

Please enter your query: Search...

Search

Team:

Jiyang Yan 64923885 Xiaoyan Qu 54848408 Shayang Zang 36326408

Content:

Page1: NDCG of our initial Search Engine against Google

Page2: NDCG of our improved Search Engine against Google

Page3: Ranking Improvements

Page4: UI Showcase

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 le	arning				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 ei					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 aff		16 . 4		DCC	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 cou	ırses				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 i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5
1				0.00	
2	0	1.00	0.00		0.00
3	0	1.58	0.00	0.00	0.00
4	0	2.00	0.00	0.00	0.00
5	0	2.32	0.00	0.00	0.00
5	0	2.58	0.00 Avg. NDCG5 =	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
-		2.50	Avg. NDCG5 =	1.00	0.51
			7.1.50 1.12 0.00		
computer ga	mes				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 le	arning				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 e	ngineering				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
		2.30	Avg. NDCG5 =	0.27	0.69
			Avg. NDCOS -		0.07
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	0.00
5	0	2.58	0.00	2.15	0.22
J	0	2.30	Avg. NDCG5 =	2.13	0.21
			Avg. NDCGJ =		0.07
student aff	airc				5
i 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	0.60
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 0.04 0.04 0.04 Crista Lopes 7 reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 0 1.00 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 8EST 8 0.00 5.00 1.00							
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 = Crista Lopes i reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 0 1.00 0.00 0.00 0.00 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 2 0 1.58 0.00 5.00 1.00 0.00 2 0 1.58 0.00 5.00 1.00 0.00 0.00 0.00	graduate co	urses					
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 = Crista Lopes i reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 0 1.00 0.00 0.00 0.00 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 4 2.00 2.00 7.00 0.70 0.00 0.00 0.00 1.00 1.00 1.00 1.00 1.00 1.00	i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5	
3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 5 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 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 5 0 2.58 0.00 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 2.32 0.00 7.00 0.71 5 0 2.58 0.00 7.00 0.71 5 0 2.58 0.00 7.00 0.71 5 0 2.58 0.00 7.00 0.71 5 0 2.58 0.00 7.00 0.07 Computer games 9 i reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 5 1.00 5.00 5.00 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 0.66 3 0 2.00 0.00 5.00 0.66 3 0 2.00 0.00 5.00 0.66 3 0 2.00 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.05 4 0 2.32 0.00 5.00 0.05 4 0 2.32 0.00 5.00 0.06 3 0 2.00 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.55 4 0 2.32 0.00 5.00 0.66 3 0 2.00 0.00 5.00 0.55 4 0 0 2.32 0.00 5.00 0.55 4 0 0 2.32 0.00 5.00 0.55 4 0 0 2.32 0.00 5.00 0.66 3 0 0 0.66 3 0 0 0.00 5.00 0.55 4 0 0 0.35 4 0 0 0.35 4 0 0 0.35 4 0 0 0.35 4 0 0 0.35 5 0 0 0.66 4 0 0 0.80 4 0 0 0.80 4 0 0 0.80 4 0 0 0.80	1	0	1.00	0.00	0.00	0.00	
A		0	1.58	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 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 5 0 2.58 0.00 0.00 0.00 0.00 REST 8 8 0.00 5.00 1.00 0.00 2 0 1.58 0.00 5.00 1.00 0.66 3 4 2.00 2.00 7.00 0.78 0.77 4 0 2.32 0.00 7.00 0.68 0.77 Avg. NDCG5 = <td col<="" td=""><td>3</td><td>0</td><td>2.00</td><td>0.00</td><td>0.00</td><td>0.00</td></td>	<td>3</td> <td>0</td> <td>2.00</td> <td>0.00</td> <td>0.00</td> <td>0.00</td>	3	0	2.00	0.00	0.00	0.00
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 2 0 2.58 0.00 0.00 0.00 2 0 1.58 0.00 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 0.68 3 i reli log(i + 1) reli / log(i + 1) DCG NDCG5 1 5 1.00 <td>4</td> <td>0</td> <td>2.32</td> <td>0.00</td> <td>0.00</td> <td>0.00</td>	4	0	2.32	0.00	0.00	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 0.00 2 0 1.58 0.00 0.00 0.00 0.00 3 0 2.00 0.00 0.00 0.00 0.00 4 0 2.32 0.00 0.00 0.00 0.00 5 0 2.58 0.00 0.00 0.00 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 0.71 5 0 2.58 0.00 7.00 0.68 4 0 2.32 0.00 5.00 1.00 </td <td>5</td> <td>5</td> <td>2.58</td> <td>1.94</td> <td>1.94</td> <td>0.19</td>	5	5	2.58	1.94	1.94	0.19	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$				Avg. NDCG5 =		0.04	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$							
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 REST 8 0.00 0.00 0.00 2 0 1.58 0.00 5.00 1.00 2 0 1.58 0.00 5.00 0.78 4 0 2.32 0.00 7.00 0.78 4 0 2.32 0.00 7.00 0.68 Avg. NDCG5 = 0.77 0.68 0.77 0.68 1 5 1.00 5.00 5.00 1.00 2 0 1.58 0.00 7.00 0.68 3 0 2.00 5.00 5.00 1.00 1 5	Crista Lope:	S				7	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5	
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 0.00 0.00 REST 8 0.00 5.00 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 0.68 Avg. NDCG5 = 0.77 computer games 9 0.70 0.68 1 5 1.00 5.00 5.00 1.00 2 0 1.58 0.00 5.00 1.00 0.66 3 0 2.00 0.00 5.00 1.00 0.66 0.00 0.50 0.55 0.50 0.55 0.50	1	0	1.00	0.00	0.00	0.00	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	2	0	1.58	0.00	0.00	0.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	3	0	2.00	0.00	0.00	0.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	4	0	2.32	0.00	0.00	0.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	5	0	2.58	0.00	0.00	0.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$				Avg. NDCG5 =		0.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$							
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	REST					8	
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5	
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 0.64 information retrieval 10 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 <	1	5	1.00		5.00	1.00	
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 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	2	0	1.58	0.00	5.00	0.66	
5 0 2.58 0.00 7.00 0.68 Avg. NDCG5 = 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	4	2.00	2.00	7.00	0.78	
Avg. NDCG5 = 0.77	4	0	2.32	0.00	7.00	0.71	
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 10 10 10 1 4 1.00 4.00 4.00 0.80 2 0 1.58 0.00 4.00 0.53	5	0	2.58	0.00	7.00	0.68	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$				Avg. NDCG5 =		0.77	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$							
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	computer g	ames				9	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	i	reli	log(i + 1)	reli / log(i + 1)	DCG	NDCG5	
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	1	5	1.00		5.00	1.00	
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	2	0	1.58	0.00	5.00	0.66	
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	5.00	0.55	
Avg. NDCG5 = 0.64	4	0	2.32	0.00	5.00	0.51	
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	5	0	2.58	0.00	5.00	0.49	
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				Avg. NDCG5 =		0.64	
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							
1 4 1.00 4.00 4.00 0.80 2 0 1.58 0.00 4.00 0.53	information	retrieval				10	
2 0 1.58 0.00 4.00 0.53	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	3	0	2.00	0.00	4.00	0.44	
4 5 2.32 2.16 6.16 0.62		5	2.32	2.16	6.16	0.62	
5 0 2.58 0.00 6.16 0.60	5	0	2.58	0.00	6.16	0.60	
Avg. NDCG5 = 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				

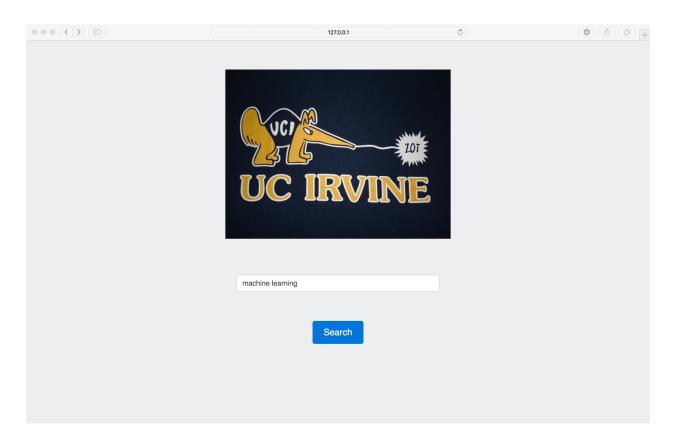
Previously, our model only used **TF-IDF** to rank all the candidate links.

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