

# In-Class Exercise, SIMS 202

September 30, 2004

**Profs. Larson and Davis**

*This exercise is based on one developed by Prof. Hearst*

## Evaluating IR Systems

Say you are an information technologist asked to decide which of three IR systems to choose for your client. Along with cost factors and assessment of the user interface, you want to assess the relative strengths of the systems' ranking algorithms (called Bear, Cardinal and Wolf).

Assume you know, for a document collection and a set of queries, what all of the relevant documents in the collection are. You only have binary (yes/no) relevance judgements for each (query, document) pair. You also know the order in which the systems rank the documents. This information appears on pages 3-6 (first the relevance judgements then the ranked results for each of the three queries in each of the three systems). This data can also be found in an Excel spreadsheet at:

<http://sims.berkeley.edu/courses/is202/f04/eval-data.xls>

if you would like to use Excel to do the computation, please do.

(a) For each system, compute the average precision (over both queries) at recall intervals of 20%. (That is, precision at 20% recall, 40% recall,..., 100% recall.) Show the results in the table below.

% Recall	Avg. Precision		
	Bear	Card	Wolf
20			
40			
60			
80			
100			

(b) Graph the results for all three systems on one plot (connect the points for each system with a smooth line).

(c) Now compute the average precision (over both queries) at three different document cutoff levels, showing the results in a table. Choose cutoff levels that help facilitate comparison of the three systems.

Cutoff Level	Avg. Precision		
	Bear	Card	Wolf

(d) Based on the results for (a-c), which ranking algorithm do you recommend?

DocId	Query0	Query1	Query2
1	1	0	0
2	1	0	0
3	1	0	0
4	1	0	0
5	1	0	0
6	0	0	0
7	0	0	0
8	0	1	0
9	0	1	0
10	0	1	0
11	0	1	0
12	0	1	0
13	0	0	0
14	0	0	1
15	0	0	1
16	0	0	1
17	0	0	1
18	0	0	1
19	0	0	0
20	0	0	0
21	0	0	0
22	0	0	0
23	0	0	0
24	0	0	0
25	0	0	0

**Relevance judgements for 25 documents for three queries. 1 indicates relevant, 0 indicates not relevant.**

Rank	Bear System								
	Query 0			Query 1			Query 2		
	DocId	Precision	Recall	DocId	Precision	Recall	DocId	Precision	Recall
1	5			12			1		
2	2			14			18		
3	1			8			11		
4	15			23			16		
5	9			9			19		
6	19			5			10		
7	3			20			23		
8	6			21			21		
9	18			4			3		
10	4			19			6		
11	12			3			22		
12	8			10			12		
13	11			7			17		
14	17			18			2		
15	7			11			13		
16	20			17			24		
17	10			24			25		
18	21			13			14		
19	16			2			5		
20	23			16			4		
21	13			6			7		
22	22			22			9		
23	24			15			15		
24	25			25			20		
25	14			1			8		

Ordering of 25 documents for each of three queries for the Bear ranking

algorithm. Documents are identified by document ID. The document shown on row 1 is the highest rank, that on row 2 is the second-highest ranked, etc. Assume there are no ties. The columns labeled Precision and Recall are there for your convenience. You don't have to fill them in, but if you put some information in them we can use this to look at your work.

Rank	Cardinal System								
	Query 0			Query 1			Query 2		
	DocId	Precision	Recall	DocId	Precision	Recall	DocId	Precision	Recall
1	7			16			5		
2	16			23			24		
3	9			3			13		
4	18			13			19		
5	4			4			17		
6	17			14			12		
7	8			17			22		
8	5			11			4		
9	11			7			1		
10	15			21			8		
11	12			18			11		
12	10			10			3		
13	6			19			16		
14	2			22			23		
15	19			20			25		
16	20			1			18		
17	21			2			2		
18	25			12			20		
19	22			15			14		
20	1			5			10		
21	23			8			6		
22	13			6			21		
23	24			9			15		
24	3			25			7		
25	14			24			9		

Ordering of 25 documents for each of three queries for the Cardinal

ranking algorithm. Documents are identified by document ID. The document shown on row 1 is the highest rank, that on row 2 is the second-highest ranked, etc. Assume there are no ties. The columns labeled Precision and Recall are there for your convenience. You don't have to fill them in, but if you put some information in them we can use this to look at your work.

Rank	Wolf System								
	Query 0			Query 1			Query 2		
	DocId	Precision	Recall	DocId	Precision	Recall	DocId	Precision	Recall
1	11			7			23		
2	2			23			9		
3	10			3			25		
4	21			13			18		
5	1			4			1		
6	5			8			2		
7	22			17			11		
8	9			25			10		
9	20			16			19		
10	3			21			16		
11	23			18			12		
12	4			22			22		
13	19			19			15		
14	6			10			8		
15	15			20			24		
16	25			5			14		
17	18			2			7		
18	16			12			13		
19	7			15			17		
20	8			1			20		
21	17			14			21		
22	12			24			4		
23	13			9			5		
24	14			11			6		
25	24			6			3		

Ordering of 25 documents for each of three queries for the Wolf ranking

algorithm. Documents are identified by document ID. The document shown on row 1 is the highest rank, that on row 2 is the second-highest ranked, etc. Assume there are no ties. The columns labeled Precision and Recall are there for your convenience. You don't have to fill them in, but if you put some information in them we can use this to look at your work.