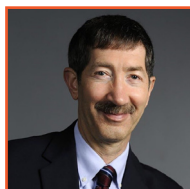


A Digital Library to Promote Use of the World's Theses and Dissertations



Edward A. Fox

Department of Computer Science, Virginia Tech, Blacksburg, VA, USA

Digital libraries (DLs) support Societies engaging in a variety of Scenarios involving various Spaces, Structures, and Streams (5S). We have been exploring how digital libraries can open up access to the millions of electronic theses and dissertations (ETDs), scattered across hundreds of institutional repositories and national libraries, that represent a key part of scholarship and research worldwide. Current technologies provide little support for long (book-length) documents, and for leveraging the useful knowledge buried within. Funding from the Institute for Museum and Library Services (USA), building on support from NSF and other sponsors, has led to an extensible DL that leverages natural language processing, deep learning, and information retrieval methods to analyze and segment ETDs, detecting important elements (e.g., chapters). With services to classify and summarize key elements, as well as identify interesting topics, and for searching, browsing, and recommending, our DL can support a variety of Societies represented by diverse personas (students, researchers, curators, administrators, and experimenters) atop a container cloud. Persona goals lead to support for their usage Scenarios, which we manifest as workflows atop services. Spaces (e.g., 2D displays) address UI/UX issues, as well as representations and algorithms involving vector (sparse and embedding) and probability spaces. Structures include metadata, genre-specific document organization, database and knowledge bases, language models, PDF page images, and XML overviews. Streams of text and images are managed with transformer and novel long-document techniques.

Biography:

Edward Fox received his BS from MIT in 1972. His M.S. and Ph.D. (1983) in Computer Science are from Cornell. He is Founder, Executive Director, and Chairman of the Board of the Networked Digital Library of Theses and Dissertations, and Professor of CS at Virginia Tech, where he directs the Digital Library Research Laboratory. He is a Fellow of both ACM (member since 1967) and IEEE, cited for his work on information retrieval and digital libraries. He has received funding from 143 grants/contracts. His publications, presentations, reports, etc. number over 1300. Over 100 graduate students have had him as advisor.