

Networked University Digital Library (NUDL): Enhancing International Collaboration in Education

Edward A. Fox
Department of Computer Science
Virginia Tech, Blacksburg, VA 24061, U.S.A.
<mailto:fox@vt.edu> <http://fox.cs.vt.edu>

1. Position Statement

In 1999 an international consortium proposed to work toward a Networked University Digital Library (NUDL), to enhance cooperation among universities in the digital library area. NSF has provided initial funding through a small award [1] to facilitate Virginia Tech collaboration with the German Dissertation Online project [2]. A larger NSF-DFG project has been recommended for funding to further develop NUDL, dealing with electronic theses and dissertations (ETDs), physics information [3], and the Open Archives initiative (OAi) [4-6]. Involving Korea in work related to NUDL could be a desirable outcome of this workshop, where the user community is primarily those attending, teaching, or benefiting (e.g., using research results) from work at colleges and universities.

1.1 Content and Collections

A number of collections of information related to university activities can fit in nicely with NUDL and US-Korea research on digital libraries. First, there are reports. For example, the Networked Computer Science Technical Reference Library (NCSTR, see www.ncstrl.org [7]) involves over a hundred institutions (including some in Korea) that have repositories of CS technical reports.

Second, there are ETDs, as collected through the Networked Digital Library of Theses and Dissertations, NDLTD, see www.ndltd.org [8-12]). Several members of NDLTD are in Korea, and some cooperation in this regard has been underway for a few years. Many students and researchers in Korea already make use of NDLTD services, and that number is expected to grow.

Third, there are educational resources. Sticking to the area of computing, it would be possible for there to be strong involvement from Korea in the ACM Journal of Educational Resources in Computing (JERIC, see <http://purl.org/net/JERIC/>), and the Computer Science Teaching Center (CSTC, see www.cstc.org [13]). The digital library supporting CSTC and JERIC could be extended to handle materials in the Korean language, for example, and such resources could be used in colleges and universities in Korea. Extending to a broader sphere, there is the NSF-funded National Science Digital Library (NSDL, where Science represents Science, Mathematics, and Technology Education) [14, 15]. If NSDL can expand into an international effort, that will be of tremendous benefit. As stated in the workshop proposal, one goal might be to

- significantly enhance the NSDL, with important content and usage from both countries.

1.2 Research and Technology

There are many related research activities that can be explored.

First, there is digital library system-building activity. The MIRAGE III system of Myaeng et al. [16, 17] has many desirable characteristics as does the MARIAN system of Fox et al. [18-21].

Cooperation could lead to:

- sharing of technology,
- making these systems interoperate, and

- supporting the use of several interesting collections by constructing a distributed environment with multiple sites running MIRAGE and MARIAN.

Second, there is research related to interoperability. Such could be set in the context of the OAI. MARIAN is already involved in studies related to OAI and interoperability, for ETDs from Germany and USA [22]. This work assumes that ontologies can be built based on repositories and their use, and that class hierarchies for nodes and links can be built from them. Then, searching and browsing can proceed at any suitable level of specificity. Gateways for federated search as well as harvesting of OAI collections can provide seamless access to distributed heterogeneous collections. Interesting research is required to ensure interoperability, and to develop filtering techniques that allow use of the contents of a variety of OAI sites for an integrated activity (e.g., searching specialized topical collections, as well as all the world's theses and dissertations).

Third, there is research related to user interaction. One line could extend prior work on visualization of search results in the ENVISION project [23-29]. The ENVISION interface has been mostly converted to Java, and so it could be tested in both countries as well as extended to help with browsing. Other related work deals with using VR environments [30, 31]. Finally, taking a rather different angle, that might be of particular appeal, we could integrate work on machine translation, ETDs, and human-computer interaction:

- Apply machine translation tools to a moderate number of complete ETDs. Translate some in English into Korean, and some in Korean into English.
- Develop special interfaces that log user interaction with translated documents, making it easy for those users to indicate areas of confusion, so that improvements in MT methods can result.
- Modify digital library system interfaces (e.g., MIRAGE III, MARIAN) to record user interactions with retrieval results that are the result of machine translation.
- Adapt the matching processes of digital library systems so that they accommodate the types of problems that might result when documents are indexed after MT processing.

Fourth, there is research related to high-performance superstorage systems. Digital libraries must be responsive, if they are to accommodate the demanding requirements of knowledge workers [32]. High-performance parallel machines with large storage capacity can increase the throughput and reduce the response time when dealing with very large inverted files [33]. VT-PetaPlex-1, a system with 2.5 terabytes of disk and 100 processors, each running Linux, is being used for research on text as well as digital video. Studies with this type of Beowulf cluster can proceed in both countries. Replication of collections in America and Asia can ensure good performance for a variety of digital library efforts while at the same time allowing experimentation with tradeoffs between storage and network bandwidth. An example of the latter relates to the use of the Local Multipoint Distribution Service (LMDS), a very high-speed wireless technology, which is being explored by the Center for Wireless Telecommunications (CWT, see www.cwt.vt.edu) at Virginia Tech.

A wide variety research studies are feasible. The ideas above may help in the selection of challenge problems and in articulating promising approaches to them.

2. Biographical Information

Edward A. Fox is Professor, Department of Computer Science, Virginia Tech and Director of the Digital Library Research Laboratory (see www.dlib.vt.edu). He has been involved in workshops related to digital libraries since 1991. He has attended conferences and workshops related to

digital libraries in many parts of the US, as well as visiting in this regard: Austria, Canada, Costa Rica, Croatia, England, Finland, Germany, Hong Kong, Japan, Mexico, Portugal, Russia, Singapore, Spain, Switzerland, and Taiwan (with other visits scheduled to Argentina, South Korea, and South Africa). His research has covered many areas related to the workshop themes, including information retrieval, computational linguistics, distributed information systems, educational technologies, multimedia education, and aspects of human-computer interaction. He directs the Networked Digital Library of Theses and Dissertations (NDLTD, www.ndltd.org), is founding co-editor-in-chief of the ACM Journal of Educational Resources in Computing (JERIC, <http://purl.org/net/JERIC/>), and hosts the Computer Science Teaching Center (CSTC, www.cstc.org), which he hopes will have many contributions from Korea. He recently has hosted one visiting professor from South Korea for a year long stay, and will soon have another visiting for a similar period – also to collaborate about digital libraries.

3. References

- [1] E. A. Fox, “NSF SGER: Core Research for the Networked University Digital Library (NUDL)”, a new NSF grant of \$79,997 for calendar years 2000, 2001; Blacksburg, VA, 2000.
- [2] K. Zimmermann, “Dissertationen Online”. Project home page. CvO University of Oldenburg: Dep. of Physics, 2000. <http://www.dissonline.org/>, http://www.educat.hu-berlin.de/diss_online/englisch/index1e.html
- [3] E. R. Hilf, “PhysDis: Physics Theses in Europe”, home page. 2000. http://elfikom.physik.uni-oldenburg.de/dissonline/PhysDis/dis_europe.html
- [4] H. Van de Sompel, “Open Archives Initiative”. WWW site. U. Ghent: OAi group, 2000. <http://www.openarchives.org>
- [5] H. Van de Sompel and C. Lagoze, “The Santa Fe Convention of the Open Archives Initiative,” *D-Lib Magazine*, vol. 6, 2000. <http://www.dlib.org/dlib/february02vandesompel-oai/02vandesompel-oai.html>
- [6] H. Van de Sompel, T. Krichel, M. L. Nelson, P. Hochstenbach, V. M. Lyapunov, K. Maly, M. Zubair, M. Kholief, X. Liu, and H. O’Connell, “The UPS Prototype: An Experimental End-User Service across E-Print Archives,” *D-Lib Magazine*, vol. 6, 2000. <http://www.dlib.org/dlib/february02vandesompel-ups/02vandesompel-ups.html>
- [7] C. Lagoze, “NCSTRL: Networked Computer Science Technical Reference Library”: Cornell University, 1999. <http://www.ncstrl.org>
- [8] E. A. Fox, R. Hall, N. A. Kipp, J. L. Eaton, G. McMillan, and P. Mather, “NDLTD: Encouraging International Collaboration in the Academy,” *Special Issue on Digital Libraries of DESIDOC Bulletin of Information Technology (DBIT)*, vol. 17, pp. 45-56, 1997. <http://www.ndltd.org/pubs/dbit.pdf>
- [9] E. Fox, “Networked Digital Library of Theses and Dissertations: An International Collaboration Promoting Scholarship,” *ICSTI Forum, Quarterly Newsletter of the International Council for Scientific and Technical Information*, vol. 26, pp. 8-9, 1997. <http://www.icsti.org/icsti/forum/fo9711.html#ndltd>
- [10] E. A. Fox, R. Hall, and N. Kipp, “NDLTD: Preparing the Next Generation of Scholars for the Information Age,” *The New Review of Information Networking (NRIN)*, vol. 3, pp. 59-76, 1997. <http://www.ndltd.org/pubs/nrin.pdf>
- [11] E. A. Fox, “Networked Digital Library of Theses and Dissertations,” in *Proceedings DLW15*. Japan: ULIS, 1999. <http://www.ndltd.org/pubs/dlw15.doc>
- [12] E. Fox, “NDLTD: Networked Digital Library of Theses and Dissertations”, 2000. <http://www.ndltd.org>
- [13] D. Knox, S. Grissom, E. A. Fox, R. Heller, and D. Watkins, “CSTC: Computer Science Teaching Center”, 2000. <http://www.cstc.org>
- [14] NSF, “National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) - Program Solicitation,” National Science Foundation NSF 00-44, 2000. <http://www.nsf.gov/cgi-bin/getpub?nsf0044>

- [15] NSF, "National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) - Program Information". National Science Foundation, 2000.
<http://www.ehr.nsf.gov/EHR/DUE/programs/nsdl/>
- [16] S. H. Myaeng, "MIRAGE: A Prototype for a Multimedia Information Retrieval and Gathering Environment," in *Proc. of International Conference on Digital Libraries and Information Services for the 21st Century*, September 10-13. Seoul, 1996.
- [17] S. H. Myaeng, "MIRAGE: A Prototype for a Multimedia Digital Library (in Korean)," *Journal Of Korea Information Science Society*, 1997.
- [18] J. Zhao, "Making Digital Libraries Flexible, Scalable, and Reliable: Reengineering the MARIAN System in JAVA," Virginia Tech Department of Computer Science, Blacksburg, VA, Master of Science Thesis, 1999. <http://scholar.lib.vt.edu/theses/available/etd-070499-204531/unrestricted/SGML-etd/>
- [19] F. Can, E. Fox, C. Snively, and R. France, "Incremental Clustering for Very Large Document Databases: Initial MARIAN Experience," *Information Systems*, vol. 84, pp. 101-114, 1995.
- [20] E. Fox, R. France, E. Sahle, A. Daoud, and B. Cline, "Development of a Modern OPAC: From REVTOLC to MARIAN," in *Proc. 16th Annual Int'l ACM SIGIR Conf. on R&D in Information Retrieval, SIGIR '93*. Pittsburgh: ACM Press, 1993, pp. 248-259.
- [21] R. K. France, "MARIAN Digital Library Information System (home page)", 2000.
<http://www.dlib.vt.edu/products/marian.html>
- [22] M. A. Gonçalves, R. K. France, E. A. Fox, E. R. Hilf, K. Zimmermann, and T. Severiens, "MARIAN: Flexible Interoperability for Federated Digital Libraries," in *Proc. ICDE'2001* (submitted). Heidelberg: IEEE, 2001.
- [23] D. Brueni, B. Cross, E. A. Fox, L. Heath, D. Hix, L. Nowell, and W. Wake, "What if there were desktop access to the Computer Science literature?," in *Proc. 21st Annual Computer Science Conference, ACM CSC '93*. Indianapolis, IN, 1993, pp. 15-22.
- [24] E. A. Fox, D. Hix, L. Nowell, D. Brueni, W. Wake, L. Heath, and D. Rao, "Users, User Interfaces, and Objects: Envision, a Digital Library," *J. American Society Information Science*, vol. 44, pp. 480-491, 1993.
- [25] E. A. Fox, N. D. Barnette, C. Shaffer, L. Heath, W. Wake, L. Nowell, J. Lee, D. Hix, and H. R. Hartson, "Progress in Interactive Learning with a Digital Library in Computer Science," in *ED-MEDIA 95, World Conference on Educational Multimedia and Hypermedia*. Graz, Austria, 1995, pp. 7-12.
- [26] L. Heath, D. Hix, L. Nowell, W. Wake, G. Averboch, and E. A. Fox, "Envision: A User-Centered Database from the Computer Science Literature," *Commun. of the ACM*, vol. 38, pp. 52-53, 1995.
- [27] L. Nowell and D. Hix, "Visualizing search results: User interface development for the project Envision database of computer science literature," in *Advances in Human Factors/Ergonomics, Proceedings of HCI International '93, 5th International Conference on Human Computer Interaction*, vol. 19B, *Human-Computer Interaction: Software and Hardware Interfaces*: Elsevier, 1993, pp. 56-61.
- [28] L. Nowell, D. Hix, R. France, L. Heath, and E. A. Fox, "Visualizing Search Results: Some Alternatives to Query-Document Similarity," in *Proceedings ACM SIGIR '96*. Zurich, Switzerland, 1996, pp. 67-75.
- [29] L. Nowell, "Graphical Encoding for Information Visualization: Using Icon Color, Shape and Size to Convey Nominal and Quantitative Data," Virginia Tech Dept. of Computer Science, Blacksburg, VA, Ph.D. Dissertation, 1997.
- [30] M. Bayraktar, C. Zhang, B. Vadapalli, N. Kipp, and E. A. Fox, "A Web Art Gallery," in *Proc. Digital Libraries '98, The Third ACM Conf. on Digital Libraries*. Pittsburgh, PA: ACM, 1998, pp. 277-278.
- [31] F. A. Das_Neves and E. A. Fox, "A study of user behavior in an immersive Virtual Environment for digital libraries," in *Proceedings of the Fifth ACM Conference on Digital Libraries: DL '00*, June 2-7, 2000, San Antonio, TX. New York: ACM Press, 2000, pp. 103-111.
- [32] R. M. Akscyn, D. McCracken, and E. Yoder, "KMS: A Distributed Hypermedia System for Managing Knowledge in Organizations," *Communications of the ACM*, 1988.
- [33] O. Sornil, "A Distributed Inverted Index for a Large-Scale, Dynamic Digital Library," Virginia Tech Computer Science, Blacksburg, Ph. D. Dissertation Draft, 2000.