Internet Information Retrieval for Enabling Student Projects

Nader Mohamed, Jameela Al-Jaroodi, and Imad Jawhar

The College of Information Technology
United Arab Emirates University
Al Ain, P.O. Box 17551, UAE

Abstract
Project-based learning (PBL) is becoming an integral part of many information technology (IT) courses. Students are guided through individual or group projects to learn and apply many of the IT concepts they learn in a course. However, a significant percentage of IT projects depend on the existence of large sets of information or live data. Therefore, it becomes difficult to complete and properly test many of the implemented projects. Some examples of such projects are financial data mining, information notification systems, weather monitoring systems, and decision support systems. While the needed information for the student projects is usually available over the Internet, there is no easy way to retrieve and reuse the required information. In this paper, we propose, describe, and discuss a tool called InetRetriever. This tool can be easily used by students to retrieve in real-time any required real information from the Internet. The students can use this tool to implement, execute, and test their projects with real life data. This tool was tested in different information technology courses to enable effective and realistic project-based learning. Using this tool we observed increased student interest in the project development and higher levels of interactions and learning.

Key Words: educational tools, project-based learning, Internet, information retrieval

1. Introduction

Project-based learning (PBL) is a model of learning that utilizes project problems and hands-on practices to enable the learning process. Projects are usually complex tasks based on challenging or real-life problems that involve students in problem analysis, design, problem solving, decision making, implementation, and documentation [1][2]. This gives students the opportunity to work over extended periods of time (and usually in groups) utilizing concepts learned from one or more courses to produce realistic solutions or products. In addition, it gives students the opportunity to practice and apply the theoretical concepts into practical solutions. In [21] the author provides a more detailed list of the benefits of PBL. On the other hand, there are several challenges for the teachers using PBL as described in [3], one of which, that is more relevant to our work, is recognizing situations that make for good projects. More particularly, in IT courses, projects are about technology and information. Yet it is hard to find realistic projects that can be implemented and evaluated effectively.

Information technology is one of the educational subjects in which the PBL model can play an important role in the learning process [3]. Most of the information technology courses at the higher educational level cannot be effectively taught without giving the students the opportunity to apply and practice the concepts learned in class and to gain more knowledge through active hands-on involvement. PBL offers a good approach to advance and motivate student-centered learning and develop higher-level thinking and judgment skills.

One of the difficulties in information technology higher education is mapping the concepts taught in courses to real-life needs and situations. Students need to practice when they can utilize the theoretical concepts they learned in class to solve practical problems around them. PBL provides an excellent approach to practice and apply this mapping skill, while learning new concepts and gaining more knowledge.

To apply PBL, one of the issues involved is choosing the right projects that match the course contents and goals. There are two types of projects in information technology courses. The first is projects based on problems that do not need real information to be developed and tested. The second is projects that cannot be developed or validated without having real data. Examples of these projects are financial data mining, information notification systems, weather monitoring systems, and decision support systems. Due to the difficulty in obtaining real data, teachers and students avoid working on this type of projects or work on limited
samples of data created specifically for a particular project.

In this paper, we propose utilizing the Internet as a source for the information needed by different student projects. The Internet contains huge amounts of information in so many different fields such as weather information, stock prices, currency exchange rates, real-estate information, statistical data in various areas and other useful and real-time information. We developed a new tool called InetRetriever to obtain and reuse internet information for student projects. This tool can be used by students to retrieve any publicly available information from the Internet, and use this information to implement and validate their projects. We have validated the usefulness of the developed tool in a number of student projects in higher education in information technology courses.

In the rest of the paper, Section 2 discusses related work. The concept of utilizing the Internet as the source of information is discussed in Section 3, while in Section 4, the InetRetriever tool is discussed. Section 5 discusses the validation of the usefulness of the developed tool and Section 6 concludes the paper.

2. Related Work

To discuss work relating to our proposed tool, we will cover two main areas. (1) Project-based learning (PBL) and the issues involved in applying the technique in IT courses. (2) Other tools and methodologies currently used in IT projects.

PBL is not a new approach in teaching; however it is currently picking up pace and becoming more popular in many disciplines. Sometimes known as problem-based or case-based learning, PBL provides an active learning environment where students are involved in solving a predefined problem. The steps to reach a solution are planned, researched, and executed by the students, while learning relevant concepts, and exploring techniques and methodologies. A review of research in PBL [19] defines PBL as “a model that organizes learning around projects.” The authors then define several criteria defining PBL projects such as projects being central to the curriculum, projects that are focused on central concepts, and that involve students in a constructive investigation. In [21], the authors state two main factors as the driving force for the increasing use of PBL. First, there has been a revolution in learning theory that confirmed that doing is much more effective than hearing or seeing. Second, the world has changed and students need skills as well as the knowledge to succeed. PBL provides the means to facilitate students’ preparation by actively involving them and giving them the chance to learn skills in addition to the knowledge. PBL is used in many disciplines, and at many levels, from grade school to graduate studies. Several applications were in medicine, science, engineering, and most recently in information technology. For example, this approach has been applied in teaching multimedia courses [18]. Here the course revolves around a multi-phase multimedia project that is assessed based on the outcomes achieved in each phase. Another example [20] describes the use of PBL in a web-based software engineering course. The authors in [22] describe applying PBL in the computer science curriculum at the University of Applied Sciences - Technikum Wien. Furthermore, in [23] there is a description of how PBL is used effectively to teach computer programming.

In addition to using PBL, there are also many tools and methodologies that support active learning and enable the learning process. Particularly many tools are currently available to support higher education in the information technology field. These tools and methodologies were developed mainly to stimulate the student learning process. One example is a new tool for web-based educational system. This tool is based on Service Oriented Architecture (SOA). It aims to achieve interoperability among student applications by utilizing reusable service logic [4]. Other examples are an authoring tool to enable videogame development and enhance gaming concepts education [5], an automatic visualization tool "PGT" for programming education [6], and an education tool for design automation of CMOS cells [7].

There are also some efforts to create a cyber infrastructure for some fields in information technology to provide education and research tools for students and researchers working in the field. nanoHUB.org is a network of universities supporting the National Nanotechnology Initiative by bringing computational tools with their education material online, making these tools easy to use [8]. This allows students to utilize the available tools for implementing their nanotechnology projects.

3. Internet Information

As we mentioned earlier, there are two types of projects in information technology courses. The first is projects based on problems that do not need real information to be developed and tested, while the second is projects that cannot be developed or validated without having real data. To solve this problem, we tried to utilize the Internet as a source of real information for student projects. The Internet provides huge information related to stock market information, currency exchange
rates, temperatures, oil prices, gold prices, interest rates, etc. Most of this information is available and updated in real-time. All this information can be reused by students to implement and validate their IT projects.

The Internet information is usually available in dynamic HTML documents [9], XML documents [10], web services, or RSS feeds [11]. If the information needed by a student project is available in web services, then the student can easily use the corresponding web service which provides the required information. Web services can provide a structured and simplified way to obtain services or specific information from the Internet. Web services provide web APIs that can be accessed over a network, such as the Internet, and executed on a remote system hosting the requested services. The main problem is that not all types of information available over the Internet are provided by web services. Most of the useful information on the Internet is still only available in semi-structured HTML documents.

Unlike XML documents, HTML documents do not have any semantics for their data. HTML documents usually contain tags, scripts, links, and user defined data. Obtaining specific data from a dynamic HTML document for reuse in other applications can be a complex task. It is very difficult to identify the required data components and dynamically use them in other applications.

Some research was conducted to benefit from the Internet HTML documents. One example is developing an approach to link the large amounts of data that are currently available in HTML documents to the Semantic Web ontology [12]. Another example is developing an approach that automatically captures the semantic hierarchies of HTML tables [13]. Some research effort was also conducted to transform HTML documents to another format to satisfy specific applications. One example of this transformation is from HTML Product Catalogues source code and images to RDF [14]. In contrast, our approach is developed to be used for generic purposes. This approach is designed to simplify the process of extracting real-time information from HTML documents. It provides several techniques to extract the information from these documents regardless of how it is formatted (e.g. in tables or as plain text). These techniques can be used to extract online information and reuse them in other applications such as student projects.

4. InetRetriever

We have recently developed a simple and efficient approach for retrieving live HTML-based Internet information [15]. This approach can be used to retrieve updated data from the Internet. The main idea of this approach is based on finding fixed titles or headers that appear in browsers for HTML documents directly or semi-directly before the needed dynamic information. These fixed titles or headers are used as references to know the position of the required dynamic information. The developed approach provides a simple and efficient technique to retrieve any required public information students need to complete and test their projects.

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>get(header)</td>
<td>To return the next field directly after the defined header. The search starts from the beginning of the page.</td>
</tr>
<tr>
<td>get(n, header)</td>
<td>To return the next field directly after the defined header appears n times. The search starts from the beginning of the page.</td>
</tr>
<tr>
<td>get(n, header, i)</td>
<td>To return the field after skipping i fields after the defined header appears n times. The search starts from the beginning of the page.</td>
</tr>
<tr>
<td>getWI()</td>
<td>To return the next field from the current read pointer position.</td>
</tr>
<tr>
<td>getWI(i)</td>
<td>To return the field after skipping i fields from the current read pointer position.</td>
</tr>
<tr>
<td>getWI(header)</td>
<td>To return the field located directly after the specified header from the current read pointer position.</td>
</tr>
<tr>
<td>getWI(n, header)</td>
<td>To return the field after the occurrence of the header n times from the current pointer position.</td>
</tr>
<tr>
<td>getWI(n, header, i)</td>
<td>To return the field after skipping i fields after the defined header appears n times from the current pointer position.</td>
</tr>
</tbody>
</table>

The proposed approach is developed as a Java class. Multiple objects can be created from this class for different Internet HTML documents that contain some of the required information. A number of techniques were developed to find this information in any HTML document. These techniques are implemented in a set of methods listed in Table 1.

All these techniques can be used to retrieve updated Internet information. As soon as the fields are identified the student can specify the arguments for the get or getWI methods, which will allow the student to retrieve the required information. Students can use any HTML
documents on the Internet to define Internet variables that they need.

The first technique is to find information directly after a specific text header. For example, in Figure 1, the Wal-Mart Stores, Inc. Stock information is displayed from the Yahoo Finance site. The user can find the last trade price, the previous closing price, and the trade volume from this page. To get the last trade price, the user needs to call the `get("Last Trade:")` method. The get method will search for the title provided, “Last Trade:”, and return the next data field after this title field, “48.10.”

The second technique is to find certain information after the appearance of a specific header for the nth time. For example, Figure 2 is web information for RAK Properties with symbol RAKPROP that is listed in Abu Dhabi Securities Market. We can see that, the bid volume amount appears after the "RAKPROP" header while that header appeared twice in the page. To get the bid volume, this technique will search for two "RAKPROP" text headers before returning the bid volume amount. This technique is implemented in another method with interface `get(n, header)`.

The third technique is to find information semi-directly after a specific fixed title or header. This finds the i^th information field after the appearance of a specific header n times. For example in the RAKPROP example, the user wants to get the current price of the stock. This price is listed 7 fields after the second appearance of "RAKPROP" header. The interface for this method is `get(n, header, i)`. Users can use any HTML document on the Internet to define their variables. More information about the methods, optimization techniques, implementation, and performance of this approach can be found in [15].

5. Student Projects

Using the proposed approach, new real-time applications can be easily developed by students. These applications can be integrated with some information from the Internet. For example, a number of projects related to stock investments were developed using the proposed approach in the Internet System Software course at the College of Information Technology, UAE University, UAE. Internet system Software is a senior course required for the undergraduate networking degree while it is an optional course for the other degrees such as Enterprise Systems, Software Engineering, and Computer Science. InetRetriever allowed students to accomplish these projects that require real information provided in real-time. Some of these projects are:

I. Notification System for Stock Price Changes in Dubai Financial Market: This system can be used by users to define certain price change criteria. The system will notify the users by emails or by any other message type mechanism whenever those criteria are met. Examples of the criteria are when a stock reaches a certain price level or when a stock price increases/decreases by a certain percentage. The system monitors stock prices through the Internet. It will generate notification messages whenever the defined criteria are met.

II. Multiple Stock Markets Monitoring System: A customer may have stocks listed in multiple financial markets. Although each market provides a software tool for live quotes there is no software tool that displays live quotes for a set of stocks that
belong to different securities markets. For example, a customer may have one stock in Abu Dhabi Securities Market, two stocks in Dubai Financial Market, and three in Kuwait stock market. He/She is only interested in monitoring the prices of these stocks. The aim of this project was to implement a networked stock monitoring system in which users can define and monitor the prices of specific stocks in different markets. This system depends on the Internet to get information about different stocks listed in different markets.

InetRetriever was also used for a number of senior exhibition projects at the college. One example is implementing a generic Internet information notification. The generic Internet notification is a framework to enable monitoring any type of information on a single web page or on multiple web pages over the Internet [16]. The students have utilized InetRetriever to implement and evaluate that framework. More information about their implementation can be found in [17].

InetRetriever helps students to complete their projects and enables effective PBL in IT courses. It opens new doors for students to develop and test new types of projects. This motivates students to be creative and innovative in some new types of applications in which it was very difficult for students to develop and validate projects in them. It also gives them the opportunity to learn new concepts and integrate different skills and knowledge in their work. The tool offers a good approach to advance and motivate student-centered learning and develop higher-level thinking and judgment skills.

6. Conclusion

PBL has multiple benefits and advantages for the students. One positive effect is that PBL enhances the quality of student learning and improves their skills. Another benefit is that students become more professional, self-reliant, and collaborative [19]. However, there are also difficulties involved such as difficulty choosing and designing problems or projects that meet the required criteria (centrality, driving question, constructive investigations, autonomy, and realism [19]). Realism is a key area to be considered in IT. As we stated earlier, many projects require some real-life data and up-to-date information to be completed and evaluated. As a result, it is necessary to provide the tools and methodologies than can help the students create realistic projects. Therefore, providing some tools that can help make projects more suitable and more realistic is important. In this paper, we introduced a new tool that enables student projects that require access to and utilization of real information. This tool stimulates student creativity and encourages them to develop solutions that cross the boundaries of diverse technologies. This tool helps to provide an opportunity in which students can utilize existing real-life information that is publicly available over the Internet to complete their projects, and thus produce innovative ideas and solutions. The tool also enables the development of interdisciplinary projects, where various areas in information technology, business and science may be integrated. This tool was successfully tested in multiple information technology courses. The students were able to complete creative projects needed in real life.

References