

ЛР2: Оценка качества поиска

Задание

Необходимо оценить качество двух поисков и сравнить их с друг другом (для Википедии можно собственный поиск по Википедии, поиск Google или Яндекса с ограничением по сайту Википедии). Как минимум, нужно измерить P, DCG, NDCG и ERR уровней @1, @3 и @5, приветствуется использование дополнительных метрик качества. Для оценки качества необходимо придумать 30 запросов, отражающих интересы пользователей или, если есть доступ к настоящим запросам пользователей, то выбрать репрезентативную подборку.

Метод решения

1. Выбрать поисковики для сравнения друг с другом (выбраны поисковики Yandex и Wikipedia).
2. Придумать запросы
3. По формулам вычислить качество результата двух данных поисковиков, посчитать средние значения и сделать вывод о лучшем поиске

Результаты выполнения

№	Текст запроса
1	economic difficulties
2	Indian Prime Minister
3	Latest news
4	Germany announces
5	Indonesian president
6	Moussaoui
7	Australian federal electoral
8	United States
9	train lines
10	Princeton report
11	A380
12	Barclays Bank credit rating
13	First Gold for NZ
14	Swimming Championships
15	Memorial Cup ring
16	UK
17	IOC lifts
18	Greek
19	NHL
20	drone strike
21	Poetry Bookshop
22	Sunderland manager
23	Houston Dynamos

24	Helicopter
25	China
26	retail development
27	Blagojevich
28	South Africa
29	torrential rains
30	ICC World Twenty20

Оценки поисковиков

№	Метрика	Яндекс			Википедия				ERR	0.97	0.99	0.99	0.50	0.77	0.79
		@1	@3	@5	@1	@3	@5								
1	P	1.00	0.67	0.80	0.00	0.33	0.20	9	P	1.00	1.00	0.80	1.00	1.00	1.00
	CG	1.00	11.0	20.0	1.00	7.00	9.00		CG	5.00	14.0	22.0	5.00	13.0	22.0
	DCG	31.0	51.1	70.3	1.00	17.1	17.9		DCG	5.00	56.0	72.0	31.0	41.0	57.0
	NDCG	6.20	3.41	2.81	1.00	1.14	0.72		NDCG	1.00	3.73	2.88	6.20	2.73	2.28
	ERR	0.97	0.96	1.03	0.50	1.13	0.60		ERR	0.97	0.99	0.99	0.97	0.98	0.98
2	P	0.00	0.60	0.85	0.00	0.33	0.60	10	P	0.00	0.33	0.60	1.00	1.00	1.00
	CG	1.00	6.00	13.0	3.00	5.00	13.0		CG	1.00	9.00	13.0	3.00	9.00	15.0
	DCG	1.00	9.13	18.0	7.00	8.13	20.4		DCG	1.00	24.1	26.6	7.00	14.9	20.7
	NDCG	1.00	0.76	0.9	2.33	0.54	1.02		NDCG	1.00	1.61	1.06	2.33	1.66	1.38
	ERR	0.50	1.07	1.02	0.88	0.50	0.33		ERR	0.50	0.74	0.75	0.88	1.18	1.20
3	P	1.00	1.00	1.00	1.00	1.00	0.80	11	P	1.00	0.33	0.60	1.00	0.33	0.40
	CG	5.00	15.0	25.0	5.00	15.0	24.0		CG	5.00	8.00	18.0	5.00	10.0	14.0
	DCG	31.0	55.5	80.9	31.0	55.5	69.3		DCG	31.0	33.4	58.8	31.0	36.9	39.4
	NDCG	6.20	3.70	3.24	6.20	3.70	2.78		NDCG	6.20	2.23	2.35	6.20	2.46	1.58
	ERR	0.97	0.97	0.97	0.97	0.97	0.97		ERR	0.97	0.97	0.97	0.97	0.97	0.99
4	P	1.00	1.00	0.80	1.00	1.00	0.80	12	P	1.00	1.00	0.20	0.00	0.33	0.60
	CG	5.00	14.0	21.0	3.00	6.00	8.00		CG	5.00	15.0	19.0	2.00	8.00	14.0
	DCG	31.0	58.1	71.4	7.00	9.40	10.2		DCG	31.0	55.5	58.9	3.00	19.1	24.9
	NDCG	6.20	3.87	2.86	2.33	1.04	0.68		NDCG	6.20	3.70	2.36	1.50	1.27	1.00
	ERR	0.97	1.00	1.02	0.88	0.90	0.94		ERR	0.97	0.97	0.96	0.75	0.98	0.99
5	P	1.00	1.00	0.80	0.00	0.33	0.40	13	P	1.00	0.67	0.60	0.00	0.00	0.00
	CG	3.00	11.0	17.0	1.00	5.00	11.0		CG	3.00	10.0	16.0	1.00	3.00	5.00
	DCG	7.00	30.1	37.2	1.00	5.93	18.4		DCG	7.00	28.1	41.9	1.00	2.13	2.95
	NDCG	2.33	2.01	1.49	0.33	0.66	1.23		NDCG	2.33	1.87	1.68	1.00	0.14	0.59
	ERR	0.88	0.89	0.89	0.50	0.77	0.79		ERR	0.88	1.11	1.13	0.50	0.58	0.58
6	P	1.00	1.00	0.80	1.00	1.00	0.80	14	P	1.00	0.33	0.60	1.00	0.67	0.40
	CG	5.00	15.0	20.0	5.00	14.0	20.0		CG	5.00	7.00	13.0	4.00	10.0	12.0
	DCG	31.0	55.5	59.7	31.0	56.0	68.4		DCG	31.0	32.1	46.1	15.0	35.2	35.9
	NDCG	6.20	3.70	2.39	6.20	3.73	2.74		NDCG	6.20	2.14	1.84	3.75	2.35	1.44
	ERR	0.97	0.97	1.30	0.97	0.99	0.99		ERR	0.97	0.98	0.98	0.94	0.97	0.97
7	P	1.00	1.00	1.00	1.00	0.67	0.80	15	P	1.00	0.33	0.60	0.00	0.00	0.20
	CG	5.00	14.0	23.0	5.00	11.0	19.0		CG	5.00	7.00	15.0	1.00	3.00	9.00
	DCG	31.0	51.1	70.3	31.0	51.1	63.4		DCG	31.0	32.1	48.2	1.00	2.13	15.9
	NDCG	6.20	3.41	3.52	6.20	3.41	2.54		NDCG	6.20	2.14	1.93	1.00	0.71	0.64
	ERR	0.97	0.98	0.98	0.97	0.99	1.01		ERR	0.97	0.98	1.03	0.50	0.58	0.74
8	P	1.00	0.40	0.80	0.00	0.40	1.00	16	P	1.00	1.00	0.80	1.00	1.00	1.00
	CG	5.00	11.0	20.0	1.00	11.0	13.0		CG	5.00	14.0	18.0	5.00	14.0	24.0
	DCG	31.0	47.1	65.6	1.00	36.1	36.9		DCG	31.0	56.0	59.1	31.0	56.0	81.4
	NDCG	6.20	3.14	2.62	1.00	2.41	1.48		NDCG	6.20	3.73	2.36	6.20	3.73	3.26
									ERR	0.97	0.98	0.99	0.97	0.98	0.98

17	P	1.00	0.67	0.80	0.00	0.00	0.00	24	P	1.00	0.33	0.60	0.00	0.33	0.20
	CG	5.00	11.0	17.0	1.00	3.00	5.00		CG	4.00	8.00	17.0	2.00	8.00	10.0
	DCG	31.0	41.0	54.7	1.00	2.13	2.95		DCG	15.0	16.1	34.6	3.00	23.1	23.4
	NDCG	6.20	2.73	2.16	1.00	0.71	0.59		NDCG	3.75	1.34	1.35	1.50	1.54	0.94
	ERR	0.97	0.98	0.98	0.50	0.58	0.58		ERR	0.94	0.97	0.98	0.75	0.83	0.83
18	P	1.00	1.00	1.00	1.00	0.67	0.80	25	P	1.00	1.00	1.00	1.00	0.67	0.40
	CG	5.00	14.0	21.0	5.00	12.0	16.0		CG	5.00	13.0	21.0	3.00	9.00	11.0
	DCG	5.00	9.53	18.4	31.0	52.1	55.5		DCG	31.0	50.9	66.0	7.00	23.1	24.0
	NDCG	1.00	0.64	0.74	6.20	3.47	2.24		NDCG	6.20	3.39	2.64	2.33	1.54	0.96
	ERR	0.97	0.98	0.98	0.97	0.99	0.99		ERR	0.97	0.98	0.99	0.88	0.91	1.06
19	P	1.00	1.00	1.00	1.00	0.67	0.60	26	P	1.00	1.00	1.00	1.00	1.00	0.80
	CG	5.00	15.0	25.0	5.00	8.00	15.0		CG	5.00	13.0	22.0	5.00	14.0	21.0
	DCG	31.0	55.5	80.9	31.0	35.9	49.7		DCG	31.0	41.0	57.0	31.0	56.0	65.2
	NDCG	6.20	3.70	3.24	2.07	2.39	1.99		NDCG	6.20	2.73	2.28	6.20	3.73	2.61
	ERR	0.97	0.97	0.97	0.97	0.97	0.97		ERR	0.97	0.98	1.08	0.97	0.99	1.00
20	P	1.00	0.67	0.40	0.00	0.67	0.40	27	P	1.00	1.00	1.00	1.00	0.33	0.20
	CG	5.00	10.0	12.0	5.00	11.0	16.0		CG	5.00	15.0	25.0	5.00	8.00	11.0
	DCG	31.0	39.1	40.0	31.0	42.0	48.2		DCG	31.0	55.5	80.9	31.0	33.1	34.8
	NDCG	6.20	2.61	1.60	6.20	3.50	2.41		NDCG	6.20	3.70	3.24	6.20	2.21	1.39
	ERR	0.97	0.99	0.99	0.50	0.77	0.79		ERR	0.97	0.97	0.97	0.97	0.98	0.98
21	P	1.00	0.33	0.60	1.00	1.00	1.00	28	P	1.00	1.00	0.80	1.00	1.00	1.00
	CG	5.00	7.00	15.0	5.00	14.0	20.0		CG	5.00	15.0	25.0	5.00	15.0	22.0
	DCG	31.0	50.9	60.1	31.0	41.0	94.7		DCG	31.0	66.1	91.5	31.0	66.1	75.3
	NDCG	6.20	3.39	2.40	6.20	2.73	3.79		NDCG	6.20	4.41	3.66	6.20	4.41	3.01
	ERR	0.97	0.98	0.98	0.97	0.98	1.07		ERR	0.97	0.97	1.09	0.97	0.97	1.14
22	P	1.00	0.33	0.40	0.00	0.00	0.20	29	P	1.00	1.00	1.00	1.00	1.00	1.00
	CG	5.00	7.00	12.0	1.00	3.00	9.00		CG	5.00	14.0	24.0	4.00	14.0	24.0
	DCG	31.0	32.1	39.0	1.00	2.13	14.6		DCG	31.0	58.1	83.5	15.0	50.1	75.5
	NDCG	6.20	2.14	1.56	1.00	0.71	0.58		NDCG	6.20	3.87	3.34	3.75	3.34	3.02
	ERR	0.97	0.98	0.98	0.50	0.58	0.60		ERR	0.97	0.98	1.16	0.94	0.97	0.97
23	P	1.00	0.67	0.80	1.00	1.00	0.80	30	P	1.00	1.00	0.80	1.00	1.00	0.60
	CG	5.00	13.0	18.0	5.00	13.0	18.0		CG	5.00	12.0	18.0	5.00	12.0	14.0
	DCG	31.0	54.1	60.4	31.0	58.1	64.4		DCG	31.0	44.0	56.4	31.0	44.0	44.8
	NDCG	6.20	3.61	2.42	6.20	3.87	2.58		NDCG	6.20	2.93	2.26	6.20	2.93	1.79
	ERR	0.97	0.99	0.99	0.97	0.98	0.98		ERR	0.97	0.98	0.98	0.97	0.98	0.99

Средние значения

Метрика	Яндекс			Википедия		
	@1	@3	@5	@1	@3	@5
P	0.93	0.79	0.73	0.63	0.62	0.60
CG	4.60	9.77	18.8	3.53	9.60	14.8
DCG	45.8	42.9	56.9	17.8	32.1	41.9
NDCG	4.96	2.87	2.21	3.66	2.29	2.11
ERR	0.93	0.88	0.97	0.78	0.84	0.74

По таблице средних значений видно, что поиск Яндекса по Википедии в целом работает лучше, чем поиск, встроенный в Википедию. Википедия же лучше ищет запросы, которые состоят либо из одного слова, либо очень близки к названиям статей.

В качестве дополнительной метрики взята метрика CG.

Выводы

В процессе выполнения данной лабораторной работы было оценено качество двух поисковиков Яндекса и Википедии и вычислены 5 метрик оценки качества для каждого запроса на трёх уровнях.