Using Open Web APIs in Teaching Web Mining

Hsinchun Chen, *Fellow, IEEE*, Xin Li, *Student Member, IEEE*, Michael Chau, *Member, IEEE*, Yi-Jen Ho, and Chunju Tseng

***Abstract*—With the advent of the World Wide Web, many business applications that utilize data mining and text mining techniques to extract useful business information on the Web have evolved from Web searching to Web mining. It is important for students to acquire knowledge and hands-on experience in Web mining during their education in information systems curricula. This paper reports on an experience using open Web Application Programming Interfaces (APIs) that have been made available by major Internet companies (e.g., Google, Amazon, and eBay) in a class project to teach Web mining applications. The instructor's observations of the students' performance and a survey of the students' opinions show that the class project achieved its objectives and students acquired valuable experience in leveraging the APIs to build interesting Web mining applications.**

***Index Terms*—Data mining, information System, open Web APIs, visualization, Web computing, Web mining.**

I. Introduction

THE World Wide Web has become an indispensable part of many business organizations. In order to effectively utilize the power of the Web, information technology (IT) professionals need to have sufficient knowledge and experience in various Web technologies and applications. In recent years, courses in Internet- and Web-related topics have been offered in many universities to equip students with such knowledge. In addition to basic courses such as Internet networking and Internet application development, more advanced topics, such as Web mining, are becoming increasingly important. Web mining [1], [2] has been frequently used in real world applications such as business intelligence [3], Website design [4], and customer opinion analysis [5]. It is imperative that students acquire knowledge and hands-on experience in applying Web mining techniques.

However, building a Web mining application from scratch is not an easy task that every student can complete in a semester. Recently, many large companies such as Google, Microsoft, Amazon, and eBay have opened access to their services and data through Application Programming Interfaces (APIs). In education, these APIs provide an ideal playground for students to gain some practical skills in Web application development and experiment with the Web mining techniques they learn in class.

This paper reports on an experience designing a Web mining class project based on open Web APIs for students in a graduate-level course at the University of Arizona. The paper discusses the instructor's observations of the student projects. Three student projects for different domains are described in detail as examples of the students' achievements. The paper also provides the students' self-assessment of their learning experience in the project.

Manuscript received August 05, 2008; revised September 14, 2008. First published June 30, 2009; current version published November 04, 2009. This work was supported in part by the NSF National Science Digital Library Program, "Intelligent Collection Services for and about Educators and Students: Logging, Spidering, Analysis and Visualization," DUE-0121741.

H. Chen, Y.-J. Ho, and C. Tseng are with the Department of Management Information Systems, The University of Arizona, Tucson, AZ 85721 USA (e-mail: hchen@email.arizona.edu; ianho823@email.arizona.edu; chunju@email.arizona.edu).

X. Li is with the Department of Information Systems, The City University of Hong Kong, Kowloon, Hong Kong (e-mail: xin.li@cityu.edu.hk).

M. Chau is with the School of Business, The University of Hong Kong, Pokfulam, Hong Kong (e-mail: mchau@business.hku.hk).

Digital Object Identifier 10.1109/TE.2008.930509

II. Background

A. Web Mining

Since the advent of the Internet, many studies have investigated the possibility of extracting knowledge and patterns from the Web, because it is publicly available and contains a rich set of resources. Many Web mining techniques are adopted from data mining, text mining, and information retrieval research [1], [2]. Most of these studies aimed to discover resources, patterns, and knowledge from the Web and Web-related data (such as Web server logs).

Web mining research can be classified into three categories: Web content mining, Web structure mining, and Web usage mining [6]. Web content mining refers to the discovery of useful information from Web contents, including text, images, audio, video, etc. Resource discovery from the Web [7]–[9], Web document categorization and clustering [10], [11], and information extraction from Web pages [12] are important Web content mining topics.

Web structure mining studies the model underlying the link structures of the Web. Such models have been widely used to infer important information about Web pages. Hyperlinks among Web pages are usually indicators of high relevance or good quality. Web structure mining has been used for search engine result ranking, with PageRank [13] and HITS [14] being the most widely used, and has also been applied to analyze online activities of different social groups [15].

Web usage mining focuses on using data mining techniques to analyze search logs or other activity logs to find interesting patterns. A Web server log contains information about every visit to the pages hosted on the server, such as files requested, user's IP address, and timestamp. By performing analysis on Web usage log data, Web mining systems can discover knowledge about a system's usage characteristics and the users' interests. Such knowledge can be used for personalized Web applications, marketing, Website evaluation, and decision support [4], [16], [17].

B. Open Web APIs

The Web no longer contains merely static pages and contents. Powered by modern database technologies, network technologies, and computational ability developments, many Websites provide sophisticated functionalities to customers. By viewing a Website as an application [18], the concept of Web APIs enables

direct access to the Website functionalities. Hundreds of Websites have published Web APIs for access to their services [19] in order to leverage third party efforts on value-adding services.

Many of the open Web APIs are constructed based on the Web services architecture. Web services are a technique developed to support interplatform function calling. Before the emergence of Web services, several similar techniques were used in different contexts, such as Remote Procedure Call (RPC), Common Object Request Broker Architecture (CORBA), Component Object Model (COM), etc. [20]. Web services are based on Extensible Markup Language (XML) and Hypertext Transfer Protocol (HTTP), which provide better platform and language independent interoperability [21]. The Web APIs make it easier to integrate a Website's functions and data into a third party's Website or software package.

Companies that open Web APIs to the public belong to several different categories, such as Web search (e.g., Google and Yahoo), geographic maps (e.g., Google Maps, Yahoo Map, and Microsoft MapPoint), e-commerce (e.g., Amazon, eBay, and PayPal), and others (e.g., BBC and Skype). We describe the three sets of open Web APIs that are most relevant to the current paper, namely, Amazon, eBay, and Google.

1) Amazon: Amazon Web Services offers Web APIs to access Amazon's product data and e-commerce functionality (*Amazon E-Commerce Service*) and sales history data (*Amazon Historical Pricing*). Amazon also provides Web services for storage (*Amazon S3*), queue (*Amazon Simple Queue Service*), and Web search (*Alexa Web Search Platform*) [22].

Amazon E-Commerce Service is based on Simple Object Access Protocol (SOAP) and Web Service Description Language (WSDL) standards [23]. In addition, Amazon provides a REST (REpresentational State Transfer) protocol for access. This set of APIs provides detailed product attributes, images, pricing information, and customer reviews for virtually all products in Amazon.com. Using this set of open Web APIs, developers can perform complex search functions on the available products in Amazon's virtual market. *Amazon E-Commerce Service* also provides an interface to access the customers' wish lists.

The *Amazon Historical Pricing* service gives developers programmatic access to the actual sales data for books, music, videos, and DVDs (as sold by third party sellers on Amazon.com) since 2002. The data includes the average, minimum, maximum, and median prices for the specified items over the given date range(s).

2) eBay: Using the *eBay Web Services API*, developers can create Web-based applications to conduct business with the eBay Platform [24]. Developers can perform functions such as sales management, item search, and user account management [25]. The *eBay Web Services API* is also based on the SOAP and WSDL standards. The *eBay Web Services API* provides wrappers of the SOAP APIs in Java, .NET, and PHP. eBay provides a testing environment—sandbox.ebay.com—to developers. Developers can add or remove products on this system just for the purpose of program testing.

3) Google: Google provides open Web APIs for accessing their advertisement service (*AdWords API*), blog service (*Blogger Atom API*), map service (*Google Maps API*), and Web search service (*Google Search API*). The most widely used are the *Google Search API* and the *Google Maps API* [26].

The *Google Search API* is implemented as a Web service using SOAP and WSDL standards. Google provides a Java library which wraps the SOAP APIs. The APIs enable developers to query all the Web page indexes on Google's server. They also allow developers to access the cached Web pages on Google's server and get suggested spellings for incorrect spellings of the search keywords [27].

The *Google Maps API* is a JavaScript API, designed mainly for data representation purposes, which can embed Google Maps in Web pages. The *Google Maps API* works in a browser interface.

C. Teaching Web Mining and Software Development

In recent years, due to the rapid development of the Web, Web mining techniques and data mining techniques are being introduced in more courses [28], [29]. Web mining courses [30]–[32] cover topics on Web search, Web usage mining, text mining, information extraction, and link analysis. Such training educates students on how to extract information from the Web and discover knowledge from such information. To help students better understand the topics and assess their performance, group projects or group programming exercises have become more popular in these courses. It has been suggested that such group activities promote cooperative learning and a positive experience among students [33], [34]. Allowing students to work on projects that can be applied to real-world problems also provides them with valuable hands-on experience [35].

Some Web mining-related courses have used Web search engines as a topic for group projects [36]. These kinds of projects provide students with an empirical understanding of spidering, parsing, indexing, and link analyzing techniques. However, creating a search engine requires students to attain sufficient mastery of programming to support their data collection. Applying Web service/Web API techniques in Web mining-related group projects enables students to view the Web as a set of applications [37]. The students thus can focus on the use of Web mining algorithms to build value-added applications. This approach also allows students to practice integrating different Web applications.

III. Practicing Web Mining Using Open Web APIs

A. Project Objectives

In the Department of Management Information Systems at the University of Arizona, a graduate-level introductory course on data structure, data mining, and Web mining has been offered for several years. The course also covers several topics related to Internet search engines, Web mining, visualization, classification, and clustering.

In previous years, the course required students to spider Web pages from the Internet and conduct Web mining projects. Given the advance of open Web APIs, a new group project was designed that required students to design innovative business applications by consolidating open Web APIs with Web mining techniques. It is believed that using Web APIs can simplify the data acquisition process and enables students to focus on the application of Web mining algorithms. This paper reports on the first experience with the group project in 2005.

B. Course Project: Web Mining Using Google, eBay, and Amazon APIs

The class project required each group of students to create a Web business with a complete Website and business functionalities for specific customers, using one or more of the three open

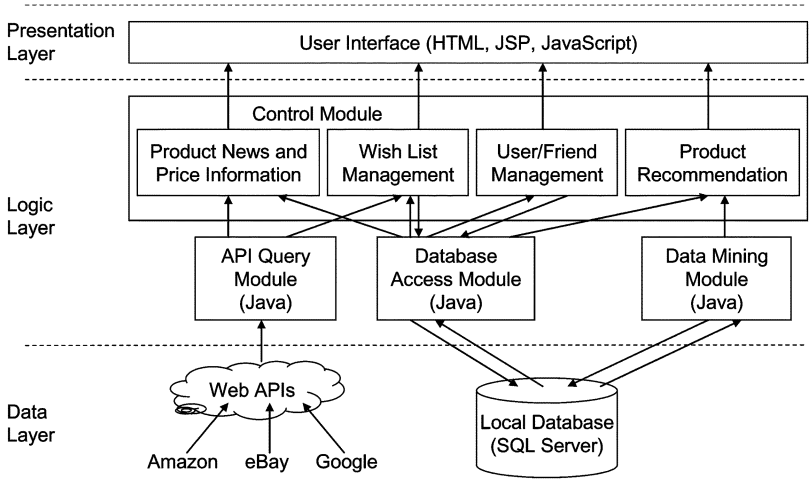

Fig. 1. WishSky system architecture.

Web APIs (Amazon, eBay, and Google) together with other data mining techniques learned in the class. Given the comprehensive data retrieval functions provided by the three Web APIs and other open source data mining software like WEKA [38], the challenge of the project was to integrate these components and design an attractive Web mining application.

Each team consisted of three members who would participate in the design, coding, implementation, and analysis of the prototype. Each team member was required to participate in all aspects of the project. Lab sessions were provided to familiarize students with Web API programming. A teaching assistant was also provided to assist the students with programming problems. After the students experimented with the APIs in the lab sessions, they designed their value-added business based on extensive team discussions. In addition to the three Web APIs, they could choose to include other data mining and Web mining algorithms and packages in their design.

The students were given three and a half months to finish the project and were required to submit a project proposal after the first month of class. At the end of the semester, each group would present their work, demonstrate the system, and submit a final report.

C. Technique Preparations for the Students

The students in the course had basic programming skills and knowledge of relational databases (a prerequisite to enroll in the class). Although the students were encouraged to use MySQL and Java in their projects, they were free to choose any database platform and programming language they were familiar with.

The class lectures provided the students with knowledge on a variety of data mining and Web mining topics. From the data mining technique perspective, the course introduced classification, clustering, knowledge representation, multiagent systems, and information visualization. From the algorithm perspective, the course introduced association rule analysis, genetic algorithms, neural network, support vector machines, self-organizing maps, and so on. The lectures also introduced resources for existing data mining packages. Students should be able to build their applications based on this class knowledge.

The lectures did not cover the technical details on open Web APIs. Thus, two simple scenarios using *Google Search API* and *Amazon E-Commerce Service* were introduced in the class lab sessions to teach students basic knowledge of XML and Web services. Step-by-step tutorials were given in the lab sessions to make sure that the students understood these two implementation methods and were able to use them. (The example source code and tutorial materials are available at http://ai.arizona.edu/hchen/chencourse/Website/WebAPI_project.htm.)

IV. Observations of the Student Projects

The instructor observed that most of the students were able to apply what they learned from the class in the projects. They also demonstrated their creativity and teamwork. The students developed several innovative business applications by incorporating the Web APIs with Web mining technologies. In this section, three sample projects with different Web mining emphases, namely WishSky, PriceSmart, and SciBubble, are described to analyze the students' achievements in the projects. (For further details of the projects, please refer to the student project reports and source code at http://ai.arizona.edu/hchen/chencourse/Website/WebAPI_project.htm.)

A. WishSky

WishSky is an integrated "wish list" management system. The system enhanced traditional wish list management by associating news and ongoing sales/auctions information with products in wish lists and introducing social relations (friendships) between wish list users/owners. WishSky can recommend alternative products and provide information related to products according to users' interests.

WishSky was implemented as traditional three-layer architecture on Tomcat (Fig. 1). The bottom layer of the system is the data layer, which includes a local Microsoft SQL Server database to store customer information, wish list data, and product recommendations. WishSky retrieved data through the Web APIs of Amazon, eBay, and Google. The middle layer of the system is the logic layer, which was written in Java. The API query module is used to retrieve wish lists, news, and price information from the three Websites. The data access module is used to manage the wish lists and customer information (of WishSky). This layer also contains a data mining module, which is mainly for product recommendation. The top layer of the system is the presentation layer. HTML, Java Server

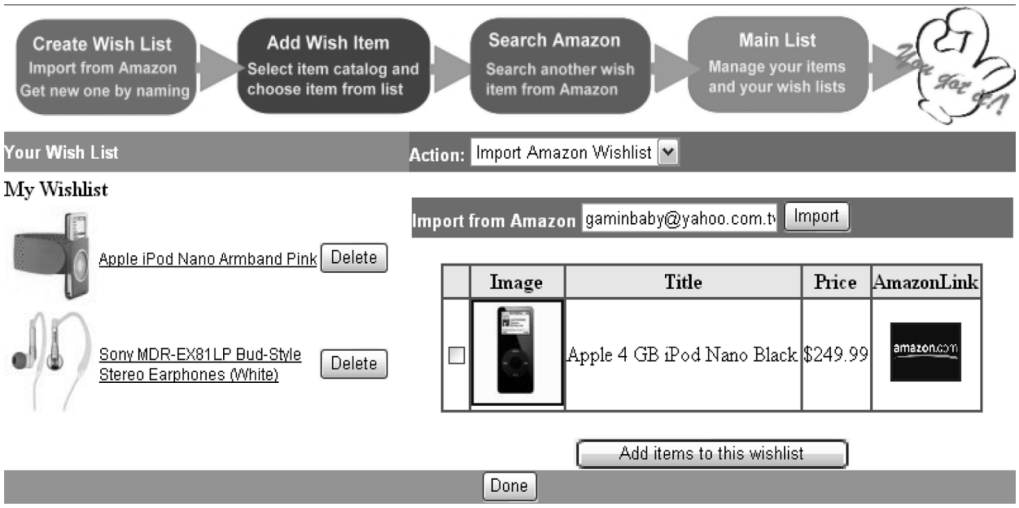

Fig. 2. The page for personalizing a wish list.

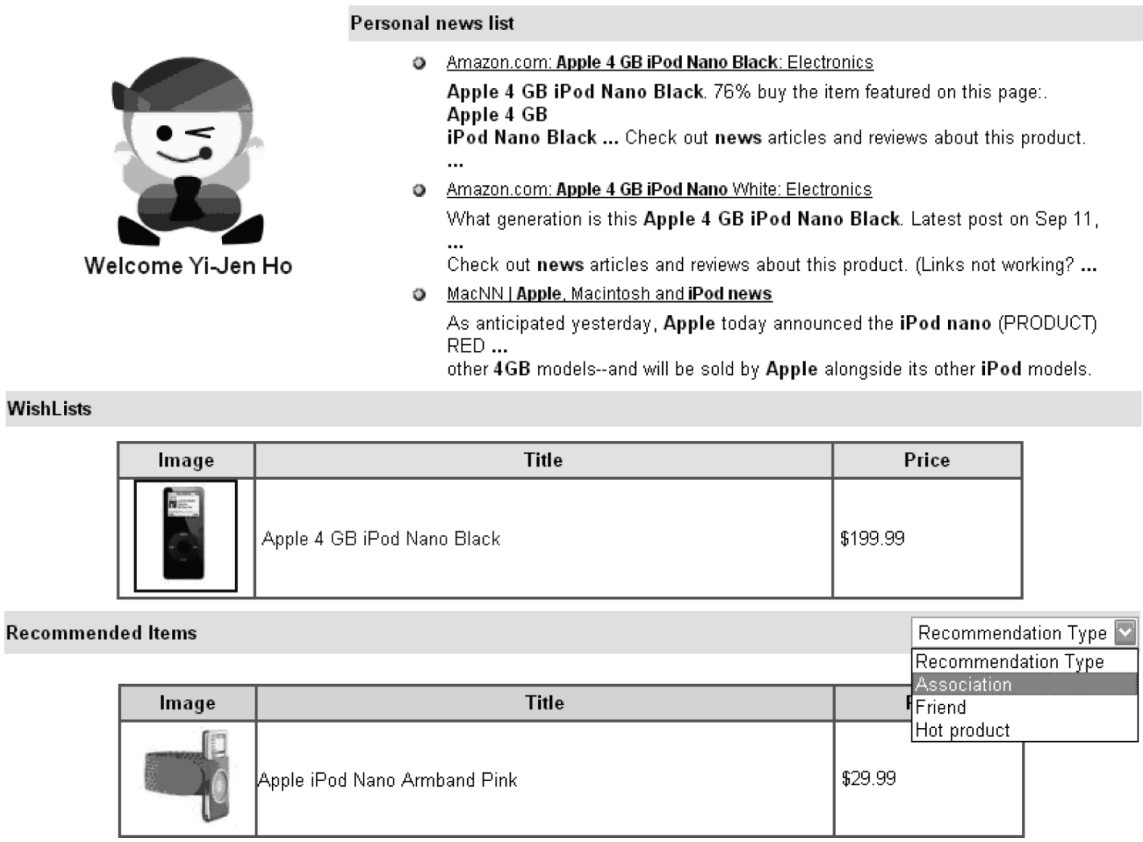

Fig. 3. WishSky displays a user's wish list, news related to the item, and recommendations.

Page (JSP), and JavaScript were used to generate friendly and intuitive user interfaces.

The wish lists on WishSky can be extracted from Amazon through the open Web APIs (Fig. 2) or by manual input. *Google Search API* was used to provide related news on users' products of interest (Fig. 3). *eBay API* provided users with the pricing information for their items of interest.

The students integrated a recommender system into wish list management (Fig. 3). The students implemented three algorithms for recommendation purposes. The first algorithm is association rule analysis. The frequent co-occur item pairs were identified using the *a priori* algorithm of the WEKA package [38] on all wish lists in WishSky. The patterns were applied to each user's wish list to provide item suggestions. The second algorithm was designed based on the social relationships embedded in WishSky. The algorithm suggests items to a user based on his/her friends' wish lists. The last algorithm, which simply recommends the most popular items to every user, is a simple algorithm that has shown good performance in some data sets [39].

Items Status Prediction

Title	Condition	Price	Status	Probability
Collins Care and Repair of Furniture	used-good	24.95	Cancelled	0.914359449074966
Fuzzy Thinking : The New Science of Fuzzy Logic	used-good	2	Closed	0.985168758945423
Homeopathy For The Soul	used-good	44.95	Closed	0.991523773017365
Body for Life : 12 Weeks to Mental and Physical Strength	used-good	15.95	Closed	0.995757609679333
Utilitarianism and Its Critics (Philosophical Topics)	used-good	19.95	Closed	0.981893212430001
Monetary Economics: Theory and Policy	used-acceptable	39.95	Closed	0.983299993773189
Study Guide for J. Gresham Machen's New Testament Greek for Beginners	used-acceptable	12	Closed	0.975593383087779
Angels Fear: Towards an Epistemology of the Sacred	used-acceptable	6.5	Closed	0.977657200722751
Flying Know-How	used-good	29.95	Closed	0.995180329629643
That Hideous Strength (Space Trilogy (Hardcover))	used-good	29.95	Closed	0.994799226651456

1 2 3 4 5 6 7 8 9 10 ...

Fig. 4. Item sales prediction.

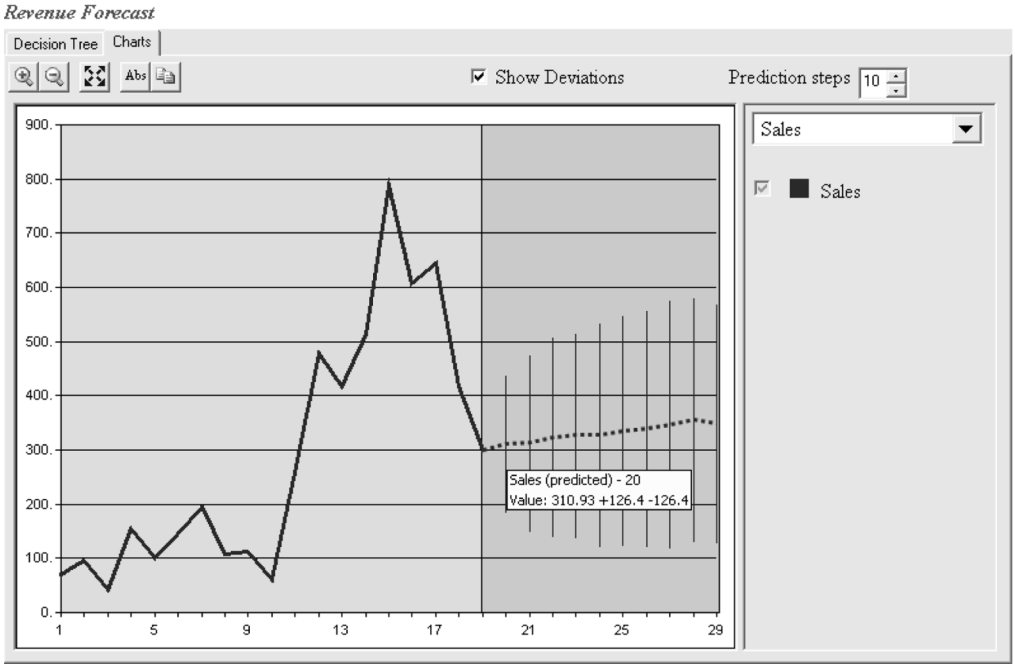

Fig. 5. Revenue forecast.

B. PriceSmart

PriceSmart is a market analysis system for Amazon sellers. The group used ASP.NET framework to implement their application. The group developed a standalone application in C# to extract sales history and product information from Amazon, and stored the data in a MS SQL database. The students then integrated several data mining algorithms in MS SQL 2005 into their Web application for market analysis.

Based on the transaction data, PriceSmart performs three types of analysis. The market analysis module clusters customers into groups according to their purchasing history for market segmentation purposes. The sales analysis module inspects the major factors in Amazon product descriptions that can differentiate sold items, open items, and canceled items using a decision tree model. The decision tree model was used to predict the probability of a product's sales (Fig. 4). In the third module, revenue analysis, the students implemented a time series analysis algorithm to predict the sellers' future profit based on their sales history (Fig. 5).

C. SciBubble

SciBubble is a science fiction book portal which features distinct visualization that helps customers find books of interest. The science fiction book data, including book details and customer reviews, were retrieved from Amazon through the *Amazon E-Commerce Service* API and were loaded into a MySQL database. The interface of the system was developed with PHP and Java Applet on Apache.

In addition to providing the function of searching books with keywords, the students' major contribution is to visualize book similarities for easier book selection. The similarity between books was defined according to the publisher, Amazon rating,

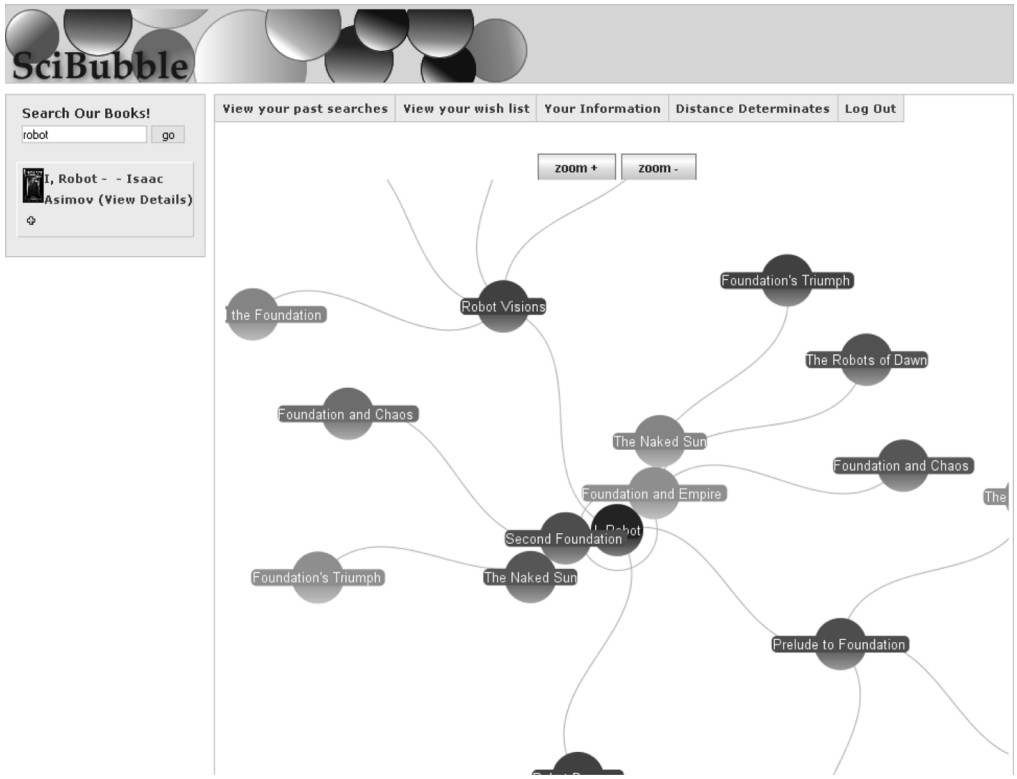

Fig. 6. The visualization interface of SciBubble.

publication year, ranking, and category. With this similarity measure, the students designed their own algorithm to visualize similar books, where each book is represented as a bubble and the similarities between books are represented by the distances and angles between the bubbles (Fig. 6). With this visualization, users can find similar books by looking at the locations and connection relations between the nodes on the graph.

D. Summary

During the implementation of the course projects, the instructor and teaching assistant observed the students' change of focus at different stages of the project. At first, most questions focused on the implementation of Web APIs. After the students became familiar with the APIs through lab sessions, their interest quickly changed to data mining algorithms. Most groups were able to implement multiple algorithms or data mining functions in their projects. Such a pattern in student implementation met the instructor's expectation of the project and showed the successful use of open Web APIs as instruments in the class.

In the projects, the students used not only the three APIs suggested by the instructor but also other APIs such as *Yahoo Geocoding API* and *Google Maps API* (Table I). Among all Web APIs, the *Amazon E-Commerce Service* was most frequently used. *Amazon E-Commerce Service* provides a lot of information on products, customer behaviors, and business transactions, which are more interesting to the students.

The students incorporated a variety of Web mining algorithms and statistical analysis in their systems (Table I). Some students integrated software packages, such as WEKA and MS SQL Server data mining component, while others implemented select data mining algorithms from scratch. Although few projects targeted developing complex visualization modules, dynamic charts and network visualization packages were widely used in data analysis and product/customer relation visualization. Table II summarizes some of the online resources suitable for the project.

In the course project, the students designed several interesting business models. One popular model is to provide product recommendations and personalize product information for customers. The second model is sales history analysis and market analysis. Because of the use of *Google Map API*, some student projects were able to use geographical information to aid business transactions. Overall, all groups successfully used Web APIs to build a Web application that involved Web mining techniques.

V. Evaluation of Student Learning Process

In addition to the inspection of the student projects from the instructor's perspective, the impact of the project on the students' learning process is also studied in this research. After the course, ten students were selected randomly and given a survey by e-mail.[1] Of the nine students who responded, four were Ph.D. students and five were Master's students.[2] The survey is a self-assessment of how effective the course project was in helping the student understand the course content. Fig. 7 highlights some of the survey results.

Although the sample size is limited, some patterns are shown in the survey results. The students' evaluation showed that the course project played a role as important as lecture, lecture

[1]Human subject approval: http://ai.arizona.edu/hchen/chencourse/Website/approval.pdf

[2]Survey results: http://ai.arizona.edu/hchen/chencourse/Website/Survey_result.doc

TABLE I
SUMMARY OF SELECTED STUDENT PROJECTS

Project name	*Business model*	*Web APIs*	*Web/data mining modules*
Cellphone Intelligent Auctioning	Cell phone market analysis	*eBay API, Yahoo Geocoding API, Google Maps API*	Market statistics; Spatial & temporal visualization
GiftChannel	Wish list management	*Amazon E-Commerce Service*	Gift recommendation; Shopping pattern visualization
MusicBox	Music album sales/news portal	*Google Search API, Amazon E-Commerce Service*	Band recommendation
PriceSmart	Market analysis/ Sales management	*Amazon E-Commerce Service*	Customer clustering analysis; Sales analysis and prediction; Revenue analysis and forecast
SciBubble	Science fiction book portal	*Amazon E-Commerce Service*	Book similarity visualization
Tucson Book Xchange	Local book flea market	*Amazon E-Commerce Service, eBay API, Google Maps API*	Book search engine
WishSky	Wish list management	*Amazon E-Commerce Service, eBay API, Google Search API*	Product recommendation; Product news and pricing information

TABLE II
ONLINE RESOURCES FOR USE IN THE PROJECTS

Type	*Name*	*URL*
Web APIs	Amazon Web Services	http://www.amazon.com/webservices/
	Google APIs	http://code.google.com/apis.html
	eBay APIs	http://developer.ebay.com/
	Yahoo APIs	http://developer.yahoo.com/
Data analysis packages	WEKA	http://www.cs.waikato.ac.nz/ml/weka/
	Yale	http://sourceforge.net/projects/yale/
	MS SQL Server 2005	http://www.microsoft.com/sql/
	IBM DB2 intelligent miner	http://www.ibm.com/software/data/iminer/
Network visualization toolkits	JUNG	http://jung.sourceforge.net/
	Graphviz	http://www.graphviz.org/

notes, and homework in the class [Fig. 7(a)]. The students felt that the project had a major impact on improving their skills and knowledge of data mining algorithms [Fig. 7(b)]. In general, the course project met its goal of helping students understand Web mining and data mining topics.

During the course project, the students felt that Web APIs were neither too difficult to use nor too easy for a graduate-level class [Fig. 7(c)]. Most students thought that the Web APIs played an important role in their projects [Fig. 7(d)]. Specifically, students felt that the Web APIs were more useful in the data collection/access part of the project [Fig. 7(e)]. They were also helpful in generating business ideas and applying the Web mining algorithms. The Web APIs showed relatively less impact on Web interface design and visualization. Such a result reflects the instructor's focus on the project as a means to help students understand Web mining algorithms rather than teaching them how to develop Web applications.

After the project, the students felt that they were more satisfied with their learning process in data collection, Web interface design, and application of data mining algorithms [Fig. 7(f)]. The instructor and teaching assistant may need to put more effort into guiding the business idea generation in the project. Teaching/training on visualization may also need to be improved.

The students provided several encouraging comments on their experience with the Web APIs and the course project, including the following.

- "The APIs are necessary to allow our application to request data from Amazon, Google, and eBay."
- "They [the Web APIs] were useful as they provided us with the information needed to be able to connect to each respective Web service provider and understand the supported features."
- "Project was very helpful to gain experience with Web Services technology and developing a Web application."
- "The project was really useful. I really liked the idea of joining three different Web services for one application."

VI. CONCLUSION AND FUTURE DIRECTIONS

This paper reports on an experience designing a class project for students in a graduate course to use open Web APIs to develop Web mining applications. These Web mining projects enabled most students building innovative Web applications

within a short period of time, which would have been impossible if the students had had to develop the systems from

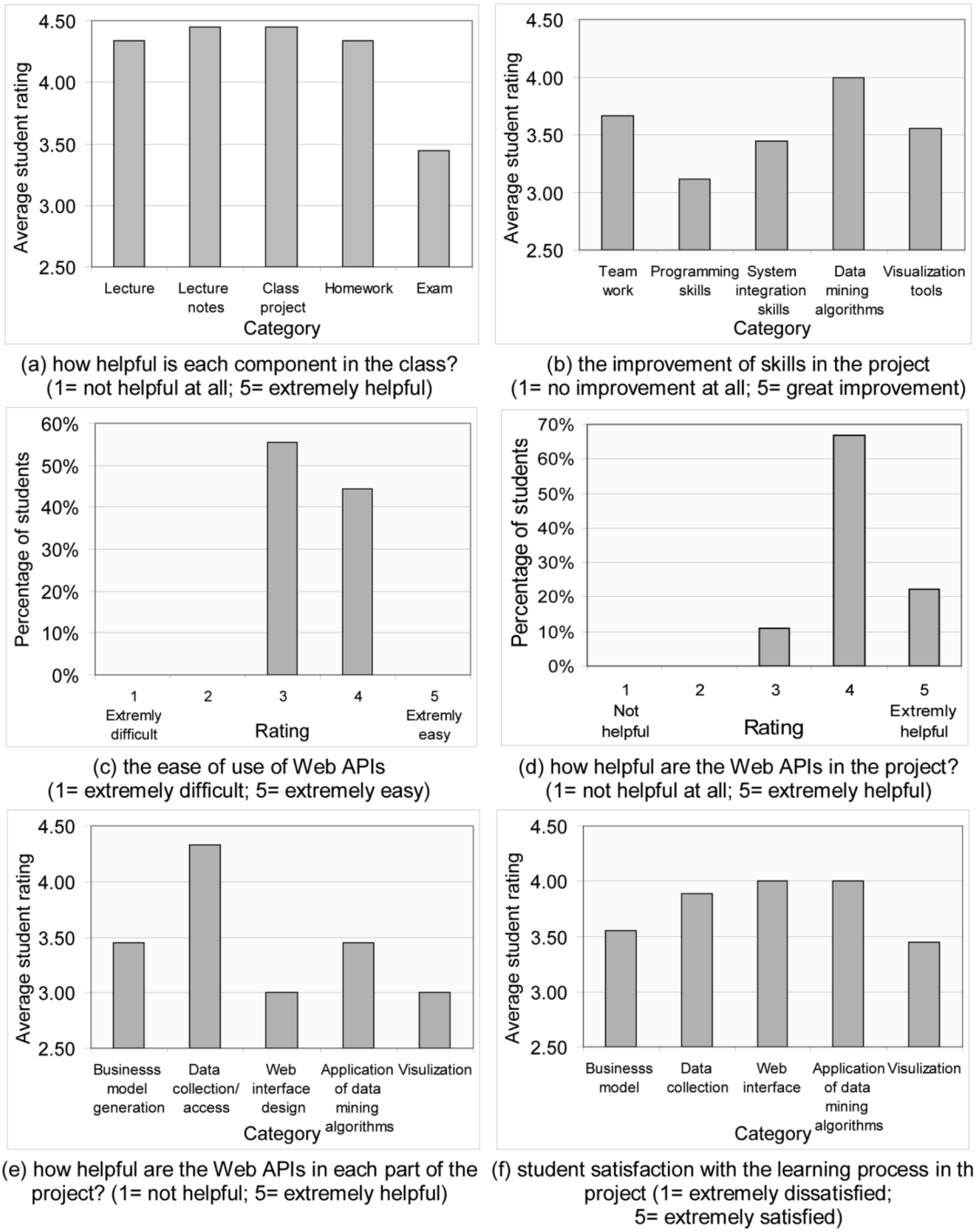

Fig. 7. Student evaluation results.

scratch. The projects allowed students to improve their understanding of Web mining algorithms and apply them in different business scenarios. The students gained great hands-on experience with Web APIs from Google, eBay, Amazon, etc. and with data mining/visualization tools.

Most students expressed the opinion that they had learned a lot during the project. They commented that the system implementation experience was much better than they thought they could achieve at the beginning of the semester. In general, using Web APIs significantly improved the student learning process in the Web mining class. In the future, more in-depth analysis will be conducted on the use of this new teaching instrument in information systems classes. It is also planned to introduce Web-based open source software in teaching Web mining topics and study its impact on students' learning.

ACKNOWLEDGMENT

The authors thank Google, Amazon, eBay, Yahoo, and WEKA for making their APIs, software, and data available to the public. The authors thank all the students who took the

course and participated in developing the Web mining projects, especially the teams of WishSky, PriceSmart, and SciBubble.

REFERENCES

[1] O. Etzioni, "The World-Wide Web: Quagmire or gold mine?," *Commun. ACM*, vol. 39, no. 11, pp. 65–68, 1996.

[2] H. Chen and M. Chau, "Web mining: Machine learning for Web applications," *Anu. Rev. Inf. Sci.*, vol. 38, pp. 289–329, 2004.

[3] M. Chau, B. Shiu, I. Chan, and H. Chen, "Redips: Backlink search and analysis on the Web for business intelligence analysis," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 58, no. 3, pp. 351–365, 2007.

[4] X. Fang, M. Chau, P. J. Hu, Z. Yang, and O. R. L. Sheng, "Web mining-based objective metrics for measuring Website navigability," in *Proc. Int. Conf. Inf. Syst.*, Milwaukee, WI, 2006.

[5] B. Liu, M. Hu, and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the Web," in *Proc. 2005 WWW Conf.*, Chiba, Japan, 2005.

[6] R. Kosala and H. Blockeel, "Web mining research: A survey," *ACM SIGKDD Explor.*, vol. 2, no. 1, pp. 1–15, 2000.

[7] J. Cho, H. Garcia-Molina, and L. Page, "Efficient crawling through URL ordering," in *Proc. 1998 WWW Conf.*, Brisbane, Australia, 1998.

[8] S. Chakrabarti, M. V. D. Berg, and B. Dom, "Focused crawling: A new approach to topic-specific Web resource discovery," in *Proc. 1999 WWW Conf.*, Toronto, Canada, 1999.

[9] M. Chau and H. Chen, "Comparison of three vertical search spiders," *IEEE Computer*, vol. 36, no. 5, pp. 56–62, 2003.

[10] O. Zamir and O. Etzioni, "Grouper: A dynamic clustering interface to Web search results," in *Proc. 1999 WWW Conf.*, Toronto, Canada, 1999.

[11] W. Y. Chung, G. Lai, A. Bonillas, W. Xi, and H. Chen, "Organizing domain-specific information on the Web: An experiment on the Spanish business Web directory," *Int. J. Hum.-Comput. St.*, vol. 66, no. 2, pp. 51–66, 2008.

[12] M. Hurst, "Layout and language: Challenges for table understanding on the Web," in *Proc. 1st Int. Workshop Web Document Analysis*, Seattle, WA, 2001.

[13] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," in *Proc. 1998 WWW Conf.*, Brisbane, Australia, 1998.

[14] J. Kleinberg, "Authoritative sources in a hyperlinked environment," in *Proc. 9th ACM-SIAM Symp. Discrete Algorithms*, San Francisco, CA, 1998.

[15] M. Chau and J. Xu, "Mining communities and their relationships in blogs: A study of online hate groups," *Int. J. Hum.-Comput. St.*, vol. 65, no. 1, pp. 57–70, 2007.

[16] H. M. Chen and M. D. Cooper, "Using clustering techniques to detect usage patterns in a Web-based information system," *J. Amer. Soc. Inf. Sci. Tech.*, vol. 52, no. 11, pp. 888–904, 2001.

[17] G. Marchionini, "Co-evolution of user and organizational interfaces: A longitudinal case study of WWW dissemination of national statistics," *J. Amer. Soc. Inf. Sci. Tech.*, vol. 53, no. 14, pp. 1192–1209, 2002.

[18] L. Baresi, F. Garzotto, and P. Paolini, "From Web sites to Web applications: New issues for conceptual modeling," in *Proc. 2000 ER Workshop Conceptual Model. Approaches for E-Business and The World Wide Web and Concept. Model.*, Salt Lake City, UT, 2000.

[19] ProgrammableWeb 2008 [Online]. Available: http://www.programmableWeb.com/apis

[20] J. Nandigam, V. N. Gudivada, and M. Kalavala, "Semantic Web services," *J. Comput. Sci. Colleges*, vol. 21, no. 1, pp. 50–63, 2005.

[21] J. Roy and A. Ramanujan, "Understanding Web services," *IT Profess.*, vol. 3, no. 6, pp. 69–73, 2001.

[22] Amazon Web Services Website Amazon, 2008 [Online]. Available: http://www.amazon.com/webservices

[23] F. Curbera, M. Duftler, R. Khalaf, W. Nagy, N. Mukhi, and S. Weerawarana, "Unraveling the Web services Web—An introduction to SOAP, WSDL, and UDDI," *IEEE Internet Comput.*, vol. 6, no. 2, pp. 86–93, 2002.

[24] J. P. Mueller, *Mining eBay Web Services: Building Applications With the eBay API*. Alameda, CA: SYBEX Inc., 2004.

[25] eBay Developers Program eBay, 2008 [Online]. Available: http://developer.eBay.com

[26] Google APIs Google, 2008 [Online]. Available: http://code.google.com/APIs.html

[27] D. C. Hoong and R. Buyya, "Guided Google: A meta search engine and its implementation using the Google distributed Web services," *Int. J. Comput. Appl.*, vol. 26, no. 3, pp. 181–187, 2004.

[28] S. Kalathur, Course Website of CS 699: Principles of Data Mining 2008 [Online]. Available: http://people.bu.edu/kalathur/courses/cs699_06_fall.htm

[29] F. Menczer, Course Website of CSCI B659: Topics in Artificial Intelligence 2008 [Online]. Available: http://informatics.indiana.edu/fil/Class/b659/

[30] B. D. Davison, Course Website of CSE 450: Web Mining Seminar 2008 [Online]. Available: http://www.cse.lehigh.edu/~brian/course/Webmining/

[31] H. Davulcu, Course Website of CSE 591: Semantic Web Mining 2008 [Online]. Available: http://www.public.asu.edu/~hdavulcu/CSE591_Semantic_Web_Mining.html

[32] Y. Xie, Course Website of CS 8990: Web Search and Mining 2008 [Online]. Available: http://science.kennesaw.edu/~yxie2/CS8990W/CS8990W.htm

[33] J. Dutt, "Cooperative lerning approach to teaching an intorductory programming course," in *Proc. 9th Int. Acad. Inf. Manage.*, Las Vegas, NV, 1994.

[34] J. McConnell, "Active learning and its use in computer science," in *Proc. SIGCSE/SIGCUE Conf. Integr. Technol. Comput. Sci. Educ.*, Barcelona, Spain, 1996.

[35] T. L. Fox, "A case analysis of real-world systems development experiences of CIS students," *J. Inf. Syst. Educ.*, vol. 13, no. 4, pp. 343–350, 2002.

[36] M. Chau, Z. Huang, and H. Chen, "Teaching key topics in computer science and information systems through a Web search engine project," *ACM J. Educ. Res. Comput.*, vol. 3, no. 3, pp. 1–14, 2003.

[37] B. Ramamurthy, "GridFoRCE: A comprehensive resource kit for teaching grid computing," *IEEE Trans. Educ.*, vol. 50, no. 1, pp. 10–16, 2007.

[38] H. I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA: Morgan Kaufmann, 2005.

[39] Z. Huang, X. Li, and H. Chen, "Link prediction approach to collaborative filtering," in *Proc. 5th ACM/IEEE-CS Joint Conf. Digit. Libraries*, 2005.

Hsinchun Chen (M'92–SM'04–F'05) received the Ph.D. degree in information systems from New York University, New York, in 1989.

He is currently the McClelland Professor of Management Information Systems at the University of Arizona, Tucson, where he is also the Director of the Artificial Intelligence Laboratory. He has been a Scientific Counselor/Advisor at the National Library of Medicine, USA; Academia Sinica, Taiwan, R.O.C.; and the National Library of China, China. He is the author or coauthor of more than 130 papers published in international journals and is the editor of several books. His current interests include biomedical informatics, homeland security, knowledge management, semantic retrieval, digital library, and Web computing.

Xin Li (S'07) received the B.Eng. degree in automation and the M.Eng. degree in control theory and control engineering from Tsinghua University, Beijing, China, in 2000 and 2003, respectively, and the Ph.D. degree in management information systems from the University of Arizona, Tuscon, in 2009.

He is currently an Assistant Professor in the Department of Information Systems at the City University of Hong Kong. His current research interests include data mining, social network analysis, business intelligence, Web mining, and text mining.

Michael Chau (S'99–M'03) received the Bachelor's degree in computer science and information systems from the University of Hong Kong, Hong Kong, in 1998, and the Ph.D. degree in management information systems from the University of Arizona, Tucson, in 2003.

He is currently an Assistant Professor with the School of Business, University of Hong Kong. His current research interests include information retrieval, Web mining, data mining, knowledge management, electronic commerce, security informatics, and intelligence agents.

Yi-Jen Ho received the B.B.A. and M.B.A. degrees from National Central University, Taiwan, R.O.C., and the M.S. degree in management information systems from the University of Arizona, Tucson.

Chunju Tseng received the B.S. degree from National Taiwan University, Taiwan, R.O.C., and the M.S. degree in management information systems from the University of Arizona, Tucson.