Improving Learning Object Schemas and Performance Support Systems through Information Retrieval Evaluation

Introduction

This study examines existent and new methods for evaluating the success of information retrieval systems. Performance support systems, knowledge management systems, and meta-data schema such as SCORM, IMS, and IEEE LOM are concerned with finding needed information quickly and effectively, whether that information is a learning object, a performance object, or a knowledge object. However, the literature of instructional design rarely examines search and retrieval. This integral part of these systems, a part that certainly affects usefulness, is overlooked. Even in IS, the theory underlying current methods of evaluation is not robust enough to effectively evaluate the web or the very targeted searches required for effective performance support.

Traditional information retrieval evaluation measures rely on judgments of whether a document is "relevant" to a particular question. A good system returns all the relevant documents and no extraneous documents. There is a rich literature questioning the efficacy of relevance judgments. Such questions as: Relevant to who? When? To what purpose? are not well answered in traditional theory.

In this study, two new measures are used in evaluating two systems, comparing these new measures with traditional measures and each other. One is derived from Spink's Information Need (Spink, 2002; Spink, Greisdorf, & Bateman, 1998) and the other derived from Cooper's Utility (Cooper, 1973a, 1973b).

Two very different systems of searching were used to search the same set of 500 documents that contained educational resources. One system, a text based system, resembled most common web search engines. The other system used a series of meta data tags for searching.

Thirty-four educators searched for information using both search engines and evaluated the information retrieved by each. The participants searched a total of four times – twice using each system. Construct measures, derived by multiplying each of the three measures (traditional, information need, and utility) by a rating of satisfaction were compared using two way analysis of variance.

Results indicated that there was a significant correlation between the three measures – so the new measures provided an equivalent method of evaluating systems and have some significant advantages – including no need for relevance judgments and easy application in situ. While the main focus of the study was on the methods of evaluation, the evaluation in this case showed that the text system was better than the tag based system.

Learning Objects – Focus on search and retrieval

Learning objects have stirred great controversy in the educational community over the past few years. Some embrace the idea as a new wave in education, allowing customization of instruction on the fly. Others flatly reject the concept of objects. If learning is constructed (either internally or socially) (Burton, Brown, & Fischer, 1999; Rogoff, 1990; Rogoff & Lave, 1984)**,** then the idea of objectified learning contradicts over two decades of research (more if we go back to the writings of Dewey and Vygotsky's original writings). Instead of the widely accepted notions of constructivist learning environments, learning objects seem to be a throwback to programmed instruction. However, learning objects are not the problem themselves. It is their intended use.

"The power of learning objects is that they can be decontextualized. This is what allows them to be used in different places. As educators, we know that context is everything. So, there is a fundamental tension between the power of learning objects and their use"(McGee, 2004). If the teacher or instructional designer provides the context, the learning object is merely a tool, a bit of information. The pedagogic problem lies in the claim that objects may be somehow strung together *automatically* based on some criteria. The problem is with the intended *use.*

If the problem is with intended use, let us look for a more useful, perhaps less explosive definition. Instead of defining by use, let us define by form. Very simply, we may define a learning object as an object that may be accessed on line by means of some kind of tags. The learning is directed by the learner and the instructor. The distinguishing feature of an information object intended to support performance or learning is that someone has added some kind of information (called meta-tags) to aid in the search and retrieval of the object. In other words, the efforts surrounding learning objects are to one end – *making it faster and easier to find and retrieve useful materials that can support instruction*.

If we focus on form, toward the end of enhancing finding and retrieving, then it behooves us to explore different methods of tagging objects and other methods to improve search and retrieval.

Performance Support Systems

Researchers in instructional systems technology have long studied more than just instruction. A particularly robust sub domain is human performance technology (HPT). Performance improvement focuses on results, not activities. Training is merely one class of intervention. The goal in any intervention is to improve performance.

So, according to the HPT view, one may have a greater impact on improved performance by providing better information during the course of performance. Rossett (Rossett, 1996) points out

that training tries to build capacity, occurring before a need arises. Performance interventions are used as needed. As information provided just in time offers a powerful tool for improving outcomes, a class of performance interventions has concerned itself with building computer-based systems that can provide just in time information. This class of interventions, performance support systems (PSS), have been in use since the 1980s.

EPSSs are "part online help, part online tutorial, part database, part application program, and part expert system…EPSSs quickly and easily provide answers to the questions workers have when performing a job, and address workers' concerns" (Carliner, 2002 p. 400).

While performance systems are designed to support the goal of improved performance, they certainly support learning, too. "In a well-designed PS (performance support) system, learning is likely, desirable, even inevitable, but it's not the point. *Performance* is the point." (Dickelman, 2000 p.8). Rosenberg makes the point even more strongly, saying that EPSS is a paradigm shift, bringing learning and work together (Rosenberg, 1995). He says the people developing EPSSs:

> ...wanted to obliterate the line between learning and work so that, in reality, learning is work, and work is learning...We see an integration of information and learning, and gain an understanding that knowledge may sometimes be more effectively and efficiently delivered...than through instruction...It is not appropriate to say that EPS will eliminate the need for education and training, for there will always be a need for new knowledge and continuous learning. But using EPS instead of training for disseminating facts and procedures frees up training resources for more sophisticated efforts in areas...better suited to the strengths of the educational model. (p. 92-93)

While there is a great diversity of functionalities across EPSS applications, one of the most common functions is a database of information that may be searched. "The primary design goal of an EPSS is that the knowledge it contains be easily retrievable by the users at the time they need it" (Cole, Fischer, & Saltzman, 1997 p. 50). A survey of readers of CBT Solutions magazine conducted

in 1996 listed “searchable reference” as the most common feature of EPSSs either built or being built for their companies. (Benson, 1997).

Much of the work concerning meta tags (Dublin core, IMS, SCORM, GEM and more) is predicated on the idea that adding meta tags can improve retrieval systems. Unfortunately, without being able to measure the effect of different search/retrieval systems, there is no way to know if efforts to improve practice are yielding results. The goal of this research is to see how the constructs and measures of information retrieval evaluation may inform the research agenda of performance support and if new constructs and measures may enrich and expand the theory.

Evaluation of information retrieval systems is a relatively mature field, with a literature dating to the mid 1960's. While this literature offers a rich foundation, the constructs and measures underlying the theory are thin. For one wishing to evaluate information retrieval within performance support systems, additional constructs and measures that reflect the unique needs of performance support are necessary.

The traditional method for evaluating information retrieval systems relies on the relevance based measures, recall and precision. To accomplish this type of evaluation requires:

- a collection of documents
- a collection of questions (queries) to be asked of the document collection
- a set of judgments of which documents are relevant to each question.

To evaluate the system, one queries the document collection (ask the questions and see what documents are retrieved) and calculates recall and precision. Recall is the measure of how many relevant documents were actually retrieved. For example, if one question was 'How many angels can dance on the head of a pin?' and 50 documents were judged relevant, and a system retrieved 20, then

the recall would be the number of relevant documents retrieved divided by the total number of relevant documents expressed as a percentage. (In this case 20/50 or 40%.) Precision is the measure of how many relevant documents were retrieved divided by the total number of retrieved documents. It is a measure of how "noisy" the results are. It would be easy to retrieve the 50 relevant documents – just retrieve all the documents in the collection. Both measures are reported in the traditional evaluation paradigm. They are generally assumed to be inversely proportional – the better the recall (the more relevant items retrieved) the lower the precision (the noisier the result).

Current Research

The current state of theory for the evaluation of information retrieval systems is not rich enough to easily map to performance systems. Some have said it does not map all that well to information retrieval systems (Cooper, 1981; Harter, 1996; Harter & Hert, 1997; Schamber, 1994). In this study, I have used a case study to examine two new constructs and their measures as well as the traditional constructs and their measures to attempt to obtain a richer view of comparative evaluation of two systems.

To accomplish this, the author constructed a document set of educational resources. 34 educators were recruited to participate. The resources were tagged with a unique meta tag set developed from interviews and analysis of potential reasons for information seeking on the web by this population. The participants were then asked to choose a task from a list of tasks developed from the interviews (attempting to provide authentic tasks for participants). Participants were then asked to seek to find information that would help them accomplish the task using either a text or a tag based system. They repeated this procedure using a text based system. They repeated the process again, for a total of four searches. They used the new constructs of Information Need and Utility to

evaluate the efficacy of the information retrieved. The traditional method of precision/recall was used as a comparison.

Results

In brief, the three measures of evaluation were very strongly correlated. All three measures were correlated within each system, significant at the .01 level. The three measures of constructs (P/R, UT, and IN) are strongly correlated within both systems.

With such strong correlations, we may hypothesize that, in this case, these are not three different constructs, but are, instead, three measures of the same construct, a construct we call System Efficacy.

Significance of Results

This research provides theoretical and technological advances. During the course of this research, three significant new online tools were developed: 1) the “tagging tool” allows fast tagging of resources; 2) the tag-based search tool; 3) the questionnaire tool used to complete the research allows in situ research on search and retrieval systems.

Theoretically, the new measures of system efficacy allow in situ evaluation, freeing evaluation from experimental settings. This is essential to allow design of systems that are “tuned” for specific populations. We can evaluate different meta data schemas for effectiveness. We can improve knowledge retrieval and performance support and can test changes to make systems work better.

References

Benson, G. (1997, January, 1997). A New look at EPSS. *Training & Development*, 43-44.

Burton, R. R., Brown, J. S., & Fischer, G. (1999). Skiing as a model of instruction. In B. Rogoff & J. Lave (Eds.), *Everyday cognition: Its Development in social context*. Cambridge: Harvard University Press.

Carliner, S. (2002). Read me first: An Introduction to this special issue. *Technical Communication, 49*(4), 399-404.

Cole, K., Fischer, O., & Saltzman, P. (1997). Just-in-time knowledge delivery. *Communications of the ACM, 40*(7), 49-53.

Cooper, W. S. (1973a). On selecting a measure of retrieval effectiveness. *Journal of the American Society for Information Science, 24*(2), 87-100.

Cooper, W. S. (1973b). On Selecting a measure of retrieval effectiveness: Part II. Implementation of the philosophy. *Journal of the American Society for Information Science, 24*(6), 413-424.

Cooper, W. S. (1981). Gedanken experimentation: An Alternative to traditional system testing? In K. Sparck Jones (Ed.), *Information Retrieval Experiment* (pp. 199-209). London: Butterworths.

Dickelman, G. J. (2000). Performance support in internet time: The State of the practice Discussion between Gloria Gery, Stan Malcom, Janet Cichelli, Hal Christensen, Barry Raybould, and Marc J. Rosenberg. *Performance Improvement, 39*(6), 7-17.

Harter, S. P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science, 47*(1), 37-49.

Harter, S. P., & Hert, C. A. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. *Annual Review of Information Science and Technology (ARIST), 32*, 3-94.

Hogan, D. M., & Tudge, J. R. H. (1999). Implications of Vygotsky's theory for peer learning. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 39-66). Mahwah, New Jersey: Lawrence Erlbaum Associates.

McGee, P. (2004). Fundamental tension between pedagogy and learning objects. In S. C. Schatz (Ed.).

Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.

Rogoff, B., & Lave, J. (Eds.). (1984). *Everyday cognition: Development in social context*. Cambridge: Harvard University Press.

Rosenberg, M. J. (1995). Performance technology, performance support, and the future of training: A Commentary. *Performance Improvement Quarterly, 8*(1), 94-99.

Rossett, A. (1996). Job aids and electronic performance support systems. In R. L. Craig (Ed.), *The ASTD training and development handbook - Fourth edition* (pp. 554-580). New York: McGraw - Hill.

Schamber, L. (1994). Relevance and information behavior. In M. E. Williams (Ed.), *Annual Review of Information Science and Technology* (Vol. 29, pp. 3 - 48). Medford: Learned Information, Inc.

Spink, A. (2002). A User-centered approach to evaluating human interaction with Web search engines: An Exploratory study. *Information Processing and Management, 38*(3), 401-426.

Spink, A., Greisdorf, H., & Bateman, J. (1998). From highly relevant to not relevant: Examining different regions of relevance. *Information Processing and Management, 34*(5), 599-621.