CS221 Project 3 Search Engine - M3 Report

Please enter your query: Search...

Search

Team:

Jiyang Yan 64923885
Xiaoyan Qu 54848408
Shayang Zang 36326408

Content:

mondego					1
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	2	1.00	2.00	2.00	0.40
2	0	1.58	0.00	2.00	0.27
3	0	2.00	0.00	2.00	0.22
4	0	2.32	0.00	2.00	0.20
5	0	2.58	0.00	2.00	0.19
			Avg. NDCG5 =		0.26

machine learning					2
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	3	1.00	3.00	3.00	0.60
2	0	1.58	0.00	3.00	0.40
3	0	2.00	0.00	3.00	0.33
4	0	2.32	0.00	3.00	0.30
5	0	2.58	0.00	3.00	0.29
			Avg. NDCG5 =		0.39

software engineering					3
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

security					4
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	2	1.00	2.00	2.00	0.40
2	0	1.58	0.00	2.00	0.27
3	0	2.00	0.00	2.00	0.22
4	0	2.32	0.00	2.00	0.20
5	0	2.58	0.00	2.00	0.19
			Avg. NDCG5 =		0.26

student affairs					5
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

graduate courses					6
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

Crista Lopes					7
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

REST					8
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	4	1.00	4.00	4.00	0.80
2	0	1.58	0.00	4.00	0.53
3	0	2.00	0.00	4.00	0.44
4	0	2.32	0.00	4.00	0.40
5	0	2.58	0.00	4.00	0.39
			Avg. NDCG5 =		0.51

computer games					9
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

information retrieval					10
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

Google Result				
i	reli	log(i + 1)	reli / log(i + 1)	DCG
1	5	1.00	5.00	5.00
2	4	1.58	2.52	7.52
3	3	2.00	1.50	9.02
4	2	2.32	0.86	9.89
5	1	2.58	0.39	10.27
		DCG5 =	10.27	41.70

Total NDCG of 10 queries =	1.41
Average NDCG of 10 queries =	0.14

mondego					1
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	5	1.00	5.00	5.00	1.00
2	0	1.58	0.00	5.00	0.66
3	0	2.00	0.00	5.00	0.55
4	0	2.32	0.00	5.00	0.51
5	0	2.58	0.00	5.00	0.49
			Avg. NDCG5 =		0.64

machine learning					2
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

software engineering					3
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	5	1.00	5.00	5.00	1.00
2	0	1.58	0.00	5.00	0.66
3	0	2.00	0.00	5.00	0.55
4	3	2.32	1.29	6.29	0.64
5	0	2.58	0.00	6.29	0.61
			Avg. NDCG5 =		0.69

security					4
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	5	2.32	2.15	2.15	0.22
5	0	2.58	0.00	2.15	0.21
			Avg. NDCG5 =		0.09

student affairs					5
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	3	1.00	3.00	3.00	0.60
2	0	1.58	0.00	3.00	0.40
3	0	2.00	0.00	3.00	0.33
4	0	2.32	0.00	3.00	0.30
5	0	2.58	0.00	3.00	0.29
			Avg. NDCG5 =		0.39

graduate courses					6
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	5	2.58	1.94	1.94	0.19
			Avg. NDCG5 =		0.04

Crista Lopes					7
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	0	1.00	0.00	0.00	0.00
2	0	1.58	0.00	0.00	0.00
3	0	2.00	0.00	0.00	0.00
4	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00	0.00	0.00
			Avg. NDCG5 =		0.00

REST					8
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	5	1.00	5.00	5.00	1.00
2	0	1.58	0.00	5.00	0.66
3	4	2.00	2.00	7.00	0.78
4	0	2.32	0.00	7.00	0.71
5	0	2.58	0.00	7.00	0.68
			Avg. NDCG5 =		0.77

computer games					9
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	5	1.00	5.00	5.00	1.00
2	0	1.58	0.00	5.00	0.66
3	0	2.00	0.00	5.00	0.55
4	0	2.32	0.00	5.00	0.51
5	0	2.58	0.00	5.00	0.49
			Avg. NDCG5 =		0.64

information retrieval					10
i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1	4	1.00	4.00	4.00	0.80
2	0	1.58	0.00	4.00	0.53
3	0	2.00	0.00	4.00	0.44
4	5	2.32	2.16	6.16	0.62
5	0	2.58	0.00	6.16	0.60
			Avg. NDCG5 =		0.60

Google Result					
i	reli	log(i + 1)		reli / log(i + 1)	DCG
1	5	1.00		5.00	5.00
2	4	1.58		2.52	7.52
3	3	2.00		1.50	9.02
4	2	2.32		0.86	9.89
5	1	2.58		0.39	10.27
		DCG5 =	0.00	10.27	41.70
Total NDCG of 10 queries =	3.85				
Average NDCG of 10 queries =	0.39				

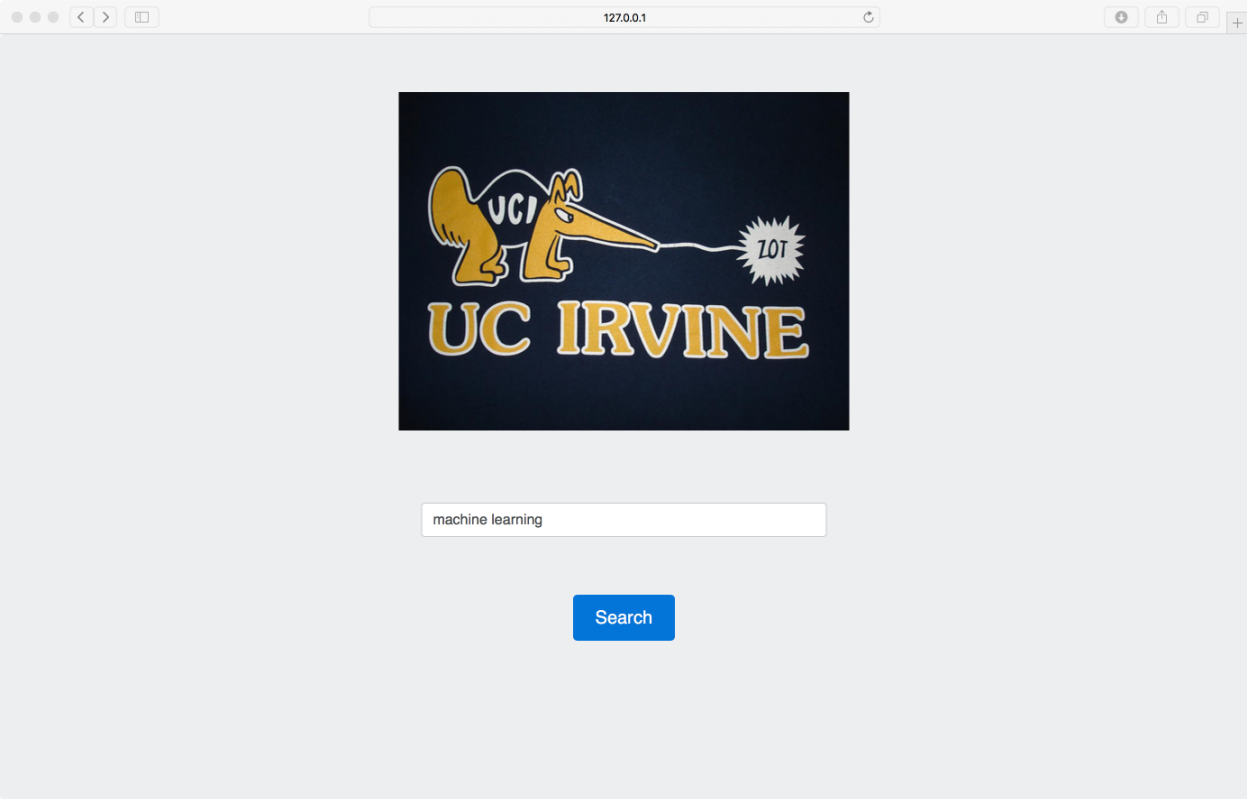
Previously, our model only used **TF-IDF** to rank all the candidate links.

To improve, we implemented the following besides **TF-IDF**:

1. include the **hubness** of selected candidate links
2. include **page rank** score of selected candidate links
3. include **title score** of those pages whose titles contain some word in the query
4. include **url score** of those pages whose urls contain some word in the query

Implementation details:

https://github.com/yjy1663/Search-Engine



Searched key word(s): machine learning

- UCI Machine Learning Repository
  www.ics.uci.edu/~mlearn/Machine-Learning.html
  We currently maintain 360 data sets as a service to the machine learning community.
- Michael Pazzani: Publications on Machine Learning, KDD and Artificial Intelligence
  www.ics.uci.edu/~pazzani/Publications/OldPublications.html
  G. Webb, Michael J. Pazzani, Daniel Billsus, (2001).Machine learning for user modeling. Daniel Billsus, Michael J. Pazzani, James Chen: A learning agent for wireless news access. Intelligent User In
- Index of /ml/machine-learning-databases/housing
  archive.ics.uci.edu/ml/machine-learning-databases/housing?C=N;O=A
  Index of /ml/machine-learning-databases/housing
- Index of /ml/machine-learning-databases/00369
  archive.ics.uci.edu/ml/machine-learning-databases/00369?C=D;O=A
  Index of /ml/machine-learning-databases/00369
- Index of /ml/machine-learning-databases/concrete/slump
  archive.ics.uci.edu/ml/machine-learning-databases/concrete/slump?C=M;O=A