

Chapter IV

Enhanced Information Retrieval Evaluation between Pseudo Relevance Feedback and Query Similarity Relevant Documents Methodology Applied on Arabic Text

Sameh Ghwanmeh

Yarmouk University, Jordan

Ghassan Kanaan

The Arab Academy for Banking and Financial Sciences, Jordan

Riyad Al-Shalabi

The Arab Academy for Banking and Financial Sciences, Jordan

ABSTRACT

Information retrieval systems utilize user feedback for generating optimal queries with respect to a particular information need. However, the methods that have been developed in IR for generating these queries do not memorize information gathered from previous search processes, and hence cannot use such information in new search processes. Thus, a new search process cannot profit from the results of the previous processes. Web Information Retrieval systems should be able to maintain results from previous search processes, thus learning from previous queries and improving overall retrieval quality. In this chapter, we are using the similarity of a new query to previously learned queries. We then expand the new query by extracting terms from documents, which have been judged as relevant to these previously

learned queries. Thus, the new method uses global feedback information for query expansion in contrast to local feedback information, which has been widely used in previous work in query expansion methods. Experimentally, we compared a new query expansion method with two conventional information retrieval methods in local and global query expansion to enhance the traditional information system. From the results gathered it can be concluded that although the traditional IR system performance is high, but we notice that PRF method increases the average recall and decreases the fallout measure.

INTRODUCTION

Information retrieval (IR) systems utilize user feedback for generating optimal queries with respect to a particular information need. In this research, we are using the similarity of a new query to previously learned queries, and then we expand the new query by extracting terms from documents, which have been classified as relevant to these previously learned queries. In addition, we explore to use two techniques to improve traditional IR system by expand the query terms using QSD technique as global query expansion and PRF as local query expansion. Query expansion is a well-known technique that has been shown to improve average retrieval performance. This technique has not been used in many operational systems because of the fact that it can greatly degrade the performance of some individual queries (Kise, Junker, Dengel & Matsumoto, 2001). Thus, the suggested method uses global feedback information for query expansion in contrast to local feedback information.

Global relevance feedback is learned from previous queries but Local relevance feedback is produced during execution of an individual query. If documents are relevant to a query, which has been issued previously by a user, then the same documents are relevant to the same query at later time, when that query is re-issued by the same or by a different user; this is the trivial case where similarity between two different queries is the highest. In the non-trivial case, a new query is similar to previously issued query only to a certain degree. Then our assumption is those documents, which are relevant to the previously

issued query, will be relevant to the new query only to a certain degree (Hust, Klink, Junker & Dengel, 2004).

Word mismatch is a common problem in information retrieval. Most retrieval systems match documents and queries on a syntactic level, the underlying assumption is that relevant documents contain exactly those terms that a user chooses for the query. However, a relevant document might not contain the query words as given by the user. Query expansion (QE) is intended to address this issue. Other topical terms are located in the corpus or an external resource and are appended to the original query, in the hope of finding documents that do not contain any of the query terms or of re-ranking documents that contain some query terms but have not scored highly (Kise, Junker, Dengel & Matsumoto, 2001).

A disadvantage of QE is the inherent inefficiency of reformulating a query. With the exception of our earlier work, these inefficiencies have largely not been investigated. In this chapter we have proposed improvements to the efficiency of QE by keeping a brief summary of each document in the collection in memory, so that during the expansion process no time-consuming disk accesses need to be made. However, expanding queries using the best of these methods still takes significantly longer than evaluating queries without expansion. When users try to search some information via internet, many of the information retrieved are not useful for the user, which might be because the poor queries that user entered or because the structure of IR system itself that used on this search site. So we try to update query itself by adding some key terms that considered to be

9 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/chapter/enhanced-information-retrieval-evaluation-between/30717

Related Content

An Adaptive CU Split Method for VVC Intra Encoding

Lulu Liu and Jing Yang (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-17).

www.irma-international.org/article/an-adaptive-cu-split-method-for-vvc-intra-encoding/322433

Design of Healthcare Lighting in Medical Centers Based on Power Carrier Communication

Yan Huang and Yongfeng Zhang (2023). *International Journal of Information Technologies and Systems Approach* (pp. 1-14).

www.irma-international.org/article/design-of-healthcare-lighting-in-medical-centers-based-on-power-carrier-communication/324748

Knowledge Acquisition on Dante Alighieri's Works

Elvira Immacolata Locuratolo and Valentina Bartalesi Lenzi (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 5067-5076).

www.irma-international.org/chapter/knowledge-acquisition-on-dante-alighieris-works/184209

Analysis of Two Phases Queue With Vacations and Breakdowns Under T-Policy

Khalid Alnowibet and Lotfi Tadj (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 1570-1583).

www.irma-international.org/chapter/analysis-of-two-phases-queue-with-vacations-and-breakdowns-under-t-policy/183872

Electronic Theses and Dissertations (ETDs)

Ralph Hartsock and Daniel G. Alemneh (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 6748-6755).

www.irma-international.org/chapter/electronic-theses-and-dissertations-etds/184370