Chapter 14

Towards a Possibilistic Information Retrieval System Using Semantic Query Expansion

Bilel Elayeb
ENSI Manouba University, Tunisia

Oussama Ben Khiroun
ENSI Manouba University, Tunisia

Ibrahim Bounhas
Faculty of Sciences of Tunis, Tunisia

Fabrice Evrard
Informatics Research Institute of Toulouse (IRIT), France

Narjès Bellamine-BenSaoud
ENSI Manouba University, Tunisia

ABSTRACT

This paper presents a new possibilistic information retrieval system using semantic query expansion. The work is involved in query expansion strategies based on external linguistic resources. In this case, the authors exploited the French dictionary "Le Grand Robert". First, they model the dictionary as a graph and compute similarities between query terms by exploiting the circuits in the graph. Second, the possibility theory is used by taking advantage of a double relevance measure (possibility and necessity) between the articles of the dictionary and query terms. Third, these two approaches are combined by using two different aggregation methods. The authors also benefit from an existing approach for reweighting query terms in the possibilistic matching model to improve the expansion process. In order to assess and compare the approaches, the authors performed experiments on the standard 'LeMonde94' test collection.

INTRODUCTION

The quasi-exponential development of the human knowledge distributed on varied interest fields led to the generation of a big mass of information increasingly difficult to manage and maintain. Within this large scale environment characterised at the same time by the great number of users and the immense mass of data, it becomes essential to conceive and develop tools allowing an effective and organized access. It is crucial to develop

automated interfaces which make it possible to formulate and satisfy users' informational needs. Information Retrieval (IR) is a branch of data processing interested to the acquisition, the organization, the storage and the research of information. We need Information Retrieval Systems (IRS) which constitute computer tools aiming to capitalize information and to locate relevant documents. Given an information requirement expressed as a query, the relevance is quantified according to a matching model between the query terms and the documents. Whatever the semantics are given to the representation of the objects (document or query) or the relevance definition; these models have an identical general behaviour (Dos Santos & Madeira, 2010; Baloglu et al., 2010). The majority of them represent the documents and the queries by lists of weighted keywords. Therefore from the concept of query/answer, the relevance of the result given by an IRS depends primarily on the query. However, the user is often unable to give some keywords which describe explicitly and clearly his intentional need what can deteriorate the quality of the awaited results. Query expansion is one of the strategies implemented in IRS to improve their performance and better satisfy users. It consists in enhancing the user's query by adding new terms to better express his need. In fact, there are two main approaches to query expansion in the literature, automatic query expansion (AQE) and interactive query expansion (IQE) (Ruthven, 2003). AQE is simpler for the user, but limits its performance because it has no user involvement. Having user involvement, IQE is more complex for the user, but means it can take more problems such as ambiguous queries. Besides, results of an IRS fail by finding too few relevant documents (low recall) or by retrieving too many irrelevant documents (low precision). Historically, AQE has better recall than IQE (Vélez & Weiss, 1997). Unfortunately, if the terms used to expand a query often changed the query's

meaning, AQE frequently decreased precision (Croft & Harper, 1979). The problem is that users typically consider just the first few results (Jansen & McNeese, 2005), which makes precision crucial to search performance. In contrast, IQE has balanced precision and recall leading to an earlier uptake within search. However, like AQE, the precision of IQE approaches needs improvement. Most recently, approaches have started to improve precision by incorporating semantic knowledge (Crabtree, 2009). This can be achieved by various techniques such as corpus analysis and classification (Chevallet & Nie, 1997; Claveau & Sébillot, 2004), user Relevance Feedback (RF) and integration of external linguistic resources (e.g., dictionaries, thesauri and ontologies). We focus in our work on this last interactive query expansion approach using a dictionary as proposed in Elayeb (2009). Several query expansion experiments were conducted for example by using the Word-Net lexical database on English IRS (Voorhees, 1994; Smeaton, 1997). The data used for query expansion in these approaches is poor, uncertain and unclear, while possibility theory is naturally appropriate for this kind of application. It allows expressing phenomena of ignorance, imprecision and uncertainty (Brini & Boughanem, 2004). Indeed, it defines two types of relevance. On the one hand, plausible relevance, quantified by the possibility trends to eliminate non-semantically similar terms (irrelevant ones); on the other hand, necessity relevance helps improve our belief in terms not eliminated by possibility measure (i.e., semantically close words useful for expansion). Ben Khiroun et al. (in press) proposed a possibilistic approach for semantic query expansion. Based on this work, we propose a new possibilistic IRS which takes advantage, combine and compare the possibilistic and the circuit-based approaches for semantic query expansion. Moreover in this paper, we propose and investigate the idea of a possibilistic network which models the dependen-

Related Content

Adaptive Neural Algorithms for PCA and ICA
Radu Mutihac (2009). *Encyclopedia of Artificial Intelligence (pp. 22-30).*
www.igi-global.com/chapter/adaptive-neural-algorithms-pca-ica/10221?camid=4v1a

Knowledge-Based Recommendation Systems: A Survey
Sarah Bouraga, Ivan Jureta, Stéphane Faulkner and Caroline Herssens (2014). *International Journal of Intelligent Information Technologies (pp. 1-19).*
www.igi-global.com/article/knowledge-based-recommendation-systems/114956?camid=4v1a

RFID and Dead-Reckoning-Based Indoor Navigation for Visually Impaired Pedestrians
Kai Li Lim, Kah Phooi Seng, Lee Seng Yeong and Li-Minn Ang (2018). *Smart Technologies: Breakthroughs in Research and Practice (pp. 1-16).*
www.igi-global.com/chapter/rfid-and-dead-reckoning-based-indoor-navigation-for-visually-impaired-pedestrians/183438?camid=4v1a

Load Balancing of Unbalanced Assignment Problem With Hungarian Method
Ranjan Kumar Mondal, Payel Ray, Enakshmi Nandi, Biswajit.Biswas, Manas Kumar Sanyal and Debabrata Sarddar (2019). *International Journal of Ambient Computing and Intelligence (pp. 46-60).*
www.igi-global.com/article/load-balancing-of-unbalanced-assignment-problem-with-hungarian-method/216469?camid=4v1a