Using Annotated Video as an Information Retrieval Interface

Andrew S. Gordon
IBM T.J. Watson Research Center
30 Saw Mill River Road
Hawthorne, NY 10532
Email: asgordon@us.ibm.com

ABSTRACT

The ability to deliver appropriate information to learners at the most appropriate time is an essential component of good instruction. In the best learning environments, this information is received in the context of the performance of the skills that are being acquired. This paper explores a technological approach to situated information retrieval by linking materials to segments of a video recording a skill performance. An interface is described where users navigate through a video performance and are presented with information relevant to the current video location. An approach to algorithmically generating interfaces of this type is then presented. The system takes as input annotations that describe a video recording of a performance, translates these annotations into subject terms used to catalog information resources, and then retrieves materials from online database servers using the Z39.50 information retrieval protocol. As an example application, the system was used to generate online teacher professional development materials by linking annotated video of classroom teaching with resources cataloged in the ERIC database.

Keywords

Computer Learning Environments, Information Retrieval, Digital Libraries, Multimedia Interfaces

1. SITUATED INFORMATION RETRIEVAL

One of the hallmarks of good teaching is delivering relevant and constructive information to learners at appropriate times. A good music teacher will tell their students aspects of music theory when it is useful in improving their performance. A good airplane flight instructor will convey information about the aerodynamics of an airplane to students when this knowledge can be used to avoid aviation hazards. In these cases of cognitive apprenticeship [4], it is the responsibility of the teacher to retrieve and deliver the information that can best advance the student's acquisition of the skills they are learning. Indeed, it is this characteristic of good teaching that reminds us that the communication of explicit knowledge plays a critical role in the acquisition of complex skills.

Technology has enabled us to develop computer learning environments that include many of the characteristics that define cognitive apprenticeship [3]. However, the ability to automatically retrieve educationally relevant materials for students engaged in skill learning has proven to be extremely challenging. In the ideal case, computers would be able to serve this function for students engaged in real-world situated learning, but realizing this goal would require systems that could perceive and assess real-world performances [9] combined with information retrieval systems that could map these assessments to relevant information. Instead, educational technology research has focused on the automated retrieval of information in the context of computer simulations [15,6,17]. The attraction of computer simulations is that the parameters of the simulation environment itself can be used as input to information retrieval systems embedded in pedagogical agents (e.g. [7,11,1]).

This paper explores a different alternative for building situated information retrieval systems. Rather than using simulation environments, we examine the use of video recordings of skill performances to provide a context for the delivery of relevant information. In section 2, we describe an interface that provides information to learners in the context of watching a video example of the performance of a skill. In this interface, users watch and navigate through a video while the system offers information that is related to the particular activities that occur in each video segment.

Following a description of the interface, an approach to algorithmically generating these interfaces from annotated video is presented. A system is described that takes as input a video recording of a performance and a set of descriptive annotations, and generates as output an interface that links

the video to relevant documents found by searching various online digital libraries. This system was designed to take advantage of new standards in library information retrieval systems, i.e. the ANSI/NISO Z39.50 information retrieval protocol [21], to allow it to be used to access a wide variety of different information sources appropriate for a wide range of skill domains.

To demonstrate the utility of this approach, section 4 describes how the system was used to generate professional development materials for primary and secondary school mathematics teachers. Using public-use video of classroom teaching obtained from the Third International Math and Science Study, a web-based interface was generated to deliver materials from the Educational Resource Information Center (ERIC) database to mathematics teachers.

2. A VIDEO-BASED RETRIEVAL INTERFACE

In order to explore the delivery of information in the context of watching a video of a skill performance, an interface was designed that would link information materials to segments of a digital video file embedded in pages accessible using a World-Wide Web browser. A screen shot of an instantiation of this interface design is given in Figure 1. In this example, the interface is being used to provide student pilots with information about aviation skills in the context of watching a video of a flight captured from inside the cockpit of a small plane. Each of the components of the interface is labeled, and described below.

A. Video of a performance of a skill

The focus of attention in the interface is a video of a performance of a skill, which can be viewed in the interface under the full control of the user. While any other time-based media (e.g. audio performances) could be substituted for video, the primary intention is to show to the user an example of someone else performing the skill that is being acquired. It is not necessary, and often not desirable, that this performance be an example of best-practices – there is much to be learned by presenting information in the context of an average or even a poor quality performance.

B. Segments

As the user watches and navigates through the video performance, the entire right side of the interface presents information that is pertinent to different segments of the video, expressed as in and out time codes. As the video progresses or is moved into a new time segment of the video, this window updates accordingly – displaying only materials that are related to the segment of video that is currently being played by the user. In order to assist the user in identifying different segments the sequential segment number and time codes can be displayed to the user, along with any descriptive annotations that interfaces developers have assigned to the segment.

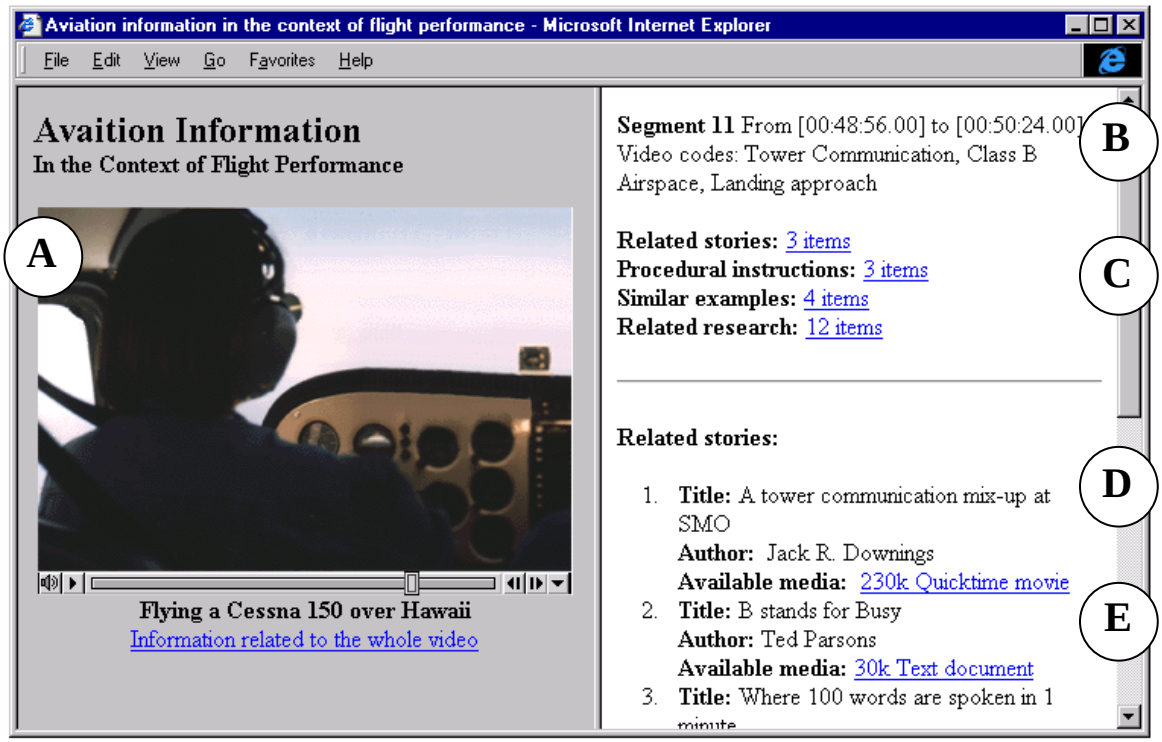


Figure 1: An interface that links a video performance with information

C. Categories

In each segment, there may be some number of educational materials that are relevant to what is currently happening in the video performance. These educational materials are grouped into categories that can be used to distinguish materials based on their genre, their educational role, their relationship to the current video segment, or any other characteristic intrinsic to the material or of its use. The primary purpose of the category heading is to direct the user to a subset of the segment's materials that they are most interested in. Clicking on the hyperlink next to the category simple scrolls this frame down to the beginning of the list of category materials.

D. Materials

The informational materials that are related to the current segment of the video performance are listed as short-form records, and typically include titles, authors, or any other identifying information that is available. The precise format of this short-form record may vary significantly according to the genre of the material that is being presented, i.e. short, textual descriptions of journal articles or full-length books will differ significantly from short descriptions of digital movie files, statistical data sets, diagrams, or navigational maps.

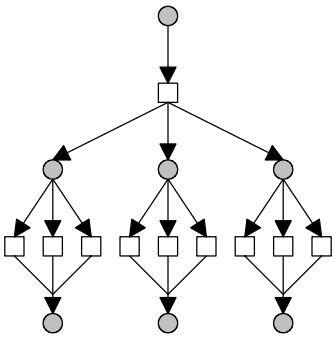
E. Available Media

If the actual media content is available online, then a hyperlink will be provided that opens the media in a new browser window for viewing by the user. In the absence of online documents, a more complete cataloging record should be provided to assist the user in locating the information off-line.

3. GENERATING INTERFACES FROM VIDEO ANNOTATIONS

Interfaces that link video to information are labor-intensive to construct manually. First, it is necessary to segment the video into smaller component parts of the larger skill performance. Second, a set of educational materials must be located that are relevant to each of the video segments. Finally, these materials must be organized on pages that can be presented to the user while they are watching and navigating through the video performance. Additionally, this work must be repeated for each performance video that is used, or when the corpus of available educational materials changes. In the ideal case, all of this work would be done algorithmically, i.e. the developer would provide only a video performance of a skill as input, and the system would generate an interface that links the video to relevant information as output. The great difficulty in realizing this generation process is the lack of adequate video analysis algorithms that could provide some meaningful information about the content of the video stream. In the absence of adequate video analysis algorithms, some amount of manual processing of the video data will be necessary.

In this research, we algorithmically generate interfaces that link video to information using video annotations, i.e. descriptions of the video consisting of skill-specific codes assigned by human experts. First, this algorithm segments the video into smaller coded segments. Then, it uses the annotations that are assigned to these segments to search various online library collections to find materials that are relevant to each segment. Finally, it collects the results of the searches into a series of web pages that are linked to the performance video. Figure 2 shows a conceptualization of the flow of information of the algorithm.



Video Codes

A set of descriptors of video content, each with in and out time codes

Segment Generator

Identifies the series of segments with unique video code sets

Segment Codes

A set of video codes for a particular video segment

Results Generators

Uses video codes to retrieve materials from online libraries; see Figure 3

Segment Web Pages

A web page with the retrieved materials organized into categories

Figure 2. Generating an interface from video annotations

The annotations of a video performance that are provided as input to the system consist of a series of descriptive codes, each assigned with a given starting time and an ending time in the video stream, or assigned to the whole video performance. While any set of codes can be used to describe the video content, the system requires a series of mapping tables which are used to translate these codes into queries sent to online library collections. The best approach, therefore, is to use a single set of codes and develop a fixed set of associated mapping tables, allowing the algorithm to be used with any video that is annotated with that set of codes.

The list of a video's codes is first passed to the Segment Generator, which identifies where segments of video are coded differently. The Segment Generator divides the video at each starting and ending time of every code, and then identifies the sets of codes that are contained within every two divisions. The result of this process is a set of segments where no adjoining segments have identical sets of codes and each segment's codes are applicable for the entire duration of the segment. Figure 2 shows the Segment Generator dividing the video into only three segments, however the actual number of generated segments is solely determined by the assignments of video codes that serve as input. In some cases, it is reasonable to assign descriptive codes to an entire video performance, i.e. without specific in and out timecodes. When whole-video codes are used, the Segment Generator gathered them up into a special set and processes them in exactly the same manner as the others.

The list of codes assigned to each segment is then passed to a number of Results Generators, one for each data sources specified to the system. An individual data source is defined by four pieces of information. First, it specifies the category of information that it is retrieving, i.e. in which category the retrieved materials will appear when presented to the interface user. Second, it gives the address of an online (Z39.50) server that catalogs the materials. Third, it selects the search strategy that will be employed to locate materials on the server. Fourth, it provides a pointer to a mapping file that is used by the Results Generator to translate the codes into the subject terms that are used to catalog materials on the server. A Results Generator is executed for each data source that has been specified, which takes as input the video codes that have been assigned to segment, and returns a set of retrieved materials to be presented to the interface user in the category defined in the data source. The operation of the Results Generator is displayed in Figure 3.

The Results Generator begins by mapping the segment codes onto subject terms that are being used to catalog the materials on the online digital library that is being searched. For example, if the Results Generator is configured to search through the books cataloged at the Library of Congress, it is necessary to first translate the segment codes into Library of Congress Subject Headings (LCSH). Likewise, if the National Library of Medicine is being searched, the Medical Subject Headings (MESH) would be the target subject vocabulary. In some cases, the video itself could be coded using the subject terms of the collection being searched, removing the need to translate the codes into a different vocabulary. However, even in these cases, there may not be a one-to-one mapping that will appropriately retrieve educationally relevant information.

The Mapping Process of the Results Generator is very straightforward. First, a single mapping table is loaded from

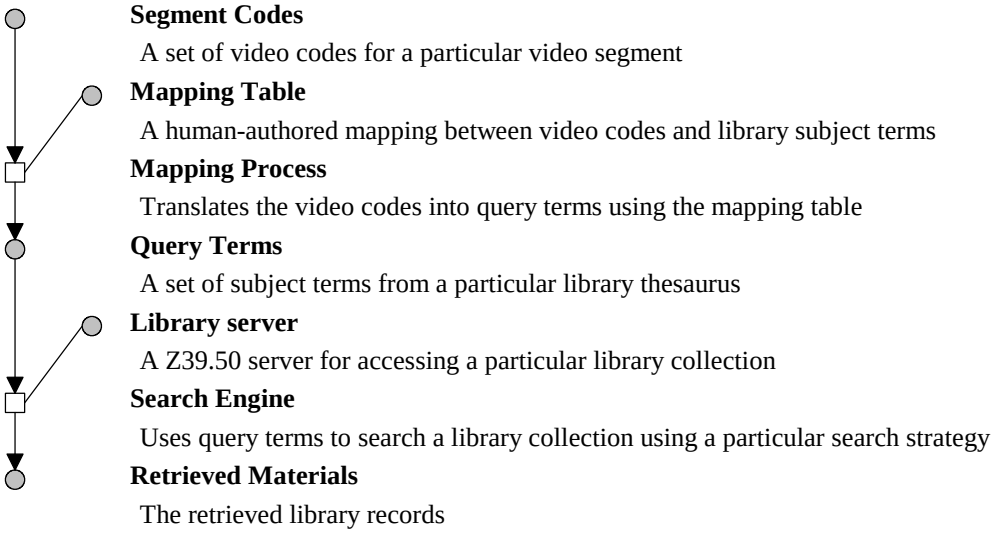


Figure 3. The Results Generator: Retrieving materials based on segment codes

a file specified by the data source. A mapping table simply pairs codes with their query term translations. One-to-one, one-to-many, and many-to-one mappings are supported. It is not required that every code in the vocabulary has a mapping – indeed, it is possible that a richly coded segment of video could produce no query terms.

The set of generated query terms are then passed to the Search Engine, which searches an online library collection for available materials. The Search Engine opens up a connection with library collection specified by the data source using the Z39.50 retrieval protocol. Then it sends a series of queries to the server that may be different depending on which search strategy is specified in the data source. Three simple search strategies have been implemented, known as Any-Match, Exact-Match, and Best-Match. Any-Match and Exact-Match strategies simply query the Z39.50 server for documents with the list of query terms combined using the OR and AND operators, respectively. The Best-Fit search strategy generates a series of queries, stopping with the first query that retrieves a positive number of materials. The strategy begins by querying for a conjunction of all of the query terms, which is equivalent to the Exact-Match strategy. Then it successively relaxes the size of the conjunction, asking for any conjunction with all but one of the terms, then all but two, etc. Finally, the system asks for any one of the query terms, which is equivalent to the Any-Match strategy. The Best-Match strategy can be further constrained by specifying the minimum percentage of query terms from the total number that must be present in the query. For example, the data source may specify that each of the retrieved items must have at least 50% of a segment's query terms as subject terms.

After all of the Results Generators for a segment have executed, the results are gathered up and presented on a web page. The category information is used to organize the retrieved materials, and a summary of number of materials retrieved in each category is presented at the top of the web page, with a hyperlink to the start of each category's list further below on the web page. In some cases, two Results Generators may have located materials that have been assigned the same category, and these result sets are simply combined. The retrieved materials are represented internally as library MARC records, as returned by the Z39.50 server. Some processing of these MARC records is required to prepare them for display to the user. Along with extracting relevant Title, Author, and other reference information, the 856 MARC field is examined to determine if the information is available in some online form (typically the URL of an electronic file). When available, a URL hyperlink will be displayed along with other information about the item.

The final step in generating the interface is to link the video to the each of the generated web pages. In this research, we used a functionality of Apple's QuickTime browser plug-in that allows video files to contain a special *HREF Track*. Playing an embedded movie with an HREF Track can cause a given frame in the browser to load a different web page depending on the current location of the video. The Segment Generator algorithm was used to create a special HREF Track containing links to each of the segment web pages at the appropriate time positions. This track was then added to the original video performance, and embedded in a web page accessible to intended users of the interface.

4. EXAMPLE: ACCESSING ERIC RECORDS USING CLASSROOM VIDEO

There is a continuing perception among classroom teachers that there is a relevance gap between their classroom activities and educational research (e.g. [18, 14]). Providers of educational databases have expressed concern about these perceptions, and have worked to improve the way that educational research is cataloged and accessed [10]. However, set in the context of traditional library-based retrieval behaviors, these improvements do little to situate information in the practice of classroom teaching. If we believe strongly that information must be provided in the context of its use, then educational databases must be more closely linked to the lives of classroom teachers.

It is our position that video of classroom teaching can provide an appropriate context for the presentation of research articles, books, and other educational materials that could be used to improve teaching practice. Classroom video is fast becoming the tool of choice for researchers and practitioners in the areas of teacher professional development and the analysis of teaching practice [13]. Continued work in this area will increase the number of in-service and pre-service teachers that analyze classroom video as part of their education, and also increase the number of classroom videos that can be widely used. These expectations create an opportunity to apply the research presented in this paper in order to provide large communities of teachers with educational research set in the context of classroom teaching.

As an example of how the system described in this paper could be used to access educational databases, we used it to generate links between classroom video collected in the TIMSS Videotape Classroom study and records in the ERIC database.

The U.S. Department of Education's Educational Resources Information Center (ERIC) began cataloging articles, books, and other documents related to education in 1966. Today, the ERIC database contains nearly one million records [5]. A variety of user groups, including researchers, teachers, students, librarians, and parents, access the ERIC database in a variety of different ways. These include searching through printed editions, online search engines, CD-ROM access

tools, as well as a personalized internet-based search service.

The TIMSS Videotape Classroom Study was conducted as a component of the Third International Mathematics and Science Study (TIMSS) between 1994 and 1996 [19]. In the main study, randomly sampled eighth-grade mathematics classrooms in the United States, Japan, and Germany were videotaped for one class period, then coded for comparative analysis. During the course of this study, 10 classroom videos (5 each from Japan and Germany) were made available as public-use for the purpose of teacher professional development. Like the restricted-use TIMSS video, this public-use collection was coded using a set of descriptors developed by the researchers involved in the video study. The codes assigned to the public-use videos included those for describing the mathematical content area of each lesson (59 codes), the objects and educational materials that were used in each classroom (15 codes), and the kind of organization and activities that are occurring at different times in each lesson (18 codes). The mathematical content and materials code sets each apply to a lesson as a whole and are represented as single terms. The activity codes are also represented as single terms, but include a start time and an end time that identifies the segment of the video to which they apply.

In order to use the software described in this paper to generate interfaces that linked ERIC materials to each of the public use videos, three necessary tasks were completed. First, the codes for each of the 10 classroom videos were entered into a format that could be read by the software. Second, an ERIC database was located on a library server supporting the Z39.50 information retrieval protocol. Third, mapping tables were created for each of the three categories of video codes. For this example, these mapping tables were designed to pair each of the 92 codes with a term or set of terms from the ERIC thesaurus whose meaning was most synonymous. After completing these three tasks, the algorithm was applied to the 10 TIMSS videos, generating 123 web pages containing a total of 1820 links to ERIC materials. Some parameters of the generated interfaces are presented in Figure 4.

To assess the effectiveness of the software in linking annotated video to relevant library materials, relevance assessments were made on a random sample of the retrieved items. One evaluator was chosen: the director of a mathematics teacher education program in California who had become familiar with two of the Japanese videos by using them in teacher discussion groups. Twenty-five items were randomly selected from those linked to these two Japanese lessons. Of the twenty-five randomly selected items, eight (32%) were judged to be relevant to the section of video to which it was linked. This level of precision discourages the use of the materials generated in this example application in teacher professional development programs. Indeed, as we look toward future work and applications of this technology, our focus must be on strategies for improving the overall relevance of the linked materials.

5. DISCUSSION AND FUTURE WORK

The use of video performances to provide appropriate contexts for information delivery has a number of attractive advantages over other types of technological alternatives. Video performances of skills are easy to produce, their description requires only a modest amount of content analysis, and techniques like the one described in this paper demonstrate that it is feasible to apply automated techniques to link videos to information resources. Their success as a learning technology, however, will greatly depend on the larger learning environment in which they are used.

For the purpose of automatically generating these interfaces from annotated video, the focus of future research must be on strategies for improving the relevance of the materials that are linked to the video recordings. There are many

Parameter of the generated interfaces for each video	**Population Mean (μ)**	**Standard Deviation (σ)**
Length of video	47.7 minutes	4.7 minutes
Number of whole-video codes	2.7 codes	0.8 codes
Number of video segment codes	13.9 codes	6.0 codes
Segments per video (including whole video segments)	12.3 segments	3.4 segments
Average segment length (excluding whole video segment)	4.6 minutes	1.5 minutes
Database searches	9.7 searches	3.6 searches
Successful database searches	8.2 searches	2.9 searches
Found database items	3518.2 items	3400.3 items
Links created (limiting links to 25 per search)	182.0 links	68.5 links

Figure 4: Parameters of the generated interfaces (N = 10 videos)

approaches that could be taken to improving the algorithm presented in this paper. First, any improvements to the richness of the coding scheme used to describe the performance would lead to increased precision of the retrieved set of materials. Second, the mapping tables used to link video codes to query terms could be improved by incorporate various pedagogical theories of information relevance, rather than the simple equivalence mappings used in this paper's example. Third, rather than using general-purpose databases of published materials, it would be better to use collections of materials that pertained directly to the improvement of skill performance, indexed by task-relevant subject terms.

Independent of these semantic improvements, the methods used to search for materials could be greatly improved. Techniques for improving precision and recall performance by expansion of query terms (e.g. [8, 20]) could serve as direct replacements for the simple search strategies used in this research. Likewise, any strategies for ranking the relevance of retrieved items based on subject terms (e.g. [16, 12]) could serve to filter or organize the materials presented to the users.

Perhaps the most exciting area of future work is to move from using video annotations to the use of unprocessed video streams as input to the information retrieval system. As computer vision and video analysis technology advance (e.g. [2]), it may become possible to extract enough task-relevant information to seed information retrieval systems like the one described in this paper for use in increasingly interesting domains. Future systems could be used not only with pre-produced videos of skill performances, but also as a means of watching and coaching learners engaged in skill performances in real-world contexts. Continued research in this area will eventually lead to systems that truly support situated information retrieval – delivering relevant information to learners engaged in the real-world performance of complex skills.

6. ACKNOWLEDGMENTS

The author would like to thank Margaret Smith at the University of Delaware for her assistance in developing the example used in this paper and Nannette Seago from the California Math Renaissance program for serving as its evaluator. The 10 public-use classroom videos were obtained from the TIMSS Video Data Center, located in the Psychology Department at the University of California Los Angeles. The software described in this paper made extensive use of the Z39.50 Java libraries provided by OCLC. Connecticut State University provided Z39.50 access to the ERIC database used in this research.

7. REFERENCES

[1] Burke, Robin (1993) Representation, Storage and Retrieval of Tutorial Stories in a Social Simulation. Ph.D. Dissertation, Northwestern University, December 1993.

[2] Christel, M., Kanade, T., Mauldin, M., Reddy, R., Stevens, S., Wactlar, H. (1996) Techniques for the creation and exploration of Digital Video Libraries. In Furht, B. (Ed) Multimedia Tools and Applications (Volume 2). Boston, MA: Kluwer Academic Publishers, 1996.

[3] Collins, A. (1991) Cognitive Apprenticeship and Instructional Technology. In L. Idol & B. F. Jones (Eds.), Educational values and cognitive instruction: Implications for reform. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

[4] Collins, A., Brown, J. S., and Newman, S. (1989) Cognitive Apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.) Kowing, learning and instruction: Essays in honor of Robert Glaser. Hillsdale, NJ: Lawrence Erlbaum Associates.

[5] Educational Resources Information Center (1998). ERIC Annual Report 1998. National Library of Education, Office of Educational Research and Improvement, U.S. Department of Education.

[6] Forbus, K. and P. Whalley (1994) Using qualitative physics to build articulate software for thermodynamics education. Proceedings of AAAI-94, Seattle, WA. Menlo Park, CA: AAAI Press.

[7] Forbus, K., Everett, J., Ureel, L., Brokowski, M., Baher, J., and Kuehne, S. (1998) Distributed Coaching for an Intelligent Learning Environment. Proceedings of the Qualitative Reasoning Workshop 1998, May 1998, Cape Cod, MA. Menlo Park, CA: AAAI Press.

[8] Fox, Edward (1980) Lexical relations: Enhancing effectiveness of information retrieval systems. ACM SIGIR Forum 15(3): 5-36.

[9] Gordon, Andrew S. (1995) Automated Video Assessment of Human Performance. In J. Greer (ed) Proceedings of AI-ED 95 - World Conference on Artificial Intelligence in Education, Washington, DC; August 16-19, 1995. Charlottesville, VA: AACE Press. pp. 541-546.

[10] Johnson, J.R.V. (1990) Educational Research and Educational Practice: Bridging the Gap. Journal of Education for Teaching; 16(1):83-90. March 1990.

[11] Jona, Menachem (1995) Representing and Applying Teaching Strategies in Computer-Based Learning by Doing Tutors. Ph.D. Dissertation, Northwestern University, June 1995.

[12] Kim, Young W. and Jin H. Kim (1990) A model of knowledge based information retrieval with hierarchical concept graph. Journal of Documentation, 46:52-59.

[13] Lampert, Magdalene and Jan Hawkins (1998). Report to the National Science Foundation: New Technologies for the Study of Teaching. Workshop held June 9-11, 1998, Ann Arbor, MI.

[14] McDonough, Jo and Steven McDonough (1990) What's the Use of Research? ELT Journal, 44(2):102-09, April 1990.

[15] Murray, Tom (1999) Authoring Intelligent Tutoring Systems: An analysis of the state of the art. International Journal of Artificial Intelligence in Education, 10:100-133, 1999.

[16] Rada, Roy an Ellen Bicknell. (1989) Ranking Documents with a Thesaurus. Journal of the American Society of Information Science, 40:304-310.

[17] Schank, Roger and Chip Cleary (1994) Engines for Education. Hillsdale, NJ: Lawrence Erlbaum Associates.

[18] Spaeth, Jeanne (1994) So What Good is Research? Teaching Music, 2(1):42-43. August 1994.

[19] Stigler, J.W., P.A. Gonzales, T. Kawanaka, S. Knoll, and A. Serrano (1999) The TIMSS Videotape Classroom Study: Methods and Findings from an exploratory research project on eighth-grade mathematics instruction in Germany, Japan, and the United States. Jessup, MD: Education Publications Center.

[20] Wang, Yih-Chen, James Vandendorpe, and Martha Evens (1985). Relational thesauri in information retrieval. Journal of the American Society of Information Science, 36:15-27.

[21] Z39.50 Maintenance Agency (1995) Information Retrieval (Z39.50): Application Service Definition and Protocol Specification (ANSI/NISO Z39.50-1995). Bethesda, MD: NISO Press.