6-2008

An Integrated Evaluation Method for Module-Based Undergraduate Information Retrieval Education

Yong Shi
Kennesaw State University, yshi5@kennesaw.edu

Lei Zhu
Armstrong Atlantic State University

Jane Wong
Armstrong Atlantic State University

Bradley Sturz
Armstrong Atlantic State University

Chun Tang
Oracle Retail Global Business Unit, Atlanta, GA

Recommended Citation

Lei Zhu, Jane Wong, Bradley Sturz, Yong Shi, and Chun Tang. 2008. An integrated evaluation method for module-based undergraduate information retrieval education. J. Comput. Sci. Coll. 23, 6 (June 2008), 148-156.

AN INTEGRATED EVALUATION METHOD FOR MODULE-BASED UNDERGRADUATE INFORMATION RETRIEVAL EDUCATION*

Lei Zhu, Jane Wong, Bradley Sturz
Armstrong Atlantic State University
11935 Abercorn Street
Savannah, GA 31419
912-921-7361
{lei.zhu, jane.wong, bradley.sturz}@ armstrong.edu

Yong Shi
Kennesaw State University
1000 Chastain Road
Kennesaw, GA 30144
770-423-6423
yshi5@kennesaw.edu

Chun Tang
Oracle Retail Global Business Unit
3353 Peachtree Rd. NE Suite 300
Atlanta, GA 30326
404-439-5355
chun.tang@oracle.com

ABSTRACT

In the Information Age, career and life-long learning increasingly depend on information-retrieval (IR) skills to utilize digital and on-line information effectively. Consequently, IR has become a valuable integral element of many disciplines. To address primary problems in the current college-level IR education, we have proposed a module-based curricular model that facilitates the development of an array of IR course modules that enable flexible adoption and integration. In addition, these modules can be designed to meet the specific needs of various disciplines and programs. This paper presents a preliminary integrated evaluation method which combines subjective and objective evaluation in order to assess the effectiveness of the IR modules.

INTRODUCTION

With the advance of modern technology, the generation of information of various kinds has proceeded at an explosive rate in many disciplines. Knowledge acquisition in this ocean of information can be a tedious and daunting task without efficient and effective methods and tools. As increasing amounts of information become available, career and life-long learning are increasingly dependent on information-retrieval (IR) skills in order to utilize digital and on-line information effectively. However, IR skills are typically overlooked in undergraduate curricula, partly because IR is listed only as elective units of Intelligent Systems (IS) and Information Management (IM) [2].

To address primary problems in current college-level IR education, a module-based approach has been proposed to integrate IR skills into undergraduate curricula of multiple disciplines by developing and adopting appropriate IR course modules based on specific needs of different disciplines and programs. Each IR course module is designed relatively independently with learning objectives, descriptions, class notes, interactive practice quizzes, suggested readings and resources. Specifically, each module contains multi-versions of class notes, including a flash version, interactive flash version, windows media version, PowerPoint version, Word document version, and PDF file version. Different versions of class notes provide convenience and flexibility for module adoption because they allow faculty and students to preferentially choose class notes. The modules also include a number of video tutorials and flash demonstrations on how to use general search engines such as Google and Yahoo along with various digital libraries search engines such as JSTOR and Lexis-Nexis. Moreover, the tutorials and demonstrations can be used as hands-on IR training materials. Presently, two kinds of IR modules have been developed: general IR modules designed primarily for IT/CS programs, and discipline-oriented IR modules designed for non-IT/CS fields.

In order to evaluate the effectiveness of the developed modules, we have developed a preliminary integrated evaluation method which combines subjective and objective evaluations of student learning. The current paper presents this method along with evaluation results. The paper is organized as follows: Section 2 briefly introduces related work; Section 3 describes the integrated evaluation method and presents preliminary results; Section 4 provides concluding remarks.

RELATED WORK

Educators have proposed a variety of methods to improve the learning environment and evaluation procedure for IR education. IR-BASE [1], a basic object oriented framework for the integration of components, documentation and services, is proposed to focus on the rapid development of prototypes for IR research and teaching. In addition, educators have created some standard information retrieval modules in a distributed manner in order to create testing systems to validate IR methods and algorithms [3]. Based on these modules, students and teachers can construct retrieval test systems by integrating different modules and manipulating the input variables of each module. Based on these previous works, we have extended the domain of module-based approach from programming to curriculum development by developing and integrating an array of IR course modules.

AN INTEGRATED EVALUATION METHOD

The focus of this evaluation method is to measure IR-related teaching and learning outcomes through the developed modules. In courses which adopt IR modules, the evaluation process is based on data collected from Pre and Post module surveys, Pre and Post module quizzes, and term papers.

PRE AND POST MODULE SURVEYS

Pre and Post course surveys are standard course assessment methods adopted by many schools and departments. In these surveys, two questionnaires with the same set of (usually 10 to 20) questions are given to students at the beginning and end of the course. Questions are generally used to quantify students self-perceived learning of key course concepts or achievement of learning objectives. For example, a typical question for an information retrieval course may be know the major components and major processes of Web search engines Generally, students are then required to indicate their agreement with this question using one of the five answers: strongly disagree, somewhat disagree, neither agree nor disagree, somewhat agree, and strongly agree.

Oftentimes, survey questions can be subjective leading to substantial biases. For example, students may either over-estimate or under-estimate their understanding about key course concepts in the absence of clearly defined measurements. To overcome limitations resulting from subjective questions, we have included both subjective and objective components in our assessment questionnaire. Each subjective question is followed by a corresponding objective question, and we explicitly inform students that the two questions are counterparts. For example, the subjective question know the major components and major processes of Web search engines is followed by its corresponding objective question lease briefly describe the major components and major processes of Web search engines Because the paired subjective and objective questions are designed to target the same key course concepts or learning objectives, the objective questions may assist students in their self-assessment.

With the newly designed questionnaire, students first answered an initial subjective question and then the corresponding objective question. Ideally, students knew to what extent they understood key course concepts through answering the objective questions and were then able to reevaluate the answer they provided to the subjective question. In the event students answered the objective questions first, they might have become better informed about their knowledge resulting in more accurate answers to the subjective questions.

In the data analysis stage, the following procedures were conducted on the collected data:

1. For each subjective question of the pre-/post- questionnaire, a 5-point scale was used for the answers to the question: 1 = Strongly Disagree; 2 = Somewhat Disagree; 3 = Neither Agree nor Disagree; 4 = Somewhat Agree; 5 = Strongly Agree. For each objective question, a priori correct answers were developed, and student answers were then compared to these answers. Students answers were assigned a value from 1-5 based upon amount of content that matched a priori correct answers (e.g., a value of 5 was given for student answers that matched a

priori correct answers, a value of 1 was given for students answers that did not contain any information in the a priori correct answer). For each question, the mean value of the pre-questionnaire (Pre Mean) and the mean value of the post-questionnaire (Post Mean) were calculated and compared. As the Pre Mean and Post Mean measures students average self-perceived performance before and after introduction of an IR module, the ideal outcome would be one that showed students average self-perceived performance above a criterion level along with performance improvement. For the purposes of this paper we define the criterion level as 65% of the optimal value (5) and performance improvement as an increase of 10%.

2. The Gain of the mean values (Post Mean - Pre Mean) and the Percentage of Gain (Gain of the mean/Pre Mean) for each question were also calculated.
3. If for a given question, the Post Mean value was statistically less than or equal to 3.25 (65% of 5), then we interpreted that as indicating that students did not learn the underlying concepts via the module learning and that these concepts may need further emphasis in the future. Results are shown in Table 1. If the Post Mean value was statistically greater than 3.25 but the Percentage of Gain was less than 10%, the question was flagged for further analysis. Specifically, if the Pre Mean value was statistically greater than 2.95 (59% of 5) suggesting students already knew the underlying concepts prior to teaching of the module, then the question was un-flagged. Resutls are shown in Table 1. Otherwise, we took those values to indicate that students did not achieve significant improvement in understanding the underlying concepts of the question and that these concepts needed more emphasis in the future.
4. For all statistical tests, an alpha value of .05 was used. As a result, a p-value > .05 for a question suggested that the changes in the data collected in pre and post surveys were not statistically different. However, such an outcome does not necessarily indicate that students did not learn via the module learning because non-significant differences may result from various factors such as sampling error or sample size. However, one factor that may have contributed to no statistical difference between Pre- and Post-Test means may have been the level of student knowledge prior to implementation of the module teaching. An a priori high level of knowledge would result in no significant changes between pre and post data.
5. The results of the above analyses were used to refine the module materials, such as inclusion and exclusion of the concepts, change in the presentation contents, styles, or time allocation, and redesigning of practice problems, assignments and quizzes.

An apparent benefit of the newly designed questionnaire is that evaluators can identify discrepancies between the answers of subjective and objective questions as well as obviously biased responses. For example, if a student has failed to describe the major components and major processes of Web search engines (resulting in a score of 1 or 2 for Question 4) but indicated trongly agree with know the major components and major processes of Web search engines (resulting in a score of 5 for Question 3), one will be able to easily detect the inconsistency.

In fall 2007, we conducted revised evaluations in two sections of the course CSCI 1150: Fundamental of the Internet and World Wide Web. These sections integrated module C7: Web Search Engines and Directories and module C8: Online IR Systems and

Digital Libraries into the curriculum. Forty students completed the pre-questionnaires and 31 students completed post-questionnaires anonymously. The questions included in the questionnaires and the data used in the statistical analyses are listed in Table 1, where the subjective questions are marked with (S) and objective questions are marked with (O).

A two-way repeated measures analysis of variance (ANOVA) on student responses with Test (pre, post) and Question (1-10) as factors revealed a main effect of Test, $F(1, 30) = 281.34$, $p < .001$, a main effect of Question, $F(9, 270) = 144.98$, $p < .001$, and a significant Test x Question interaction, $F(9, 270) = 27.91$, $p < .001$. This suggests that although performance changed across Tests ($M = 2.62$, $M = 3.76$, for Test 1 and Test 2 respectively) this change was dependent on which question was isolated for analysis. Custom contrasts comparing each question across test were performed to isolate the source of the interaction, and results are shown in Table 1. As shown, performance increased for all questions from pre- to post-test with the exception of Question 1.

The Percentage of Gain values are listed in the last column of Table 1. Overall, the Percentage of Gain (57.09%) was greater than 10%, as confirmed by a one-sample *t*-test, $t(9) = 3.13$, $p < .01$. With the exception of Question 1, the individual values were each greater than 10% suggesting that students IR knowledge and skills after learning the Module C7 and C8 improved. Although the Percentage of Gain value was 67.86% for Question 4, the post mean value was only 2.35, which was significantly less than the criterion value of 3.25 (65% of 5) as confirmed by a one-sample *t*-test, $t(30) = -3.9$, $p < .001$. This suggests that students did not learn the underlying concepts of search engines. As a result, these concepts will need more emphasis in the future. Although the Post Mean values for Questions 3, 4, 8, & 10 were not statistically greater than our criterion value of 3.25 (all *p*s > .05), the overall Post Mean value of 3.76 was statistically above the criterion level of 3.25 as confirmed by a one-tailed *t*-test, $t(9) = 2.47$, $p < .05$.

The means for Question 1 were the only means not showing a statistical difference in pre and post surveys. However, the Pre Mean value of this question was 4.13 and was statistically larger than 3.25 as confirmed by a one-sample *t*-test, $t(39) = 5.58$, $p < .001$. As a result, we conclude that students already used (or self-perceived to use) search tools a great deal prior to the teaching of the module.

For each pair of subjective and objective questions, we also calculated the mean difference (mean of subjective question minus mean of objective question) between Pre Mean value and Post Mean value, respectively. For example, the score difference between Pre Mean values of Question 1 and Question 2 was calculated as follows: 4.13 3.48 = 0.65, and listed in Table 1 as [0.65] in bold underneath the Pre Mean value of Question 2. Other score difference values were calculated in an identical manner. Two interesting observations about these score difference values emerged: first, nearly all score difference values were positive suggesting students over-estimated their IR knowledge and skills; second, score difference values obtained in the post questionnaire were smaller than those values obtained in pre questionnaire and this result was confirmed by a one-tailed *t*-test, $t(4) = -3.35$, $p < .05$ suggesting that improved IR knowledge and skills obtained through module learning also assisted students in their self-evaluations.

PRE AND POST MODULE QUIZZES

For the purpose of objective assessment, we first gave students a multiple-choice quiz containing 10 technical questions about IR before the module was taught. The quiz questions were different from the objective questions used in the questionnaire. Without publishing the answer keys to the quiz questions, we gave students the same quiz again after the module was taught. These two quizzes attempted to quantify students IR knowledge, concepts and capabilities, via questions varying in complexity. For instance, some questions asked about indexing features of Google and concepts of exact phrase search. Because we were more interested in assessing students capabilities in utilizing search engines for information search, most of the quiz questions were related to real search practices such as forming queries and evaluating search results.

Each multiple-choice question was worth 1 point (10 total points). Consistent with the analyses of the Pre and Post questionnaires, we calculated the average quiz scores achieved by students before and after the module was taught and calculated the gain. The ideal outcome would be that after module learning students average quiz scores would meet the criterion level of 65% and show at least a 10% improvement.

In fall 2007, 37 students from our two selected sections of CSCI 1150: Fundamentals of the Internet and World Wide Web, took both Quiz 1 ($M = 6.27$) and Quiz 2 ($M = 8.0$). These values were statistically different as confirmed by a paired t-test, $t(36) = -6.09$, $p < .001$, objectively suggesting that students learned from the materials regardless of their perceptions of learning. In addition, the average score gain was 1.73 and the average Percentage of Gain was 22.05%. The average Percentage of Gain was calculated using the arithmetic mean of the individual Percentage Gain values. As a result of differing values for individual denominators, this value differs from a calculation based on Average Gain/Average Pre Mean (27.59%). The Post Mean score was larger than the criterion level of 6.5 as confirmed by a one-sample t-test, $t\,(36) = 6.64$, $p < .001$, and the Percentage Gain was significantly greater than 10% as confirmed by a one-tailed t-test, $t(36) = 2.25$, $p < .05$.

IR TERM PAPERS

At the conclusion of teaching the IR modules, we gave students a term paper assignment on developing Information Retrieval and Web Search teaching materials for their individual majors/disciplines. We asked students to act as though they were professors in their disciplines wanting to teach their students basic IR concepts and skills in a one-hour workshop. This assignment gave students an opportunity to proactively summarize their information retrieval knowledge and skills.

In fall 2007, we received 30 papers covering IR educational materials for 12 disciplines, including information technology, criminal justice, political science, English, Spanish, economics, health science, pre-pharmacy, history, pre-business, psychology, and liberal studies. Although preliminary, a number of these papers have presented valuable and insightful data suggesting that students not only learned module content but also integrated and applied new IR knowledge to their specific disciplines.

SUMMARY AND CONCLUSION

To promote information retrieval education in undergraduate institutions, we have proposed a module-based approach. For the assessment purposes, this paper presented a preliminary integrated evaluation method using Pre and Post modules surveys, Pre and Post modules quizzes, and term papers. Although future studies will incorporate appropriate controls, the preliminary evaluation results suggested that the IR modules were effective in improving students IR concepts, skills and capabilities.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 0534403. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Calado, P., Cardoso-Cachopo A. and Oliveira A. L. (2007), IR-BASE: An Integrated Framework for the Research and Teaching of Information Retrieval Technologies, Proceedings of First International Workshop on Teaching and Learning of IR, Jan 10, 2007, London, UK.

[2] Computer Curricula (2001). Final Report of the Joint ACM/IEEE-CS Task Force on Computing Curricula 2001 for Computer Science. http://acm.org/education/curric_vols/cc2001.pdf, retrieved Jan 1, 2008.

[3] Ferreira J., Rodrigues A. And Delgado J. (2005), Web Services for Information Retrieval. Proceedings of International Conference on Information Technology (ITCC 2005), Vol. 2, Pages: 497-502, Apr 4-6, 2005, Las Vegas, Nevada, USA.

Table 1: Summary statistics in the pre- and post- questionnaires

Question	**Pre Mean**	**Post Mean**	**PostMean Minus PreMean**	**Percent of Gain**
1. (S) I always use search engines, web directories and digital libraries for leisure, for study and for work.	4.13***	4.35***	0.22	**5.33**
2. (O) Please list at least one your favorite search engine, web directory and digital library, respectively; and the purposes that you use them.	3.48*** **[0.65]**	4.58*** **[-0.23]**	1.10***	**31.61**
3. (S) I understand the major components and major processes of Web search engines.	2.78	3.48	0.70**	**25.18**
4. (O) Please briefly describe the major components and major processes of Web search engines.	1.40 **[1.38]**	2.35 **[1.13]**	0.95***	**67.86**
5. (S) I understand search features offered by major search engines.	3.05	3.97***	0.92***	**30.16**
6. (O) Please list at least three search features (including advanced features) provided by search engines.	1.93 **[1.12]**	3.87* **[0.10]**	1.94***	**100.52**
7. (S) I know the major differences between search engines and web directories.	2.18	4.03***	1.85***	**84.86**
8. (O) Please briefly describe the major differences between search engines and web directories.	1.25 **[0.93]**	3.26 **[0.77]**	2.01***	**160.80**
9. (S) I am good at using search engines.	3.63***	4.23***	0.60**	**16.53**
10. (O) Please construct the queries for the following questions, suppose you are going to input these queries and conduct search from Google. a) How to get pdf files which contains phrase “learn Spanish”? b) How to find pages related to giant pandas that are not in national zoo? c) How to restrict a query (e.g., information technology) to a specific website such as www.standford.edu?	2.33 **[1.30]**	3.45 **[0.78]**	1.12***	**48.07**
AVERAGES	2.62 **[1.08]**	3.76* **[0.51]**	1.14*** **[0.57]***	57.09**

Note. Single asterisk (*) indicates a significant difference at $p < .05$. Double asterisk (**) indicates a significant difference at $p < .01$. Triple asterisk (***) indicates a significant difference at $p < .001$.

The Average Percent of Gain was calculated using the arithmetic mean of the individual Percent of Gain values in the table. As a result of differing values for individual denominators, this value differs from a calculation based on Average Gain/Average Pre Mean (43.63%).