

Old Dominion University
Department of Computer Science
CS834: Introduction to Information Retrieval
Fall 2017
Assignment 5
Professor: Dr. Michael Nelson

Hussam Hallak
CS Master's Student

December 16, 2017

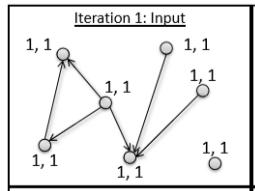
Question 1:

Exercise 10.3: Compute five iterations of HITS (see Algorithm 3) and PageRank (see Figure 4.11) on the graph in Figure 10.3. Discuss how the PageRank scores compare to the hub and authority scores produced by HITS

Answer:

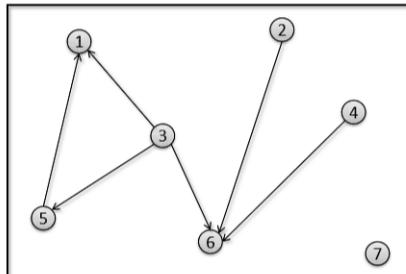
The graph below was copied from the textbook Figure 10.3.

Figure 1: Copied from the textbook Figure 10.3



I used the same node numbers in the textbook. The graph below was copied from the textbook Figure 10.4.

Figure 2: Copied from the textbook Figure 10.4



I wrote a python script and named it hp.py to calculate HITS and PageRank scores using Link Analysis.

```
1 import sys
2 import os
3 import networkx as nx
4
5 def pagerank(G, alpha=0.85, personalization=None, iterations_count=11, nstart=None, weight='weight',
6     dangling=None):
7     if len(G) == 0:
8         return {}
9
10    if not G.is_directed():
11        D = G.to_directed()
12    else:
13        D = G
14    W = nx.stochastic_graph(D, weight=weight)
15    N = W.number_of_nodes()
16    if nstart is None:
17        x = dict.fromkeys(W, 1.0 / N)
18    else:
19        s = float(sum(nstart.values()))
20        x = dict((k, v / s) for k, v in nstart.items())
21
22    if personalization is None:
23        p = dict.fromkeys(W, 1.0 / N)
```

```

23     else:
24         missing = set(G) - set(personalization)
25         if missing:
26             raise nx.NetworkXError('Personalization dictionary ,
27                                     must have a value for every node. ,
28                                     Missing nodes %s' % missing)
29         s = float(sum(personalization.values()))
30         p = dict((k, v / s) for k, v in personalization.items())
31
32     if dangling is None:
33         dangling_weights = p
34     else:
35         missing = set(G) - set(dangling)
36         if missing:
37             raise nx.NetworkXError('Dangling node dictionary ,
38                                     must have a value for every node. ,
39                                     Missing nodes %s' % missing)
40         s = float(sum(dangling.values()))
41         dangling_weights = dict((k, v/s) for k, v in dangling.items())
42         dangling_nodes = [n for n in W if W.out_degree(n, weight=weight) == 0.0]
43
44     for _ in range(iterations_count):
45         xlast = x
46         x = dict.fromkeys(xlast.keys(), 0)
47         danglesum = alpha * sum(xlast[n] for n in dangling_nodes)
48         for n in x:
49             for nbr in W[n]:
50                 x[nbr] += alpha * xlast[n] * W[n][nbr][weight]
51                 x[n] += danglesum * dangling_weights[n] + (1.0 - alpha) * p[n]
52
53     return x
54
55 def hits(G, iterations_count=11, nstart=None, normalized=True):
56     if type(G) == nx.MultiGraph or type(G) == nx.MultiDiGraph:
57         raise Exception("hits() not defined for graphs with multiedges.")
58     if len(G) == 0:
59         return {}, {}
60     if nstart is None:
61         h=dict.fromkeys(G,1.0/G.number_of_nodes())
62     else:
63         h=nstart
64         s=1.0/sum(h.values())
65         for k in h:
66             h[k]*=s
67     i=0
68     while True:
69         if i >= iterations_count: break
70
71         hlast=h
72         h=dict.fromkeys(hlast.keys(),0)
73         a=dict.fromkeys(hlast.keys(),0)
74         for n in h:
75             for nbr in G[n]:
76                 a[nbr]+=hlast[n]*G[n][nbr].get('weight',1)
77         for n in h:
78             for nbr in G[n]:
79                 h[n]+=a[nbr]*G[n][nbr].get('weight',1)
80         s=1.0/max(h.values())
81         for n in h: h[n]*=s
82         s=1.0/max(a.values())
83         for n in a: a[n]*=s
84
85         i+=1
86     if normalized:
87         s = 1.0/sum(a.values())
88         for n in a:
89             a[n] *= s
90         s = 1.0/sum(h.values())
91         for n in h:
92             h[n] *= s

```

```

93     return h,a
94
95 if __name__ == '__main__':
96     iterations_count = 5
97     G = nx.Graph()
98     G.add_nodes_from(range(1,8))
99     G.add_edges_from([(1,2), (3,1), (3,2), (3,4), (5,4), (6,4)])
100
101    pr = pagerank(G, iterations_count=iterations_count)
102
103   print 'Pagerank Algorithm ({} iterations)'.format(iterations_count)
104   print '====='
105   print 'Pagerank values = {}'.format(pr)
106
107   h, a = hits(G, iterations_count=iterations_count)
108
109   print 'HITS Algorithm ({} iterations)'.format(iterations_count)
110   print '====='
111   print 'Hubs values = {}'.format(h)
112   print 'Authorities values = {}'.format(a)
113   print ''

```

Listing 1: The content of hp.py

```

1 hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q1$ python hp.py
2 Pagerank Algorithm (5 iterations)
3 =====
4 Pagerank values = {1: 0.15415210590313183, 2: 0.15415210590313183, 3: 0.2151836555953412, 4:
   0.2730700543561289, 5: 0.08952435340504106, 6: 0.08952435340504106, 7: 0.024393371432183873}
5 HITS Algorithm (5 iterations)
6 =====
7 Hubs values = {1: 0.19870751068593528, 2: 0.19870751068593528, 3: 0.258854060655404, 4: 0.1789639731325056,
   5: 0.0823834724201099, 6: 0.0823834724201099, 7: 0.0}
8 Authorities values = {1: 0.20153417015341699, 2: 0.20153417015341699, 3: 0.25232450023245, 4:
   0.18816829381682937, 5: 0.07821943282194328, 6: 0.07821943282194328, 7: 0.0}
9
10 hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q1$
```

Listing 2: Running hp.py

Results

Pagerank Algorithm (5 iterations):

Pagerank values:

1: 0.15415210590313183
 2: 0.15415210590313183
 3: 0.2151836555953412
 4: 0.2730700543561289
 5: 0.08952435340504106
 6: 0.08952435340504106
 7: 0.024393371432183873

HITS Algorithm (5 iterations):

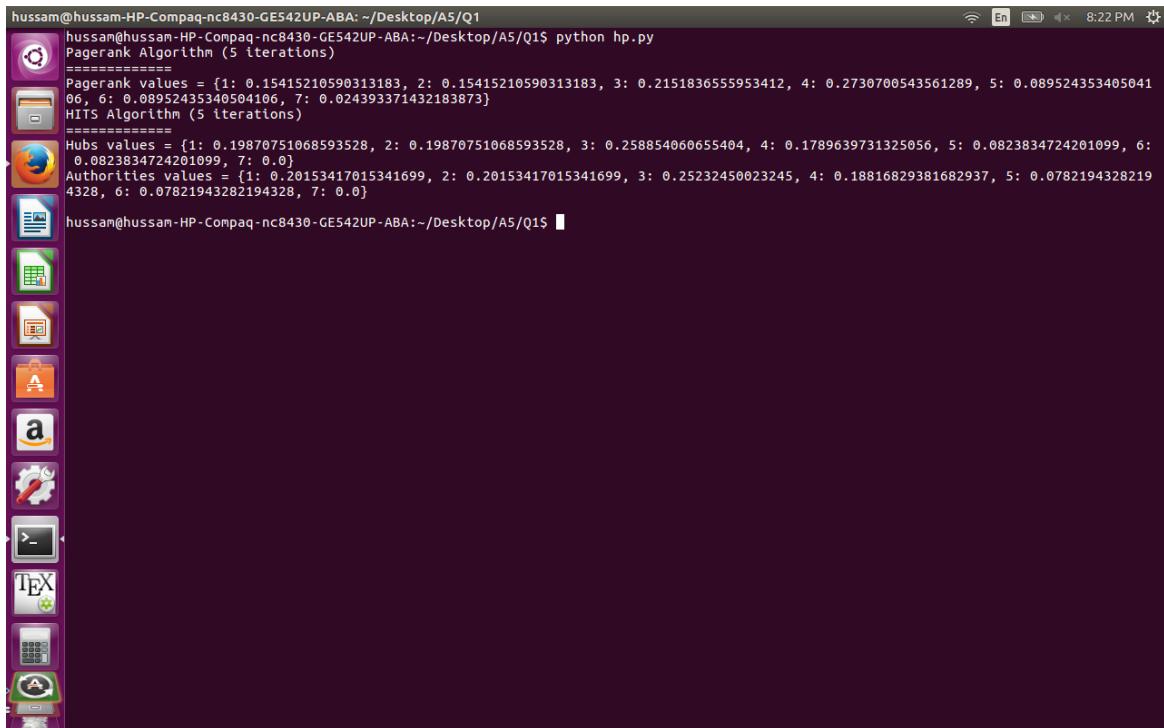
Hubs values:

1: 0.19870751068593528
 2: 0.19870751068593528
 3: 0.258854060655404
 4: 0.1789639731325056
 5: 0.0823834724201099
 6: 0.0823834724201099
 7: 0.0

Authorities values:

```
1: 0.20153417015341699
2: 0.20153417015341699
3: 0.25232450023245
4: 0.18816829381682937
5: 0.07821943282194328
6: 0.07821943282194328
7: 0.0
```

Figure 3: Running hp.py



```
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q1
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q1$ python hp.py
=====
PageRank Algorithm (5 iterations)
=====
PageRank values = {1: 0.15415210590313183, 2: 0.15415210590313183, 3: 0.2151836555953412, 4: 0.2730700543561289, 5: 0.089524353405041
6, 6: 0.08952435340504106, 7: 0.024393371432183873}
HITS Algorithm (5 iterations)
=====
Hubs values = {1: 0.19870751068593528, 2: 0.19870751068593528, 3: 0.258854060655404, 4: 0.1789639731325056, 5: 0.0823834724201099, 6:
0.0823834724201099, 7: 0.0}
Authorities values = {1: 0.20153417015341699, 2: 0.20153417015341699, 3: 0.25232450023245, 4: 0.18816829381682937, 5: 0.0782194328219
4328, 6: 0.07821943282194328, 7: 0.0}
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q1$
```

Observation:

Right off the bat, it is clear that hubs and authorities ranking is consistent for all nodes. Node 4 placed first in PageRank, but placed fourth in authorities score.

Node 3 placed second in PageRank, but placed first in authorities score.

Nodes 1 and 2 come after node 3 for both authorities score and PageRank.

Nodes 5 and 6 come after 1 and 2 for both authorities score and PageRank.

There is an obvious difference in ranking for both algorithms, but this could be due to the low number of iterations.

Question 2:

Exercise 10.5:

Find a community-based question answering site on the Web and ask two questions, one that is low-quality and one that is high-quality. Describe the answer quality of each question.

Answer:

I chose to use Yahoo! Answers and asked the following question as a high quality one:

Will Turkey leave NATO and form a coalition with Russia and Iran after US backed Syrian Kurds and moved US embassy in Tel Aviv to Jerusalem?

I consider this as a high quality question because of its correct grammar and punctuation. It also targets the kind of users that care about politics who are normally serious people. The question is about the possibility of a certain outcome as a result of current events. The answer should be definitive, either yes or no.

Below is a screen shot of the answers I got:

Figure 4: Answers for high quality question

The screenshot shows a search interface for 'YAHOO! ANSWERS' with a search bar and buttons for 'Search Answers' and 'Search Web'. Below the search area, there are four user posts:

- Best Answer:** A user named Gerald posted an answer 5 hours ago. The answer is: "its probably the idea behind Trumps statement any war 10000 miles away from America is always a money maker Korea has a nuke ouch so lets pick on the Palestinians with Israel". It has 1 like and 0 dislikes. A 'Comment' link is available.
- Asker's rating:** *****
- Get DIRECTV:** An advertisement for DIRECTV with the text: "Brave. Bold. Brilliant. Switch to DIRECTV." and "DIRECTV Sponsored".
- I hope so.**: A user named Flower posted an answer 4 hours ago. The answer is: "I hope so.". It has 0 likes and 0 dislikes. A 'Comment' link is available.
- No**: A user named Don posted an answer 5 hours ago. The answer is: "No". It has 1 like and 1 dislike. A 'Comment' link is available.
- One can dream.**: A user named Foxhole posted an answer 5 hours ago. The answer is: "One can dream.". A 'Comment' link is available.

I marked the answer that has a short analysis of why Turkey might leave NATO as best answer. The answers are high quality and meant to answer the question, but they are short, which is expected.

I asked the following question as a low quality one:

Why did you Trump to helb tha kill Saudi to Yemen? Are you forget wat hapend at 11 septemper? Saudi mans kil 3000 America at airplain atak.?

I discovered that asking a low quality question is not an easy task. I was finally able to do it by writing a question then injecting/removing pronouns, verbs, and preposition to produce wrong grammar. I also misspelled some words to make it worse.

Below is a screen shot of the answers I got:

Figure 5: Answers for low quality question

The screenshot shows a search results page for "Why did you Trump to helb tha kill Saudi to Yemen? Are you forget wat hapend at 11 septemer? Saudi mans kil 3000 America at airplain atak.?".

Answers

Flower · 5 hours ago
Source(s): <http://www.bbc.com/news/world-middle-eas...>

Asker's rating *****

Oompa Loompah · 5 hours ago

yes, this is mexican football

Comment

Add your answer

Surprisingly, I got an answer that is, to a certain extent, related to the topic of the question. I would call that a high quality answer because I do not think that it is possible to come up with a much better answer. The answer includes a source link from BBC.

For the other answer, I thought that was funny, but it turns out that Yahoo! put my question in the Mexican football category within sports, and that explains why I got this answer. It is, of course, a low quality answer, and it is definitely not related to the question.

Observation:

The results were as expected; low quality questions get low quality answers, and high quality questions get high quality answers.

Question 3:

Exercise 10.6:

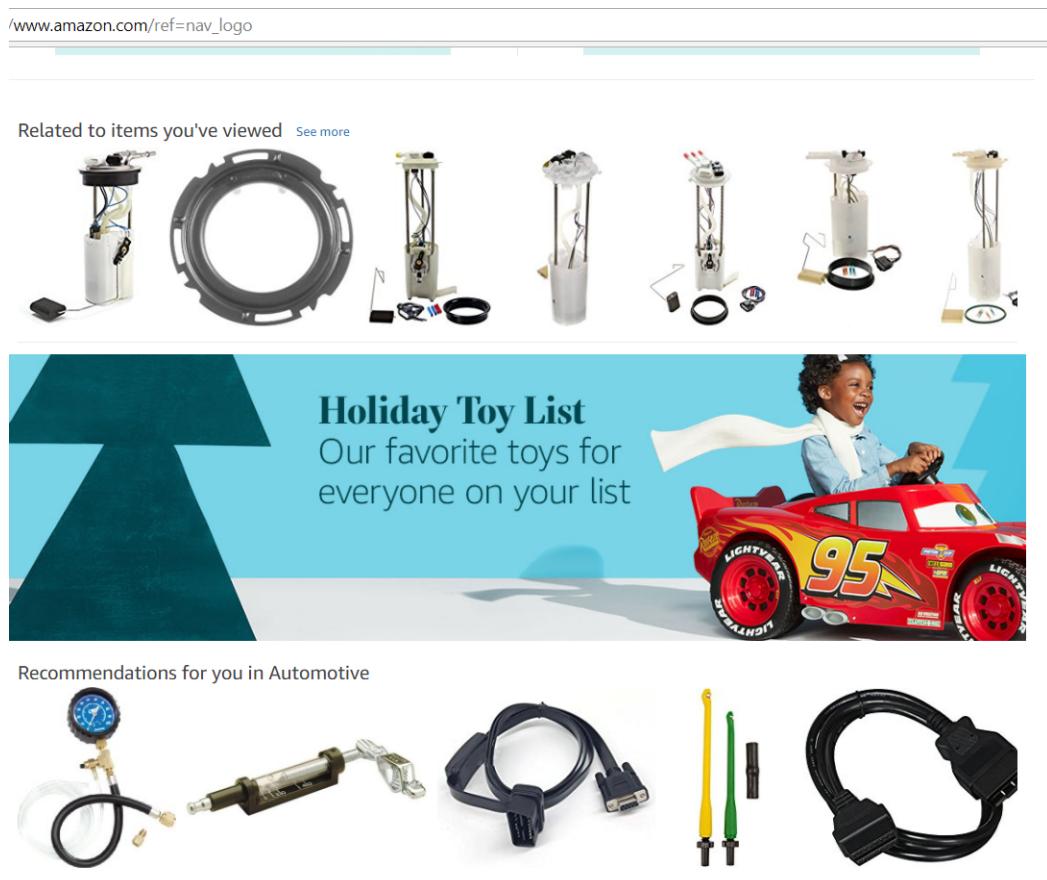
Find two examples of document filtering systems on the Web. How do they build a profile for your information need? Is the system static or adaptive?

Answer:

For my two examples I have chosen Amazon and Youtube. Both of them use an adaptive system for making recommendations.

Amazon is an e-commerce company. I use it sometimes to purchase different items including car parts, electronics, and other items. The screen shot below shows recommended items for me based on items that I viewed, fuel pumps, and items I purchased recently, a cable to connect my truck to my laptop and perform different diagnosis and scanning. Other recommended items are products that were viewed or bought by people whose purchase history is similar to mine. Amazon does such a good job as far as recommendations. I would buy all the recommended items if I had enough money to splurge.

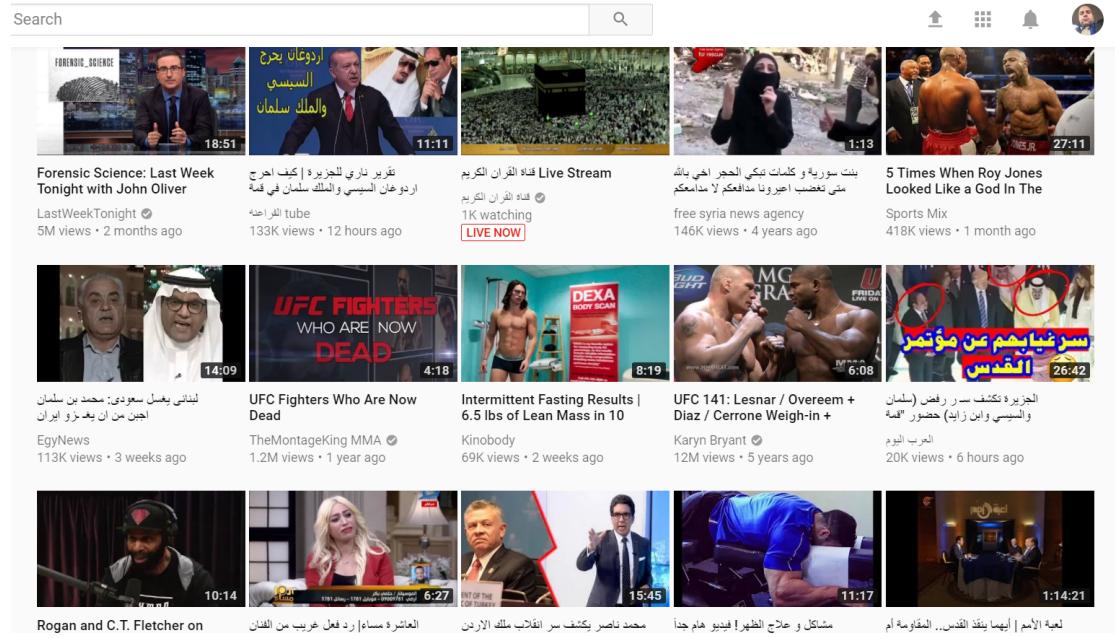
Figure 6: Recommended items by Amazon



Youtube is my favorite website. The reason is that I use it for everything including entertainment, news, music, nutrition, and fitness videos. I also watch short lectures or tutorials about concepts I do not understand in Computer Science. Youtube recommends videos for me based on other videos I watched. It also recommends channels/users for me to subscribe to based on videos I watched or other channels I am subscribed to. It also shows popular uploads from channels I am subscribed to.

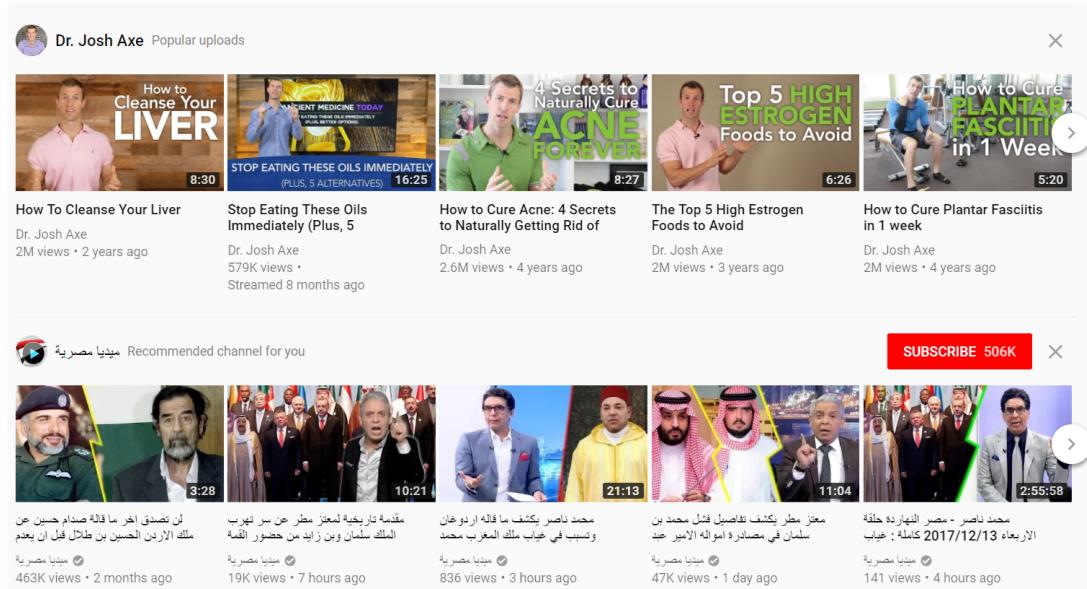
The screen shot below shows a list of recommended videos, based on videos I watched recently. The recommended videos belong to different categories including entertainment, news, music, fitness, boxing, UFC, ...etc.

Figure 7: Recommended items by Youtube



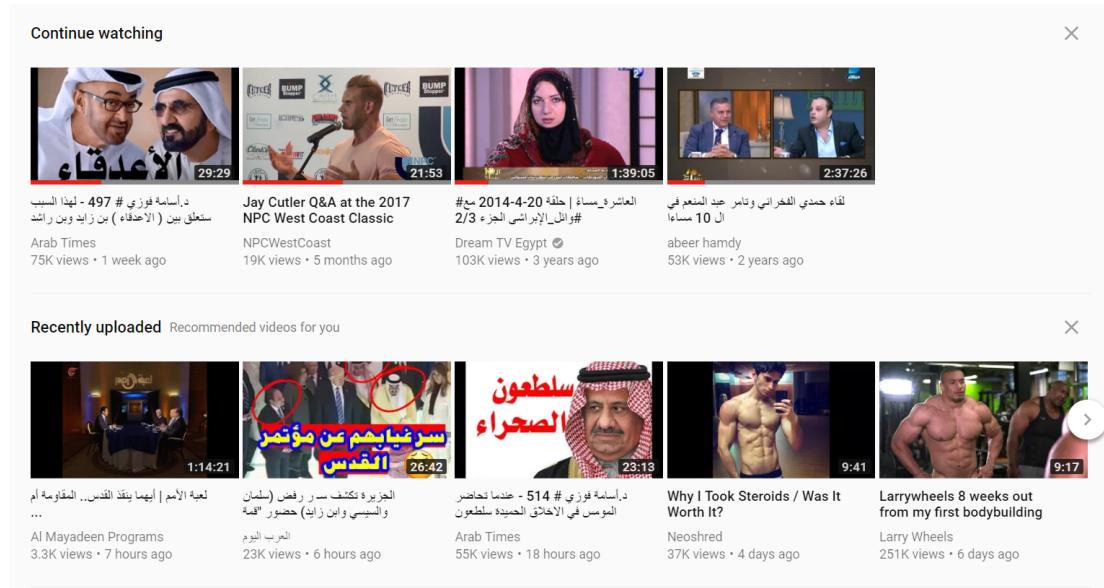
The screen shot below shows a list of popular uploads from a nutrition channel I am subscribed to, Dr. Josh Axe, as well as a recommended channel for my to subscribe to based on videos I watched and channels I am subscribed to. The recommended channel is an Arabic media channel that posts news from the middle east and other videos about politics.

Figure 8: Recommended items by Youtube



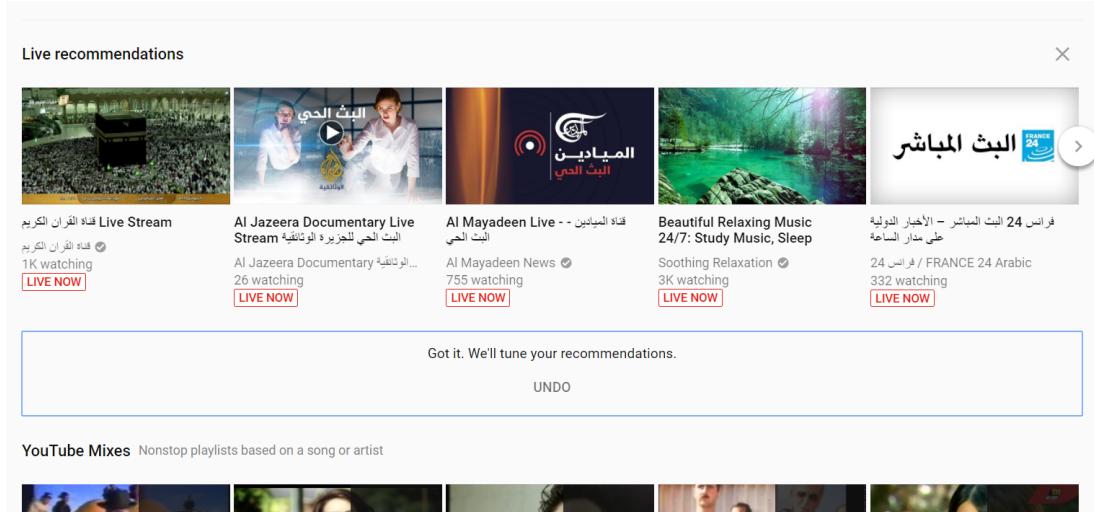
Youtube also shows a list of videos that I did not complete in case I wanted to continue to watch them. It also shows a list of recently uploaded videos that belong to categories in which I am interested based on my views history as the screen shot below shows.

Figure 9: Recommended items by Youtube



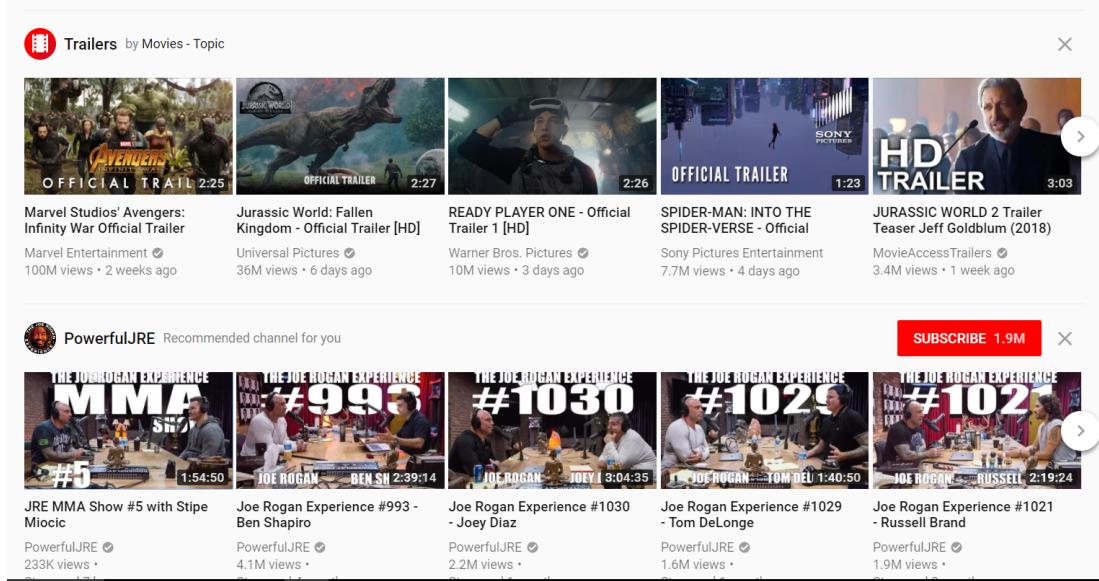
Youtube shows a list of live recommendations based on live channels I watch, mostly news. It also shows recommended a nonstop playlists ,Youtube mixes, based on a song or an artist from my views history. I turned off recommended videos from one of the channels I am not interested in, RT Arabic. Youtube says: Got it. We'll tune your recommendations. It gave me the option to undo in case my cat turned that recommendation off for me. The screen shot below shows what I described.

Figure 10: Recommended items by Youtube



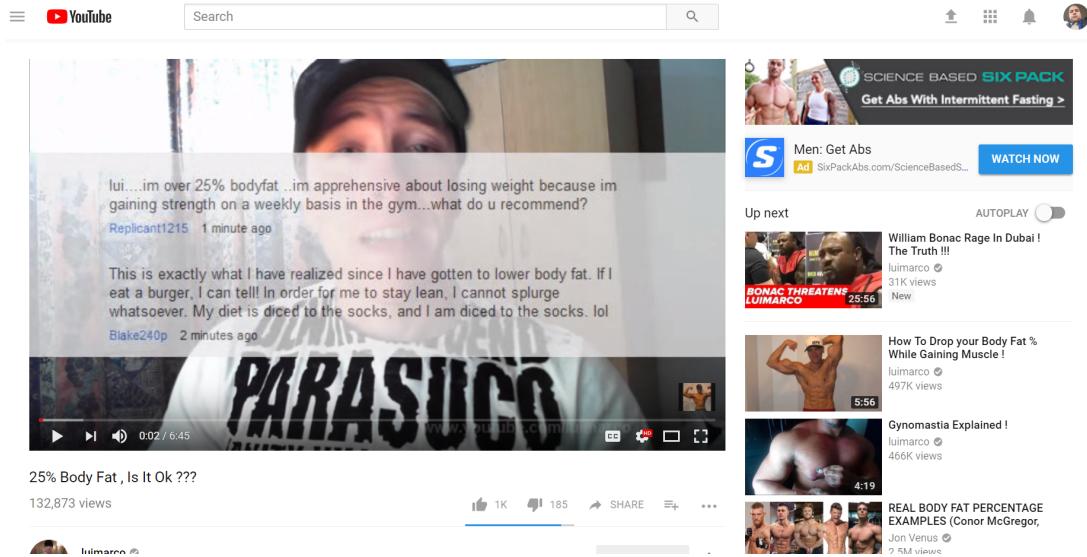
So far, it seems like Youtube is recommending the right videos for me, however, it also shows videos I am not interested in. This could be because these videos have so many views, millions, as seen in the next screen shot. I am not interested in Trailers, but Youtube shows it anyways.

Figure 11: Recommended items by Youtube



Finally, Youtube places an ad, sponsored video, in the upper right corner on video pages I watch. The users who made these videos pay Youtube to show these videos to users based on the video that is on the same page. I have AdBlock installed on my browser so I do not have to see all kinds of advertisements. I disabled AdBlock to take the screen shot below.

Figure 12: Recommended items by Youtube



Question 4:

Exercise 10.8:

Implement the nearest neighbor-based collaborative filtering algorithm. Using a publicly available collaborative filtering data set, compare the effectiveness, in terms of mean squared error, of the Euclidean distance and correlation similarity.

Answer:

I downloaded the latest MovieLens small data set, ml-latest-small, from this link:

<http://files.grouplens.org/datasets/movielens/ml-latest-small.zip>

This data set (ml-latest-small) describes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service. It contains 100004 ratings and 1296 tag applications across 9125 movies.

These data were created by 671 users between January 09, 1995 and October 16, 2016. This dataset was generated on October 17, 2016.

Users were selected at random for inclusion. All selected users had rated at least 20 movies. No demographic information is included. Each user is represented by an id, and no other information is provided.

The data are contained in the files links.csv, movies.csv, ratings.csv and tags.csv under Q4 directory.

I implemented the nearest neighbor-based collaborative filtering algorithm in python and saved it in the file nn.py. It takes one command line argument, which is the ratings file in csv format. The program computes MSE of Euclidean distance and correlation similarity.

```
1 import sys
2 import io
3 from numpy import *
4 from sklearn.linear_model import *
5 from sklearn.metrics import *
6 from sklearn.neighbors import *
7
8 if len(sys.argv) != 2:
9     print "Usage: python nn.py <ratings_csv_file>"
10    print "Example: python nn.py ratings.csv"
11    exit()
12
13 ratings = genfromtxt(sys.argv[1], delimiter=',', skip_header=1, usecols=range(3))
14 random.shuffle(ratings)
15 train, test = ratings[:len(ratings)/10, :], ratings[:len(ratings)/10, :]
16 neighbors = KNeighborsClassifier(n_neighbors=1)
17 neighbors.fit(train, train[:, 0])
18 distances, neighbors_indices = neighbors.kneighbors(test)
19 euclidean_neighbors = array([train[ns[0]] for ns in neighbors_indices])
20 linear_reg = LinearRegression()
21 linear_reg.fit(train, train[:, 0])
22 neighbors_idx = linear_reg.predict(test)
23 pearson_neighbors = array([train[int(idx)] for idx in neighbors_idx])
24 y_true = []
25 for user, movie, rating in test:
26     y_true.append(rating)
27 y_pred_euc = []
28 for user, movie, rating in euclidean_neighbors:
29     y_pred_euc.append(rating)
30 y_pred_pear = []
31 for user, movie, rating in pearson_neighbors:
32     y_pred_pear.append(rating)
33
34 print 'Euclidean Distance MSE:'
35 print mean_squared_error(y_true, y_pred_euc)
```

```
36 | print '-----',
37 | print 'Correlation Similarity MSE:'
38 | print mean_squared_error(y_true, y_pred_pear)
```

Listing 3: The content of nn.py

Results:

Figure 13: Running nn.py

```
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA: ~/Desktop/A5/Q4
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~$ cd Desktop/A5/Q4
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q4$ python nn.py ratings
.csv
Euclidean Distance MSE:
1.39875
-----
Correlation Similarity MSE:
2.21875
hussam@hussam-HP-Compaq-nc8430-GE542UP-ABA:~/Desktop/A5/Q4$ █
```

Euclidean Distance MSE = 1.39875

Correlation Similarity MSE = 2.21875

Observation:

The MSE of Euclidean Distance is smaller than the MSE of the Correlation Similarity. Therefore the Euclidean Distance is more effective than Correlation Similarity.

Question 5:

Exercise 11.5: How many papers dealing with term dependency can you find in the SIGIR proceedings since 2000? List their citations.

Answer:

To answer this question, I ran the query “term dependency source:SIGIR” in Google scholar.

Figure 14: Results returned by Google scholar for the query “term dependency source:SIGIR”

The screenshot shows a Google Scholar search results page. The search query is "term dependency source:SIGIR". The results are filtered to show "Articles" and there are approximately 319 results found in 0.03 seconds. The results are listed in descending order of relevance. Each result includes the title, author(s), publication year, abstract, and citation information. Some results have PDF links available. The interface includes navigation buttons for previous and next pages, and a sidebar with filtering options like "Any time", "Sort by relevance", and "Include patents/citations".

Result Number	Title	Author(s)	Year	Abstract/Citation Info
1	Incorporating term dependency in the DFR framework	J Peng, C Macdonald, B He, V Plachouras	2007	Abstract: Term dependency, or co-occurrence, has been studied in language modelling, for instance by Metzler & Croft who showed that retrieval performance could be significantly enhanced using term dependency information. In this work, we show how term dependency can be used in a term dependency framework. We propose a new framework called DFR (Dependency Feature Representation) which uses term dependency information to improve the performance of term ranking. The proposed framework is based on a feature representation of terms which takes into account their dependencies. The proposed framework is evaluated on a standard test collection and compared with other term dependency frameworks. The results show that the proposed framework outperforms other term dependency frameworks.
2	Query term ranking based on dependency parsing of verbose queries	JH Park, WB Croft	2010	Abstract: Query term ranking approaches are used to select effective terms from a verbose query by ranking items. Features used for query term ranking and selection in previous work do not consider grammatical relationships between terms. To address this issue, we use a dependency parser to parse a query and extract its dependency structure. Then, we use the dependency structure to rank the query terms. Finally, we select the top-k terms from the ranked list. The experimental results show that our approach outperforms state-of-the-art query term ranking approaches.
3	Score-safe term-dependency processing with hybrid indexes	M Petri, A Moffat, JS Culpepper	2014	Abstract: Score-safe index processing has received a great deal of attention over the last two decades. By pre-calculating maximum term impacts during indexing, the number of scoring operations can be minimized, and the top-k documents for a query can be located efficiently. However, this approach is not suitable for processing terms with high impact values. In this paper, we propose a hybrid approach that combines score-safe processing with a fast search algorithm for terms with high impact values. The proposed approach is called score-safe term-dependency processing with hybrid indexes (SDPHI). The experimental results show that SDPHI is faster than state-of-the-art score-safe processing approaches.
4	Random walk term weighting for information retrieval	R Blanco, C Llorente	2007	Abstract: Random walk term weighting is a new term weighting scheme for information retrieval. It is based on a random walk process that starts at a query term and moves to other terms in the document. The weight of a term is proportional to the probability of reaching it. This approach is able to capture the context of a term and its relationships with other terms in the document. The experimental results show that random walk term weighting outperforms state-of-the-art term weighting schemes.
5	A comparison of various approaches for using probabilistic dependencies in language modeling	Perez	2003	Abstract: Probabilistic dependencies have been used in language modeling for several years. In this paper, we compare different approaches for using probabilistic dependencies in language modeling. We evaluate the performance of these approaches on several datasets. The results show that the proposed approach outperforms state-of-the-art language modeling approaches.
6	Cited by 56			
7				
8				
9	Query term ranking based on dependency parsing of verbose queries	JH Park, WB Croft	2010	Abstract: Query term ranking approaches are used to select effective terms from a verbose query by ranking items. Features used for query term ranking and selection in previous work do not consider grammatical relationships between terms. To address this issue, we use a dependency parser to parse a query and extract its dependency structure. Then, we use the dependency structure to rank the query terms. Finally, we select the top-k terms from the ranked list. The experimental results show that our approach outperforms state-of-the-art query term ranking approaches.
10	Score-safe term-dependency processing with hybrid indexes	M Petri, A Moffat, JS Culpepper	2014	Abstract: Score-safe index processing has received a great deal of attention over the last two decades. By pre-calculating maximum term impacts during indexing, the number of scoring operations can be minimized, and the top-k documents for a query can be located efficiently. However, this approach is not suitable for processing terms with high impact values. In this paper, we propose a hybrid approach that combines score-safe processing with a fast search algorithm for terms with high impact values. The proposed approach is called score-safe term-dependency processing with hybrid indexes (SDPHI). The experimental results show that SDPHI is faster than state-of-the-art score-safe processing approaches.
11				
12				
13				
14	Cited by 28			
15				
16				
17				
18				
19				

The query returned 319 results, however, some of the results are papers dealing with term dependency prior to 2000. I tried to scrape the results page, but I ran into problems and was running out of time. I found a python script that scrapes Google Scholar results page, but the script did not return any results, nor did it return any errors. I abandoned the idea when I realized that I only have one hour to finish this assignment and submit it on time.

I filtered the results and cleaned them up by hand and saved them in the file `sigir2kplus.txt`. Below is the content of the file:

```
1 Incorporating term dependency in the DFR framework
2 J Peng, C Macdonald, B He, V Plachouras - international ACM SIGIR, 2007 - dl.acm.org
3
4
5
6 Cited by 56
7
8
9 Query term ranking based on dependency parsing of verbose queries
10 JH Park, WB Croft - Proceedings of the 33rd international ACM SIGIR, 2010 - dl.acm.org
11
12
13
14 Cited by 28
15
16
17 Score-safe term-dependency processing with hybrid indexes
18 M Petri, A Moffat, JS Culpepper - of the 37th international ACM SIGIR, 2014 - dl.acm.org
19
```

20 Cited by 10
21
22
23
24 Random walk term weighting for information retrieval
25 R Blanco, C Lioma – of the 30th annual international ACM SIGIR ª, 2007 – dl.acm.org
26
27
28
29 Cited by 35
30
31
32 A comparison of various approaches for using probabilistic dependencies in language modeling
33 P Bruza, D Song – of the 26th annual international ACM SIGIR ª, 2003 – dl.acm.org
34
35
36
37 Cited by 16
38
39
40 Modeling higher-order term dependencies in information retrieval using query hypergraphs
41 M Bendersky, WB Croft – of the 35th international ACM SIGIR conference ª, 2012 – dl.acm.org
42
43
44
45 Cited by 54
46
47
48 Dependence language model for information retrieval
49 J Gao, JY Nie, G Wu, G Cao – of the 27th annual international ACM SIGIR ª, 2004 – dl.acm.org
50
51
52
53 Cited by 294
54
55
56 Refining term weights of documents using term dependencies
57 H Kim, I Choi, M Kim – of the 27th annual international ACM SIGIR ª, 2004 – dl.acm.org
58
59
60
61 Cited by 9
62
63
64 Integrating word relationships into language models
65 G Cao, JY Nie, J Bai – of the 28th annual international ACM SIGIR ª, 2005 – dl.acm.org
66
67
68
69 Cited by 184
70
71
72 Modelling term dependence with copulas
73 C Eickhoff, AP de Vries, T Hofmann – of the 38th international ACM SIGIR ª, 2015 – dl.acm.org
74
75
76
77 Cited by 8
78
79
80 MRF based approach for sentence retrieval
81 K Cai, C Chen, K Liu, J Bu, P Huang – annual international ACM SIGIR ª, 2007 – dl.acm.org
82
83
84
85 Cited by 6
86
87
88 Automatic Phrase Indexing for Document Retrieval: An Examination of Syntactic and Non-Syntactic Methods
89 JL Fagan – ACM SIGIR Forum, 2017 – dl.acm.org

90
91
92
93 Cited by 169
94
95
96 Improving Search using Proximity-Based Statistics
97 X Lu – àÃ¢ 38th International ACM SIGIR Conference on Research àÃ¢, 2015 – dl.acm.org
98
99
100
101 Cited by 1
102
103
104 Exploiting term dependence [while](#) handling negation in medical search
105 N Limsopatham, C Macdonald, R McCreadieàÃ¢ – àÃ¢ international ACM SIGIR àÃ¢, 2012 – dl.acm.org
106
107
108
109 Cited by 21
110
111
112 A proximity language model [for](#) information retrieval
113 J Zhao, Y Yun – Proceedings of the 32nd international ACM SIGIR àÃ¢, 2009 – dl.acm.org
114
115
116
117 Cited by 69
118
119
120 Mining dependency relations [for](#) query expansion in passage retrieval
121 R Sun, CH Ong, TS Chua – àÃ¢ of the 29th annual international ACM SIGIR àÃ¢, 2006 – dl.acm.org
122
123
124
125 Cited by 68
126
127
128 Social annotation in query expansion: a machine learning approach
129 Y Lin, H Lin, S Jin, Z Ye – àÃ¢ of the 34th international ACM SIGIR àÃ¢, 2011 – dl.acm.org
130
131
132
133 Cited by 39
134
135
136
137 Dependency relation matching [for](#) answer selection
138 R Sun, H Cui, K Li, MY Kan, TS Chua – àÃ¢ annual international ACM SIGIR àÃ¢, 2005 – dl.acm.org
139
140
141 Cited by 20
142
143
144
145 Exploiting proximity feature in bigram language model [for](#) information retrieval
146 SH Na, J Kim, IS Kang, JH Lee – àÃ¢ the 31st annual international ACM SIGIR àÃ¢, 2008 – dl.acm.org
147
148
149
150 Cited by 6
151
152
153 Exploiting semantics [for](#) improving clinical information retrieval
154 A Babashzadeh, J Huang, M Daoud – àÃ¢ the 36th international ACM SIGIR àÃ¢, 2013 – dl.acm.org
155
156
157
158 Cited by 21
159

160
161 Exploring and measuring dependency trees for information retrieval
162 C Liu – àÁe of the 31st annual international ACM SIGIR conference àÁe, 2008 – dl.acm.org
163
164
165
166 Cited by 1
167
168
169 Learning to reweight terms with distributed representations
170 G Zheng, J Callan – Proceedings of the 38th International ACM SIGIR àÁe, 2015 – dl.acm.org
171
172
173
174 Cited by 35
175
176
177 Query term ranking based on search results overlap
178 W Song, Y Zhang, Y Xie, T Liu, S Li – àÁe of the 34th international ACM SIGIR àÁe, 2011 – dl.acm.org
179
180
181
182 Cited by 1
183
184
185
186 Positional language models for information retrieval
187 Y Lv, CX Zhai – Proceedings of the 32nd international ACM SIGIR àÁe, 2009 – dl.acm.org
188
189
190
191 Cited by 198
192
193
194 Event tracking based on domain dependency
195 F Fukumoto, Y Suzuki – àÁe of the 23rd annual international ACM SIGIR àÁe, 2000 – dl.acm.org
196
197
198
199 Cited by 38
200
201
202 Question answering passage retrieval using dependency relations
203 H Cui, R Sun, K Li, MY Kan, TS Chua – àÁe annual international ACM SIGIR àÁe, 2005 – dl.acm.org
204
205
206
207 Cited by 232
208
209
210 Parameterized concept weighting in verbose queries
211 M Bendersky, D Metzler, WB Croft – àÁe of the 34th international ACM SIGIR àÁe, 2011 – dl.acm.org
212
213
214
215 Cited by 100
216
217
218
219 Sigir 2014 workshop on semantic matching in information retrieval
220 J Gonzalo, H Li, A Moschitti, J Xu – àÁe of the 37th international ACM SIGIR àÁe, 2014 – dl.acm.org
221
222
223
224 Cited by 4
225
226
227 Compact query term selection using topically related text
228 KT Maxwell, WB Croft – Proceedings of the 36th international ACM SIGIR àÁe, 2013 – dl.acm.org
229

230
231 Cited by 38
232
233
234
235
236 Incorporating query term dependencies in language models [for](#) document retrieval
237 M Srikanth, R Srihari – [of the 26th annual international ACM SIGIR](#), 2003 – dl.acm.org
238
239
240 Cited by 19
241
242
243
244 Building a web test collection using social media
245 CJ Lee, WB Croft – [Proceedings of the 36th international ACM SIGIR](#), 2013 – dl.acm.org
246
247
248 Cited by 3
249
250
251 Exploring reductions [for long](#) web queries
252 N Balasubramanian, G Kumarana – [international ACM SIGIR](#), 2010 – dl.acm.org
253
254
255 Cited by 74
256
257
258
259 Learning user reformulation behavior [for](#) query auto-completion
260 JY Jiang, YY Ke, PY Chien, PJ Cheng – [the 37th international ACM SIGIR](#), 2014 – dl.acm.org
261
262
263
264 Cited by 40
265
266
267 Modeling term dependencies with quantum language models [for](#) IR
268 A Sordoni, JY Nie, Y Bengio – [of the 36th international ACM SIGIR](#), 2013 – dl.acm.org
269
270
271
272 Cited by 21
273
274
275 An alignment-based pattern representation model [for](#) information extraction
276 S Kim, M Jeong, GG Lee – [of the 31st annual international ACM SIGIR](#), 2008 – dl.acm.org
277
278
279
280 Cited by 3
281
282
283
284 [PDF] Extracting adverse drug reactions from forum posts and linking them to drugs
285 A Yates, N Goharian, O Frieder – [of the 2013 ACM SIGIR](#), 2013 – pdfs.semanticscholar.org
286
287
288 Cited by 11
289
290
291 Copulas [for](#) information retrieval
292 C Eickhoff, AP de Vries – [international ACM SIGIR](#), 2013 – dl.acm.org
293
294
295
296 Cited by 19
297
298
299

300 | Hybrid word–subword decoding [for](#) spoken term detection
301 | I Szoke, M Fapso, L Burget – àÃ© International ACM SIGIR àÃ©, 2008 – researchgate.net
302
303
304 | Cited by 42
305
306
307
308 | Investigation of web query refinement via topic analysis and learning with personalization
309 | L Bing, W Lam – àÃ© of the 2nd SIGIR Workshop on Query àÃ©, 2011 – researchgate.net
310
311
312 | Cited by 2
313
314
315 | An exploration of proximity measures in information retrieval
316 | T Tao, CX Zhai – àÃ© of the 30th annual international ACM SIGIR àÃ©, 2007 – dl.acm.org
317
318
319 | Cited by 245
320
321
322 | Feature selection [for](#) automatic taxonomy induction
323 | H Yang, J Callan – Proceedings of the 32nd international ACM SIGIR àÃ©, 2009 – dl.acm.org
324
325
326 | Cited by 6
327
328
329 | Query type classification [for](#) web document retrieval
330 | IH Kang, GC Kim – àÃ© of the 26th annual international ACM SIGIR àÃ©, 2003 – dl.acm.org
331
332
333 | Cited by 332
334
335
336 | Automatic refinement of patent queries using concept importance predictors
337 | P Mahdabi, L Andersson, M Keikha – àÃ© international ACM SIGIR àÃ©, 2012 – dl.acm.org
338
339
340 | Cited by 34
341
342 | Efficient cost–aware cascade ranking in multi–stage retrieval
343 | RC Chen, L Gallagher, R Blanco – àÃ© International ACM SIGIR àÃ©, 2017 – dl.acm.org
344
345
346 | Cited by 3
347
348
349 | Using key concepts in a translation model [for](#) retrieval
350 | JH Park, WB Croft – Proceedings of the 38th International ACM SIGIR àÃ©, 2015 – dl.acm.org
351
352
353 | Cited by 8
354
355
356 | Enhancing ad–hoc relevance weighting using probability density estimation
357 | X Zhou, JX Huang, B He – àÃ© of the 34th international ACM SIGIR àÃ©, 2011 – dl.acm.org
358
359
360 | Cited by 9
361
362
363
364 | Efficient & Effective Selective Query Rewriting with Efficiency Predictions
365 | C Macdonald, N Tonellotto, I Ounis – àÃ© of the 40th International ACM SIGIR àÃ©, 2017 – dl.acm.org
366
367
368 | Cited by 0
369

370
371 Revisiting the dependence language model [for](#) information retrieval
372 L Maisonnasse, E Gaussier, JP Chevallet – àÃ© international ACM SIGIR àÃ©, 2007 – dl.acm.org
373
374
375 Cited by 24
376
377 Flat vs. hierarchical phrase–based translation models [for](#) cross–language information retrieval
378 F Ture, J Lin – àÃ© 36th international ACM SIGIR conference on Research àÃ©, 2013 – dl.acm.org
379
380
381 Cited by 3
382
383 Learning [for](#) search result diversification
384 Y Zhu, Y Lan, J Guo, X Cheng, S Niu – àÃ© the 37th international ACM SIGIR àÃ©, 2014 – dl.acm.org
385
386
387 Cited by 42
388
389
390
391
392 Cross–language text classification
393 JS Olsson, DW Oard, J HajiÃIJ – àÃ© the 28th annual international ACM SIGIR àÃ©, 2005 – dl.acm.org
394
395
396 Cited by 50
397
398 On the cost of phrase–based ranking
399 M Petri, A Moffat – Proceedings of the 38th International ACM SIGIR àÃ©, 2015 – dl.acm.org
400
401
402 Cited by 1
403
404
405 Selecting indexing strings using adaptation
406 Y Takeda, K Umemura – àÃ© of the 25th annual international ACM SIGIR àÃ©, 2002 – dl.acm.org
407
408
409 Cited by 11
410
411
412 Efficient in–memory top–k document retrieval
413 JS Culpepper, M Petri, F Scholer – àÃ© of the 35th international ACM SIGIR àÃ©, 2012 – dl.acm.org
414
415
416 Cited by 35
417
418 Automatic term mismatch diagnosis [for](#) selective query expansion
419 L Zhao, J Callan – Proceedings of the 35th international ACM SIGIR àÃ©, 2012 – dl.acm.org
420
421
422 Cited by 17
423
424 Structured retrieval [for](#) question answering
425 MW Bilotti, P Ogilvie, J Callan, E Nyberg – àÃ© international ACM SIGIR àÃ©, 2007 – dl.acm.org
426
427
428 Cited by 106
429
430
431 The document as an ergodic markov chain
432 E Hoenkamp, D Song – àÃ© of the 27th annual international ACM SIGIR àÃ©, 2004 – dl.acm.org
433
434
435 Cited by 8
436
437 Non–compositional term dependence [for](#) information retrieval
438 C Lioma, JG Simonsen, B LarsenàÃ© – àÃ© International ACM SIGIR àÃ©, 2015 – dl.acm.org
439

440 |
441 | Cited by 7
442 |
443 | Probabilistic model [for](#) contextual retrieval
444 | JR Wen, N Lao, WY Ma – àÃ© of the 27th annual international ACM SIGIR àÃ©, 2004 – dl.acm.org
445 |
446 |
447 | Cited by 65
448 |
449 |
450 | Resolving query translation ambiguity using a decaying co–occurrence model and syntactic dependence relations
451 | J Gao, M Zhou, JY Nie, H He, W Chen – àÃ© annual international ACM SIGIR àÃ©, 2002 – dl.acm.org
452 |
453 |
454 | Cited by 106
455 |
456 |
457 | Major topic detection and its application to opinion summarization
458 | LW Ku, LY Lee, TH Wu, HH Chen – àÃ© 28th annual international ACM SIGIR àÃ©, 2005 – dl.acm.org
459 |
460 |
461 | Cited by 76
462 |
463 | [PDF] Retrieval Techniques [for](#) Contextual Learning.
464 | N Weingart, C Eickhoff – SAL@ SIGIR, 2016 –
465 |
466 |
467 | Cited by 0
468 |
469 |
470 | A study of statistical models [for](#) query translation: finding a good unit of translation
471 | J Gao, JY Nie – àÃ© of the 29th annual international ACM SIGIR àÃ©, 2006 – dl.acm.org
472 |
473 |
474 |
475 | Cited by 40
476 |
477 |
478 | Based on Term Frequency Ranks
479 | IJJ Aalbersberg – àÃ© Seventeenth Annual International ACM–SIGIR àÃ©, 2012 – books.google.com
480 |
481 |
482 | Cited by 0
483 |
484 | Report on the ad–hoc track of the INEX 2005 workshop
485 | M Lalmas, G Kazai – ACM Sigir Forum, 2006 – dl.acm.org
486 |
487 |
488 |
489 | Cited by 13
490 |
491 |
492 | Automatic summarization focusing on document genre and text structure
493 | Y Seki – ACM SIGIR Forum, 2005 – dl.acm.org
494 |
495 |
496 |
497 | Cited by 11
498 |
499 | Context aware query classification using dynamic query window and relationship net
500 | N Goharian, SSR Mengle – àÃ© of the 33rd international ACM SIGIR àÃ©, 2010 – dl.acm.org
501 |
502 |
503 | Cited by 4
504 |
505 |
506 | Pseudo test collections [for](#) training and tuning microblog rankers
507 | R Berendsen, M Tsagkias, W WeerkampàÃ© – àÃ© international ACM SIGIR àÃ©, 2013 – dl.acm.org
508 |
509 |

510 Cited by 17
511
512 Evolutionary document summarization [for](#) disaster management
513 D Wang, L Zheng, T Li, Y Deng – àÁe of the 32nd international ACM SIGIR àÁe, 2009 – dl.acm.org
514
515
516 Cited by 9
517
518
519
520 [PDF] Learning to rank documents [for](#) ad-hoc retrieval with regularized models
521 G Cao, JY Nie, L Si, J Bai – SIGIR 2007 workshop on Learning to àÁe, 2007 – academia.edu
522
523
524
525 Cited by 8
526
527 Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization
528 F Wei, W Li, Q Lu, Y He – àÁe of the 31st annual international ACM SIGIR àÁe, 2008 – dl.acm.org
529
530
531
532 Cited by 99
533
534 Learning [for](#) Efficient Supervised Query Expansion via Two-stage Feature Selection
535 Z Zhang, Q Wang, L Si, J Gao – àÁe of the 39th International ACM SIGIR àÁe, 2016 – dl.acm.org
536
537
538
539 Cited by 2
540
541
542 Modeling Document Novelty with Neural Tensor Network [for](#) Search Result Diversification
543 L Xia, J Xu, Y Lan, J Guo, X Cheng – àÁe of the 39th International ACM SIGIR àÁe, 2016 – dl.acm.org
544
545
546 Cited by 10
547
548 Hierarchical classification [for](#) automatic image annotation
549 J Fan, Y Gao, H Luo – àÁe of the 30th annual international ACM SIGIR àÁe, 2007 – dl.acm.org
550
551
552
553 Cited by 93
554
555
556 Open Relation Extraction [for](#) Support Passage Retrieval: Merit and Open Issues
557 A Kadry, L Dietz – Proceedings of the 40th International ACM SIGIR àÁe, 2017 – dl.acm.org
558
559
560
561 Cited by 1
562
563
564 Citation sentences [for](#) semantic analysis of bioscience text
565 PI Nakov, AS SchwartzàÁe – àÁe of the SIGIR, 2004 – ai2-s2-pdfs.s3.amazonaws.com
566
567
568
569 Cited by 122
570
571
572 Properties of Extended Boolean Models
573 JH Lee – àÁe of the Seventeenth Annual International ACM-SIGIR àÁe, 2012 – books.google.com
574
575
576 Cited by 0
577
578
579 Generic summarization and keyphrase extraction using mutual reinforcement principle and sentence clustering

580 H Zha – àÃ© of the 25th annual international ACM SIGIR àÃ©, 2002 – dl.acm.org
581
582
583
584 Cited by 303
585
586
587 Translation of Natural Language Query Into Keyword Query Using a RNN Encoder–Decoder
588 HJ Song, A Kim, SB Park – àÃ© of the 40th International ACM SIGIR àÃ©, 2017 – dl.acm.org
589
590
591 Cited by 0
592
593 A model for expert finding in social networks
594 E Smirnova – àÃ© 34th international ACM SIGIR conference on Research àÃ©, 2011 – dl.acm.org
595
596
597
598 Cited by 14
599
600
601 Structured translation for cross–language information retrieval
602 R Sperer, DW Oard – àÃ© of the 23rd annual international ACM SIGIR àÃ©, 2000 – dl.acm.org
603
604
605
606 Cited by 63
607
608
609 A unified and discriminative model for query refinement
610 J Guo, G Xu, H Li, X Cheng – àÃ© of the 31st annual international ACM SIGIR àÃ©, 2008 – dl.acm.org
611
612
613
614 Cited by 137
615
616
617
618 An improved markov random field model for supporting verbose queries
619 M Lease – Proceedings of the 32nd international ACM SIGIR àÃ©, 2009 – dl.acm.org
620
621
622
623 Cited by 56
624
625
626 Information retrieval and applications of graphical models (IRGM 2007)
627 JM Fernández-Luna, B Piwowarski, JF Huete – ACM SIGIR Forum, 2007 – dl.acm.org
628
629
630 Cited by 2
631
632 Integrating logical operators in query expansion in vector space model
633 JY Nie, F Jin – SIGIR workshop on Mathematical and Formal àÃ©, 2002 – cir.dcs.uni–pannon.hu
634
635
636
637 Cited by 21
638
639
640 Deep classification in large–scale text hierarchies
641 GR Xue, D Xing, Q Yang, Y Yu – àÃ© the 31st annual international ACM SIGIR àÃ©, 2008 – dl.acm.org
642
643
644
645 Cited by 164
646
647
648 A location–based news article recommendation with explicit localized semantic analysis
649 JW Son, A Kim, SB Park – àÃ© of the 36th international ACM SIGIR àÃ©, 2013 – dl.acm.org

650
651
652
653 Cited by 27
654
655
656
657 Hierarchical classification of Web content
658 S Dumais, H Chen – àÁe of the 23rd annual international ACM SIGIR àÁe, 2000 – dl.acm.org
659
660
661
662 Cited by 1006
663
664
665 TSCAN: a novel method for topic summarization and content anatomy
666 CC Chen, MC Chen – àÁe of the 31st annual international ACM SIGIR àÁe, 2008 – dl.acm.org
667
668
669
670 Cited by 34
671
672
673
674 Collective entity linking in web text: a graph-based method
675 X Han, L Sun, J Zhao – Proceedings of the 34th international ACM SIGIR àÁe, 2011 – dl.acm.org
676
677
678
679 Cited by 286
680
681
682
683 Statistical cross-language information retrieval using n-best query translations
684 M Federico, N Bertoldi – àÁe of the 25th annual international ACM SIGIR àÁe, 2002 – dl.acm.org
685
686
687
688 Cited by 79
689
690
691 Aspect-based opinion mining from online reviews
692 S Moghaddam, M Ester – Tutorial at SIGIR Conference, 2012 – cs.sfu.ca
693
694
695
696 Cited by 25
697
698
699 Enhancing web search by mining search and browse logs.
700 D Jiang, J Pei, H Li – SIGIR, 2011 – sigir2011.org
701
702
703 Cited by 2
704
705
706 On ranking the effectiveness of searches
707 V Vinay, IJ Cox, N Milic-Frayling, K Wood – àÁe international ACM SIGIR àÁe, 2006 – dl.acm.org
708
709
710
711 Cited by 73
712
713
714
715 Block-based web search
716 D Cai, S Yu, JR Wen, WY Ma – àÁe the 27th annual international ACM SIGIR àÁe, 2004 – dl.acm.org
717
718
719 Cited by 234

720
721
722 An event extraction model based on timeline and user analysis in latent dirichlet allocation
723 B Tsolmon, KS Lee – Proceedings of the 37th international ACM SIGIR àÁ, 2014 – dl.acm.org
724
725
726
727 Cited by 8
728
729
730 Report on ACM SIGIR workshop on mathematical/formal methods in information retrieval
731 S Dominich, M Lalmas, CJ van Rijsbergen – ACM SIGIR Forum, 2002 – dl.acm.org
732
733 Cited by 0
734
735
736 A scalability analysis of classifiers in text categorization
737 Y Yang, J Zhang, B Kisiel – àÁ of the 26th annual international ACM SIGIR àÁ, 2003 – dl.acm.org
738
739
740
741 Cited by 217
742
743
744 Distributed Representations [for](#) Semantic Matching in non-factoid Question Answering.
745 P Molino, LM Aiello – SMIR@ SIGIR, 2014 – ceur-ws.org
746
747
748 Cited by 1
749
750 Multi-labelled classification using maximum entropy method
751 S Zhu, X Ji, W Xu, Y Gong – àÁ of the 28th annual international ACM SIGIR àÁ, 2005 – dl.acm.org
752
753
754
755 Cited by 203
756
757
758 Contextual search and name disambiguation in email using graphs
759 E Minkov, WW Cohen, AY Ng – àÁ of the 29th annual international ACM SIGIR àÁ, 2006 – dl.acm.org
760
761
762
763 Cited by 184
764
765
766 Opinion holder extraction from author and authority viewpoints
767 Y Seki – àÁ of the 30th annual international ACM SIGIR àÁ, 2007 – dl.acm.org
768
769
770
771 Cited by 17
772
773
774 Context-aware point-of-interest recommendation using tensor factorization with social regularization
775 L Yao, QZ Sheng, Y Qin, X WangàÁ – àÁ ACM SIGIR conference àÁ, 2015 – dl.acm.org
776
777
778
779 Cited by 26
780
781
782
783
784
785 Generative modeling of persons and documents [for](#) expert search
786 P Serdyukov, D Hiemstra, M FokkingaàÁ – àÁ international ACM SIGIR àÁ, 2007 – dl.acm.org
787
788
789

790 Cited by 10
791
792
793 Taily: shard selection using the tail of score distributions
794 R Aly, D Hiemstra, T Demeester – à  of the 36th international ACM SIGIR à , 2013 – dl.acm.org
795
796
797
798 Cited by 31
799
800
801 [PDF] DIR 2010: the tenth Dutch–Belgian information retrieval workshop.
802 W Kraaij, S Verberne, M Hinne, M van der Heijdenà  – SIGIR Forum, 2010 – researchgate.net
803
804
805 Cited by 0
806
807 That's not my question: Learning to weight unmatched terms in CQA vertical search
808 B Petersil, A Mejer, I Szpektor, K Crammer - à  International ACM SIGIR à , 2016 - dl.acm.org
809
810
811
812 Cited by 2
813
814
815 Exploiting sequential dependencies for expert finding
816 P Serdyukov, H Rode, D Hiemstra - à  31st annual international ACM SIGIR à , 2008 - dl.acm.org
817
818
819
820 Cited by 17
821
822
823 Iterative translation disambiguation for cross-language information retrieval
824 C Monz, BJ Dorr - à  of the 28th annual international ACM SIGIR à , 2005 - dl.acm.org
825
826
827
828 Cited by 68
829
830
831 ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews
832 S Moghaddam, M Ester - à  of the 34th international ACM SIGIR à , 2011 - dl.acm.org
833
834
835
836 Cited by 119
837
838
839
840
841 Discovering key concepts in verbose queries
842 M Bendersky, WB Croft - à  of the 31st annual international ACM SIGIR à , 2008 - dl.acm.org
843
844
845
846 Cited by 262
847
848
849 Proximity-based opinion retrieval
850 S Gerani, MJ Carman, F Crestani - à  of the 33rd international ACM SIGIR à , 2010 - dl.acm.org
851
852
853
854 Cited by 72
855
856
857 Automatic query refinement using lexical affinities with maximal information gain
858 D Carmel, E Farchi, Y Petruschka, A Soffer - à  international ACM SIGIR à , 2002 - dl.acm.org
859

860
861 Cited by 104
862
863
864
865 AIR 2006: First international workshop on adaptive information retrieval
866 H Joho, J Urban, R Villa, JM Jose - ACM SIGIR Forum, 2008 - dl.acm.org
867
868
869
870 Cited by 4
871
872
873 Translating unknown queries with web corpora for cross-language information retrieval
874 PJ Cheng, JW Teng, RC Chen, JH Wang - ACM SIGIR conference, 2004 - dl.acm.org
875
876
877
878 Cited by 120
879
880
881 [PDF] When is a query a question? Reconstructing wh-requests from ad hoc-queries
882 M van der Heijden, M Hinne, S Verberne - Development SIGIR, 2010 - lands.let.ru.nl
883
884
885 Cited by 0
886
887
888 Improving query translation for cross-language information retrieval using statistical models
889 J Gao, JY Nie, E Xun, J Zhang, M Zhou - international ACM SIGIR, 2001 - dl.acm.org
890
891
892
893 Cited by 154
894
895
896 Integrating phrase inseparability in phrase-based model
897 L Shi, JY Nie - 32nd international ACM SIGIR conference on research, 2009 - dl.acm.org
898
899
900
901 Cited by 10
902
903
904 Cross-Device User Linking: URL, Session, Visiting Time, and Device-log Embedding
905 MC Phan, A Sun, Y Tay - of the 40th International ACM SIGIR, 2017 - dl.acm.org
906
907
908
909 Cited by 0
910
911
912 A new probabilistic retrieval model based on the dirichlet compound multinomial distribution
913 Z Xu, R Akella - of the 31st annual international ACM SIGIR conference, 2008 - dl.acm.org
914
915
916
917 Cited by 31
918
919
920
921 Linear discriminant model for information retrieval
922 J Gao, H Qi, X Xia, JY Nie - of the 28th annual international ACM SIGIR, 2005 - dl.acm.org
923
924
925
926 Cited by 118
927
928
929 High performance question/answering

930 MA Pasca, SM Harabagiu - àÁe of the 24th annual international ACM SIGIR àÁe, 2001 - dl.acm.org
931
932
933
934 Cited by 208
935
936
937 DUMPLING: A Novel Dynamic Search Engine
938 AJ Zhou, J Luo, H Yang - àÁe of the 38th International ACM SIGIR àÁe, 2015 - dl.acm.org
939
940
941
942 Cited by 2
943
944
945 Bundle recommendation in ecommerce
946 T Zhu, P Harrington, J Li, L Tang - àÁe of the 37th international ACM SIGIR àÁe, 2014 - dl.acm.org
947
948
949
950 Cited by 26
951
952
953 Document representation with generalized latent semantic analysis
954 I Matveeva, A Farahat, C Royer - àÁe in Information Retrieval (SIGIR 2005), 2005 - parc.com
955
956
957
958 Cited by 2
959
960
961 Unifying user-based and item-based collaborative filtering approaches by similarity fusion
962 J Wang, AP De Vries, MJT Reinders - àÁe annual international ACM SIGIR àÁe, 2006 - dl.acm.org
963
964
965
966 Cited by 740
967
968
969
970 Learning phonetic similarity for matching named entity translations and mining new translations
971 W Lam, R Huang, PS Cheung - àÁe 27th annual international ACM SIGIR àÁe, 2004 - dl.acm.org
972
973
974 Cited by 25
975
976
977 Latent concepts and the number orthogonal factors in latent semantic analysis
978 G Dupret - àÁe of the 26th annual international ACM SIGIR àÁe, 2003 - dl.acm.org
979
980
981
982 Cited by 33
983
984
985
986 [PDF] Cross lingual issues in patent retrieval
987 L SarasÁza, G Corremans - àÁe of the ACM SIGIR 2000 Workshop on àÁe, 2000 - research.nii.ac.jp
988
989
990
991 Cited by 10
992
993
994
995
996 Analysis of structural relationships for hierarchical cluster labeling
997 M Muhr, R Kern, M Granitzer - àÁe of the 33rd international ACM SIGIR àÁe, 2010 - dl.acm.org
998
999

1000
1001 Cited by 28
1002
1003
1004 Topic themes for multi-document summarization
1005 S Harabagiu, F Lacatusu - \ddagger of the 28th annual international ACM SIGIR \ddagger , 2005 - dl.acm.org
1006
1007
1008
1009 Cited by 150
1010
1011
1012 Improving QA retrieval using document priors
1013 J Mayfield, P McNamee - \ddagger of the 29th annual international ACM SIGIR \ddagger , 2006 - dl.acm.org
1014
1015
1016 Cited by 0
1017
1018
1019 Coarse-to-fine review selection via supervised joint aspect and sentiment model
1020 Z Hai, G Cong, K Chang, W Liu, P Cheng - \ddagger 37th international ACM SIGIR \ddagger , 2014 - dl.acm.org
1021
1022
1023
1024 Cited by 11
1025
1026
1027 To personalize or not to personalize: modeling queries with variation in user intent
1028 J Teevan, ST Dumais, DJ Liebling - \ddagger 31st annual international ACM SIGIR \ddagger , 2008 - dl.acm.org
1029
1030
1031 Cited by 263
1032
1033
1034
1035
1036 Learning to respond with deep neural networks for retrieval-based human-computer conversation system
1037 R Yan, Y Song, H Wu - Proceedings of the 39th International ACM SIGIR \ddagger , 2016 - dl.acm.org
1038
1039
1040
1041 Cited by 31
1042
1043
1044 Social context summarization
1045 Z Yang, K Cai, J Tang, L Zhang, Z Su, J Li - \ddagger 34th international ACM SIGIR \ddagger , 2011 - dl.acm.org
1046
1047
1048
1049 Cited by 85
1050
1051 Information-based models for ad hoc IR
1052 S Clinchant, E Gaussier - \ddagger of the 33rd international ACM SIGIR \ddagger , 2010 - dl.acm.org
1053
1054
1055
1056 Cited by 112
1057
1058
1059 Set-based model: A new approach for information retrieval
1060 B Pățas, N Ziviani, W Meira Jr \ddagger international ACM SIGIR \ddagger , 2002 - dl.acm.org
1061
1062
1063
1064 Cited by 38
1065
1066
1067 Multi-label informed latent semantic indexing
1068 K Yu, S Yu, V Tresp - \ddagger of the 28th annual international ACM SIGIR \ddagger , 2005 - dl.acm.org
1069

1070
1071
1072 Cited by 215
1073
1074
1075 Personalized query expansion for the web
PA Chirita, CS Firan, W Nejdl - àÃ© the 30th annual international ACM SIGIR àÃ©, 2007 - dl.acm.org
1076
1077
1078
1079
1080 Cited by 342
1081
1082
1083
1084 The role of knowledge in conceptual retrieval: a study in the domain of clinical medicine
J Lin, D Demner-Fushman - àÃ© of the 29th annual international ACM SIGIR àÃ©, 2006 - dl.acm.org
1085
1086
1087
1088
1089 Cited by 73
1090
1091
1092
1093 Modularity-based query clustering for identifying users sharing a common condition
MGO Harel, E Yom-Tov - àÃ© of the 38th International ACM SIGIR àÃ©, 2015 - dl.acm.org
1094
1095
1096
1097 Cited by 1
1098
1099
1100 Modeling multi-query retrieval tasks using density matrix transformation
Q Li, J Li, P Zhang, D Song - àÃ© of the 38th International ACM SIGIR àÃ©, 2015 - dl.acm.org
1101
1102
1103
1104 Cited by 4
1105
1106
1107 Analysis of lexical signatures for finding lost or related documents
ST Park, DM Pennock, CL Giles, R Krovetz - àÃ© international ACM SIGIR àÃ©, 2002 - dl.acm.org
1108
1109
1110
1111 Cited by 38
1112
1113
1114 Robustness of adaptive filtering methods in a cross-benchmark evaluation
Y Yang, S Yoo, J Zhang, B Kisiel - àÃ© 28th annual international ACM SIGIR àÃ©, 2005 - dl.acm.org
1115
1116
1117
1118 Cited by 55
1119
1120
1121 Context sensitive stemming for web search
F Peng, N Ahmed, X Li, Y Lu - àÃ© the 30th annual international ACM SIGIR àÃ©, 2007 - dl.acm.org
1122
1123
1124
1125 Cited by 107
1126
1127
1128
1129 Summarizing the differences from microblogs
D Wang, M Ogihara, T Li - àÃ© of the 35th international ACM SIGIR àÃ©, 2012 - dl.acm.org
1130
1131
1132
1133
1134 Cited by 6
1135
1136
1137 A method of rating the credibility of news documents on the web
R Nagura, Y Seki, N Kando, M Aono - àÃ© annual international ACM SIGIR àÃ©, 2006 - dl.acm.org
1138
1139

1140
1141
1142 Cited by 39
1143
1144
1145
1146 Simultaneously modeling semantics and structure of threaded discussions: a sparse coding approach and its applications
1147 C Lin, JM Yang, R Cai, XJ Wang - â€¢ international ACM SIGIR â€¢, 2009 - dl.acm.org
1148
1149
1150
1151 Cited by 40
1152
1153
1154 Collaborative filtering via gaussian probabilistic latent semantic analysis
1155 T Hofmann - â€¢ of the 26th annual international ACM SIGIR â€¢, 2003 - dl.acm.org
1156
1157
1158
1159 Cited by 429
1160
1161
1162 In situ insights
1163 Y Lv, A Fuxman - â€¢ 38th International ACM SIGIR Conference on Research â€¢, 2015 - dl.acm.org
1164
1165
1166
1167 Cited by 2
1168
1169
1170
1171 [PDF] Combining text and linguistic document representations for authorship attribution
1172 A Kaster, S Siersdorferâ€¢ - SIGIR workshop: stylistic â€¢, 2005 - domino.mpi-inf.mpg.de
1173
1174
1175
1176 Cited by 32
1177
1178
1179 Information retrieval as card playing: A formal model for optimizing interactive retrieval interface
1180 Y Zhang, C Zhai - Proceedings of the 38th International ACM SIGIR â€¢, 2015 - dl.acm.org
1181
1182
1183
1184 Cited by 16
1185
1186
1187 Evaluating evaluation measure stability
1188 C Buckley, EM Voorhees - ACM SIGIR Forum, 2017 - dl.acm.org
1189
1190
1191
1192 Cited by 707
1193
1194
1195 Applying Information Extraction for Patent Structure Analysis
1196 M Okamoto, Z Shan, R Orihara - â€¢ of the 40th International ACM SIGIR â€¢, 2017 - dl.acm.org
1197
1198
1199 Cited by 0
1200
1201 -----
1202
1203
1204 Fast context-aware recommendations with factorization machines
1205 S Rendle, Z Gantner, C Freudenthalerâ€¢ - â€¢ ACM SIGIR conference â€¢, 2011 - dl.acm.org
1206
1207
1208 Cited by 249
1209

1210
1211 Joint Learning of Response Ranking and Next Utterance Suggestion in Human-Computer Conversation System
1212 R Yan, D Zhao - Proceedings of the 40th International ACM SIGIR â€¢, 2017 - dl.acm.org
1213
1214
1215 Cited by 0
1216
1217 Synthesizing high utility suggestions for rare web search queries
1218 A Jain, U Ozertem, E Velipasaoglu - â€¢ of the 34th international ACM SIGIR â€¢, 2011 - dl.acm.org
1219
1220
1221 Cited by 26
1222
1223
1224 A syntactic tree matching approach to finding similar questions in community-based qa services
1225 K Wang, Z Ming, TS Chua - â€¢ of the 32nd international ACM SIGIR â€¢, 2009 - dl.acm.org
1226
1227
1228
1229 Cited by 195
1230
1231
1232 [PDF] Experiments and evaluation of link discovery in the wikipedia
1233 D Huang, A Trotman, S Geva - â€¢ of the SIGIR 2008 â€¢, 2008 - pdfs.semanticscholar.org
1234
1235
1236
1237 Cited by 7
1238
1239
1240 Automatic boolean query suggestion for professional search
1241 Y Kim, J Seo, WB Croft - â€¢ of the 34th international ACM SIGIR conference â€¢, 2011 - dl.acm.org
1242
1243
1244
1245 Cited by 44
1246
1247
1248 On statistical analysis and optimization of information retrieval effectiveness metrics
1249 J Wang, J Zhu - â€¢ 33rd international ACM SIGIR conference on Research â€¢, 2010 - dl.acm.org
1250
1251
1252
1253 Cited by 25
1254
1255
1256
1257 Learning online discussion structures by conditional random fields
1258 H Wang, C Wang, CX Zhai, J Han - â€¢ of the 34th international ACM SIGIR â€¢, 2011 - dl.acm.org
1259
1260
1261
1262 Cited by 45
1263
1264
1265
1266 Deep learning for information retrieval
1267 H Li, Z Lu - â€¢ 39th International ACM SIGIR conference on Research â€¢, 2016 - dl.acm.org
1268
1269
1270
1271 Cited by 8
1272
1273
1274 Query representation for cross-temporal information retrieval
1275 M Efron - â€¢ 36th international ACM SIGIR conference on Research â€¢, 2013 - dl.acm.org
1276
1277
1278
1279 Cited by 5

1280
1281
1282 Modeling user's receptiveness over time for recommendation
1283 W Chen, W Hsu, ML Lee – àĘ of the 36th international ACM SIGIR àĘ, 2013 – dl.acm.org
1284
1285
1286
1287 Cited by 26
1288
1289
1290 Report on the INEX 2005 interactive track
1291 S Malik, B Larsen, A Tombros – ACM SIGIR Forum, 2007 – dl.acm.org
1292
1293
1294
1295 Cited by 10
1296
1297
1298
1299 Learning temporal-dependent ranking models
1300 M Costa, F Couto, M Silva – àĘ of the 37th international ACM SIGIR àĘ, 2014 – dl.acm.org
1301
1302
1303
1304 Cited by 22
1305
1306
1307 Table extraction using conditional random fields
1308 D Pinto, A McCallum, X Wei, WB Croft – àĘ annual international ACM SIGIR àĘ, 2003 – dl.acm.org
1309
1310
1311
1312 Cited by 425
1313
1314
1315 Query expansion using path-constrained random walks
1316 J Gao, G Xu, J Xu – Proceedings of the 36th international ACM SIGIR àĘ, 2013 – dl.acm.org
1317
1318
1319
1320 Cited by 25
1321
1322
1323
1324 Investigating users' query formulations for cognitive search intents
1325 MP Kato, T Yamamoto, H Ohshima – àĘ international ACM SIGIR àĘ, 2014 - dl.acm.org
1326
1327
1328
1329 Cited by 12
1330
1331
1332
1333
1334 Web page segmentation with structured prediction and its application in web page classification
1335 L Bing, R Guo, W Lam, ZY Niu, H Wang – àĘ 37th international ACM SIGIR àĘ, 2014 - dl.acm.org
1336
1337
1338
1339 Cited by 3
1340
1341
1342 Efficient top-k querying over social-tagging networks
1343 R Schenkel, T Crecelius, M Kacimi, S Michel – àĘ ACM SIGIR conference àĘ, 2008 - dl.acm.org
1344
1345
1346
1347 Cited by 173
1348
1349

1350 [PDF] Quickumls: a fast, unsupervised approach for medical concept extraction
1351 L Soldaini, N Goharian - MedIR workshop, sigir, 2016 - ir.cs.georgetown.edu
1352
1353
1354
1355 Cited by 15
1356
1357
1358 Content-enriched classifier for web video classification
1359 B Cui, C Zhang, G Cong - *of the 33rd international ACM SIGIR*, 2010 - dl.acm.org
1360
1361
1362
1363 Cited by 18
1364
1365
1366 Click-through prediction for news queries
1367 AC Käünig, M Gamon, Q Wu - *of the 32nd international ACM SIGIR*, 2009 - dl.acm.org
1368
1369
1370
1371 Cited by 65
1372
1373 Identifying relevant temporal expressions for real-world events
1374 N Kanhabua, S Romano, A Stewart - *Proceedings of The SIGIR 2012*, 2012 - 13s.de
1375
1376
1377
1378 Cited by 18
1379
1380
1381 Personalized time-aware tweets summarization
1382 Z Ren, S Liang, E Meij, M de Rijke - *of the 36th international ACM SIGIR*, 2013 - dl.acm.org
1383
1384
1385
1386 Cited by 55
1387
1388
1389 Large-scale graph mining and learning for information retrieval
1390 B Gao, T Wang, TY Liu - *of the 35th international ACM SIGIR conference*, 2012 - dl.acm.org
1391
1392
1393
1394 Cited by 1
1395
1396
1397 Ranking with multiple hyperplanes
1398 T Qin, XD Zhang, DS Wang, TY Liu, W Lai - *international ACM SIGIR*, 2007 - dl.acm.org
1399
1400
1401
1402 Cited by 106
1403
1404
1405
1406 Methods and metrics for cold-start recommendations
1407 AI Schein, A Popescul, LH Ungar - *international ACM SIGIR*, 2002 - dl.acm.org
1408
1409
1410
1411 Cited by 1519
1412
1413
1414 A Probabilistic Model for Information Retrieval Based on Maximum Value Distribution
1415 JH Paik - *38th International ACM SIGIR Conference on Research*, 2015 - dl.acm.org
1416
1417
1418
1419 Cited by 3

1420
1421
1422 On feature distributional clustering for text categorization
1423 R Bekkerman, R El-Yaniv, N Tishby - à  international ACM SIGIR à , 2001 - dl.acm.org
1424
1425
1426
1427 Cited by 156
1428
1429
1430 Unsupervised query segmentation using clickthrough for information retrieval
1431 Y Li, BJP Hsu, CX Zhai, K Wang - à  of the 34th international ACM SIGIR à , 2011 - dl.acm.org
1432
1433
1434
1435 Cited by 46
1436
1437
1438 Frontiers, challenges, and opportunities for information retrieval: Report from SWIRL 2012 the second strategic workshop
1439 J Allan, B Croft, A Moffat, M Sanderson - ACM SIGIR Forum, 2012 - dl.acm.org
1440
1441
1442
1443 Cited by 107
1444
1445
1446 [PDF] Semi-supervised statistical inference for business entities extraction and business relations discovery
1447 RYK Lau, W Zhang - SIGIR 2011 workshop, July, 2011 - pdfs.semanticscholar.org
1448
1449
1450
1451 Cited by 6
1452
1453
1454 Modeling click-through based word-pairs for web search
1455 J Jagarlamudi, J Gao - Proceedings of the 36th international ACM SIGIR à , 2013 - dl.acm.org
1456
1457
1458
1459 Cited by 3
1460
1461
1462 Ranking related news predictions
1463 N Kanhabua, R Blanco, M Matthews - à  the 34th international ACM SIGIR à , 2011 - dl.acm.org
1464
1465
1466
1467 Cited by 40
1468
1469
1470 Evaluating non-deterministic retrieval systems
1471 GK Jayasinghe, W Webber, M Sanderson - à  ACM SIGIR conference à , 2014 - dl.acm.org
1472
1473
1474
1475 Cited by 3
1476
1477
1478
1479 Modeling action-level satisfaction for search task satisfaction prediction
1480 H Wang, Y Song, MW Chang, X He, A Hassan - à  ACM SIGIR conference à , 2014 - dl.acm.org
1481
1482
1483
1484 Cited by 20
1485
1486
1487 Content-based retrieval for heterogeneous domains: domain adaptation by relative aggregation points
1488 MP Kato, H Ohshima, K Tanaka - à  of the 35th international ACM SIGIR à , 2012 - dl.acm.org
1489

1490
1491 Cited by 9
1492
1493
1494
1495 A nonparametric hierarchical bayesian framework for information filtering
1496 K Yu, V Tresp, S Yu - à€ of the 27th annual international ACM SIGIR à€, 2004 - dl.acm.org
1497
1498
1499
1500 Cited by 68
1501
1502
1503 Contextual bandits in a collaborative environment
1504 Q Wu, H Wang, Q Gu, H Wang - à€ of the 39th International ACM SIGIR à€, 2016 - dl.acm.org
1505
1506
1507 Cited by 4
1508
1509
1510
1511 Modeling subset distributions for verbose queries
1512 X Xue, WB Croft - Proceedings of the 34th international ACM SIGIR à€, 2011 - dl.acm.org
1513
1514
1515
1516 Cited by 11
1517
1518
1519 Explicit factor models for explainable recommendation based on phrase-level sentiment analysis
1520 Y Zhang, G Lai, M Zhang, Y Zhang, Y Liu - à€ international ACM SIGIR à€, 2014 - dl.acm.org
1521
1522
1523
1524 Cited by 119
1525
1526
1527 What reviews are satisfactory: novel features for automatic helpfulness voting
1528 Y Hong, J Lu, J Yao, Q Zhu, G Zhou - à€ the 35th international ACM SIGIR à€, 2012 - dl.acm.org
1529
1530
1531
1532 Cited by 26
1533
1534
1535 Analyzing user's sequential behavior in query auto-completion via markov processes
1536 L Li, H Deng, A Dong, Y Chang, H Zha - à€ ACM SIGIR à€, 2015 - dl.acm.org
1537
1538
1539 Cited by 11
1540
1541
1542 Topic segmentation with an aspect hidden Markov model
1543 DM Blei, PJ Moreno - à€ of the 24th annual international ACM SIGIR à€, 2001 - dl.acm.org
1544
1545
1546
1547 Cited by 183
1548
1549
1550 Information storage and retrieval : a survey and functional description
1551 J Minker - ACM SIGIR Forum, 1977 - dl.acm.org
1552
1553
1554
1555 Cited by 21
1556
1557
1558 Improving Exploratory Search Experience through Hierarchical Knowledge Graphs
1559 B Sarrafzadeh, E Lank - à€ of the 40th International ACM SIGIR à€, 2017 - dl.acm.org

1560
1561
1562 Cited by 0
1563
1564
1565
1566 Learning to suggest: a machine learning framework [for](#) ranking query suggestions
1567 U Ozertem, O Chapelle, P Donmez – àÃ© international ACM SIGIR àÃ©, 2012 – dl.acm.org
1568
1569
1570
1571 Cited by 52
1572
1573
1574 Fast First–Phase Candidate Generation [for](#) Cascading Rankers
1575 Q Wang, C Dimopoulos, T Suel – àÃ© of the 39th International ACM SIGIR àÃ©, 2016 – dl.acm.org
1576
1577
1578
1579 Cited by 1
1580
1581
1582 Explainable user clustering in [short](#) text streams
1583 Y Zhao, S Liang, Z Ren, J Ma, E YilmazàÃ© – àÃ© International ACM SIGIR àÃ©, 2016 – dl.acm.org
1584
1585
1586
1587 Cited by 13
1588
1589
1590 Searching by talking: Analysis of voice queries on mobile web search
1591 I Guy – àÃ© 39th International ACM SIGIR conference on Research àÃ©, 2016 – dl.acm.org
1592
1593
1594
1595 Cited by 21
1596
1597
1598 Exploring the limits of single–iteration clarification dialogs
1599 J Lin, P Wu, D Demner–Fushman, E Abels – àÃ© international ACM SIGIR àÃ©, 2006 – dl.acm.org
1600
1601
1602
1603 Cited by 5
1604
1605
1606 Novelty based ranking of human answers [for](#) community questions
1607 A Omari, D Carmel, O RokhlenkoàÃ© – àÃ© International ACM SIGIR àÃ©, 2016 – dl.acm.org
1608
1609
1610
1611 Cited by 9
1612
1613
1614 Accurately interpreting clickthrough data as implicit feedback
1615 T Joachims, L Granka, B Pan, H HembrookeàÃ© – àÃ© international ACM SIGIR àÃ©, 2005 – dl.acm.org
1616
1617
1618
1619 Cited by 1236
1620
1621
1622 Faster BlockMax WAND with Variable–sized Blocks
1623 A Mallia, G Ottaviano, E Porciani, N TonellottoàÃ© – àÃ© ACM SIGIR àÃ©, 2017 – dl.acm.org
1624
1625
1626 Cited by 0
1627
1628
1629 Structural relationships [for](#) large–scale learning of answer re–ranking

1630 A Severyn, A Moschitti — Proceedings of the 35th international ACM SIGIR à Ä, 2012 — dl.acm.org
1631
1632
1633
1634 Cited by 60
1635
1636
1637 [PDF] Introduction to Probabilistic Models for Information Retrieval
1638 V Lavrenko — à Ä of the 33rd International ACM SIGIR à Ä, 2010 — homepages.inf.ed.ac.uk
1639
1640
1641
1642 Cited by 4
1643
1644
1645 SPOT: Selecting occuPations frOm Trajectories
1646 P Li, J Yao, L Wang, X Lin — à Ä of the 40th International ACM SIGIR à Ä, 2017 — dl.acm.org
1647
1648
1649
1650
1651
1652 Cited by 0
1653
1654
1655 Entity query feature expansion using knowledge base links
1656 J Dalton, L Dietz, J Allan — à Ä of the 37th international ACM SIGIR à Ä, 2014 — dl.acm.org
1657
1658
1659
1660 Cited by 84
1661
1662
1663 Uncovering task based behavioral heterogeneities in online search behavior
1664 R Mehrotra, P Bhattacharya, E Yilmaz — à Ä of the 39th International ACM SIGIR à Ä, 2016 — dl.acm.org
1665
1666
1667
1668 Cited by 4
1669
1670
1671 When does Relevance Mean Usefulness and User Satisfaction in Web Search?
1672 J Mao, Y Liu, K Zhou, JY Nie, J Song, M Zhang à Ä of the ACM SIGIR conference à Ä, 2016 — dl.acm.org
1673
1674
1675
1676 Cited by 10
1677
1678
1679 Modeling annotated data
1680 DM Blei, MI Jordan — à Ä of the 26th annual international ACM SIGIR à Ä, 2003 — dl.acm.org
1681
1682
1683
1684 Cited by 1198
1685
1686 Aggregated search interfaces in multi-session tasks
1687 M Bron, J Van Gorp, F Nack à Ä of the SIGIR 2013: 36th à Ä, 2013 — dspace.library.uu.nl
1688
1689
1690
1691 Cited by 1
1692
1693
1694
1695
1696
1697
1698 Local Ranking Problem on the BrowseGraph
1699 M Trevisiol, LM Aiello, P Boldi, R Blanco — à Ä of the 38th International ACM SIGIR à Ä, 2015 — dl.acm.org

1700
1701
1702
1703 Cited by 1
1704
1705
1706
1707 Jointly Learning Word Embeddings and Latent Topics
1708 B Shi, W Lam, S Jameel, S Schockaert – àĂę International ACM SIGIR àĂę, 2017 – dl.acm.org
1709
1710
1711 Cited by 0
1712 Quads: question answering for decision support
1713 Z Yang, Y Li, J Cai, E Nyberg – àĂę of the 37th international ACM SIGIR àĂę, 2014 – dl.acm.org
1714
1715
1716
1717 Cited by 7
1718
1719
1720 Utilizing marginal net utility for recommendation in e-commerce
1721 J Wang, Y Zhang – Proceedings of the 34th international ACM SIGIR àĂę, 2011 – dl.acm.org
1722
1723
1724
1725 Cited by 49
1726
1727
1728 Relevance scores for triples from type-like relations
1729 H Bast, B Buchhold, E Haussmann – àĂę of the 38th International ACM SIGIR àĂę, 2015 – dl.acm.org
1730
1731
1732
1733 Cited by 25
1734
1735
1736
1737
1738 Learning for interaction extraction and verification from biological full articles
1739 E Kim, Y Song, GG Lee, BK Yi – àĂę of the ACM SIGIR 2004 àĂę, 2004 – pdfs.semanticscholar.org
1740
1741
1742
1743 Cited by 1
1744
1745
1746 Generating reformulation trees for complex queries
1747 X Xue, WB Croft – Proceedings of the 35th international ACM SIGIR àĂę, 2012 – dl.acm.org
1748
1749
1750
1751 Cited by 9
1752
1753
1754 Predicting quality flaws in user-generated content: the case of wikipedia
1755 M Anderka, B Stein, N Lipka – àĂę of the 35th international ACM SIGIR àĂę, 2012 – dl.acm.org
1756
1757
1758
1759 Cited by 56
1760
1761
1762
1763
1764
1765 Effective sentiment stream analysis with self-augmenting training and demand-driven projection
1766 IS Silva, J Gomide, A Veloso, W Meira Jr – àĂę international ACM SIGIR àĂę, 2011 – dl.acm.org
1767
1768
1769

1770 Cited by 43
1771
1772
1773 A generalized hidden markov model with discriminative training for query spelling correction
1774 Y Li, H Duan, CX Zhai – Proceedings of the 35th international ACM SIGIR àĘ, 2012 – dl.acm.org
1775
1776
1777
1778 Cited by 26
1779
1780
1781 Aggregated search interface preferences in multi-session search tasks
1782 M Bron, J Van Gorp, F Nack, LB BaltussenàĘ – àĘ ACM SIGIR conference àĘ, 2013 – dl.acm.org
1783
1784
1785
1786 Cited by 21
1787
1788
1789 An IR-based evaluation framework for web search query segmentation
1790 R Saha Roy, N Ganguly, M ChoudhuryàĘ – àĘ international ACM SIGIR àĘ, 2012 – dl.acm.org
1791
1792
1793
1794 Cited by 15
1795
1796
1797 A sequential decision formulation of the interface card model for interactive ir
1798 Y Zhang, C Zhai – Proceedings of the 39th International ACM SIGIR àĘ, 2016 – dl.acm.org
1799
1800
1801
1802 Cited by 2
1803
1804
1805 Random walks on the click graph
1806 N Craswell, M Szummer – àĘ of the 30th annual international ACM SIGIR àĘ, 2007 – dl.acm.org
1807
1808
1809
1810 Cited by 459
1811
1812
1813 User variability and IR system evaluation
1814 P Bailey, A Moffat, F Scholer, P Thomas – àĘ 38th International ACM SIGIR àĘ, 2015 – dl.acm.org
1815
1816
1817
1818 Cited by 27
1819
1820
1821 Learning to rank short text pairs with convolutional deep neural networks
1822 A Severyn, A Moschitti – Proceedings of the 38th International ACM SIGIR àĘ, 2015 – dl.acm.org
1823
1824
1825
1826 Cited by 195
1827
1828
1829 Accounting for the Correspondence in Commented Data
1830 R Cai, C Wang, H Wang – àĘ of the 40th International ACM SIGIR àĘ, 2017 – dl.acm.org
1831
1832
1833 Cited by 0
1834
1835
1836 Predicting searcher frustration
1837 H A Feild, J Allan, R Jones – àĘ of the 33rd international ACM SIGIR àĘ, 2010 – dl.acm.org
1838
1839

1840 Cited by 131
1841
1842
1843 Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings
1844 I VuliÄG, MF Moens – Proceedings of the 38th International ACM SIGIR â€, 2015 – dl.acm.org
1845
1846
1847
1848 Cited by 72
1849
1850
1851 Fusion helps diversification
1852 S Liang, Z Ren, M De Rijke – â€ of the 37th international ACM SIGIR â€, 2014 – dl.acm.org
1853
1854
1855
1856 Cited by 33
1857
1858
1859 Predicting user satisfaction with intelligent assistants
1860 J Kiseleva, K Williams, A Hassan Awadallahâ€ – â€ ACM SIGIR conference â€, 2016 – dl.acm.org
1861
1862
1863
1864 Cited by 12
1865
1866 Bayesian graphical models **for** adaptive filtering
1867 Y Zhang – SIGIR Forum, 2005 – www-2.cs.cmu.edu
1868
1869
1870
1871 Cited by 9
1872
1873
1874 Injecting user models and time into precision via Markov chains
1875 M Ferrante, N Ferro, M Maistro – â€ of the 37th international ACM SIGIR â€, 2014 – dl.acm.org
1876
1877
1878
1879 Cited by 12
1880
1881 Interaction pool: Towards a user-centered test collection
1882 H Joho, R Villa, J Jose – SIGIR'07 Web Information-Seeking and â€, 2007 - academia.edu
1883
1884
1885
1886 Cited by 4
1887
1888
1889 Language intent models **for** inferring user browsing behavior
1890 M Tsagkias, R Blanco – Proceedings of the 35th international ACM SIGIR â€, 2012 - dl.acm.org
1891
1892
1893
1894 Cited by 8
1895
1896
1897 Exploiting user feedback to learn to rank answers in q&a forums: a case study with stack overflow
1898 DH Dalip, MA GonÃ‡alves, M Cristoâ€ – â€ international ACM SIGIR â€, 2013 - dl.acm.org
1899
1900
1901
1902 Cited by 37
1903
1904
1905 Exploiting hybrid contexts **for** tweet segmentation
1906 C Li, A Sun, J Weng, Q He – â€ of the 36th international ACM SIGIR â€, 2013 - dl.acm.org
1907
1908
1909

Listing 4: The content of sigir2kplus.txt

Question 6:

Extra Credit: SVMlight, 10 points extra credit:

see: http://www.cs.cornell.edu/People/tj/svm_light/

* 1 point:

work through the "Inductive SVM" example, discuss in detail the steps and resulting output

Answer:

I downloaded and installed the SVM light from:

http://www.cs.cornell.edu/people/tj/svm_light/

I followed the installation instructions and downloaded the Inductive SVM example.

I also followed the instructions to run svm_learn and passed the training data file train.dat as a command line argument. It produced the model file, which I used as an input for the program svm_classify.

I saved the entire session in the file svm.txt under Q6 folder.

Below is the content of the file.

```
1 hhallak@atria:~$ wget http://download.joachims.org/svm_light/current/svm_light.tar.gz
2 --2017-12-15 03:32:51-- http://download.joachims.org/svm_light/current/svm_light.tar.gz
3 Resolving download.joachims.org (download.joachims.org)... 81.88.34.174, 81.88.42.187
4 Connecting to download.joachims.org (download.joachims.org)|81.88.34.174|:80... connected.
5 HTTP request sent, awaiting response... 302 Found
6 Location: http://osmot.cs.cornell.edu/svm_light/current/svm_light.tar.gz [following]
7 --2017-12-15 03:32:52-- http://osmot.cs.cornell.edu/svm_light/current/svm_light.tar.gz
8 Resolving osmot.cs.cornell.edu (osmot.cs.cornell.edu) ... 128.253.51.182
9 Connecting to osmot.cs.cornell.edu (osmot.cs.cornell.edu) |128.253.51.182|:80... connected.
10 HTTP request sent, awaiting response... 200 OK
11 Length: 51026 (50K) [application/x-gzip]
12 Saving to: Šsvm_light.tar.gzŠ
13
14 100%[=====] 51,026 203KB/s in 0.2s
15
16 2017-12-15 03:32:52 (203 KB/s) - Šsvm_light.tar.gzŠ saved [51026/51026]
17
18 hhallak@atria:~$ ls
19 532 git mid779 svm_light
20 779 Gorgon Music svm_light.tar.gz
21 cs476Drwahab_hhallak_cert_request Output.odt Templates
22 cs772 hhallak_privatekey.pem Pictures Videos
23 cs772A2.pdf hhallakPrivateKey.pem Public WINDOWS
24 Desktop hhallak_publickey.pem public_html win_user_profile
25 Documents hhallakPublicKey.pem python workspace
26 Downloads hussam setenvwahab
27 final M1 sqlnet.log
28 hhallak@atria:~$ cp svm_light.tar.gz ./svm_light
29 hhallak@atria:~$ cd svm_light/
30 hhallak@atria:~/svm_light$ ls
31 svm_light.tar.gz
32 hhallak@atria:~/svm_light$ gunzip -c svm_light.tar.gz | tar xvf --
33 LICENSE.txt
34 Makefile
35 svm_learn.c
36 kernel.h
37 svm_learn.h
38 svm_learn_main.c
39 svm_classify.c
40 svm_loqo.c
41 svm_common.c
42 svm_common.h
43 svm_hideo.c
44 hhallak@atria:~/svm_light$ make all
45 gcc -c -O3 svm_learn_main.c -o svm_learn_main.o
```

```

46 gcc -c -O3           svm_learn.c -o svm_learn.o
47 gcc -c -O3           svm_common.c -o svm_common.o
48 svm_common.c: In function `read_model':
49 svm_common.c:600:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
50   fscanf(modelfl,"SVM-light Version %s\n",version_buffer);
51
52 svm_common.c:605:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
53   fscanf(modelfl,"%d*[\n]", &model->kernel_parm.kernel_type);
54
55 svm_common.c:606:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
56   fscanf(modelfl,"%d*[\n]", &model->kernel_parm.poly_degree);
57
58 svm_common.c:607:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
59   fscanf(modelfl,"%f*[\n]", &model->kernel_parm.rbf_gamma);
60
61 svm_common.c:608:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
62   fscanf(modelfl,"%f*[\n]", &model->kernel_parm.coef_lin);
63
64 svm_common.c:609:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
65   fscanf(modelfl,"%f*[\n]", &model->kernel_parm.coef_const);
66
67 svm_common.c:610:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
68   fscanf(modelfl,"%[^#]*[\n]", model->kernel_parm.custom);
69
70 svm_common.c:612:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
71   fscanf(modelfl,"%d*[\n]", &model->totwords);
72
73 svm_common.c:613:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
74   fscanf(modelfl,"%d*[\n]", &model->totdoc);
75
76 svm_common.c:614:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
77   fscanf(modelfl,"%d*[\n]", &model->sv_num);
78
79 svm_common.c:615:9: warning: ignoring return value of `fscanf', declared with attribute warn_unused_result [-Wunused-result]
80   fscanf(modelfl,"%f*[\n]", &model->b);
81
82 svm_common.c:623:10: warning: ignoring return value of `fgets', declared with attribute warn_unused_result [-Wunused-result]
83   fgets(line,(int)ll,modelfl);
84
85 gcc -c -O3           svm_hideo.c -o svm_hideo.o
86 gcc -O3             svm_learn_main.o svm_learn.o svm_common.o svm_hideo.o -o svm_learn -L.
     -lm
87 gcc -c -O3           svm_classify.c -o svm_classify.o
88 gcc -O3             svm_classify.o svm_common.o -o svm_classify -L. -lm
89 hhallak@atria:~/svm_light$ ls
90 kernel.h      svm_classify.c  svm_common.o  svm_learn.c    svm_learn.o
91 LICENSE.txt   svm_classify.o  svm_hideo.c   svm_learn.h   svm_light.tar.gz
92 Makefile      svm_common.c   svm_hideo.o   svm_learn_main.c  svm_loqo.c
93 svm_classify  svm_common.h   svm_learn     svm_learn_main.o
94 hhallak@atria:~/svm_light$ wget http://download.joachims.org/svm_light/examples/example1.tar.gz
95 --2017-12-15 03:38:15-- http://download.joachims.org/svm_light/examples/example1.tar.gz
96 Resolving download.joachims.org (download.joachims.org)... 81.88.34.174, 81.88.42.187
97 Connecting to download.joachims.org (download.joachims.org)|81.88.34.174|:80... connected.
98 HTTP request sent, awaiting response... 302 Found
99 Location: http://osmot.cs.cornell.edu/svm_light/examples/example1.tar.gz [following]
100 --2017-12-15 03:38:15-- http://osmot.cs.cornell.edu/svm_light/examples/example1.tar.gz
101 Resolving osmot.cs.cornell.edu (osmot.cs.cornell.edu)... 128.253.51.182
102 Connecting to osmot.cs.cornell.edu (osmot.cs.cornell.edu) |128.253.51.182|:80... connected.

```

```

103 | HTTP request sent, awaiting response... 200 OK
104 | Length: 1223606 (1.2M) [application/x-gzip]
105 | Saving to: Šexample1.tar.gzŠ
106 |
107 | 100%[=====] 1,223,606 1020KB/s in 1.2s
108 |
109 | 2017-12-15 03:38:17 (1020 KB/s) – Šexample1.tar.gzŠ saved [1223606/1223606]
110 |
111 hhallak@atria:~/svm_light$ ls
112 example1.tar.gz  svm_classify.c  svm_hideo.c  svm_learn_main.c
113 kernel.h        svm_classify.o  svm_hideo.o  svm_learn_main.o
114 LICENSE.txt     svm_common.c   svm_learn   svm_learn.o
115 Makefile        svm_common.h   svm_learn.c  svm_light.tar.gz
116 svm_classify   svm_common.o   svm_learn.h  svm_logo.c
117 hhallak@atria:~/svm_light$ gunzip -c example1.tar.gz | tar xvf -
118 example1/
119 example1/train.dat
120 example1/test.dat
121 example1/words
122 hhallak@atria:~/svm_light$ ./svm_learn example1/train.dat example1/model
123 Scanning examples...done
124 Reading examples into memory
125 ...100..200..300..400..500..600..700..800..900..1000..1100..1200..1300..1400..1500..1600..1700..1800..1900..2000..
126 Setting default regularization parameter C=1.0000
127 Optimizing ..... done. (425 iterations)
128 Optimization finished (5 misclassified, maxdiff=0.00085).
129 Runtime in cpu-seconds: 0.10
130 Number of SV: 878 (including 117 at upper bound)
131 L1 loss: loss=35.67674
132 Norm of weight vector: |w|=19.55576
133 Norm of longest example vector: |x|=1.00000
134 Estimated VCdim of classifier: VCdim<=383.42791
135 Computing XiAlpha-estimates...done
136 Runtime for XiAlpha-estimates in cpu-seconds: 0.00
137 XiAlpha-estimate of the error: error<=5.85% (rho=1.00,depth=0)
138 XiAlpha-estimate of the recall: recall=>95.40% (rho=1.00,depth=0)
139 XiAlpha-estimate of the precision: precision=>93.07% (rho=1.00,depth=0)
140 Number of kernel evaluations: 45954
141 Writing model file ... done
142 hhallak@atria:~/svm_light$ ./svm_classify example1/test.dat example1/model example1/predictions
143 Reading model...OK. (878 support vectors read)
144 Classifying test examples ..100..200..300..400..500..600.. done
145 Runtime (without IO) in cpu-seconds: 0.00
146 Accuracy on test set: 97.67% (586 correct, 14 incorrect, 600 total)
147 Precision/recall on test set: 96.43%/99.00%

```

Listing 5: The content of svm.txt

The results were printed on the screen:

Accuracy on test set: 97.67

Precision/recall on test set: 96.43

The results show that these scores are very high for these measures.

Figure 15: Inductive SVM Example

References

- [1] Stackoverflow. <https://stackoverflow.com/questions/tagged/python>.
- [2] http://www.cs.cornell.edu/people/tj/svm_light/