Methods for Measuring Search Engine Performance over Time

Judit Bar-Ilan

The Hebrew University of Jerusalem

School of Library, Archive and Information Studies

P.O. Box 1255, Jerusalem, 91904, Israel

judit@cc.huji.ac.il

ABSTRACT

This study introduces methods for evaluating search engine performance over a time period. Several measures are defined, which as a whole, describe search engine functionality over time. The use of these measures is illustrated though a specific example.

Keywords

search engines, Web, performance, stability, measures

1. INTRODUCTION

The Web is extremely dynamic: new pages are published; old pages are removed while the content of other pages is changed. Such behavior is not encountered in classical information systems. Web search engines are expected to be able to handle these dynamic changes. The aim of the current study is to introduce methods for evaluating search engine performance during a time period.

2. DEFINING THE MEASURES

Any study monitoring search engine performance over time, must examine its behavior periodically. The length of the period between consecutive searches has to be set. We call each search point a *search round*. At each search round the same query or set of queries is presented to each of the search engines under examination. For each query, in each search round, all the results from all the search engines under examination must be collected.

A document is defined to be *technically relevant* if it fulfills all the conditions posed by the query: all search terms and phrases that suppose to appear in the document do appear, and all terms and phrases that are supposed to be missing from the document, i.e. terms preceded by a minus sign or a NOT operator, do not appear in the document. *Technical precision* is defined as the percentage of technically relevant retrieved documents out of the total number of retrieved documents.

The set of *well-handled URLs* is defined as the URLs that are either continuously retrieved by the engine from the first time they was discovered by it, or if the URLs are not retained by the search engine, it is because they either disappeared from the Web or cease to be technically relevant.

The set of *mishandled URLs* is the set of URLs that disappeared from the search results during the search period (such URLs are counted with multiplicity one even if they were "forgotten" several times), even though the URLs continued to exist and continued to be technically relevant. The set of mishandled URLs can be further partitioned into the set of *mishandled reappeared URLs* (these are URLs which disappeared from the search results at one stage, but reappeared later) and the set of *mishandled disappeared URLs*, which are mishandled URLs that never reappeared at a later stage.

3. CALCULATING THE MEASURES FOR THE SPECIFIC EXAMPLE

As our example, we picked a single word query, which has no stems or extensions (to avoid problems with search engines handling multiple word queries and stemming/extension in a different manner). Our choice was "aporocactus". The searches were carried out once a month over a period of ten months between January and October, 2000. The six largest search engines as of January 2000 were queried in each search round: AltaVista, Excite, Fast, Google, Hotbot and NorthernLight. Figure 1 displays the number of URLs each search engine retrieved in each search round. The graph shows that the number of pages containing the search term grows almost monotonously. The number of pages Google retrieved peaked in September (probably due to Google's expansion), but had a steep drop in October. The other engines showed a more consistent retrieval pattern.

The technical precision of the search engines was rather high between 76.3% and 99.5%. Excite and Hotbot had the highest technical precision. Note that besides pages not containing the search term, unreachable URLs and documents not found (404 errors) were also categorized as technically irrelevant.
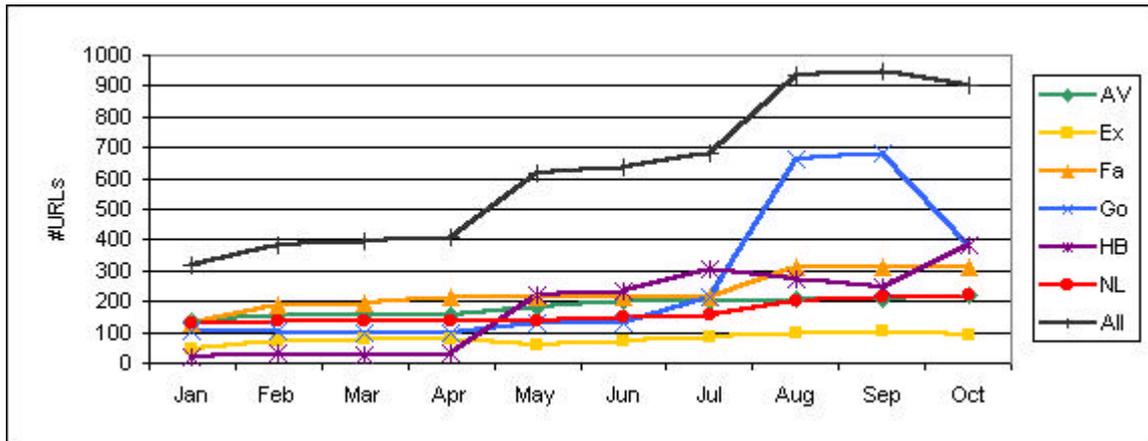
Figure 1: URLs per search round and search engine

Figure 2 depicts the extent to which each search engine mishandled URLs during the search period.

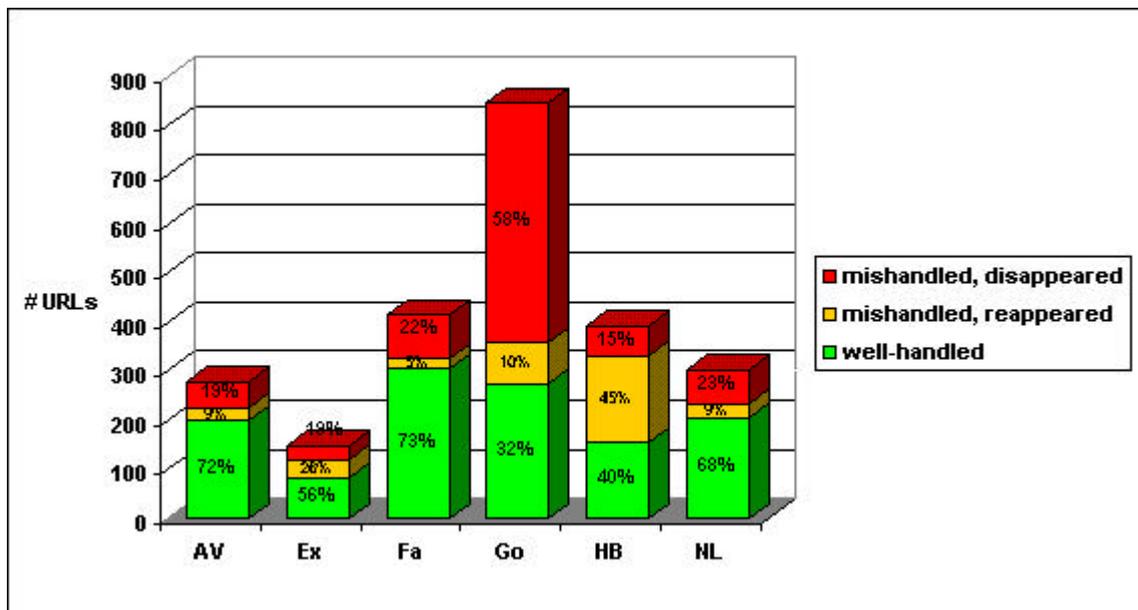Even the "best" engines mishandled more than 25% of the URLs.

Figure 2: Well handled and mishandled URLs over the search period

4. CONCLUSIONS

The example illustrates that there is a need to study search engine stability (or rather instability) over time.