Chapter XXIX
Towards a Model for Evaluating Web Retrieval Systems in Non–English Queries

Fotis Lazarinis
University of Sunderland, UK

ABSTRACT

As the Web population continues to grow, more non-English users will be amassed online. The purpose of this chapter is to describe the methods and the criteria used for evaluating search engines and to propose a model for evaluating the searching effectiveness of Web retrieval systems in non-English queries. The qualities and weaknesses related to the handling of Greek and Italian queries are evaluated based on this method. The fundamental purpose of the methodology is to establish quality measurements on search engine utilization from the perspective of end users. Application of the proposed evaluation methodology aids users to select the most effective search engine and developers to identify some of the modules of their software that need improvements.

INTRODUCTION

Since its conception in 1992 (Berners Lee, Caillau, Groff, & Pollermann, 1992), the World Wide Web (WWW or Web) has rapidly become one of the most widely-used services of the Internet along with e-mail. Its friendly interface and its hypermedia features attract a significant number of users around the globe. As a result, the Web has become a pool of various types of data, dispensed in a measureless number of locations. Finding information that satisfies specific criteria is a regular daily activity of almost every Web user. Web search engines provide searching services through their uncomplicated interfaces.

Some recent statistical estimations claim that 64.2% of the online population are non-English speakers (Global Reach, 2004). This makes the Web a multicultural and multilingual information space. Therefore, the preferences and requests of non-English-speaking users should undoubtedly be taken into account in the design of any Web information system and especially in Web retrieval systems, since they are utilized on a daily basis by virtually every Web surfer.
Towards a Model for Evaluating Web Retrieval Systems in Non-English Queries

Even though several Web search engines exist, most of their features and virtues are catered for the English language only. For example, the query “Book Agatha Christie” in Google retrieves pages mentioning the word “books” as well. This is easily understood as the matching terms are emboldened. In contrast, the queries “Livre Agatha Christie” in French, “Libro Agatha Christie” in Italian and Spanish and “Βιβλίο Αγκάθα Κρίστι” in Greek, retrieve only pages which include exactly the query terms as they are typed in the query. This query could be more problematic in the case of German where the word Book changes from “Buch” in singular to “Bücher” in plural. Both diacritics and endings change in this case. In general, in natural languages with conjugations and intonation like Greek and in languages with non-Latin alphabets, like Greek, Russian, Arabic, Asian, and African languages, searching the Web imposes additional difficulties which should be taken into account so as to design search systems of high quality and effectiveness. In addition to Web search engines, several e-shops maintain local search systems to facilitate their customers searching their catalogs. In these systems, it is imperative to comprehend and assimilate the characteristics of the local natural languages.

The majority of the search engine evaluation studies focus on precision (relevance) of the top-ranked pages returned in specific queries. However, since Web retrieval systems are utilized by several people with medium or low technical expertise, and since the natural language of the queries affect the process, other factors should additionally influence their development and evaluation.

In this chapter, we focus on creating and testing a generalized evaluation methodology which combines interface issues, for example, adaptation to the local language, with searching effectiveness, for example, case insensitivity or effect of the removal of common words (stopwords). The model is presented and analyzed, and it is then applied to evaluating the capabilities of Greek- and Italian-supporting Web search engines. This framework can serve as the basis for evaluating the effectiveness of Web retrieval systems and of local e-shop search engines in non-English text retrieval. The fundamental purposes of the methodology are:

- To establish quality measurements at every stage of search engine utilization from the perspective of end users;
- To help non-English users in selecting the most effective and user-friendly search engine; and
- To aid developers of search engines in identifying some of the modules of their software that need improvements, so as to satisfy the needs of searchers of various ethnicities.

Search Engine Evaluation

Evaluation is an important aspect in an information retrieval (IR) system (Cleverdon, Mills, & Keen, 1966; Robertson, 1969). Cleverdon et al. (1966) listed six criteria that could be used to evaluate IR systems: (i) coverage, (ii) time lag, (iii) recall, (iv) precision, (v) presentation, and (vi) user effort. Of these criteria, recall and precision have most frequently been applied in assessments of IR software tools. Information retrieval on the Web is fairly different from retrieval in traditional indexed databases. This difference arises from the high degree of dynamism of the Web, its hyperlinked character, the absence of a controlled indexing vocabulary, the heterogeneity of document types and authoring styles, and the easy access that different types of users may have to it (Gwizdka & Chignell, 1999). Therefore, the six evaluation criteria proposed reshaped to fit in this environment. Chu and Rosenthal (1996) evaluated the capabilities of AltaVista, Excite, and Lycos, and proposed a methodology for evaluating search engines in terms of five aspects: