DOCUMENT RESUME

ED 476 978 IR 021 706

AUTHOR	Brusilovsky, Peter
TITLE	Web-Based Interactive Visualization in an Information Retrieval Course.
PUB DATE	2002-06-00
NOTE	7p.; In: ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications. Proceedings (14th, Denver, Colorado, June 24-29, 2002); see IR 021 687.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	Computer Assisted Instruction; Computer Science Education; Computer Uses in Education; Higher Education; *Information Retrieval; Online Systems; *Search Strategies; *Visualization; World Wide Web
IDENTIFIERS	*Interactive Communication; *Interactive Courseware; Interactive Systems; University of Pittsburgh PA

ABSTRACT

Interactive visualization is a powerful educational tool. It has been used to enhance the teaching of various subjects from computer science to chemistry to engineering. In computer science education, this powerful tool is used almost exclusively in programming and data structure courses. This paper suggests that visualization could be very helpful in teaching a larger variety of computer science courses and also presents several visualization tools developed at the University of Pittsburgh (Pennsylvania) that have been used in the context of an information retrieval course. Highlights include visualization for information retrieval; interactive visualization of the Boolean Information Retrieval model; interactive visualization environments for other models; and implementation issues. Includes four figures in color. (Contains 14 references.) (Author/AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Web-based Interactive Visualization in an Information Retrieval Course

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Peter Brusilovsky
School of Information Sciences
University of Pittsburgh
Pittsburgh PA 15260
peterb@mail.sis.pitt.edu

ED 476 978

Abstract: Interactive visualization is a powerful educational tool. It has been used to enhance the teaching of various subjects from computer science to chemistry to engineering. Surprisingly enough, in computer science education, this powerful tool is used almost exclusively in programming and data structure courses. This paper suggests that visualization could be very helpful in teaching a larger variety of computer science courses and also presents several visualization tools that have been used in the context of an information retrieval course.

Introduction

Interactive visualization is a powerful educational tool. Visualization can provide a clear visual metaphor for understanding complicated concepts and uncovering the dynamics of important processes that are usually hidden from the student's eye. Visualization has been used to enhance the teaching of various subjects ranging from chemistry (Yaron et al., 2001) to mechanics (Hampel, Keil-Slawik & Ferber, 1999) to physics (McKenna & Agogino, 1997). Computer science is one of the most active application areas for educational visualization research. In computer and information science (CIS) education, visualization is used almost exclusively in programming and data structure courses. We can name dozens of papers devoted to visualization of program execution on several levels from machine-level languages (Butler & Brockman, 2001) to high-level languages (Domingue & Mulholland, 1998; Haajanen et al., 1997; Tung, 1998) to algorithms and data structures (Brown & Najork, 1997). Our claim is that a number of other traditional CIS courses could benefit from this powerful technology.

This paper explores the opportunities for using interactive visualization in the context of an information retrieval course. Information retrieval has been in the program of many computer, information, and library science departments for more than 30 years. With the maturity of the World Wide Web, information retrieval became an important practical subject. Elements of information retrieval are now taught to students of many different specialties. We think, that information retrieval provides an interesting and important application area for exploring the power of interactive visualization. In the following sections we discuss the use of visualization in teaching information retrieval and present several visualization tools developed at the University of Pittsburgh.

Visualization for Information Retrieval

The core of a traditional information retrieval (IR) course is a set of models, algorithms and technologies for processing, storing and retrieving textual information. This core has been already explored by now and placed on a solid mathematical foundation. Traditional presentation of this core usually starts with several IR models (such as Boolean, vector, probabilistic and several variations of them) and then follows by explaining how the information is organized and retrieved in each of these models (Baeza-Yates & Ribeiro-Neto, 1999; Korfhage, 1997).

The process of retrieving the information in different models is one of the hardest topics in an IR course. Despite being formalized and well understood by the IR research community, it is still very hard for students to grasp. We have observed that even the Boolean information retrieval, the simplest of the models, is difficult for students. At the same time, traditional educational tools – research or commercial IR systems - offer little educational help. The process of retrieving the information has several steps, from getting the query in to matching the query to the documents to prioritizing the results. In an IR system all these steps are hidden

IR021706

BEST COPY AVAILABLE

from a user – the only thing that a user can observe are the final results – a list of ordered documents. In that sense it is similar to a non-visualized execution of a computer program. A user can see input data and observe final results, but it offers no help in understanding how these results were computed.

Naturally, similar contexts encourage the use of similar remedies. So, the first thing we have decided to visualize is the process of retrieving the information in several known models. For the moment, we have developed and explored interactive visualization environments for several models – Boolean, fuzzy, vector, and extended Boolean (see (Baeza-Yates & Ribeiro-Neto, 1999; Korfhage, 1997) for the description of these classic models). Since space is limited and our environments are reasonably similar we present here in more details the one for exploring Boolean IR.

Interactive Visualization of Boolean Information Retrieval Model

The Boolean IR model is the oldest and the simplest of known IR models. In this model, a query is written as a set of elementary queries (usually *keywords*) connected by Boolean operators such as OR, AND, NOT. The mechanism of this model is set theoretical. Every query is associated with a set of *matching documents*. For an elementary query such as a keyword the set of matching documents is simply all documents indexed by this keyword. To obtain the set of matching documents for two queries connected by a Boolean operator one has simply to perform the corresponding set operation on their matching sets (i.e., set intersection for AND, complement for NOT, etc.). Thus in several steps, a matching set for any complex Boolean query can be found.

While it all sounds quite simple and clear, we have found that many of our students have problems understanding how the Boolean matching works. Our talks with students have indicated that one of the sources of their troubles is the failure to perceive Boolean operators as operations on sets of matching documents. Naturally, our students have good programming background and have been routinely using Boolean operators for writing conditional expression in their programs. Still, many of them have problems transferring their knowledge of these operators to the set context.

In developing an interactive visualization environment for the Boolean IR model we were trying to achieve two goals: to provide a helpful visual metaphor and to visualize the process of Boolean IR step by step. Figure 1 presents an interface of our environment. The core of this interface is a set of all documents visualized in a table (one document per row). Note that in our system, documents are textbooks since it is one of the most traditional kinds of documents.

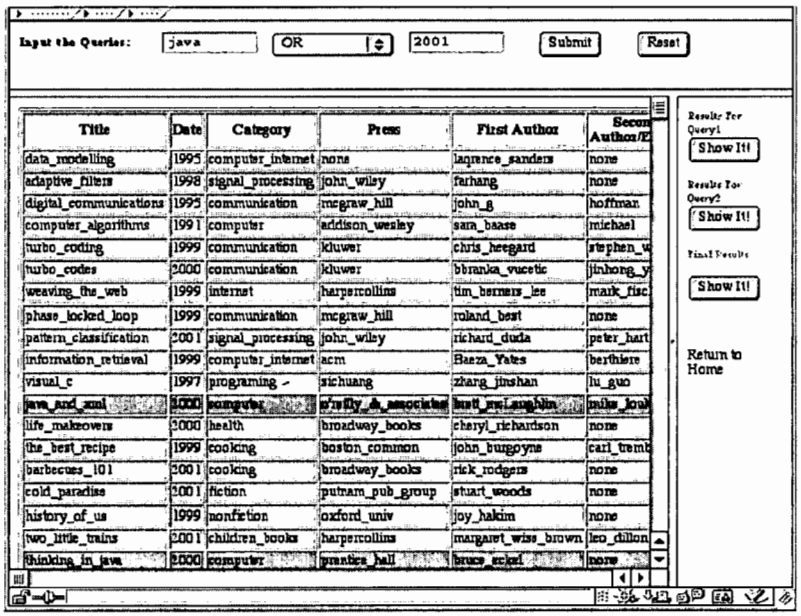


Figure 1: Boolean Model Environment. Visualization of matching for a simple Boolean OR query. Documents matching the first elementary query are highlighted.

BEST COPY AVAILABLE

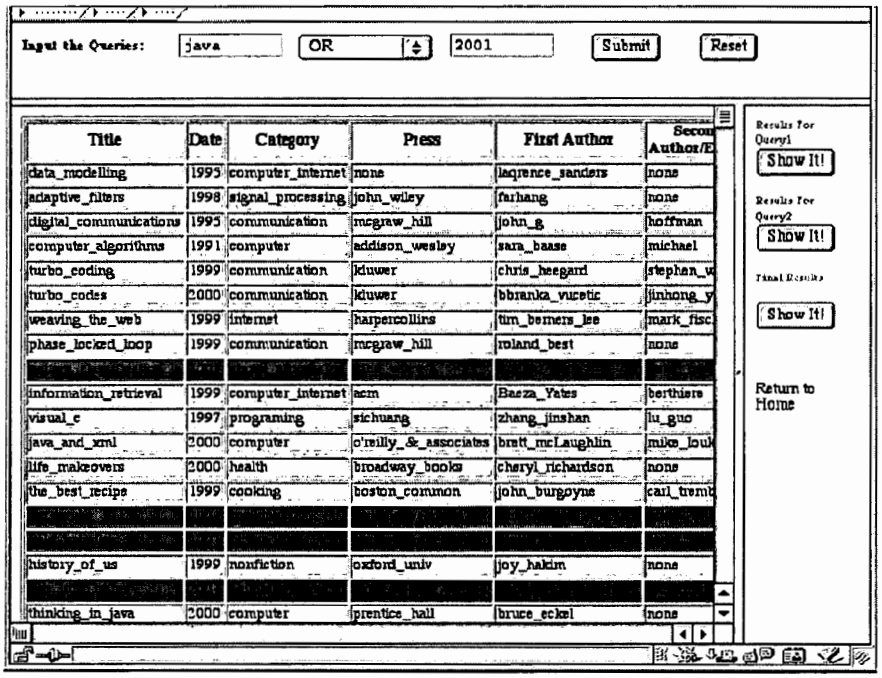


Figure 2: Visualization of matching for a simple Boolean OR query. Documents matching the second elementary query are highlighted.

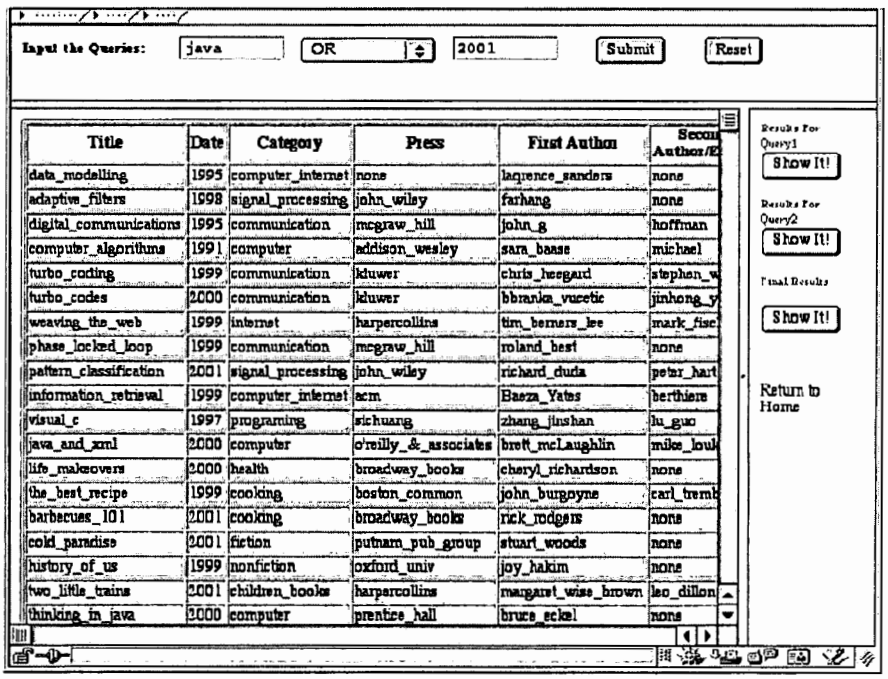


Figure 3: Visualization of matching for a simple Boolean OR query. Documents matching the whole Boolean OR query are highlighted.

BEST COPY AVAILABLE

The goal of this representation is to help the student to understand the core principle of this model – every query is associated with a particular *subset* of all documents. Showing the set of all documents on the screen makes it easy to demonstrate different subsets of the whole set as sets of differently colored table rows. The goal of the whole environment is to help the student to understand (a) the process of matching an elementary query to the set of the documents and (b) how different set theoretic operations work in obtaining a new subset from contributing sets.

The student can explore Boolean matching by writing simple Boolean queries (two terms connected by one or two operators) and observing the matching process in these steps by clicking on each of the three buttons on the right panel.

The first button highlights the subset of documents matching the first elementary query (Figure 1), the second highlights the subset matching the second query (Figure 2) and the third, the results of the chosen set operation on the contributing sets (Figure 3). We choose to have three buttons to enable the student to explore the matching process several times forwards and backwards.

Beyond the component shown on Figures 1-3 the Boolean IR environment has several other components. In particular, to help the student transfer the understanding of Boolean IR from classic IR to the database context, we have provided a very similar exploration interface where elementary queries are constructed not from keywords as in classic IR but from restrictions on various fields of a database record (i.e., year = 2000 and publisher != kluwer). There is also a registration screen and an interface for a teacher to edit the collection of documents. The environment works on the Web and is implemented as a set of CGI scripts.

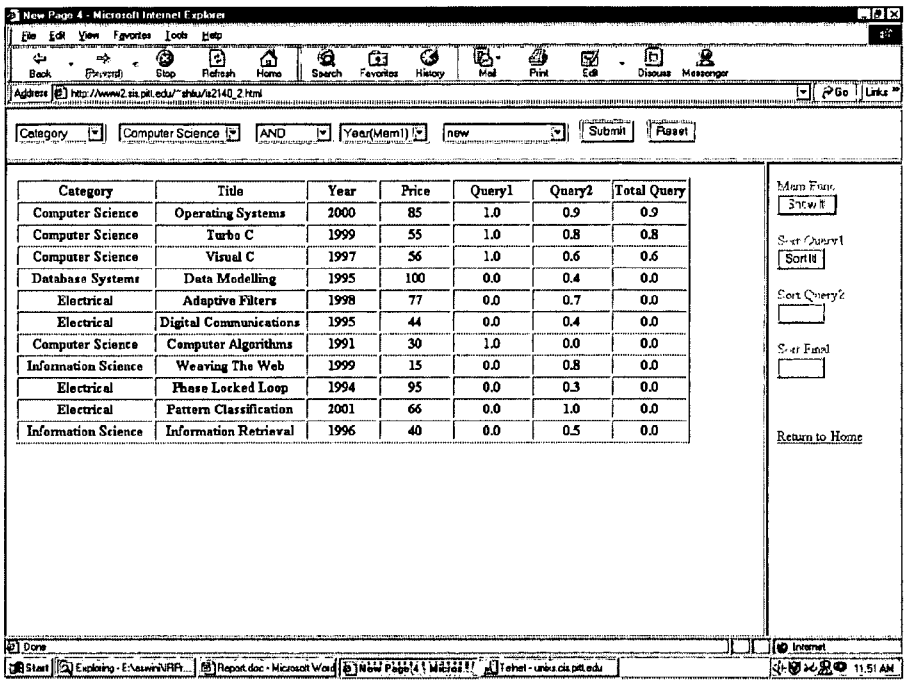
Other Learning Environments

As we have mentioned, the Boolean matching environment is just one of several interactive visualization environments that we have developed and explored. The environments for other models are reasonably similar. In all of them we have tried to center the visualization of the matching process on a visual representation of the whole set of documents in some form. The environments for other models are a bit more complicated (since the models themselves are more complicated) but they also provide the students with more opportunities for interactive exploration. For example, in the fuzzy matching model, an elementary query corresponds to a fuzzy subset. A fuzzy subset can not be shown by simply highlighting its documents – we need to show the "membership" of each document in a set. Thus an environment for fuzzy matching (Figure 4) has to provide extra fields for every document to show their memberships in all contributing queries (three rightmost columns on Figure 4) and provide the student an opportunity to re-sort the documents in the set by the value of each of these memberships.

Altogether our environments provide a very useful suite of tools for teaching information retrieval. We anticipate the use of these environments as both teaching and learning tools. First, we have found that these environments provide an excellent tool for a teacher to explain complex topics of IR models. Yes, it assumes that a teacher uses a computer and a projector in the classroom, but this is currently the standard context in most Computer and Information Science Departments. Still, computer projectors are most often used in classes to show the same static slides as in the age of overhead projectors and blackboards. By using the power of interactive visualization, our environments go well beyond traditional whiteboard and slides. At the same time, their use requires almost zero preparation time (just to plan which examples to show in order to cover the main set of ideas). The teacher can easily accommodate very different audiences by adjusting the number of examples to show, the speed, and the granularity of presentation. Even in a department with "computerless classes" the environments can serve as a powerful teaching tool to handle the questions and the problems of a troubled student in a one-to-one "office" context.

Using these environments as learning tools provides even better value. They let the students to switch from passive learning-by-reading to active and interactive exploratory learning. By exploring a number of different examples with an interactive visualization environment they should be able to achieve a better understanding of complex IR topics.

We have already performed an informal evaluation of several environments as teaching and learning tools in Summer 2001 semester and have got a very positive feedback from our students. Now we are planning an extensive formal evaluation during the Summer and Fall 2002 semesters.



Category	Title	Year	Price	Query1	Query2	Total Query
Computer Science	Operating Systems	2000	85	1.0	0.9	0.9
Computer Science	Turbo C	1999	55	1.0	0.8	0.8
Computer Science	Visual C	1997	56	1.0	0.6	0.6
Database Systems	Data Modelling	1995	100	0.0	0.4	0.0
Electrical	Adaptive Filters	1998	77	0.0	0.7	0.0
Electrical	Digital Communications	1995	44	0.0	0.4	0.0
Computer Science	Computer Algorithms	1991	30	1.0	0.0	0.0
Information Science	Weaving The Web	1999	15	0.0	0.8	0.0
Electrical	Phase Locked Loop	1994	95	0.0	0.3	0.0
Electrical	Pattern Classification	2001	66	0.0	1.0	0.0
Information Science	Information Retrieval	1996	40	0.0	0.5	0.0

Figure 4: Fuzzy Model Environment. Visualization of matching and ordering for a fuzzy AND query. The documents are sorted according to the value of the membership function of the compound query.

Implementation Issues

It was a critical design decision for us to implement all environments as Web-based tools. Web interface ensures that our tools can be accessed anytime anywhere. It lets us forget about platform differences and avoid troublesome process of installation. This is very important in the context of college education where neither teachers nor students have full control over the computers used in the classrooms and labs. Besides, it offers an extra benefit in the context of Web-based and distance education enriching the learning experience of a remote student.

In fact, most of our environments were implemented with server-side CGI scripts. In the age of Java this choice is harder to advocate. Indeed, the first "pure Web" learning environments for computer science subjects were developed with CGI technology. Examples include both algorithm animation (Campbell et al., 1995; Ibrahim, 1994), and program visualization (Brusilovsky, Schwarz & Weber, 1996). However nowadays, Java is becoming a dominant technology for developing Web-based interactive learning environments.

We also use Java when it is necessary due to the highly graphical interface demands (for example, for set diagrams), however we were trying to avoid it when it is not necessary for several reasons. First, we have found that Java applets are not always working on the browsers installed on various campus computers. Second, one of our projects involves wireless access to learning tools using handheld computers (that have browsers with no Java support). Third, a CGI-based tool provides an easier platform for centralized collection of interaction logs, which provide the most useful source of data for our educational research. Fourth, we are working on adaptive interactive visualization (Brusilovsky, 1994) that requires the presence of a centralized server-side student model. Overall, we think that in the cases when advanced graphics or smooth visualization is not important the old CGI-based technology is simply better.

We should add that the suite of our tools for teaching information retrieval is available for teachers and students of any IR courses. The home page of this project is http://www2.sis.pitt.edu/~ir/Projects/. Currently, all tools are running on our servers and could be used by anyone who is interested to teach or learn information retrieval and have access to the Internet. We are also working on packaging these tools as public domain software to be installed wherever someone wants to use it.

BEST COPY AVAILABLE

Currently we are working on developing several other environments to support teaching and learning of IR. In these environments we were exploring different visualization metaphors (such as set diagrams) to demonstrate matching process as well as the use of interactive visualization to support other traditionally hard topics of an information retrieval course.

References

Baeza-Yates, R. and Ribeiro-Neto, B. (1999) Modern information retrieval. N.Y.: ACM Press.

Brown, M. H. and Najork, M. A. (1997) Collaborative Active Textbooks. *Journal of Visual Languages and Computing* 8 (4), 453-486.

Brusilovsky, P. (1994) Explanatory visualization in an educational programming environment: connecting examples with general knowledge. In: B. Blumenthal, J. Gornostaev and C. Unger (eds.) *Human-Computer Interaction*. Lecture Notes in Computer Science, Vol. 876, (Proceedings of 4th International Conference on Human-Computer Interaction, EWHCI'94, St. Petersburg, Russia, August 2-5, 1994) Berlin: Springer-Verlag, pp. 202-212.

Brusilovsky, P., Schwarz, E., and Weber, G. (1996) ELM-ART: An intelligent tutoring system on World Wide Web. In: C. Frasson, G. Gauthier and A. Lesgold (eds.) *Intelligent Tutoring Systems*. Lecture Notes in Computer Science, Vol. 1086, (Proceedings of Third International Conference on Intelligent Tutoring Systems, ITS-96, Montreal, June 12-14, 1996) Berlin: Springer Verlag, pp. 261-269.

Butler, J. E. and Brockman, J. B. (2001) A Web-based learning tool that simulates a simple computer architecture. *SIGCSE Bulletin - inroads* 33 (2), 47-50.

Campbell, J. K., Hurley, S., Jones, S. B., and Stephens, N. M. (1995) Constructing educational courseware using NCSA Mosaic and the World-Wide Web. *Computer Networks and ISDN Systems* (Proceedings of Third World-Wide Web Conference, Darmstadt, 10-14 April, 1995) 27 (6), 887-895.

Domingue, J. and Mulholland, P. (1998) An Effective Web Based Software Visualization Learning Environment. *Journal of Visual Languages and Computing* 9 (5), 485-508.

Haajanen, J., Pesonius, M., Sutinen, E., Tarhio, J., Teräsvirta, T., and Vanninen, P. (1997) Animation of user algorithms on the Web. In: Proceedings of VL '97, IEEE Symposium on Visual Languages, , IEEE, pp. 360-367, also available at http://www.cs.helsinki.fi/research/aaps/Jeliot/vl.ps.gz.

Hampel, T., Keil-Slawik, R., and Ferber, F. (1999) Explorations - A New Form of Highly Interactive Learning Materials. In: P. D. Bra and J. Leggett (eds.) Proceedings of WebNet'99, World Conference of the WWW and Internet, Honolulu, HI, Oct. 24-30, 1999, AACE, pp. 463-468.

Ibrahim, B. (1994) World-Wide Algorithm Animation. *Computer Networks and ISDN Systems* (Proceedings of First World-Wide Web Conference) 27 (2), 255-265.

Korfhage, R. R. (1997) Information storage and retrieval. N.Y.: Wiley Computer Publishing.

McKenna, A. and Agogino, A. (1997) Engineering for middle schools: A Web-based module for learning and designing with simple machines. In: Proceedings of FIE'97, Frontiers in Education Conference, Pittsburgh, PA, November 5-8, 1997, Stipes Publishing L.L.C., pp. 1496-1501.

Tung, S.-H. S. (1998) Visualizing Evaluation in Scheme. *Lisp and Symbolic Computation* 10 (3), 201-222, ftp://140.125.81.71/pub/tungsh/lasc.ps.Z.

Yaron, D., Freeland, R., Lange, D., and Milton, D. J. (2001) Using Simulations to Transform the Nature of Chemistry Homework, Carnegie Mellon University, Pittsburgh. http://www.ched-ccce.org/confchem/2000/b/yaron/Default.htm

Acknowledgements

Most of the tools in the original suite of learning environments were developed under the guidance of the author by M.S. students of the School of Information Sciences at the University of Pittsburgh. We are fortunate to have many bright students and the author thanks them for their devotion and creativity.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").