

A Fuzzy Algorithm for Optimizing Semantic Documental Searches: A Case Study with Mendeley and IEEExplore

Sara Paiva, Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal

ABSTRACT

Search for documents is a common and pertinent task lots of organizations face every day as well as common Internet users in their daily searches. One specific document search is scientific paper search in reference manager systems such as Mendeley or IEEExplore. Considering the difficult task finding documents can sometimes represent, semantic search is currently being applied to improve this type of search. As the act of deciding if a document is a good result for a given search expression is vague, fuzziness becomes an important aspect when defining search algorithms. In this paper, the author present a fuzzy algorithm for improving documental searches optimized for specific scenarios where we want to find a document but don't remember the exact words used, if plural or singular words were used or if a synonym was used. The author also present the application of this algorithm to a real scenario comparing to Mendeley and IEEExplore results.

Keywords: Documental Search, Fuzzy Algorithms, IEEExplore, Mendeley, Semantic Search

1. INTRODUCTION

Search for information continues to be a current matter to address and to improve. Search necessities keep increasing as they apply to several and distinct domains such as finding documents in a company's intranet, scientific articles in specific sites or the most common broad general information search on the World Wide Web. In all these cases, users always want the same thing: quickly find what they are looking for. However, traditional IR techniques have recognized limitations in some type (majority) of search needs. A common task users perform almost every day is to "google" something. As

an example, the search for "Michael Jackson" obviously return as first hyperlinks pages referring to the famous music pop singer, what will satisfy the majority of people making this specific search. However, for users belonging to the academic field, it would be more relevant to them the page of a recognized teacher with the same name. But how easy will they find that page or even any kind of information about that teacher? This example brings up questions about the efficiency when refining a search, as mentioned in Mislove, Gummadi, and Druschel (2006): "refining a search is possible but can be complicated". The reality regarding searches is well expressed in Collier and Arnold (2003)

DOI: 10.4018/ijwp.2014010104

where the authors refer that “the majority of users know the frustration of searching on the web: zero results or a million of them”. This happens because the majority of web content is made to be read by humans and not to be manipulated by computers. This reality is what the Web 3.0 or Semantic Web (SW) intends to change. In SW era, page contents will have a structure and well defined meaning. The traditional search will then extend to a semantic search—a search with meaning. The search for “papers written by Sara Paiva” is pretty specific to a human but not for a machine that doesn’t know what a paper is or who Sara Paiva is. Currently, without semantic search, instead of getting papers written by Sara Paiva, probably we would get several documents where the word Sara or the word Paiva appear (that is what we can expect if traditional search techniques based on keyword are used).

So the content of resources placed on the web have a meaning, SW relies on metadata, defined in Berners-Lee (1997) as data over data. For each resource placed on the web, informations about it should be provided so searches are made over those informations and not the content of the resource. For the search “papers written by Sara Paiva” to evolve to a semantic search, resources should have at least information of type and author. In the moment the search is performed, returned hyperlinks will be those whose type is equal to “paper” and author equal to “Sara Paiva”.

We have dedicated some investigation work on this theme, namely when developing the systems PRECISION and GSSP. PRECISION S. Paiva, Ramos-Cabrer, Gil-Solla, Fernandez-Vilas, and Diaz-Redondo (2011), and Sara Paiva, Ramos-Cabrer, and Gil-Solla (2010) stands for “guided and Personalized Expression Construction with Semantic validation” and is a guided-based search system with two main characteristics: semantic validation and personalized natural language generation of search expressions. Additionally, the system, which is oriented to comparative searches, supports 1:N ontology class relations and also the notion of search and auxiliary classes which

gives each of these type of class different roles in the query construction process. GSSP (Sara Paiva, Ramos-Cabrer, & Gil-Solla, 2012) stands for “Generic Semantic Search Platform” and its main goal is to provide a platform that allows a given search system to incorporate semantics in its search process. GSSP is built on top of PRECISION and was designed to suite any scenario where searches are helpful and with few configuration needs. It is optimized to documental searches and four criteria were defined for handling free searches.

Following this line of investigation, we believe some degree of fuzziness is missing to the already defined criteria as deciding that a given resource satisfies a given search expression is not entirely black and white as it is a vague decision. As a response to that, in this paper we present a fuzzy algorithm for optimizing semantic documental searches. In concrete, we try to address the following scenario:

Bob is a master degree student currently writing his dissertation on Education Sciences. For this task, he feels the need to use several scientific papers digital repositories such as Mendeley or IEEEExplore. Two days ago, during his searches, he found an interesting paper but it was time to go home and he closed the Internet browser. Late that night, he made a routine cleanup task on his computer erasing internet temporary files. Yesterday, he picked up where he left and remembered that paper he saw and went back to Mendeley to search for the paper but was having some troubles finding it. He remembered the paper had somewhere the words “students” and “learn” in the title but he didn’t know for sure if those were the exact words used, or if synonyms were used instead, other verb conjugation (“learning”), etc. It took him several search expressions to finally find the paper, which was ranked too low.

The rest of this paper is organized as follows: the next section introduces a literature review on search techniques. Section 3, to justify the need and demonstrate the pertinence

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/a-fuzzy-algorithm-for-optimizing-semantic-documental-searches/110887

Related Content

Educational Data Mining and Recommender Systems Survey

M. Premalatha, V. Viswanathan, G. Suganya, M. Kaviyaand Aparna Vijaya (2018). *International Journal of Web Portals* (pp. 39-53).

www.irma-international.org/article/educational-data-mining-and-recommender-systems-survey/198443

Commerce Web-Portal Redesigning Based on Usability Evaluation

Sergey Sakulin, Alexander Alfimtsev, Yuri Kalginand Vladimir Devyatkov (2021). *International Journal of Web Portals* (pp. 1-19).

www.irma-international.org/article/commerce-web-portal-redesigning-based-on-usability-evaluation/287386

A Case Studies Approach to the Analysis of Profiling and Framing Structures for Pervasive Information Systems

José Eduardo Fernandes, Ricardo J. Machadoand João Á. Carvalho (2012). *International Journal of Web Portals* (pp. 1-18).

www.irma-international.org/article/case-studies-approach-analysis-profiling/73912

Countermeasures for Protecting Legally Sensitive Web-Powered Databases and Web Portals

Theodoros Evdoridis (2007). *Encyclopedia of Portal Technologies and Applications* (pp. 188-191).

www.irma-international.org/chapter/countermeasures-protecting-legally-sensitive-web/17868

E = Mportfolios2?: Challenges and Opportunities in Creating Mobile Electronic Portfolio Systems for Lifelong Learning

Ian Weberand Peter Evans (2011). *International Journal of Web Portals* (pp. 1-13).

www.irma-international.org/article/mportfolios2-challenges-opportunities-creating-mobile/55107