USING AN ORIENTEERING STRATEGY TO BROWSE SEMANTICALLY-ENHANCED EDUCATIONAL WIKI PAGES

Luciano T. E. Pansanato
Univ. Tecnológica Federal do Paraná – Campus de Cornélio Procópio
Av. Alberto Carrazzi, 1640, 86300-000 Cornélio Procópio, PR, Brazil
luciano@utfpr.edu.br

Renata P. M. Fortes
Departamento de Ciências de Computação, Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo – Campus de São Carlos
Caixa Postal 668, 13560-970 São Carlos, SP, Brazil
renata@icmc.usp.br

ABSTRACT
In recent years, Wikis have been used aiming to provide an easy and simple support for keeping the information on the Internet up-to-date and making possible efficient collaborative updating and authoring. The Wiki goals have been achieved, but as the number of pages and corresponding content increases wiki users face difficulties when browsing for wiki pages. This paper presents an orienteering strategy to browse semantically-enhanced educational wiki pages. We present a prototype developed to demonstrate our approach. The results of an experiment to evaluate the prototype are also discussed.

KEYWORDS
Information Seeking, Interactive Information Searching, Semantic Web, Educational Metadata, Wiki.

1. INTRODUCTION

Wikis (Leuf and Cunningham 2001) have been widely used, making efficient collaborative updating and authoring possible. Moreover, the Wikis have stimulating an increasing number of pages and corresponding contents; consequently, wiki users face difficulties when browsing for wiki pages. In Wikis, the support for information seeking activities is important to enable users to find the information they need. However, browsing for information in Wikis may be difficult when users are supported only by the traditional keyword search paradigm.

Although the keyword search paradigm is attractive since it provides a mean to quickly identify the pages containing specific information, it is not efficient to find information in several situations. For example: (a) a user may be uncertain of what he is looking for until the available options are presented; (b) the target cannot be expressed by keywords or (c) the exact terminology used in the pages is unknown, and (d) in cases where a great deal of information and context should be obtained along with the pages (and not only the final page). In such situations, other paradigms such as traditional browsing may be more useful than keyword search (Teevan et al. 2004).

The problem of poor aid for searching information in Wikis has been observed with the CoTeia (Arruda Jr et al. 2002), a Computer Supported Collaborative Learning (CSCL) tool used to complement face-to-face lectures with collaborative learning activities at Institute of Mathematical Sciences and Computing / University of São Paulo (ICMC/USP). CoTeia (available at http://incubadora.fapesp.br/projects/coteia/) is a wiki-based asynchronous collaborative tool analogous to CoWeb (Guzdial 1999). CoTeia users, in particular faculty, frequently look for wiki pages to reuse their material. CoTeia has been used since 2001 and contains

11
a significant amount of teaching material. Although CoTeia provides a keyword search tool, we have observed many cases where users engage in an exhaustive browsing to find what they are seeking.

In fact, typical wiki-based environments, such as CoTeia, can be enhanced by metadata, which may strongly influence information searching techniques and tools. Several Semantic Wikis are mentioned in the literature, e.g. Platypus Wiki (Tazzoli et al. 2004), WikSAR (Aumüller 2005), SemperWiki (Oren 2005), and Semantic MediaWiki (Vökel et al. 2006), and they use semantic information (metadata) to offer improved navigation, browsing, and searching. In general, these Wikis focused on semantic navigation (e.g. providing additional information on the relation a link describes) and semantic search (e.g. allowing a semantic query on the underlying knowledge base). SemperWiki has also provided faceted browsing (Oren et al. 2006); however, it does not integrate this technique with other ones.

The facility in using metadata to find wiki pages (or other kind of information objects) is related to the existence of appropriate support to information seeking strategies. The main objective of our research is to investigate and develop an interactive support inspired by an orienteering strategy (Teevan et al. 2004; O’Day and Jeffries 1993) to browsing for wiki pages, which uses an infrastructure of Semantic Web (W3C 2001). Orienteering denotes a strategy in which people found particular information need through a sequence of small steps (or actions) to narrow the focus of the goal. At each step, prior information and local context are used to decide the next step.

To support orienteering strategy we have developed a prototype integrating a variety of tools related to information seeking. We argue that an information seeking environment could make available several categories of tools (including keyword search), enabling users to choose the appropriate tool or the best combination of tools (that is, the best strategy) in agreement with different levels of users’ ability, background, preferences, and kind of information that they are looking for at moment. It enables users to prioritize different ways their choices of tools to each step in an orienteering strategy. These tools may present complementary advantages in an information seeking task. We expected that users take advantage of this kind of environment to carry out information seeking tasks as an alternative to the one-size-fits-all approach of the keyword search paradigm.

Section 2 describes the orienteering strategy and the techniques we have explored to support information seeking activities. Sections 3 and 4 present the prototype and an experiment that was performed to show the prototype effectiveness. Section 5 is dedicated to conclusions and future work.

2. STRATEGIES AND TOOLS

We take advantage of the levels of information seeking activities as described by Bates (1990). Based on empirical studies of the information seeking behavior of experienced library users, Bates distinguishes four levels of activity:

1. **Move.** A move is an identifiable thought or action that is a part of information seeking. For example, locating some portion of text in a wiki page.

2. **Tactic.** A tactic is one or a handful of moves made to further an information seeking activity. For example, broadening or narrowing a query to retrieve a larger or smaller number of wiki pages.

3. **Stratagem.** A stratagem is a complex set of moves and/or tactics, and generally involves both a particular information domain anticipated being productive by the user, and a mode of tackling the particular organization of that domain. For example, finding all wiki pages which are related to a particular course.

4. **Strategy.** A strategy is a plan for satisfying specific information need, and may include combinations of all the previously mentioned types of information seeking activities. For example, performing a keyword search to retrieve wiki pages related to specific terms describing a subject, browsing through those considered relevant and then finding references in the text to books and articles.

To help users to interactively find wiki pages at CoTeia, we have designed a support for *moves* and *tactics* implemented as a set of integrated tools. Combining the tools users may compose *stratagems* and *strategies*. Subsection 2.1 describes the underlying orienteering strategy used to integrate the tools which has influenced the choice of the tools as well as some extensions. The tools are briefly discussed in Subsection 2.2.
2.1 Orienteering Strategy

Orienteering is a sport of finding one’s way across country on foot using a map and a compass. The main strategy consists in using the information on the current position to make better choices about the way to reach the next checkpoint or the final target. Furthermore, the orienteer must constantly concentrate, make decisions, and keep track of the path covered.

A similar strategy has been described in studies of information seeking behavior in the literature (Teevan et al. 2004; O’Day and Jeffries 1993). Orienteering denotes a strategy (or a class of strategies) in which people found a particular information need through a sequence of small steps (moves/tactics/strategies, according to Bates, or simply actions) to narrow down for the goal. At each step, prior information and local context are used to decide the next step. For example, first one can submit a query to a search engine to get into the proximity of the information that satisfy the target information need, and then explore the links retrieved to find the desired information. Other works also described this strategy (Navarro-Prieto et al. 1999; Hölscher and Strube 2000), though not under that name.

We have elected the following categories of techniques that could help the user to engage in an orienteering strategy: combining browsing and searching; showing context; previewing content; keeping interaction history; and narrowing toward the goal. The latter is a key feature of an orienteering strategy. It consists in starting with a general query and using small steps to narrow down the information space until the user find what they are looking for. Next subsection outlines the functionality and briefly discusses the support for each of the foregoing categories of techniques.

2.2 Orienteering Tools

We have implemented common tools and extended others to incorporate metadata in their mechanism, i.e. a Resource Description Framework (RDF) model (W3C 2004) about a collection of wiki pages used in educational activities. The associated metadata record for each wiki page of CoTeia is an application profile (Duval and Hodgins 2003) based on the IEEE Learning Object Metadata (LOM) standard (IEEE LTSC 2002) and Dublin Core (DC) standard (DCMI 2005). Furthermore, the application profile of CoTeia was augmented with five metadata elements to support our specific application domain (educational wiki pages at ICMC/USP). With appropriate support in this context, metadata can better assist CoTeia users in finding/discovering wiki pages (educational resources). Methods for automatic metadata generation were explored to populate the RDF model as described in previous work (Pansanato and Fortes 2005a, b). The nine implemented tools were: keyword search; facet browse; highlight; flag; work memory; sort; group; restart and undo.

**Keyword search**: This tool is a search engine which searches previously indexed wiki pages for specified keywords and returns a list of those where the keywords were found. In spite of this approach being straightforward, it works for its purpose: to obtain a list of wiki pages relating to the keywords the user entered. Some search engines can use relevance ranking that many users find disconcerting: some pages can be ranked high even if they do not contain all the keywords.

**Facet browse**: Faceted browsing is one way to use faceted metadata to allow users to find information. Metadata can have several facets: attributes in various orthogonal sets of categories. For example, in the domain of educational wiki pages, possible facets might be authors (professors or students). This tool (Facet browse) allows users to filter a set of items (e.g. wiki pages) by progressively selecting from only valid values (instances) of facets. The list of valid values is filtered to show only those that have results available. Thus, it is impossible to get an empty result. The combination of facets and hierarchy can help the user to decide how to start and to explore the collection.

**Highlight**: This tool displays the occurrences of query terms within the context of the document retrieved. It is useful to support local exploration because most users do not read pages carefully when they scan text for what they are looking for. We extend this technique to highlight also names, locations, email addresses, and phone numbers. The knowledge about people’s names and locations is harvested from ICMC/USP website and added inside the RDF model.

**Flag**: This tool implements a preview technique that consists in automatically flagging a result which contains certain content so it can be found easily among other results. For example, results of documents that
contain names, locations, email addresses, dates, or times. It is useful for users to be able to see at a glance whether the results they get in response to an action have a particular content.

**Work memory:** The underlying technique of this tool consists in providing a special memory resource which can store results and some operations to handle them. Basically, we implemented the same memory functions of a typical calculator: \( M+ \) (sum), \( M- \) (difference), \( MR \) (recall), and \( MC \) (clear). This support is important in cases where the goal consists of a collection of results that should be obtained from a variety of different strategies, not just at the end of a strategy. For example, collecting links to wiki pages which contains educational material (for reuse) may involve the storage of intermediary results to compose the final list of those considered relevant.

**Sort:** This tool sorts the results by relevance, as determined by the number and location of matched words in the wiki page, and by any metadata of wiki pages in the RDF model. While all search engines sort results by relevance as default, a few of them can sort by other option (e.g. Ask, http://www.ask.com, allows to sort saved results by date or title).

**Group:** Some search engines attempt to classify results automatically into concepts (or domains), such as Vivisimo (http://vivisimo.com). The drawback of this technique is that the categories automatically generated are not always well organized. This tool (Group) implements a straightforward improvement by exploiting the available metadata in the RDF model to organize the results according to the same category layout that is used by the tool for faceted browsing.

**Undo and restart:** The reverse of actions is supported by means of undoing and starting over the task. The Undo tool allows user to cancel any performed action and consequently, the results will reflect on the new sequence. The Restart tool allows user to begin again from scratch. These tools help to keep the interaction process under user control.

3. **Prototype**

We have developed a prototype (Figure 1) that has been used to browse wiki pages stored in two CoTeia repositories (see Table 1). The interface is divided into three parts: the set of tools on the left, the history of actions on the top right, and the results and content area on the bottom right.

**Table 1. CoTeia repositories**

<table>
<thead>
<tr>
<th>Main period</th>
<th>URL</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004-2006</td>
<td><a href="http://coteia.icmc.usp.br/coteia/">http://coteia.icmc.usp.br/coteia/</a></td>
<td>1,731</td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td></td>
<td><strong>4,234</strong></td>
</tr>
</tbody>
</table>

Using a tool often corresponds to a move, tactic, or stratagem, and this interaction is called action in the interface. The sequence of performed actions during a strategy is showed in the history in the top right. The history serves as important aid to reduce working memory load. Any particular information seeking activity may include other behaviors and cognitions that cannot be captured by the prototype.

When the user performs an action, the results (i.e. a list of links to wiki pages in this case) to that action are presented. The next action has effect on the current set of results. This is an important aspect in the interaction process of the user with the prototype: the tools are integrated so that the results obtained in one are used as the input to the next. Furthermore, the user may cancel (undo) any one of the performed actions and the results will reflect the new resultant sequence.

Figure 1 shows the result of a hypothetical interaction in which the user has chosen different tools to look for wiki pages with material (slides) used in a particular course. First, the user selected the term “SCER0225 Hipermitia” (A) in the facet Course to refine the entire collection of all wiki pages to a set of those related with the course. Second, the user performed a keyword search (B) to narrow the current set of wiki pages (from previous tool) to those that contain the word “aula” (lecture). Next, the user sorts the results alphabetically by Title (C) and follows a link to the wiki page entitled “Aula a Aula” (D). The wiki page is showed in the content area (E). Next, the highlight tool is used to highlight the word “Aula” (F) in the
content of the wiki page (the highlight tool does not appear in the figure). Finally, the flag tool is used to put a mark (a red lozenge) on the results which contain one or more people’s names.

The prototype was built using the scripting language PHP (http://www.php.net/) and the database manager MySQL (http://www.mysql.com/). We have used AJAX (Asynchronous JavaScript and XML) programming techniques to allow better user interactions in the interface. The RDF model is manipulated and searched using RAP (RDF API for PHP, http://sourceforge.net/projects/rdfapi-php/), a Semantic Web toolkit for PHP.

Figure 1. Screenshot of the prototype

4. EXPERIMENT

This section presents an experimental study comparing the prototype to a keyword search approach. In previous work (Pansaneto and Fortes 2007) we conducted a user-based evaluation to test and evaluate the prototype, and, through the users’ interaction with the prototype, to get knowledge about that. Hence this section presents the results of a new study to answer the research question: Is the prototype approach better than the traditional keyword search approach?

4.1 Methodology

A total of 34 undergraduate students in their fourth year at ICMC-USP were recruited as participants. Seventeen participants were randomly assigned to each experimental condition in a blocked subject-object study (Table 2). The participants were all experienced users of search engines, searching for information daily. They were regular users of the CoTeia with 3-4 years of experience and weekly they access wiki pages for reading.
The participants completed a set of tasks that involved browsing for information in CoTeia repositories. We were interested in covering the largest number of scenarios using the smallest number of tasks. Thus, we defined tasks in four different task scenarios (types), based on Shneiderman’s definition (Shneiderman 1997): specific fact-finding, extended fact-finding, open-ended browsing and exploration of availability (Table 3).

The evaluation was comprised of four sessions as follows: (a) a survey session, (b) a training session, (c) a task session, and (d) a perceptive evaluation session. The survey was designed to collect demographic information and participant’s experience with search tools and experience with CoTeia. A training session was given, including an overview of the system approach (prototype or keyword search) so that the participant could become familiar with the environment. After the training session, participants completed tasks. After each task participants answered the questions (concerning that task) about relevance of the results obtained, usefulness and easy of use of the system approach. Also after performing the tasks, they were asked for their perceptions of the system approach that they had experienced.

Table 2. Experiment design (seventeen participants per object study)

<table>
<thead>
<tr>
<th>Groups</th>
<th>Prototype</th>
<th>Keyword search</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group I</td>
<td>Group II</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 3. Four tasks in four scenarios

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Scenarios</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: Look for information about data types of two particular programming languages.</td>
<td>specific fact-finding</td>
</tr>
<tr>
<td>T2: Look for names and e-mails of course monitors of a particular course.</td>
<td>extended fact-finding</td>
</tr>
<tr>
<td>T3: Look information about how to develop applications to people with special needs.</td>
<td>open-ended browsing</td>
</tr>
<tr>
<td>T4: Look for the learning material available to a particular course.</td>
<td>exploration of availability</td>
</tr>
</tbody>
</table>

4.2 Measurement

The experience of the participants was defined by means of a survey applied at the first session. These data provides the input to the characterization of the participants, and hence are the independent variables in the experiment.

The dependent variables of performance are the time to accomplish tasks and relevance of the answers by mean of recall and precision measures. Recall is the fraction of the relevant documents which has been retrieved, and precision is the fraction of the retrieved documents which is relevant (Baeza-Yates and Ribeiro-Neto, 1999). These variables in each task may compete with each other. For example, long completion time may be consequence of high relevance, or low relevance may be consequence of short completion time.

The results from task questions and perceptive evaluation also were investigated. The judgments were made on 7-point Likert scales, with 1 meaning “none”, 4 meaning “neutral”, and 7 meaning “extreme”.

4.3 Results and Discussions

As a first step in analyzing the data, we use descriptive statistics to visualize the data collected. The mean completion time and relevance (recall and precision) for each of the experimental group are shown in Table 4. From Table 4, we see that participants from the Group I seem to have a higher performance. To gain an even better understanding of the data, a box plot for each measure is done (Figure 2). The box plot for the time measure is more interesting: it has the highest and lowest value besides larger variability, but it has no outliers. Moreover, from Figure 2, we can see that there is a clear pattern within Group II participants: all of them had a lower performance. Thus, it may be possible to prove the difference in a hypothesis test.

An unpaired t-test was conducted to examine whether there were any performance differences between groups. According to the analysis, there was a significant difference in time to accomplish tasks ($t (32) = -0.59$, $p = .557$) and relevance of the answers regarding the measures of recall ($t (32) = 2.41$, $p = .022$) and precision ($t (32) = 1.35$, $p = .187$). All these tests provide statistical indications that participants had a better performance using the prototype than keyword search.
Table 4. Statistical results

<table>
<thead>
<tr>
<th>Group</th>
<th>Time</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SD</td>
<td>Mean SD</td>
<td>Mean SD</td>
</tr>
<tr>
<td>I</td>
<td>29:42 9:46</td>
<td>0.31 0.12</td>
<td>0.66 0.14</td>
</tr>
<tr>
<td>II</td>
<td>31:24 6:41</td>
<td>0.22 0.10</td>
<td>0.58 0.17</td>
</tr>
</tbody>
</table>

Figure 2. Box plots of time, recall, and precision for the two groups

In the last session of the experimental study, a set of questions was presented to the participants to rate their experience with the system that they have used. The questionnaire consisted of six questions (see Table 5 for a summary of the questions and results).

Table 5. Summary of the perceptive evaluation questions and results

<table>
<thead>
<tr>
<th>Questions</th>
<th>Group I</th>
<th></th>
<th>Group II</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1. Indicate the understandability.</td>
<td>5.41 5.00 0.94</td>
<td></td>
<td>5.59 6.00 0.87</td>
<td></td>
</tr>
<tr>
<td>Q2. Indicate the ease of use.</td>
<td>5.24 5.00 0.97</td>
<td></td>
<td>5.47 6.00 0.87</td>
<td></td>
</tr>
<tr>
<td>Q3. Indicate the usefulness.</td>
<td>5.29 5.00 1.16</td>
<td></td>
<td>4.47 5.00 1.18</td>
<td></td>
</tr>
<tr>
<td>Q4. How often did you get lost?</td>
<td>2.82 3.00 0.95</td>
<td></td>
<td>2.88 3.00 1.05</td>
<td></td>
</tr>
<tr>
<td>Q5. Would you consider future re-use?</td>
<td>4.47 6.00 1.33</td>
<td></td>
<td>4.41 5.00 1.58</td>
<td></td>
</tr>
<tr>
<td>Q6. Which your satisfaction in using the system?</td>
<td>5.00 5.00 0.71</td>
<td></td>
<td>4.29 5.00 1.45</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

Browsing for information in Wikis can become difficult as the number of pages and corresponding contents increases. We present a prototype developed to demonstrate our initial ideas combining searching tools to compose an orienteering strategy to browse semantically-enhanced Educational wiki pages. It is worth to observe that users may choose an appropriate tool or a combination of tools to take steps along the way to satisfy their information needs. An experiment was performed to answer the research question: *Is the prototype approach better than the traditional keyword search approach?* The results provide statistical indications that participants had a better performance using the prototype than keyword search.

As future work, we intend to make an effort to address other experiments with different types of user, in particular novice students, to investigate their interactions using the prototype. We also plan to recognize repetitive strategies that can yield to behavioral patterns.

ACKNOWLEDGEMENT

Our thanks to FINEP, FAPESP and UTFPR for funding parts of this research.
REFERENCES


18