On the Interactive Search with Web Documents

Authors : Mario Kubek, Herwig Unger

Abstract : Due to the large amount of information in the World Wide Web (WWW, web) and the lengthy and usually linearly ordered result lists of web search engines that do not indicate semantic relationships between their entries, the search for topically similar and related documents can become a tedious task. Especially, the process of formulating queries with proper terms representing specific information needs requires much effort from the user. This problem gets even bigger when the user's knowledge on a subject and its technical terms is not sufficient enough to do so. This article presents the new and interactive search application DocAnalyser that addresses this problem by enabling users to find similar and related web documents based on automatic query formulation and state-of-the-art search word extraction. Additionally, this tool can be used to track topics across semantically connected web documents