Chapter 20
Information Retrieval Model Using Uncertain Confidence's Network

Fatiha Naouar
MARS Research Unit, University of Monastir, Tunisia

Lobna Hlaoua
MARS Research Unit, University of Monastir, Tunisia

Mohamed Nazih Omri
MARS Research Unit, University of Monastir, Tunisia

ABSTRACT

This paper proposes a new relevance feedback approach to collaborative information retrieval based on a confidence's network, which performs propagation relevance between annotations terms. The main contribution of our approach is to extract relevant terms to reformulate the initial user query considering the annotations as an information source. The proposed model introduces the concept of necessity that allows determining the terms that have strong association relationships. The authors estimated the association relationship to a measure of a confidence. Another contribution consists on determining the relevant annotations for a given evidence source. Since the user is over whelmed by a variety of contradictory annotations on even one which are far from the original subject, the authors' model proceed filtering these annotations to determine the relevant one and then it classify them by grouping those related semantically. The experimental study conducted on different queries gives promoters results. They show very encouraging results that could reach an improvement rate.

1. INTRODUCTION

Collaborative retrieval allows increasing the number of relevant information found and to reduce the time needed to search. Working collaboratively can also share historical, with others. The collaborative retrieval can reduce the retrieval time performed by the users of the same profile. The optimization of retrieval time may be caused by the formulation of collaborative queries the through dialogue and

the mutual consultation of queries sent and the search results received by everyone. It also enables the sharing of search history by displaying the search results put in order by relevance. One of the most popular tools for sharing results and personal judgments is the annotations. Several problems exist with respect to Collaborative Retrieval (hereafter CR). There is, in particular, the problem of the relevance of information. Indeed, a user always finds problems in meeting his/her needs in relevant information.

As in classical Information Retrieval (hereafter IR), the Collaborative Information Retrieval (hereafter CIR) is designed to return and display a set of documents to a user according to his need. On the other hand, the users of a retrieval system, are not always specialists in this field (Lin & Wang, 2006), they can make a bad choice of terms to express its information needs. A reformulation of query has been necessary, since the initial user query can return unsatisfactory results. It's a question then of amending the original user query and this happens by adding meaningful terms to improve the initial result returned. Three types of approaches use different techniques to select the terms to be added to a query. The difference between these approaches is based on the choice of terms to be used to reformulate the query. In the first approach, the mechanism used allows the user to select the terms that are related to the original query terms. In the second approach, the re-injecting of the terms is to be selected from a terminological resource (like semantic network, thesaurus, ontology, et.). While in the third approach the terms are selected for a reformulation from the results of previous research (also called relevance feedback).

The first type of approach is based on the overall analysis of the collection of documents considered (Baeza-Yates & Ribeiro-Neto., 1999). This approach consists of analyzing the entire set of documents, in the collection, to extract relevant terms to add to the initial query. Two techniques are used: the similarity thesaurus and the statistical thesaurus. The most prevalent technique is based on statistical analysis of query logs (Cui, Wen, Nie & Ma, 2002). The objective is to automatically reformulate a query by adding terms of the documents that are correlated with the query terms. The correlation is based on a conditional probability. The second type of approach, in the literature, uses terminology resources such as the anthologies or the thesaurus containing the vocabulary to enrich queries (Abderrahim, 2013; Harb, Khaled & Nagdy, 2011). This type of approach uses anthologies with the equivalence relations and the subsumption (Navigli & Velardi, 2003) to find the similar terms to the original query. In (Boughammoura, Omri & Hlaoua L., 2011, 2012, 2013) the authors propose a new rendering approach of deep Web forms which is easy to interpret by user and reflects the exact meaning of query. The final type of approach (omri, 1994) is based on the principle of relevance feedback which also aims to reformulate an initial query to better reflect the contents of the collection. To do this, the user begins by submitting its initial application to the system that renders a first set of documents that the user must judge as relevant or irrelevant. The judgment of the relevance of the documents initially returned is used to select the terms to add to the original query (Lin & Wang, 2006).

It is in this context that we suggest to improve the research performance, using the relevance feedback to extend the initial query. This technique consists of extracting terms from documents deemed relevant and considered in a new extended query. This technique has been used in classical Information Retrieval (Rocchio. 1971) and semi-structured Information Retrieval (Hlaoua, 2006; Schenkel & Thbobald, 2005) based on only relevant document and has shown its interest. In collaborative context, we propose to use the annotations as new source of information, because an annotation has a close relationship with the annotated document. It also enriches the document in question with simple terms and nearby terms to the user better that the vocabulary used by the author. However, the research work, which is based on the annotations, can give relatively relevant resulting since the annotations are diverse and can be performed by experts or non-specialist users.

Related Content

Towards a Statistical Approach to the Analysis, the Indexing, and the Semantic Search of Medical Videoconferences
Ameni Yenguiand Mahmoud Neji (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 1274-1299).*
www.irma-international.org/chapter/towards-a-statistical-approach-to-the-analysis-the-indexing-and-the-semantic-search-of-medical-videoconferences/198599

Next Generation Search Engine for the Result Clustering Technology
Lin-Chih Chen (2012). *Next Generation Search Engines: Advanced Models for Information Retrieval (pp. 274-290).*
www.irma-international.org/chapter/next-generation-search-engine-result/64429

Investigations on implementation of e-ATM Web Services based on .NET technique
Subhash Medhiand Tulshi Bezboruah (2014). *International Journal of Information Retrieval Research (pp. 41-56).*
www.irma-international.org/article/investigations-on-implementation-of-e-atm-web-services-based-on-net-technique/126328

Researching Information Seeking in Digital Libraries Through Information-Seeking Models
Zeinab Zaremohzzabieh, Seyedali Ahrari, Bahaman Abu Samahand Jusang Bolong (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 2157-2180).*
www.irma-international.org/chapter/researching-information-seeking-in-digital-libraries-through-information-seeking-models/198641

Comparative Study Between Two Swarm Intelligence Automatic Text Summaries: Social Spiders vs. Social Bees
Mohamed Amine Boudia (2018). *Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management (pp. 276-302).*
www.irma-international.org/chapter/comparative-study-between-two-swarm-intelligence-automatic-text-summaries/197706