu/ugrad/**  http://www.ics.uci.edu/ugrad/index.php | **http://www.ics.uci.edu/about/annualreport/2006-07/sao.php**  **http://www.ics.uci.edu/about/annualreport/2005-06/sao.php**  http://www.ics.uci.edu/prospective/en/contact/student-affairs/  http://www.ics.uci.edu/ugrad/qa/index.php  http://www.ics.uci.edu/ugrad/qa/ |

|  |  |  |
| --- | --- | --- |
| graduate courses | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/grad/courses/  http://www.ics.uci.edu/grad/courses/listing.php?year=2014&level=ALL&department=CS&program=1COMPSCI  http://www.ics.uci.edu/grad/courses/listing.php?year=2014&level=ALL&department=EECS&program=0CSSS  http://www.ics.uci.edu/grad/courses/details.php?id=69  http://www.ics.uci.edu/grad/courses/details.php?id=48  http://www.ics.uci.edu/grad/courses/details.php?id=59  http://www.ics.uci.edu/grad/courses/details.php?id=89  http://www.ics.uci.edu/grad/courses/details.php?id=72  http://www.ics.uci.edu/grad/courses/details.php?id=51  http://www.ics.uci.edu/grad/courses/details.php?id=56 | http://www.ics.uci.edu/ugrad/qa/index.php  http://www.ics.uci.edu/ugrad/qa/  http://www.ics.uci.edu/~jcarlson/gradschool.html  http://www.ics.uci.edu/community/news/notes/notes\_2007.php  http://www.ics.uci.edu/~pattis/ICS-33/lectures/overview/lecture.html | http://www.ics.uci.edu/grad/courses/index.php  **http://www.ics.uci.edu/grad/courses/**  http://www.ics.uci.edu/grad/courses/index  http://www.ics.uci.edu/grad/policies/GradPolicies\_Grading.php  http://www.ics.uci.edu/~welling/teaching/courses.html |

|  |  |  |
| --- | --- | --- |
| crista lopes | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/~lopes/  http://www.ics.uci.edu/~lopes/patents.html  http://www.ics.uci.edu/~lopes/publications.html  http://www.ics.uci.edu/~tdebeauv/  http://mondego.ics.uci.edu/  http://luci.ics.uci.edu/blog/?tag=crista-lopes&paged=2  http://www.ics.uci.edu/~lopes/teaching/cs221W12/  http://luci.ics.uci.edu/blog/?p=416  http://luci.ics.uci.edu/blog/?feed=rss2&tag=crista-lopes  http://luci.ics.uci.edu/blog/?tag=crista-lopes | **http://www.ics.uci.edu/~lopes/publications.html**  http://www.ics.uci.edu/~hsajnani/  http://www.ics.uci.edu/community/news/notes/notes\_2007.php  http://sourcerer.ics.uci.edu/publications.html  http://www.ics.uci.edu/community/news/features/view\_feature?id=67 | **http://www.ics.uci.edu/~lopes/publications.html**  **http://www.ics.uci.edu/~lopes/patents.html**  **http://www.ics.uci.edu/~lopes/**  http://fano.ics.uci.edu/cites/Author/Cristopher-Moore.html  http://fano.ics.uci.edu/cites/Author/Cristian-Sorin-Mata.html |

|  |  |  |
| --- | --- | --- |
| rest | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm  http://www.ics.uci.edu/~fielding/pubs/dissertation/rest\_arch\_style.htm  http://www.ics.uci.edu/~fielding/pubs/dissertation/evaluation.htm  http://www.ics.uci.edu/~fielding/pubs/dissertation/abstract.htm  http://www.ics.uci.edu/~fielding/pubs/dissertation/introduction.htm  http://www.ics.uci.edu/~fielding/pubs/dissertation/conclusions.htm  http://www.ics.uci.edu/~fielding/talks/webarch\_9805/  http://asterixdb.ics.uci.edu/documentation/api.html  http://www.ics.uci.edu/~fielding/pubs/dissertation/software\_arch.htm  http://www.ics.uci.edu/~fielding/ | **http://www.ics.uci.edu/~fielding/pubs/dissertation/rest\_arch\_style.htm**  **http://www.ics.uci.edu/~fielding/pubs/dissertation/evaluation.htm**  http://www.ics.uci.edu/~kay/courses/i41/answers.html  http://www.ics.uci.edu/~thornton/ics184/MidtermSolutions.html  http://archive.ics.uci.edu/ml/datasets/SPECTF+Heart | **http://www.ics.uci.edu/~fielding/pubs/dissertation/rest\_arch\_style.htm**  **http://www.ics.uci.edu/~fielding/pubs/dissertation/evaluation.htm**  http://www.ics.uci.edu/~kay/courses/i41/answers.html  http://www.ics.uci.edu/~thornton/ics184/MidtermSolutions.html  http://www.ics.uci.edu/~eppstein/pix/ontheroad/BwilRestArea.html |

|  |  |  |
| --- | --- | --- |
| computer games | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/prospective/en/degrees/computer-game-science/  http://cgvw.ics.uci.edu/  http://frost.ics.uci.edu/cs113/  http://cgvw.ics.uci.edu/affiliated-faculty/  https://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset  http://www.ics.uci.edu/community/events/gameday2015/  http://www.ics.uci.edu/~jwross/courses/ics60/  http://cgvw.ics.uci.edu/author/venita/  http://www.ics.uci.edu/~eppstein/gina/vidgames.html  http://cgvw.ics.uci.edu/author/admin/ | http://www.ics.uci.edu/~wscacchi/  http://fano.ics.uci.edu/cites/Organization/  http://www.ics.uci.edu/~eppstein/bibs/eppstein.html  http://www.ics.uci.edu/~pattis/quotations.html  http://www.ics.uci.edu/~kay/courses/i42/hw/labA.html | **http://www.ics.uci.edu/prospective/en/degrees/computer-game-science/**  http://www.ics.uci.edu/ugrad/degrees/degree\_cgs.php  http://frost.ics.uci.edu/cs113/DesignDoc.html  http://www.ics.uci.edu/ugrad/degrees/degree\_cs.php  http://fano.ics.uci.edu/cites/Document/Subdivision-Surfaces-in-Computer-Games.html |

|  |  |  |
| --- | --- | --- |
| information retrieval | | |
| Results from Google (top 10) | Before improvement | After improvement |
| http://www.ics.uci.edu/~djp3/classes/2014\_01\_INF141/calendar.html  http://www.ics.uci.edu/~djp3/classes/2009\_01\_02\_INF141/  http://www.ics.uci.edu/~djp3/classes/2010\_01\_CS221/  http://www.ics.uci.edu/~lopes/teaching/cs221W12/  http://www.ics.uci.edu/~djp3/classes/2014\_01\_INF141/structure.html  http://www.ics.uci.edu/~lopes/teaching/cs221W13/  http://www.ics.uci.edu/~djp3/classes/2009\_01\_02\_INF141/calendar.html  http://www.ics.uci.edu/~kay/courses/i141/w15.html  http://www.ics.uci.edu/~lopes/  http://www.ics.uci.edu/~kay/courses/i141/refs.html | http://www-db.ics.uci.edu/pages/research/mars/  http://www-db.ics.uci.edu/pages/research/mars.shtml  http://www-db.ics.uci.edu/pages/research/mars/index.shtml  http://www.ics.uci.edu/~gbowker/converge.html  http://isg.ics.uci.edu/events.html | http://www.ics.uci.edu/~gmark/Home\_page/Research.html  **http://www.ics.uci.edu/~djp3/classes/2009\_01\_02\_INF141/**  http://www.ics.uci.edu/~djp3/classes/2008\_01\_01\_INF141/  http://www.ics.uci.edu/~djp3/classes/2010\_01\_CS221/index.html  **http://www.ics.uci.edu/~lopes/teaching/cs221W12/index.html** |