Decision Support Systems 61 (2014) 106–114

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Knowledge map-based method for domain knowledge browsing

Jia Hao *[,1], Yan Yan, Lin Gong, Guoxin Wang, Jianjun Lin

School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China
Department of Mechanical Engineering, University of Michigan, MI, USA

ARTICLE INFO

Article history:
Received 8 June 2013
Received in revised form 24 December 2013
Accepted 6 February 2014
Available online 15 February 2014

Keywords:
Information overload
Knowledge browsing
Knowledge map
Social network analysis

ABSTRACT

The exponential growth of available information and the deployment of knowledge management systems delivers excessive information to the end users that they cannot manage at once. This problem has led to an increased emphasis on solutions to information overload. Searching and browsing are two methods to locate information. Many studies have focused on solving the information overload problem in the searching process, but the methods to alleviate information overload in browsing process have not been adequately studied. Hence, a method that addresses information overload in the browsing process is presented in this paper. The aim is to reduce the information overload during browsing domain knowledge for new knowledge users who have little understanding of the information. In this method, a knowledge map and social network analysis are utilized to navigate the knowledge users. Technologies first construct a knowledge map from text mining and the important knowledge that includes more information about the domain is then identified via social network analysis. Based on this process, the knowledge user can browse the domain knowledge starting from the important knowledge and navigate via the knowledge map. We applied the method to assist new knowledge users in browsing the Computer Numerical Control (CNC) domain knowledge base to validate the method. The results indicate that the method can identify the important knowledge at a highly acceptable level, the constructed knowledge map can efficiently navigate the knowledge users, and the information overload can be significantly decreased.

1. Introduction

Domain knowledge has become increasingly critical to develop competitive mechanical products [1]. In domain knowledge management systems, knowledge users usually employ one of two strategies to obtain domain knowledge: searching and browsing [2,3]. In the searching strategy, users have a topic or some keywords in mind, while they do not search for a specific thing in the browsing strategy, which is characterized by the absence of planning. Although the users can obtain information or domain knowledge via the two strategies, the information overload phenomenon is becoming increasingly frequent during the information acquisition process due to the exponentially increasing amount of information in the knowledge base. Information overload indicates that more information is provided than users can process in a period of time [4–7], which makes them spend more time on information acquisition. Therefore, new methods and tools need to be developed to assist the users in addressing information overload.

The information overload originates from the internet. Therefore, most studies have focused on information overload on the internet. This focus hinders the application of existing methods to domain knowledge browsing, because content of domain knowledge is more complicated (unstructured) than information on the internet and no hyperlinks exist between the content. In this study, we attempted to decrease the information overload when new knowledge users browse domain knowledge by providing a knowledge browsing aided method. The method takes the domain knowledge as the only input, and generates a group of domain knowledge as the starting points. The knowledge users will then be navigated by a constructed knowledge map to browse the domain knowledge. This study focuses on the reduction of information overload during domain knowledge browsing and provides three contributions to the research community. First, a knowledge map based domain knowledge browsing method is proposed. Second, a content-based domain knowledge map construction method is proposed and studied. Finally, we adopted techniques from complex networks to investigate the internal structure of the knowledge map and extract relative important knowledge for the domain.

The rest of the paper is structured as follows. The next section provides some background information for this study. The following three sections explain the proposed method in detail. Section six shows an example of implementing the proposed method. Section 7 discusses the research results. Section eight summarizes the work and outlines the possible future studies.

* Corresponding author at: 3635 Green Brier Blvd, Apt 120A, Ann Arbor, MI 48105–2611, USA. Tel.: +1 734 707 9263.
E-mail addresses: haojia632@gmail.com (J. Hao), yanyan331@bit.edu.cn (Y. Yan), gonglin@bit.edu.cn (L. Gong), wangguoxin@bit.edu.cn (G. Wang), bitlin@sina.cn (J. Lin).
[1] 5 South Zhongguancun Street, Haidian District, Beijing 100081, China.

http://dx.doi.org/10.1016/j.dss.2014.02.001
0167-9236/

2. Background

Several studies have been conducted to decrease the information overload that users are facing. Many studies have aimed to ease the information overload when obtaining information via the searching strategy. A search engine is the main tool to implement a searching strategy. Depending on the techniques employed, search engines can be classified into hierarchical search engines, agent-based search engines and Meta search engines [17]. To improve the search engine, several methods have incorporated the interest profile of subjects into the search engine [6], adopted a new result display method to replace the one-dimensional result list [6,7], provided a query-aided application to the subjects to help them compose important query sentences [8] and processed the information via social tags [9], such as Yahoo! Delicious.

However, this same problem has not been adequately researched for the browsing strategy. In our opinion, information overload in the browsing strategy is a more serious problem than in the searching strategy, because the search engine can remove a large amount of unrelated domain knowledge when searching knowledge, while this removal is not possible when browsing the knowledge base. Conversely, the interest information of knowledge users is helpful to refine the searching result. However, the knowledge users may lack a specific vocabulary to represent their interest when they browse information [3]. Therefore, knowledge users can easily become lost when browsing strategy. The site map and navigation page seem to be common methods to ease the information overload in the browsing strategy. However, the site maps generally cover only the main structure of information, which is insufficient to alleviate the information overload. To this end, the hierarchical subject category (like Yahoo) has been incorporated to guide the browsing process [2]. The recommendation system has also been adopted to alleviate the information overload; collaborative filtering is most likely the most widely implemented technique [10]. Other recommendations systems, including content-based recommenders [11], utility-based recommenders [12] and demographic recommenders [13], have also been widely studied. The obvious limitation of the above methods is the effort requirement of collecting extra information to implement the recommendation. Information visualization is another type of method that is used to alleviate the information overload. Chen [15] indicates that a Kohonen self-organizing map (SOM) -based method can categorize a large information space into manageable sub-spaces that the user can easily browse. Therefore, Yang [3] suggests that visual overload may occur when the size of visualization increases. The visual overload means knowledge users cannot easily locate the information from a large pool of information if the information is packed very densely in the visualization map. They proposed a fisheye view and fractal view to support information visualization. The evaluation result shows that the method increases the effectiveness of information visualization. See reference [14] for the information visualization method. The information visualization is indeed useful to alleviate the information overload. However, as stated by Yang [3] visual overload will emerge as the amount of information increases. Different from visualization methods, Wan [5] developed a new research tool called Citation-Sensitive In-Browser Summarizer (CSIBS) to help literature browsing tasks. This method alleviates the information overload by presenting information about a citation when the knowledge user encounters it, and this information can help the reader determine whether to invest the time to explore the citation. Kruk [16] presented the MultiBeeBrowser (MBB), which is a faceted navigation tool for content browsing. Faceted navigation means that the information space is partitioned using orthogonal conceptual dimensions of the data. These dimensions are called facets and represent important patterns of the information elements. Wang [17] presented a navigation graph-based recommendation system in which the browsing track information of previous knowledge users is utilized to provide the contents to new knowledge users.

In this paper, we propose a method to alleviate the information overload in design knowledge browsing. This method provides the knowledge users with a small number of important knowledge items at the beginning of browsing the knowledge base, and all other knowledge can be found very quickly starting from the important knowledge. The important knowledge will be defined and calculated in the following sections. The method incorporates a knowledge map [18] and social network analysis (SNA) [19,20]. The knowledge map will be constructed via latent semantic analysis (LSA) [21], and the important knowledge will be defined and identified using SNA algorithms. The method includes two main stages. The first stage takes a domain knowledge base as an input and gives the corresponding domain knowledge map as the output, and the domain knowledge base under consideration is a natural (Chinese) language document database. The knowledge map is represented by a mathematical graph in which vertices indicate the domain knowledge and edges indicate the semantic similarity between domain knowledge. Technologies determine the semantic similarities from text-mining. The second stage identifies the important knowledge from the constructed knowledge map. Inspired by social network analysis (SNA), we used a graph-theory to define and calculate important knowledge. Our approach differs from others because the only required input is the content of the domain knowledge itself. This character is suitable for domain knowledge browsing, because it is unstructured and lacks connections.

3. Proposed method

Our method aims to alleviate the information overload when browsing domain knowledge base. Fig. 1 shows an overall illustration of the entire method.

As shown in Fig. 1, the method includes two steps: Knowledge Map Construction (KMC) and Important Knowledge Identification (IKI). The domain knowledge map can be determined from its corresponding domain knowledge base in the KMC step. Based on this map, the important knowledge can be identified from the constructed knowledge map in the IKI step. The proposed method contains three data objects: domain knowledge base, domain knowledge map and important knowledge, as shown in Fig. 1. Knowledge users can more easily browse the knowledge base with these three data objects. At the beginning of browsing the knowledge base, the important knowledge is presented to the knowledge users and the domain knowledge map is then used to navigate the knowledge users to the relevant domain knowledge. The knowledge base can provide the detail of the domain knowledge. The following two sections explain the two steps in detail.

4. Knowledge map construction

The goal of the knowledge map construction is to convert a knowledge base into a corresponding knowledge map. We first illustrate the meaning of the knowledge map in our work and then detail the three steps of constructing the domain knowledge map.

4.1. Domain knowledge map

The knowledge map is a knowledge representation technology that reveals the underlying relationships between knowledge sources [22]. It can be used to find sources of knowledge, implement knowledge creation and increase knowledge sharing [23]. Furthermore, the knowledge map is helpful to many fields, such as information visualization [15], information retrieval [6], strategic decision-making support [23] and business process re-engineering [25]. However, few studies have focused on the method to build a knowledge map [23,24], likely because the definitions and uses vary by situations. In this work, we treat the knowledge map as a mathematical graph and provide the construction method. In mathematics, a graph is a representation of a set of objects in which some pairs of the objects are connected by links. The interconnected objects are represented by mathematical abstractions called