



The concept exploration model and an application



Yin Zhang^{a,c}, Kening Gao^{b,c}, Bin Zhang^{a,c,*}

^a College of Information Science and Engineering, Northeastern University, Shenyang, China

^b Computing Center, Northeastern University, Shenyang, China

^c Key Laboratory of Medical Image Computing of Ministry of Education, Northeastern University, Shenyang, China

HIGHLIGHTS

- The mechanism for users learning concepts in exploratory search is observed.
- The concept exploration model is proposed to capture the mechanism.
- An application example is introduced to demonstrate how to apply the model.
- The performance of the concept exploration is evaluated using the example.

ARTICLE INFO

Article history:

Received 23 April 2014

Received in revised form 15 October 2014

Available online 7 November 2014

Keywords:

Exploratory search

Query recommendation

Concept exploration model

Probabilistic information retrieval

ABSTRACT

For a user who is unfamiliar with a target domain, the first step to conduct an exploratory search task is to go over a learning phrase, which means to learn from the search results to acquire basic domain knowledge. Since lots of search results could be returned by a search engine, and usually only a small portion of all the results contain valuable knowledge to the current search task, the user usually needs to read lots of documents and could only learn limited knowledge. This makes the learning phrase a low efficiency, time consuming and easy to fail process. In order to support the learning phrase of the exploratory search process, this paper proposes the concept exploration model which describes how a user reads search results and figures out interesting concepts. The model focuses on how does a user explore related concepts during the learning phrase, and factorizes the concept exploration process as a production of the probability that concepts form a specific relation structure, and the probability that a user is attracted by a concept. In an application example, the concept exploration model is used in a query recommendation task to support exploratory search. We demonstrate how to determine the two probabilistic factors and evaluate the model with a set of metrics. The experiment results show that the application example could help users explore domain concepts more effectively.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Search engines have become the most important tools for people to solve problems and to find answers. When a user is familiar with a problem domain and could clearly specify his information need, then search engines could usually yield search results with high quality. However, when the user lacks the knowledge or the contextual awareness needed to properly describe his information need, search engines will usually fail to return search results that could directly meet

* Corresponding author at: College of Information Science and Engineering, Northeastern University, Shenyang, China. Tel.: +86 24 83687775; fax: +86 24 23893138.

E-mail addresses: zhangyin@ise.neu.edu.cn (Y. Zhang), gkn@cc.neu.edu.cn (K. Gao), zhangbin@ise.neu.edu.cn (B. Zhang).

the user's information need. In this case, the user will have to carry out a more complicated search process which is known as “exploratory search” [1]: submitting a tentative query, taking things from there, exploring the retrieved information, selectively seeking and passively obtaining cues about where the next steps lie [2].

In an exploratory search process, one of the most important works for a user is to carry out a learning phrase, which means to learn something from the search results so as to plan the next search direction [3]. Mostly, this means that the user will need to read the documents returned by the search engine, figuring out some concepts that he is interested in, and then studies the meanings of these concepts to decide whether they could help. For example, an unconscious laptop buyer may Google “laptop”, randomly reading some search results and gets attracted by a previously unknown concept “ultrabook”. After this, the user may try to learn what is “ultrabook” and if he needs an “ultrabook”. In this example we could see that during a learning phrase, a user would carry out three operations: (1) discovering interesting concepts, (2) learning the meaning of these concepts, (3) determining whether these concepts are related to his information need. In this paper, we focus on supporting the first operation namely “discovering interesting concepts”, and we use “concept exploration” to refer to the process that a user in an unfamiliar problem domain reads search results and discovers interesting concepts after submitting a tentative query. Here a “concept” refers to a formal representation of a “thing”, and occurs as words in search results. For example, concept “Laptop” may occur as “laptop”, “notebook”, “portable pc” in various documents.

Concept exploration is a form of information discovery that points the user towards interesting additional facts relevant for an information need based on the search already undertaken or the selections made [4]. In real practice, concept exploration could usually be painful. On one side, a large amount of documents could be returned by a search engine with interesting concepts distributed in probably any of them. While on the other side, a tentative query may usually corresponds to broad semantics and leads to lots of useless search results. These discrepancies force a user to put lots of time and energy to read many useless documents for fear of missing valuable ones. This fact makes concept exploration inefficient, arduous and easy to fail.

We introduce the *concept exploration model* to model the process that a user reads search results and discovers interesting concepts. For a user conducting a concept exploration process, the model calculates two probabilities about how does a concept distribute in the search results: the probability that the concept has valuable relations with the query submitted by the user (i.e. the *relation probability*), and the probability that the user would be attracted by a concept (i.e. the *interest probability*). With an application example we demonstrate how to determine the relation probability and the interest probability using domain knowledge and a frequency based method. We then consider the probability that a user will explore a concept as the product of the two probabilities: a concept is explored only when it has valuable relations with the query, and it could get a user attracted. Such an assumption means that a concept will not be explored if it fails to attract a user's attention even when it is closely related to the query, and if a user finds out that an interesting concept has nothing to do with his information need, he will also give up exploring the concept.

The main contribution of this paper is to model the process that a user explores concepts. This makes it possible to support the complicated, time consuming concept exploration process. The key points that we concern are the reasons that a concept occurs in search results and the chance that a user would be attracted by a concept: these two factors affect which concept will be explored by a user. The concept exploration model is an actionable, aggregated representation of the two factors. It is highly extensible. The factorization of the concept exploration process supports different kinds of analyzing methods, which makes it easy to leverage the latest information retrieval facilities. We also provide an application example to show how to use the model to support concept exploration.

We use undirected acyclic graphs to describe the valuable relations between concepts and queries with nodes corresponding to concepts and weights of edges giving the quantified measures of relations as probabilities. Such a representation enables generating concept relations using domain knowledge and determining the relation probability with formal knowledge about relations. Based on this, the interest probability is analyzed using statistical semantic approaches. This means that we combine formal and statistical semantic methods in the concept exploration model. The combination allows us to take the advantages of both formal and statistical manners while avoiding shortages of the two ways: formal knowledge could provide high quality concepts and explain them but is unable to weight them, which could be done using statistical semantic analysis.

We demonstrate how to use the proposed model to support concept exploration processes. In the application example, the probability that a user explores a concept is calculated and is used to provide query recommendations. We showcase how to determine the relation probability and the interest probability, together with the concept exploration model based query recommendation method and the algorithms. Experimental results show that the application example could help users explore domain concepts more effectively.

2. Related works

Exploratory search refers to the specific searching mode that users would take under situations such as lacking domain knowledge: submitting a tentative query, taking things from there, exploring the retrieved information, selectively seeking and passively obtaining cues about where the next steps lie [2]. We study concept exploration in this paper, which is a form of information discovery. Information discovery plays an important role in exploratory search. Singer et al. argue that information discovery points the user towards interesting additional facts relevant to an information need based on the search already undertaken or selections made [4]. They also point out that most of the features that an exploratory search

Download English Version:

<https://daneshyari.com/en/article/977724>

Download Persian Version:

<https://daneshyari.com/article/977724>

[Daneshyari.com](https://daneshyari.com)