Easy Retrieving Knowledge from Web Databases for Elementary School Students

CHIEN-I LEE
Department of Information and Learning Technology,
National University of Tainan,
33, Sec. 2, Su-Lin St., Tainan,
TAIWAN, R.O.C.

Abstract: - Through the Internet, students can conveniently get any information what they want. However, it is not quite easy for elementary school students to get what they just want through the Internet by using some keywords in order to do the current full-text searching. In this paper, we propose a new retrieval interface for elementary school students, called *leading-question retrieval interface*, which applies a checklist of questions to investigate and analyze the answers from the users to understand their queried intentions. From the experiments, we observed that it could promote the computer attitude, the retrieval precision and retrieval recall of the students by using the leading-question retrieval interface.

Key-Words: - E-learning, Elementary School Students, Internet, Information Retrieval, Query Language

1. Introduction

With the rapid data accumulation on the WWW, it becomes an urgent challenge for all Internet learning resource centers to support an efficient information retrieval interface to help users acquire easily and quickly their really desired data from the resource centers. Most of them adopt the conventional full-text retrieval with keyword-based techniques, which requires users to provide some keywords and combine them to form a Boolean expression for information retrieval. However, it is much difficult for most users, especially for elementary school students, to choose some appropriate keywords and construct them with Boolean expressions. Such a task imposes indeed a heavy burden upon elementary school students whose knowledge concept and semantic configuration is still infantile [4][15].

By browsing and searching the useful data on WWW, elementary school students can independently act on their own learning activities, but the students cannot master them, which has a negative impact on managing their own learning activities [13]. Consequently, it is very important to guide properly the students in the complicated linked structures of homepages to avoid going astray, and at the same time, also leave the free-learning activities of students. Therefore, in this paper, we propose a new efficient retrieval interface by using a leading-question strategy to guide the students to the correct searching direction. Moreover, homepages are always composed in the form of a tree-like structure via the hyperlinks. Users have to follow the linked structures through the hyperlinks to browse the homepages they need, and the retrieved homepages will be shown out in a predetermined linking manner. Users cannot browse randomly the related homepages that are newly created or modified without reorganization by the system managers.

In this paper, however, we will also propose a mechanism, which can dynamically support referential connections of relative information while the retrieved results are presented. Then, we will study the effect on students' learning achievements by supporting the referential connections. By using the proposed leading-question retrieval interface, students can easily construct their real retrieval processes and implement an Internet learning database system with referential connections.

1.1 Related work

Bruner [2][3] advocates that learners should explore knowledge on their own initiative, and discover the structure of every kind of knowledge. The Learning theory of Constructivism supports that knowledge consists of the initiative of those individuals with ability of recognition. Thus, we have to offer a real and interactive learning environment. While the learners are exploring new knowledge, they can also adjust their cognitive structure. Furthermore, the role of a teacher should be changed from a knowledge transmitter into a learning promoter [17].

The leading-question teaching is a learning activity in which teachers ask students one by one, and through those relative questions-asking teachers guide their students from the surface of the question into the deeper point, and let students figure out new knowledge and some universal principles from their own experiences. This method coheres with the basic spirit of the Learning theory included in Constructivism. Actually, the theory of Constructivism emphasizes on learning methods that guide students to find by themselves (Guided Discovery Approach) [10].

In the conventional full-text searching interface, users have to submit a set of alphabetic strings as keywords, which may be combined with Boolean operators, to form a Boolean expression [9]. However, it is so difficult for most users to construct such a Boolean expression without being well trained. Ideally, a good information retrieval interface should not only return what the users really want but also help users to easily develop their searching process [11][16][18].

Therefore, in this paper, we will apply the leading-question manner to the user interface for information retrieval in order to not only make the users kept with the basic spirit of Constructivism but also help the users to get easily their really wanted data.

Moreover, Bruner advocates that a learning activity is like a classification or an arrangement of learner's code systems. A code means a general idea or a rule. When learning the ideas of higher stage that contains many wide meanings, the learners try to figure out the new relationship between two or among more similar ideas. This categorizing form of ideas will be built under the fact that learners have interaction between their own knowledge as well as the environment or situation, then they can go on to the next step and develop deeper ideas [1][6][7][12].

Therefore, to apply this characteristic of learning to the design of our proposed retrieval interface, we offer dynamic referential connections between the relative information, which can provide the learners with much relative information besides the retrieval results to build their perfect code systems. That is, we can help the learners to promote their learning achievements with the referential connections.

2. Method for the study

2.1. Investigative design

We applied the nonequivalent-control group design of quasi-experimental research dividing randomly those who received the experiment into four groups to process the information retrieval activity according to two independent variables. To avoid the original computer attitudes and science learning achievements of those who received the experiment to interrupt the dependent variables, we took the results of the pretest of their original computer attitudes and science learning achievements as the covariates, applying the method of statistical control to exclude their influences.

2.1.1. Independent variable

(1) Retrieval interface: We implement two retrieval interfaces respectively, which are "the leading-question retrieval interface" and "full-text retrieval interface".

(2) Relative retrieval result: Two results gained, which present the referential connections of relatively retrieval results

but not present the referential connections.

2.1.2. Dependent variable

(1)Science learning achievement: The score of one who receive the posttest in the science learning achievement test.
(2)Computer attitude: The score of one who receive the posttest of computer attitude test.
(3)Retrieval precision: The retrieval precision means the proportion of correct results, replied by the retrieval system. (Note that the correct results mean the home pages actually meet the user's request.) The measure to the method: the retrieval precision for per search = (the amount of correct results in the search) / (the amount of total results in the search). For example, there are two users who perform two times of searching and get a piece of data they want in the second time of searching.

(a) For User A, there are three pieces of result existed in the retrieval results of searching for the first time; however, none of them are useful. Moreover, in searching of the second time, there are five pieces of result in the retrieval results and suppose only one of them matches the retrieval intention. Thus, the average retrieval precision for User A is (0/3+1/5)/2 = 1/10.
(b) For User B, there are five results in the retrieval results of the first time searching; however, none of them are the correct one. Moreover, there are three results in the retrieval results of the second time searching, and only one of the results matches the retrieval intention. Thus, the average retrieval precision is (0/5+1/3)/2 = 1/6.

Suggested from the two examples above, we can observe that the retrieval precision of User B is better than the one of User A even thought these two users have the same times for searching and the same amount of retrieved results. The reason is that the amount of retrieval results in the second time of searching for User B (3 pieces of result) is larger than the one for User A (5 pieces of result) while they have only one right result.

(4)Retrieval recall: The retrieval recall means the proportion of the amount of the correct results in the amount of entire correct results that assume the users retrieval intention replied by the retrieval system. The measure to the method: the retrieval recall for per search = the amount of correct result / the amount of the entire correct result. For example, there are three pieces of retrieved results in the searching of first time, and none of them are useful. Then, in the searching for the second time, there are five pieces of retrieval results and one of them matches the retrieval intention. In this case, the average retrieval recall is (0/1+1/1)/2 = 1/2.

2.2. Investigative objects

We take the students from sixth grade of the elementary school as samples for this study. To avoid students' culture background affecting the results of the study, we sample the students from four schools where geographical distance and learning environment are far away from each other, and have many differences. Totally we choose 160 students as our experimental sample.

2.3 Investigative implements

2.3.1 Database system

We take vertebrates as the investigating subject, because that the vertebrates, such as cats, dogs, birds, and fishes, are usually seen in the daily life of elementary school students. They can arouse students' interests in understanding what they are by searching information related to them in the Internet. The following is the general introduction of its major contents and functions:

1. Retrieval interface

(1) The retrieval system offers both the "leading-question" retrieval interface as shown in Figure 1 and the typical "full-text" retrieval interfaces.
(2) Users can use the leading-question retrieval interface to input their retrieval intentions while answering the questions on the referential interface; they do not have to key any keyword.

For each answer of the questions, there will be one or some attributes to be determined. The basic principle of the leading-question retrieval interface is to use a set of questions to apply questions to users in order to collect users' retrieval intentions. After knowing users' intentions, the system will check, automatically, and construct the retrieval condition with the collected attributes related to the answers.

(3) The retrieval system offers many sets of leading questions designed by some professionals; users can choose one set of the leading questions by themselves.

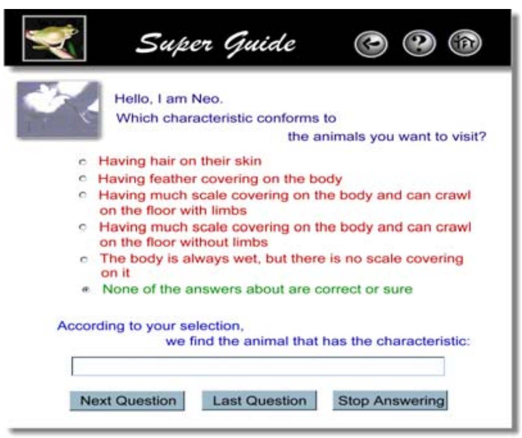

Fig. 1. Leading-question retrieval interface.

(4) There could exist a conflict during the question-answer process, where a conflict denotes that two values of answers made by the users are opposite. For example as shown in Table 1, there are two questionnaires (*i*, *j*) (*i* < *j*, that is, the order of the presentation of questionnaire *j* is after that of questionnaire *i*). Supposed that in the group of questionnaire *i*, users select the answer with value "having hair on skin" which is a characteristic only belonging to mammal. Then in the group of questionnaire *j*, they select the answer with value "having a pair of wing" which is a characteristic only belonging to birds. Obviously, there is a conflict between these two values of answers because there do not exist an animal, which is a mammal and is also a bird, and no retrieved results will be returned. The reason why a conflict occurs is that the users may have wrong judgment on the observation or wrong recognition about the vertebrates. To avoid this conflict occurring, our retrieval system can eliminate automatically such a situation by disabling the conflicting values in the following process of answering, to help users not to make the mistake above.

Table 1. The rejection between two answers in the pre and post questionnaires

Answers in questionnaire *i*	Answers in questionnaire *j*
◆ Having hair on skin	◆ Having a pair of membrane, like wings, at two sides of its body
◆ Having feather covering on the body	◆ Having a pair of wings
◆ Having much scale covering on the body and can crawl on the floor with limbs.	
◆ Having much scale covering on the body and can crawl on the floor without limbs.	
◆ The body is always wet, but there is no scale covering on it.	

(5)While users are selecting the characteristics of vertebrate, this retrieval system will also present the vertebrate category that conforms to the users' demand, and achieve the effect on "learning by doing" as shown in Figure 2. For instance, when users select the answer "Having feather covering on the body", then, the system will present the category—Aves-- that conforms to the answer at the same time.

(6)After users answer the whole set of questions, or stop answering, the system will start checking the answers, and present a list of retrieved results.

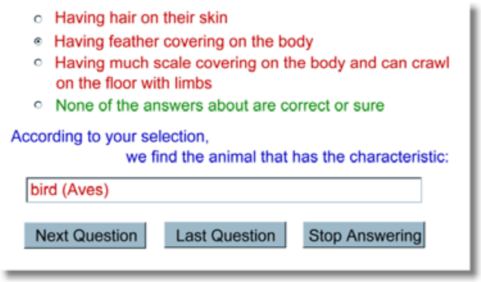

Fig. 2. Presenting the vertebrate category that conforms to users' demand.

2. Outcome presentation

While presenting the information about the vertebrate, the system will also offer the connection to its relative terms as shown in Figure 3. Via these referential connections, users can understand other vertebrates that have the similar attributes. We expect to achieve the effect on learning more from one event.

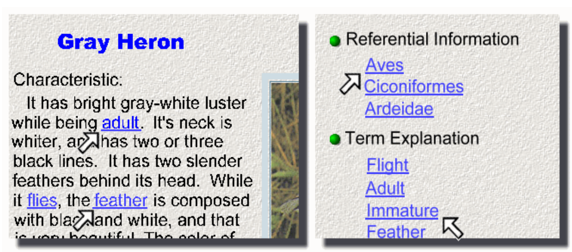

Fig. 3. Automatically connect to related retrieval results

If we can use the method we have discussed above to show the interpretation of every vertebrate in the web site, introducing vertebrates; also offer the explanation of special terms as well as the referential information with the method of hyperlink, we will support more extra knowledge for our users. However, for the system managers who maintain the web site, whenever adding a piece of information about vertebrates, they not only have to submit the data, but also make much efforts to embed the connections to the referential information among all the vertebrates in the database and the reason is that in a typical web site about vertebrates, the numbers of vertebrates included in the database will be more than thousands or ten thousands. Experiencing this case, if we want to add an explanation of the term "marsh" for every kind of birds, living in marsh, and the number of the relative birds may be 500, then the system managers have to update the 500 homepages of birds.

To reduce the burden of maintenance for the system managers, and to keep the current update information, this retrieval system will automatically embed the relative hyperlink nodes by the dynamic connections. It means that the hyperlink paths of related information do not be connected statically in the interpretations for each vertebrate. Therefore, the web site managers only have to pay attention to the edition of update data with no worry about the change of hyperlink paths. Take the term, "marsh" as the example, after submitting the data then the system will hyperlink only this new added term to the retrieved results when they are presented to the users rather than that the system managers have to embed the hyperlinks into the homepages of all 500 birds. Supposed that there are 5 birds retrieved according to the users' answers, the hyperlink corresponding to the term "marsh" will be on-line embedded into the five homepages only when they are viewed by the users. In other words, there is no hyperlink for the added term "marsh" in the original hypertext for each bird, and no maintenance cost is needed. This technique we support is combined by the techniques of ASP, HTML, and VBScript. Before the content of the homepage sent, the system will check the content of relative information and embed the link node into the document, and then send out.

3. System maintenance

The number of vertebrates in the world is very large. Putting the relative information in order and putting them on line is really a great job. However, in order to richen the content of database, we have implemented the editing interface for vertebrate data and the leading-question sets respectively in order to let users have chances to participate in the edition for vertebrates through the browser. Certainly, to guarantee the accuracy of input data, the professionals in vertebrates filed will assist in identifying every piece of new-added information before being retrieved by other users.

2.3.2. Science learning achievement test

(1) Selecting retrieval targets

For this study, there are 206 kinds of common vertebrates in the system. There are 8 kinds of vertebrates, which are chosen randomly from 206 kinds as the retrieval targets for the retrieval activity and science learning achievement test. They are Koala, Asiatic Elephant, Duck-billed Platypus, Formosan Blue Magpie, Red-eared Guenon, Black-faced Spoonbill, Formosan Whistling Thrush, and Crab-eating Mongoose. Of which some are familiar to the elementary school students, such as Koala and Asiatic Elephant. Some are easy to be identified from their appearance, such as, Duck-billed Platypus, Formosan Blue Magpie, Red-eared Guenon and Black-faced Spoonbill. And the others are hard for students to guess their names or partial names, inclusive of Formosan Whistling Thrush and Crab-eating Mongoose.

(2) Designing the subject

According to the content of learning database, we take the characteristics – the appearance of vertebrates, the position of sorting, and the custom of finding food – as the topics to design the questions for science learning achievement test, including 45 items of questions in each the pretest and posttest.

(3) Pretest

We focused on 118 six-grade elementary school students to process the pretest of science learning achievement, and then teach them a course about the common sense of vertebrates. This course lasted 60 minutes and after the test finishing, we processed the posttest of science learning achievement on the same group, and then, went on the next step of item analysis.

(4) Choosing items for exam paper

According to Ebel's criteria for indices of discrimination [8] and Chase's suggestion [5] for difficulty of an item, we chose the items, which discrimination index is above .35, and the difficulty index is between .40 to .80 as the formal items for the achievement pretest and posttest. There are 30 items in the pretest, and 30 items in the posttest. The average item difficulty index of the pretest is .61, and the posttest is .66.

(5) Validity analysis

After the science teachers who teach high grade identified the exam paper of learning achievement of pretest and posttest, the content validity will be affirmed from the two-way specification table.

(6) Reliability analysis

After choosing the items from the exam paper of the science learning achievement of pretest and posttest, we selected other 122 six-grade elementary school students to conduct the test to measure the reliability of this exam paper. After computation and analysis, the interior accordance Cronbach's α coefficient of the pretest is .83, and the interior accordance Cronbach's α coefficient of the pretest is .80.

2.3.3. The measure for computer attitude

We apply the measure for computer attitude, designed by Lin [14]. This measure can be divided into three minor measures: the confidence in computer, the computer application in education and the utilization of computer. There are 24 questions contained in the whole measure, which includes 14 positive questions and 10 negative ones. The interior accordance Cronbach's α coefficient is .81.

2.4 The procedure of carrying out the experiment

The procedure is:

(1) Dividing those students who receive the experiment into 4 groups by random sampling.

(2) Letting those students receive the pretest of science learning achievement and that of the measure for computer attitude.

(3) Training those students how to use the Internet for about 10 minutes, and then starting the information retrieval activity on line for about 40 minutes. During the retrieval process, the system will record the retrieval results and calculate the retrieval precision and recall.

(4) Letting those students receive the posttest

of science learning achievement and that of the measure for computer attitude.

3. Results of the study

3.1. The statistic of valid sample

During the process of retrieval activity, we find that some students are not familiar with computer. To avoid affecting the statistic results by those students, we will regard them as invalid samples. After excluding those invalid samples, we have 145 valid samples left. During the pretest and posttest of the science learning achievement, all students are serious about the test, so all of them are valid samples of assumptive test of the science leaning achievement, and they can go on the next step to receive the assumptive test of computer attitude presentation. Nevertheless, if those students have the tendency of choosing the same scale, or not answering the question completely during the test, they will be also regarded as invalid samples. And excluding those invalid ones, we still have 112 valid samples of the assumptive test of computer attitude presentation.

3.2. The assumptive test of computer attitude presentation

We take the pretest results of the measure for computer attitude as the covariate. And, we take groups as independent variable, where Groups A, B use the leading-question retrieval interface and Groups C, D use the full-text retrieval interface, taking the posttest results of the measure of computer attitude as dependent variable. And, we are going to process one way analysis of covariance of independent samples to test investigative (i.e., the computer attitude of the elementary school students who use leading-question retrieval interface to retrieval information will be better than the one of those who use full-text retrieval interface.). Before doing the analysis of covariance, ANCOVA, we have to process the test of homogeneity of within-class regression coefficient (F=1.853, p>.05), not meeting obvious class, conforming to the basic assumption of ANCOVA, then we can go on the step of ANCOVA.

From Table 2 and Table 3, we know that the groups that use the leading-question retrieval interface (i.e., Groups A and B) have better presentation at the posttest of the measure for computer attitude than the groups that use the full-text retrieval interface (i.e., Groups C and D).

Table 2. Summery of ANCOVA of the posttest of the measure of computer attitude (*p<.05)

Resource of Change	*SS*	*df*	*MS*	*F*
Retrieval Interface	1462.444	1	1462.444	39.84*
Error	4001.2	109	36.708	

Table 3. The average number after adjusting the results of the posttest of the measure of computer attitude

	leading-question retrieval interface	full-text retrieval interface
Average number after adjustment	98.460	91.222

4. Conclusion and Future Work

Indicated from this study, we find that the computer attitude of elementary school students who use the leading-question retrieval interface is obviously better than that of students who use the full-text retrieval interface. The reason could be that most experimental samples, using the full-text retrieval interface, feel puzzled while keying in the retrieval keywords, especially while facing those retrieval targets whose names are unknown or hard to guess. They could also feel frustrated when the system returns the useless results by submitting the wrong keywords again and again. This situation will also affect the positive computer attitude indirectly. However, there are still some experiments needed to be performed and the related issues needed to be studied the effects on the learning achievement in our future work.

Acknowledgement

This research was supported in part by the National Science Council of Republic of China under Grant No. NSC 95-2520-S-024-002.

References:

[1] J. D. Brinkerhoff, J. D. Klein and C. M. Koroghlanian, Effects of overviews and computer experience on learning from hypertext, *Journal of Educational Computing Research*, Vol.25, No.4, 2001, pp. 427-440.

[2] J. S. Bruner, The act of discovery, *Harvard Educational Review,* Vol. 31, No.1, 1961, pp. 21- 32.

[3] J. S. Bruner, Education as social invention, *Journal of Social Issues*, Vol.39 No.4, 1983, pp.129-141.

[4] K. E. Chang, Y. T. Sung and C. L. Lee, Web-based collaborative inquiry learning, *Journal of Computer Assisted Learning*, Vol. 19, No.1, 2003, pp. 56-69.

[5] C. I. Chase, *Measurement for Educational Evaluation*, Reading, Massachusetts: Addison-Wesley, 1978.

[6] E. Cloete, Electronic education system model, *Computers & Education*, Vol.36, No.2, 2001, pp. 171-182

[7] R Driver and V. Oldham, A constructivist approach to curriculum development in science, *Studies in Science Education,* Vol.13, 1986, pp.105-122.

[8] R. L. Ebel, *Essentials of Educational Measurement*. (3rd ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1979.

[9] P. A. Howard-Jones and R. J. Martin, The effect of questioning on concept learning within a hypertext system, *Journal of Computer Assisted Learning*, Vol.18, No.1, 2002, pp. 10-20.

[10] S. Hsi, A study of user experiences mediated by nomadic web content in a museum, *Journal of Computer Assisted Learning*, Vo.19, No.3, 2003, pp. 308-319.

[11] E. M. John, Does hypertext disadvantage less able readers? *Journal of Educational Computing Research*, Vo.29, No.1, 2003, pp. 1-12.

[12] A. W. Lazonder, Exploring novice users' training needs in searching information on the WWW, *Journal of Computer Assisted Learning*, Vol.16, No.4, 2000, pp 326-335.

[13] S. S. Liaw, Understanding user perceptions of World-wide web environments, *Journal of Computer Assisted Learning*, Vo.18, No.2, 2002, pp. 139-150.

[14] W. Lin, *The Effects of CAI on Elementary Students Mathematics Achievement of Different Cognitive Styles/Locus of Control*. Unpublished master's thesis, National Tainan Teachers College, Tainan, Taiwan, 1994.

[15] D. S. Niederhauser, R. E.Reynolds, D. J. Salmen and P. Skolmoski, The Influence of Cognitive Load on Learning from Hypertext, *Journal of Educational Computing Research*, Vol. 23, No.3, 2000, pp. 237-255.

[16] C. T. Panagiotakopoulos and G. S. Ioannidis, Assessing children's understanding of basic time concepts through multimedia software, *Computers & Education*, Vol.38, No.4, 2002, pp. 331-349.

[17] J. J. Pear and D. E. Crone-Todd, A social constructivist approach to computer-mediated instruction, *Computers & Education,* Vol.38, No.1-3, 2002, pp. 221-231.

[18] N. Stanton, A. P. Correia and P. Dias, Efficacy of a map on search, orientation and access behavior in a hypermedia system, *Computers & Education*, Vol. 35, No.4, 2000, pp.263-279.