#Tianya Chen 59359881, Lewis Liu 24427400, Kaiyi Ma 62838370, Jianyu Zheng 33062456

Search Engine

1-mondego
http://mondego.ics.uci.edu/datasets/wikipedia-events/files/?C=M%3BO%3DA
http://mondego.ics.uci.edu/datasets/?C=D%3BO%3DA
http://mondego.ics.uci.edu/datasets/?C=S%3BO%3DA
http://www.ics.uci.edu/~lopes/datasets/sourcerer-maven-aug12.html
http://www.ics.uci.edu/~lopes/datasets/index.html

2-machine learning
http://www.ics.uci.edu/~pazzani/Publications/APubs.html
http://www.ics.uci.edu/~qliu1/MLcrowd_ICML_workshop/index.html
http://www.ics.uci.edu/~pazzani/Research.html
http://www.ics.uci.edu/~rickl/rickl-patent-5526281.html
http://sli.ics.uci.edu/Classes/2012F-178

3-software engineering
http://www.ics.uci.edu/~emilyo/SimSE/publications.html
http://www.ics.uci.edu/~taylor/ICS221/221_FQ_02.html
http://www.ics.uci.edu/~taylor/classes/211/syllabusFQ06.html
http://www.ics.uci.edu/~taylor/ICS223/syllabus.html
http://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_cv_2000.htm

4-security
http://www.ics.uci.edu/~gts/pubs.html
http://drzaius.ics.uci.edu/~swirl/impromptu-0.30/apidocs/index-all.html
http://www.ics.uci.edu/~kobsa/privacy/israel.htm
http://www.ics.uci.edu/~ics54/w00/doc/security/pkhistory.html
http://www.ics.uci.edu/~goodrich/pubs/index.html

5-student affairs
http://www.ics.uci.edu/ugrad/qa/index.php
http://www.ics.uci.edu/~kobsa/privacy/israel.htm
http://www.ics.uci.edu/grad/sao/
http://www.ics.uci.edu/ugrad/policies/index.php/index.php
http://www.ics.uci.edu/ugrad/degrees/degree_ics.php

6-graduate courses
http://www.ics.uci.edu/ugrad/qa/index.php

http://www.ics.uci.edu/grad/qa/
http://www.ics.uci.edu/grad/sao/
http://www.ics.uci.edu/grad/degrees/degree_cs.php
http://www.ics.uci.edu/ugrad/degrees/degree_ics.php

7-informatics
http://www.ics.uci.edu/faculty
http://www.ics.uci.edu/ugrad/qa/index.php
http://www.ics.uci.edu/~kay/courses/131/su13.html
http://www.ics.uci.edu/~kay/courses/131/su11.html
http://www.ics.uci.edu/faculty/index.php?department=Computer%20Science

8-REST
http://www.ics.uci.edu/~kay/courses/i41/answers.html
http://www.ics.uci.edu/~thornton/ics184/MidtermSolutions.html
http://vcp.ics.uci.edu/content/extending-representational-state-transfer-rest-architectural-style-decentralized-systems-0
http://www.ics.uci.edu/~kay/courses/141/hw/hw1.html
http://www.ics.uci.edu/~eppstein/pix/sunsets/ontheroad-bra.html

9-computer games
http://www.ics.uci.edu/~magda/cs620/announceOG.html
http://www.ics.uci.edu/~eppstein/cgt/bib.html
http://www.ics.uci.edu/community/news/articles/view_article?id=77
http://www.ics.uci.edu/~eppstein/cgt/sylver.html
http://cgvw.ics.uci.edu/affiliated-faculty/

10-information retrieval
http://www-db.ics.uci.edu/pages/research/mars/index.shtml
http://www.ics.uci.edu/~lopes/teaching/cs221W12/
http://www.ics.uci.edu/~djp3/classes/2008_01_01_INF141/calendar.html
http://www.ics.uci.edu/~ejw/authoring/www6/tsld030.htm
http://www.ics.uci.edu/~djp3/classes/2008_01_01_INF141/materials.html

Extra Credit 1: Improved Term-Document Scoring

i)

1. Improvement1: At first, we tried to calculate scores for documents by simply adding up tf-idf scores for terms appeared in query. And then we normalized documents' scores by dividing their norm (the square root of so-call magnitude in our code; Please refer to function calculate_tfidf_original). Then we implemented efficient cosine ranking. By using that way we can efficiently retrieve the documents which have high similarity to query (Please refer to

function calculate_doc_score).

2. Improvement2: The first time when we were calculating term frequency, we just count one if a term appears once in a document. Then we treated terms in title and body differently. If the term appears in title once, we count 2.5, and if the term appears in body, we count 1. So we give more weight to the terms appearing in the title. It turned out that this method worked pretty well in the ics.uci.edu domain.

ii)

1. These two scores are not on the same scale, but the ranks for documents changed.
   Original scores for top five (didn't adjust score for terms in title & didn't use cosine similarity)

   Example format: (docID, score) url

   ('8884', 8.71215684281459)
   http://www.ics.uci.edu/~pazzani/Publications/APubs.html

   ('6473', 7.058056835015305)
   http://www.ics.uci.edu/~pazzani/Research.html

   ('29030', 6.905039287505364)
   http://isg.ics.uci.edu/events.html

   ('3083', 6.86461169602303)
   http://www.ics.uci.edu/~interfac/all-sessions.html

   ('21896', 6.138426915511356)
   http://sli.ics.uci.edu/Classes/2010W-178

   Following are the improved scores for top five (give for weight to terms in title & use cosine similarity)

   ('8884', 274.71883496443786)
   http://www.ics.uci.edu/~pazzani/Publications/APubs.html

   ('30615', 235.9828634243867)
   http://www.ics.uci.edu/~qliu1/MLcrowd_ICML_workshop/index.html

   ('6473', 217.94006708632867)
   http://www.ics.uci.edu/~pazzani/Research.html

   ('15702', 204.0650101069769)

http://www.ics.uci.edu/~rickl/rickl-patent-5526281.html

('29364', 201.01335036074227)
http://sli.ics.uci.edu/Classes/2012F-178

Now, documents having query terms in title and having high cosine similarity to query rank higher than other documents.

2. For query “machine learning”
   Original:
   http://www.ics.uci.edu/~pazzani/Publications/APubs.html
   http://www.ics.uci.edu/~pazzani/Research.html
   http://isg.ics.uci.edu/events.html
   http://www.ics.uci.edu/~interfac/all-sessions.html
   http://sli.ics.uci.edu/Classes/2010W-178
   http://www.ics.uci.edu/~smyth/courses/cs274
   http://sli.ics.uci.edu/pmwiki/pmwiki.php?n=Classes.Classes
   http://computableplant.ics.uci.edu/papers
   http://cml.ics.uci.edu/?page=events&subPage=dss_schedule_0607
   http://www.ics.uci.edu/~pattis/ICS-33/lectures/overview/lecture.html
   Improved:
   http://www.ics.uci.edu/~pazzani/Publications/APubs.html
   http://www.ics.uci.edu/~qliu1/MLcrowd_ICML_workshop/index.html
   http://www.ics.uci.edu/~pazzani/Research.html
   http://www.ics.uci.edu/~rickl/rickl-patent-5526281.html
   http://sli.ics.uci.edu/Classes/2012F-178
   http://www.ics.uci.edu/~dramanan
   http://www.ics.uci.edu/~qliu1/nips13_workshop/index.html
   http://www.ics.uci.edu/~rickl/courses/cs-171/2014-wq-cs171/CS-171-WQ-2014.htm
   http://sli.ics.uci.edu/pmwiki/pmwiki.php?n=Classes.Classes
   http://isg.ics.uci.edu/events.html

Extra Credit 2: Web UI

Show at demo.