Expert Systems with Applications 36 (2009) 10952–10961

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Formal concept analysis for an e-learning semantic web

Ghassan Beydoun

School of Information Systems and Technology, University of Wollongong, Wollongong, Australia

ARTICLE INFO

Keywords:
Semantic Web
E-learning
Formal concept analysis

ABSTRACT

We capture student interactions in an e-learning community to construct a semantic web (SW) to create a collective meta-knowledge structure guiding students as they search the existing knowledge corpus. We use formal concept analysis (FCA) as a knowledge acquisition tool to process the students virtual surfing trails to express and exploit the dependencies between web-pages to yield subsequent and more effective focused search results. We mirror social navigation and bypass the cumbersome manual annotation of webpages and at the same time paralleling social navigation for knowledge.

We present our system KAPUST2 (Keeper and Processor of User Surfing Trails) which constructs from captured students trails a conceptual lattice guiding student queries. We use KAPUST as an e-learning software for an undergraduate class over two semesters. We show how the lattice evolved over the two semesters, improving its performance by exploring the relationship between 'kinds' of research assignments and the e-learning semantic web development. Course instructors monitored the evolution of the lattice with interesting positive pedagogical consequences. These are reported as well in this paper.

1. Introduction

Building a semantic web (SW) is often hindered by the following two factors: users do not agree on an annotation standard that can be used to extricate web inter-dependencies and/or users feel burdened by the annotation effort. We proposed an alternative in Beydoun, Kultchitsky, and Manasseh (2007), based on indirect social navigation (Dieberger, 1997), to exploit web-pages dependencies using surfing trails left behind. These are not beneficial if individuals surfing the net have different interests, but in a given interest group individuals produce trails that are of interest to the whole group. Our earlier experiments in e-learning in Beydoun et al. (2007) indicated that processing surfing trails left by students using formal concept analysis (FCA) is possible. But lack of crossings between surfing trails lowered the usability of the resultant lattice. Our hypothesis in this paper is that the effectiveness of the resultant conceptual lattice depends on a sufficient complexity of the conceptual lattice itself on the one hand and the following factors on the other hand: the way assignments are set to ensure trails crossings, a reflective learning setting, and a sufficient number of trails. In this paper, we employ our FCA-based system for two consecutive semesters in an American University and using comparative assignments to ensure higher number of trail crossings. The conceptual lattice resultant from the first semester is used as a starting point for further collective development by students in the second semester. The number of crossings is much higher and the experiments confirm that indeed social navigation can be simulated in e-learning settings, and moreover that FCA is suitable to process the trails into a usable semantic web which supports the learning process of students.

This paper is organized as follows: Section 2 overviews related work, Section 3 presents the conceptual basis of our FCA approach in e-learning, Section 4 presents our system KAPUST, Section 5 presents our experimental results, Section 6 discusses and analyses these results from both perspectives, technical and pedagogical concluding with a discussion of future work.

2. Related work

To effectively use web-surfing experience of people, a user friendly exposure of the experience that all people can understand is required. In Beydoun et al. (2007), we advocated building a SW incrementally by the users themselves, where no intermediate expert is needed, and ontologies are not predetermined. In capturing and organizing user trails according to the browsing topic, and later allowing intelligent search of the web, users provide their topic of interest, within their interest group, and begin browsing web pages. Submitted topics of interest and trails left behind by the users form the raw information to constitute the SW structure, and determine how it evolves. We process this raw information using machine learning. Our approach expanded the work of Wexelblat and Maes at the MIT Media Lab, which highlighted the importance of tracking user trails and interactions to benefit from interaction history for information navigation (Wexelblat, 1999). We applied formal concept analysis (FCA) (Ganter & Wille, 1999) to reason about the traces instead of only displaying and browsing

E-mail address: beydoun@uow.edu.au

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.023

trails as in Wexelblat (1999). The reasoning component has a retrieval knowledge base (the actual SW) which integrates users knowledge scattered in their left-behind surfing traces. Our FCA-based approach imposes very little effort on users and bypasses manual marking up of current tools to build the SW. Our system, KAPUST, captures users behaviors and stores them for analysis and reasoning. Capturing user traces is similar to Laus (2001) and to Wexelblat (1999) and Wexelblat and Maes (1999) which store interactions history on a user basis. The kind of information of interest to us in this paper is the internet parallel of *social navigation*, which is human interactions in pursuit of information gathering (Dieberger, 1997).

Formal concept analysis has been used in various classification tasks, e.g. to classify software engineering activities (Tilley, Cole, Becker, & Eklund, 2005) and to impose structure on semi-structured texts (e.g. Eklund & Wormuth, 2005; Kim & Compton, 2001). It has been found efficient when applied to document retrieval systems for domain specific browsing (Kim & Compton, 2001). Our use of FCA is unique in that the lattice is built in a collaborative manner reflecting the collective meaning of *user trails*, and modelling *unintended* and *indirect* social navigation over the web: users are not intentionally helping each other (e.g. following footsteps in a forest) and they do not directly communicate. We build a complex information space (the SW), where we analyze traces left by the users. Our approach is similar to Footprints (Wexelblat, 1999; Wexelblat & Maes, 1999), where a theory of interaction history was developed and a series of tools were built to allow navigation of history-rich pages and to contextualize web pages during browsing. However, unlike our approach, it does not use history to make recommendations nor does it have a reasoning technique embedded in it.

Creating an online collaborative learning environment is a necessary aspect for e-learning which is the systematic use of networked information and communications technology in teaching and learning (Horton & Horton, 2003). E-Learning is flexible, relatively cheap and supplies "just in time" learning opportunities. E-learning is directly underpinned by the development of and access to information and communications technology infrastructure. Creating a sense of community and understanding the online behaviors of the participants are also crucial (Blunt & Ahearn, 2000). Distant academic learning is one important application. E-learning techniques in the corporate world are also often used for residential workshops and staff-training programs. Several efforts have been made to create e-learning environments. Notably, in George Mason University under the Program on Social and Organizational Learning (PSOL), research is being done to create and maintain a Virtual Learning Community for the participants in the program. The purpose of that research is studying the learning of the community within the developed environment and a better understanding of the dynamics of collaborative dialogue to enable more informed and sound decision making (Blunt & Ahearn, 2000).

3. Constructing the semantic web with user trails and formal concept analysis

In our semantic web, search ontologies are evolved from the keywords that users enter to name their browsing sessions at the start. These keywords relate to users search domains. How these naming keywords are transformed into a semantic web is described in this section.

A group of students sharing an assignment problem usually discuss the assignment topic meeting face to face every day at the university. Our approach creates an interactive environment to simulate those discussions and turn the collective knowledge generated from such discussions into a comprehensible semantic web that can be accessed not only by the students of the current class, but also by the students who will take the same class in the following semesters and develop it further. The web is a conceptual lattice developed using formal concept analysis which is a machine learning technique. In our approach, individual traces are the data points that the FCA algorithm uses for learning. Traces are stored as a sequence of URL of pages that users visit in a *browsing session* when a user from a particular interest group is searching for a specialized topic. For example, in e-learning, users are students who provide one or more keywords to identify their search domain at the beginning of each session (Fig. 1). Their trails consist of a sequence of URLs annotated by the session title word(s), entered at the beginning of each web session. Web page addresses and session title keywords are the building blocks for our SW. Initially entered words are checked against dictionary of existing set of keywords in the database. This minimizes the redundancy of keywords (e.g. synonyms) and corrects any syntactical errors by users. The evolved SW structure gives authenticated users recommendations in the form of categorized web page links, based on session keywords. In addition, they can browse any notes added previously by authorized users. As user trails are accumulated, browsing sessions begin to intersect one another to create new knowledge. For example, while a student *A* searches web course notes for pages about the "*Public Sector*", she comes across a webpage *p1*, which has been visited by student *X* but it was related to "*IT*, *E-Learning*" in his session. This creates new knowledge relating "*IT*, *E-Learning*" and "*Public Sector*" due to intersection in the corresponding trails.

FCA (Ganter & Wille, 1999) reasoning turns user traces into structured knowledge, the conceptual lattice. This involves two steps: a matrix table is constructed showing keywords that each page satisfies, a conceptual lattice is then assembled from the matrix table. FCA starts with a context $K = (G; M; I)$, where G is a set whose elements are called objects, M is a set whose elements are called attributes, and I is a binary relation between G and M [$(g; m) \in I$ is read "*object g has attribute m*"]. Formalized concepts reflect a relation between objects and attributes. A *formal concept*, C, of a *formal context* $(G; M; I)$ is a pair (A, B) where $A \subseteq G$ is the set of objects (extent of C) and $B \subseteq M$ is the set of attributes (intent of C). The set of all formal concepts of a context K together with the order relation $\leqslant$ is a *complete* lattice, $F(G, M, I)$: For each subset of concepts there is always a unique greatest common subconcept and a unique least common superconcept. In our approach, web page URLs form G, the set of objects. Keywords of session names form M, the set of attributes. A concept in the resulting conceptual lattice is formed of a set of page URLs as the extents and a set of keywords as the intents. Concepts can be a result of either a single user session or multiple sessions that intersect each other.

Fig. 2 displays an example of three different user-sessions that share some common web pages in their trails. For example, "*Web Page1*" is visited by users *A* and *C* having different keywords identifying their session. This indicates the creation of a new concept in

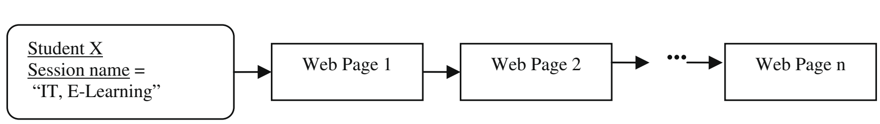

Fig. 1. *User trail*: A student *X* searching for webpages related to *IT* and *E-Learning*: she logs in, enters session title, and browses for related articles. Web pages *1* to *n* will be given under any session with title keywords *IT* and *E-Learning*.