8-15-1997

Emerging Web-Based Knowledge-Bases: Categories, Trends and Implications for the IS Research Community

Alexander Hars
University of Southern California, hars@rcf.usc.edu

Arvind Malhotra
University of Southern California, amalhot@rcf.usc.edu

Emerging Web-Based Knowledge-Bases: Categories, Trends and Implications for the IS Research Community

Alexander Hars **Arvind Malhotra**
hars@rcf.usc.edu **amalhot@rcf.usc.edu**
Department of Information and Operations Management
Marshall School of Business
University of Southern California
Los Angeles, CA 90089-1421

The number of electronic knowledge bases on the world wide web is exploding. Organizations and individuals are investing time and resources to build and maintain knowledge bases. However, they have currently have minimal guidance in this effort. Identifying the underlying characteristics of existing knowledge bases and different paths they take for the creation, synthesis and organization of knowledge, is the first step towards determining the best model for designing knowledge bases in the future. In this paper we present an analysis of electronic knowledge bases which have a presence on the WWW and are related to information technology. This work is a preliminary effort in the development of a knowledge base design theory. A total of 38 such knowledge bases were identified and then categorized along two dimensions: "Mode of Interaction" and "Approach for Knowledge Organization". The results show that the majority of current knowledge bases focus on low levels of interaction and give little thought to the organization of their knowledge. At the same time, several knowledge bases with innovative approaches are identified. Taken together, they can serve as models for building more effective web-based knowledge bases which not only capture and disseminate knowledge but leverage information technology to improve the process of knowledge creation.

Surveying existing knowledge bases

Currently the predominant approach to publishing such knowledge is through electronic journals, most of which have been modeled on paper journals. However, as more and more electronic knowledge bases better exploit the capabilities of hypertext, databases and other information technologies significantly different types of electronic knowledge bases emerge on the WWW. Some of the new approaches change the focus from knowledge storage and dissemination to incorporate collaborative knowledge creation and synthesis of knowledge [Engelbart 1995, El Sawy et al. 1996].

The authors conducted a survey of scholarly resources related to information technology on the world wide web to identify the unique characteristics of available knowledge bases. The sites were identified through an extensive search of lists of scholarly resources, electronic journals and other types of link-lists available on the WWW. This was complemented with the use of search engines and through personal references. Among the many potential sites thus identified, those that were not related to information technology or that only publish abstracts were excluded. In addition, list servers and newsgroups were not included in the study a) because their characteristics are well known and b) because of their ad-hoc focus. This reduced the number of electronic knowledge bases to 38. Both authors evaluated each of the knowledge bases and listed a set of characteristics which best described the notable features of each knowledge base. Thus two sets of characteristics to describe current electronic knowledge bases were obtained. After reviewing and integrating both sets, two main dimensions around which to differentiate electronic knowledge bases were identified:

Mode of interaction

The knowledge bases differ in possible modes of interaction. In paper journals the primary mode of interaction in which the user engages is reading and - to a lesser degree - browsing. Information technology permits more advanced levels of interaction such as annotation and on-line discussion of contributions. Only a small number of current knowledge bases provide this capability but we expect this number to increase significantly in the near future because of the availability of robust web-enabled database

technology. For authors, the primary type of interaction are electronic submission and review of contributions. Most electronic journals allow electronic submission of contributions, although some journals which are an electronic copy of a paper journal still ask for the submission of paper. Some electronic knowledge bases have mechanisms to improve the efficiency of the review process [Barua, Chellapa and Whinston 1996]. Other journals have started to reengineer the review process and parallelize publishing and review. At ISWorldNet, for example, new submissions for a topic are published immediately [Ives and Zmud, 1994]. Reviewers as well as the audience are invited for comments. If the review panel comes to a negative decision, the submission is simply removed from the site. We consolidate these different characteristics into three different interaction modes: The first mode is restricted to read/browse. The second mode combines read/browse with support for annotation and comments, while the third mode adds immediate publishing. Immediate publishing does not imply that all contributions have to be ranked equally or that contributions can not be removed from the knowledge base if they are below quality standards. Note that in the presence of near-zero publishing costs, such a removal strategy is only optimal for the research community as a whole if we ensure that rejected papers have a value of zero for the audience. That position would be very difficult to achieve. An alternative is to use information technology to mark papers as rejected and to provide selective views on the knowledge base. An "edited view", for example, would provide only those contributions which have been accepted; an "in-progress view" would included accepted and unreviewed contributions while a "complete view" would include the rejected contributions. For someone who writes a paper on a narrow topic, for example, it might be valuable to be able to access rejected contributions for the same topic. The point we are advancing here is not that review is not necessary. It is rather that information technology allows us to separate out three elements of the review process: quality control, filtering (selection of interesting contributions) and publishing (dissemination).

		Mode of interaction-			
		Read / browse	+ online annotation by readers	+ immediate publishing of contributions	Total,%
	By publishing date (usually as volumes and issues)	18 (47%) **Journal of Artificial Intelligence Research**	3 (8%) **Electronic Journal on Networks and Distributed Processing**	1 (3%) **(ISWorld research working paper series)**[1]	22 (58%)
Approach for organizing knowledge	+ subject trees or other forms of category networks	8 (21%) **EC World Frontiers**[2]	3 (8%) **Journal of Computer Mediated Communication**	1 (3%) **(ISWorldNet)**[1]	12 (32%)
	+ active dependency maintenance	0 (0%)	3 (8%) **The Free On-line Dictionary of Computing**	1 (3 %) **Stanford ontolingua server**	4 (11%)
	Total, percent	26 (68%)	9 (24%)	3 (9%)	38 100%[3]

[1] Currently has no annotation capabilities. Publishes instantaneously but on temporary basis, Once approved then published permanently.

[2] Allows comments in a separate discussion forum; they are, however, not visible in the journal.

[3] Difference of 1% results from rounding.

Figure 1: Types of electronic knowledge bases: frequency and examples

Approach for organizing knowledge

This second dimension refers to the problem of organizing the knowledge base and ensuring that over time a consistent, integrated body of knowledge emerges. It is surprising how many electronic journals adopt the traditional paper-centric structuring into volumes and issues. Contributions within an issue may have nothing in common except the publication date. These journals lack a meta-structure for knowledge. More advanced electronic knowledge bases replace or complement this approach through a subject tree or a network of categories which the reader can browse. This enables the reader to easily identify contributions belonging to similar topics. A third approach improves the organization of knowledge by introducing procedures and mechanisms to actively monitor the consistency of the knowledge base and to identify dependencies between different contributions. This approach is often found in electronic reference works which originate from dictionaries and in knowledge bases which focus on the development of formal representations of a domain such as the Ontolingua system at Stanford.

The combination of both dimensions leads to a matrix with 9 cells where each cell corresponds to one type of an electronic knowledge base. Figure 1 shows how the number of knowledge bases for each type as well as the relative weight of each type. Where possible, an example for the type is included. The data shows that the largest number of electronic knowledge bases belong to the type which is characterized by the lowest level of interaction and the most basic approach for knowledge organization. These knowledge bases are modeled on paper journals; they replicate their structure of volumes and issues and focus on the dissemination of knowledge rather than on its creation and synthesis. Almost fifty percent of all knowledge bases fall into this category. Our data shows a clear trend of improving the levels of interaction and organization. The three adjacent types are the second and third-largest by number of knowledge bases. 21% of the knowledge bases focus on read/browse as main interaction but have improved their organization through the introduction of subject areas and category networks. A smaller percentage has included facilities for annotation. With increasing distance from the upper left cell, the number of knowledge bases becomes significantly smaller. Only 11% of the knowledge bases actively manage the dependencies between different elements in the knowledge base. Only 9% provide the basis for an increased level of collaboration by immediately publishing an author's contribution.

Implications

Our study shows two paths, an increase in interactive capabilities and improved organization of knowledge, along which current web-based knowledge bases can grow to become more effective in the creation, synthesis and application of knowledge. Many other important issues still need to be researched. Methods for database design, knowledge representation and hypermedia design [Isakowitz, Stohr and Balasubramanian, 1995] may provide a basis for developing the meta-structure of scientific knowledge bases. Research is needed into the types of processes and interactions which determine success or failure of a knowledge base. The potentials and pitfalls of reader-based review need to be determined. It needs to be investigated to which extend an increase in interactivity and open access leads to a decrease in the signal to noise ratio [Malhotra, Gosain and Hars, 1997]. These are only some of the many issues which need to be addressed. As all fields of science move from paper as the primary medium of storage to the Internet, there is an opportunity for the IS community to take the initiative and to provide guidance to other fields. Current initiatives such as ISWorldNet and ECWorld are not only excellent fields for experimentation and learning, they also have the potential of drawing the IS community closer together.

References

A. Barua, R. Chellappa and A.B. Whinston (1995): "Creating a collaboratory in cyberspace: Theoretical foundation and implementation." In: Journal of Organizational Computing, Vol 5, No.4, pp.417-442.

O. El Sawy et. al: "Reengineering the IS Research Process for the 21st Century: From Cottage Craft to Community Cybrariums." IOM Working Paper 1996/02, University of Southern California, Los Angeles, 1996.

D. Engelbart: Toward augmenting the human intellect and boosting our collective IQ. In: Communications of the ACM, Vol. 38, No. 8 (Aug. 1995), pp.30-33.

T. Isakowitz, T. Stohr and P. Balasubramanian (1995): "RMM a methodology for structured hypermedia design". In: Communications of the ACM, Vol. 38, No. 8 (Aug. 1995), pp.34-44.

B. Ives and S.L. Jarvenpaa (1996): "Will the internet revolutionize business education and research?" In: Sloan Management Review, Spring 1996, pp.33-41.

B. Ives and R. Zmud (1994): "ISWorldNet: Scholarly infrastructure for information systems". In: MIS Quarterly, December 1994.

A. Malhotra, S. Gosain and A. Hars: "Evolution of Virtual Communities: A Design Study Using Action Research," IOM Working Paper #IOM-97-03, Marshal School of Business, University of Southern California.