

New challenges of Search Engines

Katarzyna Wegrzyn-Wolska

ESIGETEL

`katarzyna.wegrzyn@esigetel.fr`

Abstract. Human history shows us the importance of access to information. At present, mankind produces and publishes an enormous amount of content, mostly in electronic format. The Web with its on-line access to information replaces the role of the postal services and print libraries so important in the past. Access to fresh and relevant information is and will be increasingly important from both a sociological and an economical point of view. At present we should develop a new systems that could hold far more data and search through it far faster than the previous could before. The Web created new perspectives for information retrieval. The amount of information (also multimedia) on the Web is growing rapidly, as well as the number of new users inexperienced in the art of Web research. General people are likely to surf the Web using Search Engines. Modern Search Engine is designed to crawl and index the Web efficiently and to produce much more satisfying search results than existing systems. To engineer such a Search Engine is a challenging task. Search Engines index more than billion of Web pages involving a comparable number of distinct terms. They answer about hundreds millions of queries every day. Fast crawling technology is needed to collect the fresh Web documents and keep them up to date. Storage space must be used very efficiently to save indexed data and, optionally, the documents themselves. The indexing system must process thousands of gigabytes of data efficiently and the queries must be handled very quickly. These tasks are becoming increasingly difficult as the Web grows. At the same time, users have come to expect that Search Engine can scan through all the data and find what they are looking for, with just a few words in their question. Our main goal is to look at some problems of present and future Search Engines and to discuss how to model and develop a practical system which can exploit the all of the information present on the Web and how to improve the quality of Web Search Engines. We look also at the problem of how to effectively deal with uncontrolled content where everybody can publish everything without any censure and how to combat the Web Spam. Another question, is search personalisation and how to understand what people think about when searching the information? Is it better to provide new information or to display pages that have stood the test of time and are more likely to be of higher quality? For example, recently, a search for French Revolution returned too many sites about the recent French presidential election campaign in which candidates opined on various policy revolutions rather than the ouster of King Louis XVI. This work introduces a variety of issues, applications

and challenges with regard to the design of future Search Engines. Our reflection analyse the most important question due to the future Search Engine concern with; search quality, combating Web Spam, multimedia search, the economics of Internet search, personalization and on the other hand the privacy and intellectual property.

Key words: World Wide Web, Search Engines, Information Retrieval