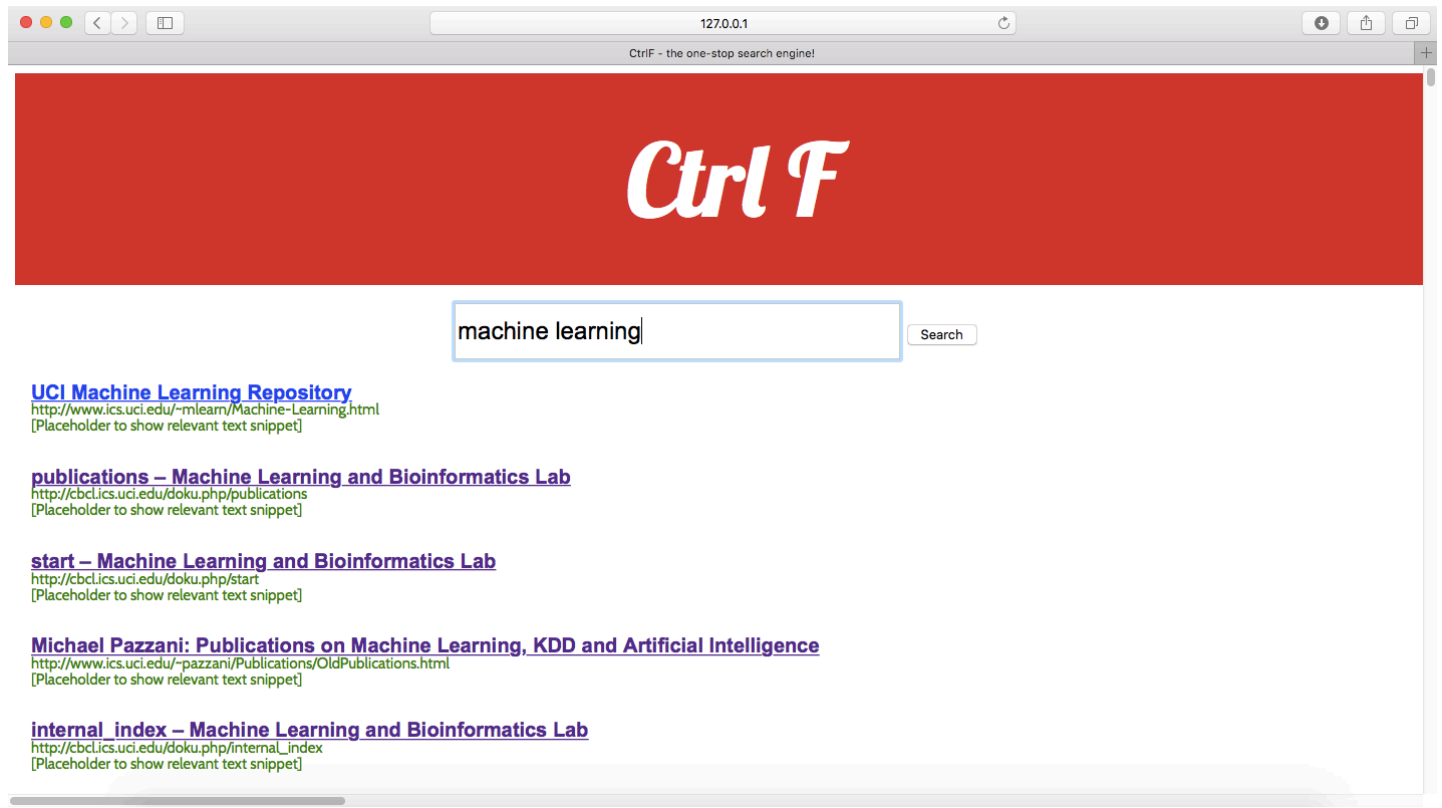


CS 221 – Project 3 Milestone 2

TEAM: Priyanka Ravi (33246700) and Rahul Sridhar (41608676)

APPROACH: As far as this milestone is concerned, we have primarily taken a four-pronged approach for ranking documents after tokenizing and stemming/lemmatizing the user's query – score documents based on the document url, title, each query term's tf-idf and finally PageRank. A Flask-based web interface serves as a conduit for delivering results. Included below are a snapshot of the interface and the top five search results for the queries listed in the problem statement.

INTERFACE:



QUERY RESULTS:

Query 1: mondego

1. <http://mondego.ics.uci.edu>
2. <http://mondego.ics.uci.edu/datasets?C=S;O=A>
3. <http://mondego.ics.uci.edu/datasets?C=M;O=A>
4. <http://mondego.ics.uci.edu/datasets?C=D;O=A>
5. <http://mondego.ics.uci.edu/datasets/wikipedia-events?C=M;O=A>

Query 2: machine learning

1. <http://www.ics.uci.edu/~mlearn/Machine-Learning.html>
2. <http://cbcl.ics.uci.edu/doku.php/publications>
3. <http://cbcl.ics.uci.edu/doku.php/start>
4. <http://www.ics.uci.edu/~pazzani/Publications/OldPublications.html>
5. http://cbcl.ics.uci.edu/doku.php/internal_index

1. http://www.ics.uci.edu/~ziv/ooad/intro_to_se/sld008.htm
2. http://www.ics.uci.edu/~ziv/ooad/intro_to_se/sld022.htm
3. http://www.ics.uci.edu/~icgse2016/4_01tutorial.html
4. http://www.ics.uci.edu/~ziv/ooad/intro_to_se/sld007.htm
5. http://www.ics.uci.edu/~ziv/ooad/intro_to_se/tsld008.htm

Query 4: security

1. <http://sli.ics.uci.edu/PmWiki/Security>
2. <http://www.ics.uci.edu/computing/linux/security.php>
3. <http://www.ics.uci.edu/computing/linux/file-security.php>
4. <http://fano.ics.uci.edu/cites/Location/Sequences-II-Communication-Security-and-Computer-Science.html>
5. <http://sli.ics.uci.edu/Category/Security>

Query 5: student affairs

1. <http://www.ics.uci.edu/prospective/en/contact/student-affairs>
2. http://www.ics.uci.edu/ugrad/sao/SAO_Events.php
3. <http://www.ics.uci.edu/prospective/en/contact/student-affairs/%e2%80%8b>
4. <http://www.ics.uci.edu/prospective/en/contact/student-affairs/contact/student-affairs>
5. <http://www.ics.uci.edu/prospective/en/contact/student-affairs/contact/student-affairs/%E2%80%8B>

Query 6: graduate courses

1. <http://www.ics.uci.edu/grad/courses/index.php>
2. <http://www.ics.uci.edu/grad/courses>
3. <http://www.ics.uci.edu/ugrad/courses/index>
4. <http://www.ics.uci.edu/grad/courses/listing.php>
5. <http://www.ics.uci.edu/grad/courses/index/faculty/highlights>

Query 7: Crista Lopes

1. <http://www.ics.uci.edu/~lopes>
2. <http://www.ics.uci.edu/~lopes/patents.html>
3. http://www.ics.uci.edu/community/news/view_news.php?id=1032
4. http://www.ics.uci.edu/community/news/view_news?id=1032
5. http://www.ics.uci.edu/community/news/view_news.php?id=1033

Query 8: REST

1. http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
2. <http://cbcl.ics.uci.edu/doku.php/publications>
3. <http://www.ics.uci.edu/~kay/courses/i41/answers.html>
4. <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
5. <http://www.ics.uci.edu/~fielding/pubs/dissertation/evaluation.htm>

Query 9: computer games

1. <http://www.ics.uci.edu/prospective/en/degrees/computer-game-science>
2. <http://cgvw.ics.uci.edu/oc-weekly-game-on>
3. <http://www.ics.uci.edu/>
4. <http://vcp.ics.uci.edu/content/exploring-potential-computer-games-decentralized-command-and-control>
5. <http://cgvw.ics.uci.edu/the-future-of-research-in-computer-games-and-virtual-worlds>

Query 10: information retrieval

1. <http://www.ics.uci.edu/~lopes/teaching/cs221W12/index.html>
2. <http://www.ics.uci.edu/~lopes/teaching/cs221W13/index.html>
3. <http://www.ics.uci.edu/~lopes/teaching/cs221W14/index.html>
4. <http://www.ics.uci.edu/~lopes/teaching/cs221W15/index.html>
5. <http://www.ics.uci.edu/~lopes/teaching/cs221W16/index.html>

ENHANCEMENTS TO BE PERFORMED:

Our next few steps with respect to ranking and retrieval are (in the order of priority):

1. Parameter tuning
2. Score documents based on other special tags (such as meta, anchor, h1, h2, etc.) and bigrams
3. Acronym expansion
4. Include relevant text snippets from the documents as part of the results
5. Latent Semantic Indexing
6. Enhance aesthetics of the interface,

And measure the impact of the above steps on NDCG@5 and improve the search engine's performance