Comparing Two Popular Search Engines

Ehsun Darody (e.darody@ece.ut.ac.ir)
Under supervision of
Dr. Orumchian (orumchian@ut.ac.ir)
University of Tehran, 4 Jan, 2003

Project Summary

Objective:

Comparing two popular search engines with various evaluation methods

Search engines:

- Google (www.google.com)
- AlltheWeb (www.alltheweb.com)

Number of Queries: 7

Queries:

1. buy a electronic book about the UNIX written by "Harley Hahn" published in 2001
2. find "travel agency" for Mashhad and Tehran and Isfahan and Tabriz and Kashan in Iran
3. dj bobo "Freedom.mp3" for free download
4. "Do you know you make me die"
5. How to "estimate recall" for WEB collections
6. "The future of the Turing Test" robots
7. "next generation search engine" with Conceptual capabilities

Evaluation methods:

1. Precision-Recall diagram
2. R-Precision
3. Average precision at seen relevant documents
4. Precision Histogram

Evaluation Results:

See next pages.

For a PDF format of this report please visit:
http://www.geocities.com/darody/mir.html

Project Report

To carry out this project, the first step was to select two search engines to compare. This process was a rather difficult task for some reasons:

1. Most search engines vary much in the number of returned results. For example for a query, Google returns 1000 web pages, while AskJeeves returns nothing! I tried to find two search engines that are relatively equal in the number of returned results.
2. Most search engines don't show the number of total retrieved pages!

The following search engines were examined initially:

- Google
- HotBot
- InfoSeek
- AskJeeves
- AllTheWeb
- And some others

Finally, I selected **Google** and **AlltheWeb**.

Then, I decided to build some good queries. Again, it wasn't an easy task. In my opinion a "good query" is:

1. A query that is similar to common queries of typical search engine users.
2. A query that yields a manageable number of results.

My selected queries are:

1. buy a electronic book about the UNIX written by "Harley Hahn" published in 2001
2. find "travel agency" for Mashhad and Tehran and Isfahan and Tabriz and Kashan in Iran
3. dj bobo "Freedom.mp3" for free download
4. "Do you know you make me die"
5. How to "estimate recall" for WEB collections
6. The future of the Turing Test" robots
7. "next generation search engine" with Conceptual capabilities

Giving queries to search engines and saving results was a routing task, but the challenge was to determine which page is relevant or not. I had to open each result page and read it to find out its relevancy.
For each query in Google and AlltheWeb, I logged the following information:

- **Retrieved Docs:** the number of retrieved documents.
- **Relevant:** the number of relevant retrieved documents.
- **Total Relevant:** the number of total relevant documents. I estimated this number with union of all relevant-retrieved documents in both search engines. I found this method of estimation in a paper about the efficiency of IR systems (I don't remember the title of that paper).
- **Docs Relevancy:** Shows my judgement about relevancy of retrieved documents. A "1" shows that the corresponding documents is relevant and a "0" shows it isn't relevant.
- **Raw P-R:** Non-standard Precision-Recall values.
- **Standard P-R:** Standard Precision-Recall values.

For each search engine, I computed the **average Precision-Recall** values at the end. The final Precision-Recall diagram was plotted by Microsoft Excel 97.

R-Precision, **Average Precision at Seen Relevant Documents** and **Precision Histogram** are computed according to the comments in our textbook.

Special notes:

- Google is faster that AlltheWeb in generating results.
- Google is restricted to only 10 terms per query.
- Google returns feedback about stop words omitted from the query to the user.

Conclusion:

Google is better than AlltheWeb in all criteria, except for "Average Precision at Seen Relevant Documents". My personal opinion is that both these search engines suffer from using just term-based indexing. These search engines (and all similar search engines) are only useful when you know exactly the specifications of your information need. You may need lots of browsing to find what you want. This implies that one day you may find a specific web page and another day you wouldn't. I'm looking ahead for the next generation of search engines, where the search engines read a web page and understand it. They think about the query and answer it intelligently, just like you and me or maybe better!

In the next pages, you'll see the search results and my analysis.

Search Results

Total relevant documents is estimated by the union of the relevant retrieved documents from both search engines

Google 1 (first query in Google)

Retrieved Docs	10										
Relevant	4										
Total Relevant	7										
Docs Relevancy	0	0	1	0	0	1	1	0	0	1	(1 means Doc is relevant)
Raw P-R											
Recall	14	28	42	57							
Precision	33	33	42	40							
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	42	42	42	42	42	40	0	0	0	0	0

Google 2

Retrieved Docs	7										
Relevant	5										
Total Relevant	5										
Docs Relevancy	0	0	1	1	1	1	1				
Raw P-R											
Recall	20	40	60	80	100						
Precision	33	50	60	66	71						
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	71	71	71	71	71	71	71	71	71	71	71

Google 3

Retrieved Docs	8										
Relevant	2										
Total Relevant	2										
Docs Relevancy	1	0	0	0	0	0	1	0			
Raw P-R											
Recall	50	100									
Precision	100	28									
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	100	28	28	28	28	28

Google 4

Retrieved Docs	8										
Relevant	5										
Total Relevant	6										
Docs Relevancy	1	1	0	0	1	0	1	1			
Raw P-R											
Recall	16	33	50	66	83						
Precision	100	100	60	57	62						
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	62	62	62	62	62	0	0

Google 5

Retrieved Docs	7										
Relevant	3										

Total Relevant	3										
Docs Relevancy	1	1	0	1	0	0	0				
Raw P-R											
Recall	33	66	100								
Precision	100	100	66								
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	100	100	66	66	66	66

Google 6											
Retrieved Docs	9										
Relevant	8										
Total Relevant	9										
Docs Relevancy	1	1	1	1	1	1	0	1	1		
Raw P-R											
Recall	11	22	33	44	55	66	77	88			
Precision	100	100	100	100	100	100	87	88			
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	100	100	88	88	0	0

Google 7											
Retrieved Docs	8										
Relevant	5										
Total Relevant	7										
Docs Relevancy	1	1	1	1	0	1	1	0			
Raw P-R											
Recall	14	28	42	57	71	85					
Precision	100	100	100	100	83	85					
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	100	85	85	85	0	0

Google											
Average P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	88	88	88	88	82	82	64	57	57	24	24

AlltheWeb 1																	
Retrieved Docs	17																
Relevant	4																
Total Relevant	7																
Docs Relevancy	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
Raw P-R																	
Recall	14	28	42	57													
Precision	100	100	37	30													
Standard P-R																	
Recall	0	10	20	30	40	50	60	70	80	90	100						
Precision	100	100	100	37	30	30	0	0	0	0	0						

AlltheWeb 2											
Retrieved Docs	5										
Relevant	2										
Total Relevant	5										

Docs Relevancy	1	0	1	0	0						
Raw P-R											
Recall	20	40									
Precision	100	66									
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	66	66	0	0	0	0	0	0

AlltheWeb 3

Retrieved Docs	1										
Relevant	0										
Total Relevant	2										
Docs Relevancy	0										
Raw P-R											
Recall	0										
Precision	0										
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	0	0	0	0	0	0	0	0	0	0	0

AlltheWeb 4

Retrieved Docs	8										
Relevant	3										
Total Relevant	6										
Docs Relevancy	1	1	0	0	0	1	0	0			
Raw P-R											
Recall	16	33	50								
Precision	100	100	66								
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	66	0	0	0	0	0

AlltheWeb 5

Retrieved Docs	1										
Relevant	1										
Total Relevant	3										
Docs Relevancy	1										
Raw P-R											
Recall	33										
Precision	100										
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	100	100	100	100	100	100	100

AlltheWeb 6

Retrieved Docs	7										
Relevant	5										
Total Relevant	9										
Docs Relevancy	1	1	1	0	1	1	0				
Raw P-R											
Recall	11	22	33	44	55						
Precision	100	100	100	80	83						
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	100	100	83	83	0	0	0	0	0

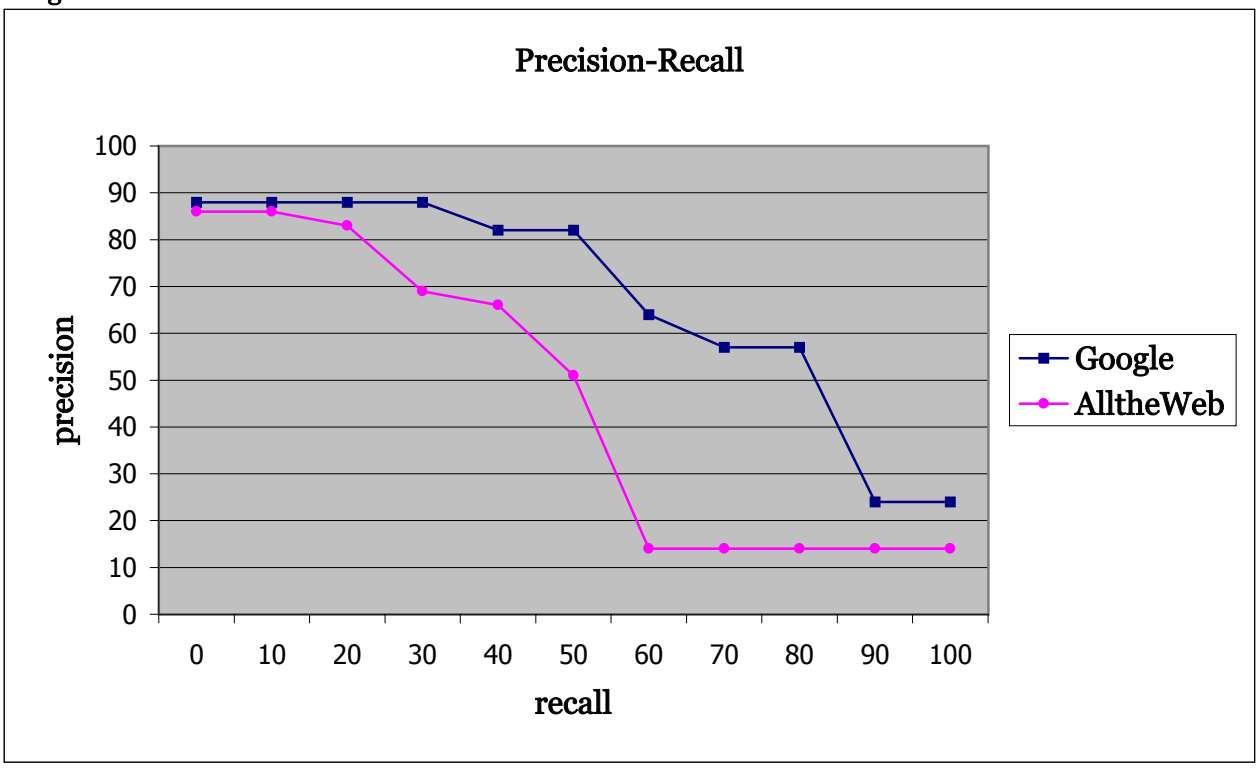
AlltheWeb 7

Retrieved Docs	6										
Relevant	4										
Total Relevant	7										
Docs Relevancy	1	0	1	1	1	0					
Raw P-R											
Recall	14	28	42	57							
Precision	100	66	75	80							
Standard P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	100	100	80	80	80	80	0	0	0	0	0

AlltheWeb

Average P-R											
Recall	0	10	20	30	40	50	60	70	80	90	100
Precision	86	86	83	69	66	51	14	14	14	14	14

Diagram



2) R-Precision Measurement

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Google	0.428	0.6	0.5	0.5	0.666	0.888	0.857
AlltheWeb	0.285	0.4	0	0.5	0.333	0.555	0.571

	Average
Google	0.634
AlltheWeb	0.377

3) Average Precision at Seen Relevant Documents Measurement

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Google	0.37	0.56	0.64	0.758	0.886	0.968	0.946
AlltheWeb	0.667	0.83	0	0.886	1	0.926	0.802

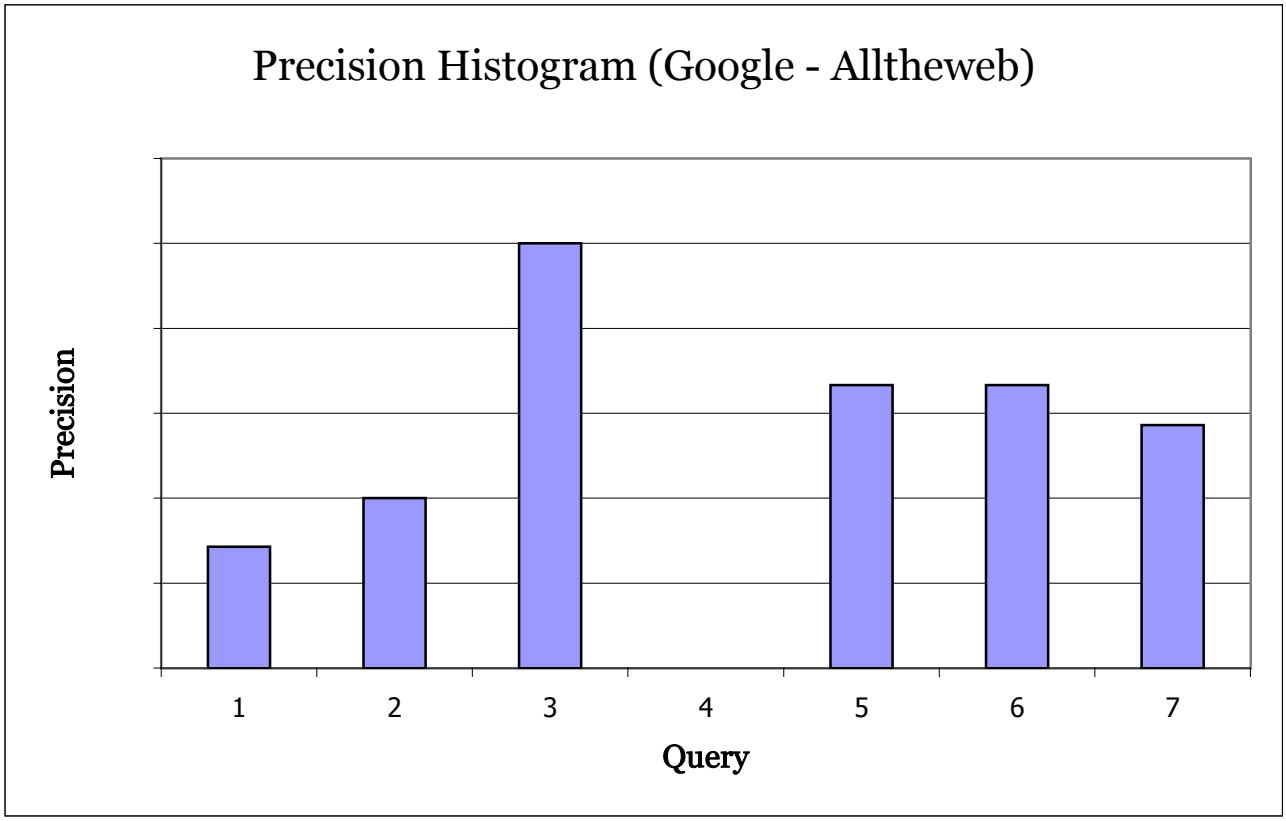
	Average
Google	0.732
AlltheWeb	0.73

4) Precision Histogram

Using differences between R-Precision values of Google and AlltheWeb for all queries:
(a positive value implies that Google has better r-precision for that query)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Difference	0.143	0.2	0.5	0	0.333	0.333	0.286

Diagram in the next page


Precision Histogram (Google - Alltheweb)
Precision
1
2
3
4
5
6
7
Query