“A Digital Library to Promote Use of the World’s Theses and Dissertations”


Keynote by Edward A. Fox, Ph.D., Professor

- fox@vt.edu  http://fox.cs.vt.edu
- Dept of Computer Science (& ECE by courtesy)
- Virginia Tech, Blacksburg, VA 24061 USA
- NDLTD: Exec. Director, Chairman of the Board
Presentation Outline

• Acknowledgments
• NDLTD  (Networked Digital Library of Theses and Dissertations: ndltd.org=theses.org)
• Intelligent Information Systems, 5S
• Digital Libraries
• Piloting: IMLS, CS5604 PBL instances
• Your To Do List
Acknowledgements (selected)

• Mentors (Licklider, Kessler, Salton); IMLS, NSF, and other sponsors
Related Funded Grants


3. NSF IIS-0086227: Open Archives: Distributed services for physicists and graduate students (OAD): 2001-2004; PD Fox; German DFG PI E. Hilf


5. SOLINET (Southeastern Library Network, USA): Networked Digital Library of Theses and Dissertations: 2000. Project director Fox


Selected ETD-related VT ETDs

• William Ingram, Digital Libraries of Electronic Theses and Dissertations, 2025
• Satvik Chekuri, Scholarly Information System for Long Documents and their Elements, 2025
• Bipasha Banerjee. Improving access to ETD elements through chapter categorization and summarization, Summer 2024
• Aman Ahuja. Analyzing and Navigating Electronic Theses and Dissertations", 2023
• Dhanush Dinesh, Utilizing Docker and Kafka for Highly Scalable Bulk Processing of Electronic Theses and Dissertations (ETDs), 2023
• Javaid Manzoor, Segmenting Electronic Theses and Dissertations by Chapters, 2022
• Prashant Chandrasekar, Continuously Extensible Information Systems: Extending the 5S Framework by Integrating UX and Workflows, 2021
• Sampanna Kahu, Figure extraction from scanned electronic theses and dissertations, 2020
• Palakh Mignonne Jude, Increasing Accessibility of Electronic Theses and Dissertations (ETDs) Through Chapter-level Classification, 2020
• Sung Hee Park, Discipline-Independent Text Information Extraction from Heterogeneous Styled References Using Knowledge from the Web, 2013
• W. Ryan Richardson, Using Concept Maps as a Tool for Cross-Language Relevance Determination, 2007
• Douglas Gorton, Practical Digital Library Generation into DSpace with the 5S Framework, 2007
• Hussein Suleman, Open Digital Libraries, 2002
NDLTD: Mission

The Networked Digital Library of Theses and Dissertations (NDLTD) is an international organization dedicated to promoting the adoption, creation, use, dissemination, and preservation of electronic theses and dissertations (ETDs). We support electronic publishing and open access to scholarship in order to enhance the sharing of knowledge worldwide. Our website includes resources for university administrators, librarians, faculty, students, and the general public. Topics include how to find, create, and preserve ETDs; how to set up an ETD program; legal and technical questions; and the latest news and research in the ETD community.
J-ETD.org, j-etd@ndltd.org
Journal of Electronic Theses and Dissertations

• Open-access launch 1/1/2021. Please support!
• Managing Editor: Charles J. Greenberg
• Executive Editor: Edward A. Fox; Associate Editors: Suzanne Lorraine (Suzie) Allard (USA), Ramesh C. Gaur (India), Charles J. Greenberg (USA), Libio Huaroto (Peru), William A. Ingram (USA), Ana Sofia de Sousa Machado Mota (Portugal), Prashant Pandey (Australia), Ana Pavani (Brazil), Joachim Schöpfel (France), Janette Wright (UAE)
Global ETD Search

Search the 6,480,478 electronic theses and dissertations contained in the NDLTD archive:

Type something to start searching...

advanced search tips ➔ how to contribute records ➔
Scenarios of Future Use of ETD DLs

1. Open problem -> plan for research
2. Problem -> list of references, related ETDs
3. Bibliography -> clusters -> lit. review chapter
4. Course (e.g., seminar) units based on ETDs
5. Final defense -> told missing cites of related ETDs
6. Promotion: impact of candidate’s students’ ETDs
7. Research trends: classification, topic modeling
8. Analysis & Assessment -> logs -> use by:
   - Local grad students, faculty, undergrads
   - Graduate School, Registrar, Research Division
Scenarios of Future Use:
Example: Open problem -> plan for research

1. Student volunteers to pilot test the new DL
2. Goal: find problem to solve
3. Explains her interest and background
4. Receives extracts from related ETDs:
   – open problems, planned future work
5. Selects top 5
6. Receives related ETD list, with chapter summaries
7. Fetches and studies top 2 ETDs from the list
8. Meets advisor to devise research plan
Intelligent Information Systems

- Digital libraries, repositories
- Search engines, recommenders
- Chatbots
- Bots
- Smart homes, smart cities
- Robots
- (Semi)autonomous vehicles, UAVs
  - Land: Cars, Trucks
  - Air: Planes, Drones
  - Sea: Boats, UUVs
Theoretical Foundations for Digital Libraries

The 5S (Societies, Scenarios, Spaces, Structures, Streams) Approach

Edward A. Fox
Marcos André Gonçalves
Rao Shen

Key Issues in Digital Libraries
Integration and Evaluation

Rao Shen
Marcos André Gonçalves
Edward A. Fox

Synthesis Lectures on Information Concepts, Retrieval, and Services
Gary Marchionini, Series Editor
Digital Libraries (DLs) have evolved since their launch in 1991 into an important type of information system, with digital and print formats. For more information visit www.morganclaypool.com

Digital Library Technologies
Complex Objects, Annotation, Ontologies, Classification, Extraction, and Security

Edward A. Fox
Ricardo da Silva Torres

Digital Libraries Applications
CBIR, Education, Social Networks, eScience/Simulation, and GIS

Edward A. Fox
Jonathan P. Leidig

Synthesis Lectures on Information Concepts, Retrieval, and Services
Gary Marchionini, Series Editor
Informal 5S & DL Definitions

DLs are complex systems that

• help satisfy info needs of users (societies)
• provide info services (scenarios)
• organize info in usable ways (structures)
• present info in usable ways (spaces)
• communicate info with users (streams)
Digital Libraries: Content

Content Types

- Text Documents: Articles, Reports, Books
- Video Audio: Speech, Music
- Geographic Information: (Aerial) Photos
- Software, Programs: Models Simulations
- Bio Information: Genome Human, animal, plant
- Images and Graphics: 2D, 3D, VR, CAT
### Supporting Services across the Lifecycle

<table>
<thead>
<tr>
<th>Infrastructure Services</th>
<th>Repository-Building</th>
<th>Add Value</th>
<th>Information Satisfaction Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creational</td>
<td>Conserving</td>
<td>Annotating</td>
<td>Browsing</td>
</tr>
<tr>
<td></td>
<td>Converting</td>
<td>→ Classifying</td>
<td>Collaborating</td>
</tr>
<tr>
<td></td>
<td>Copying/Replicating</td>
<td>→ Clustering</td>
<td>Customizing</td>
</tr>
<tr>
<td></td>
<td>Emulating</td>
<td>→ Evaluating</td>
<td>Filtering</td>
</tr>
<tr>
<td></td>
<td>Renewing</td>
<td>→ Extracting</td>
<td>Providing access</td>
</tr>
<tr>
<td></td>
<td>Translating (format)</td>
<td>→ Indexing</td>
<td>Recommending</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measuring</td>
<td>Requesting</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publicizing</td>
<td>Searching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rating</td>
<td>Translating (format)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reviewing (peer)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Surveying</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Visualizing</td>
<td></td>
</tr>
</tbody>
</table>

- **Acquiring**
- **Cataloging**
- **Crawling (focused)**
- **Describing**
- **Digitizing**
- **Federating**
- **Harvesting**
- **Purchasing**
- **Submitting**

- **Conserving**
- **Converting**
- **Copying/Replicating**
- **Emulating**
- **Renewing**
- **Translating (format)**
# Quality Dimensions

<table>
<thead>
<tr>
<th>DL Concept</th>
<th>Dimensions of Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital object</td>
<td>Accessibility</td>
</tr>
<tr>
<td></td>
<td>Pertinence</td>
</tr>
<tr>
<td></td>
<td>Preservability</td>
</tr>
<tr>
<td></td>
<td>Relevance</td>
</tr>
<tr>
<td></td>
<td>Similarity</td>
</tr>
<tr>
<td></td>
<td>Significance</td>
</tr>
<tr>
<td></td>
<td>Timeliness</td>
</tr>
<tr>
<td>Metadata specification</td>
<td>Accuracy</td>
</tr>
<tr>
<td></td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Conformance</td>
</tr>
<tr>
<td>Collection</td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Impact Factor</td>
</tr>
<tr>
<td>Catalog</td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Consistency</td>
</tr>
<tr>
<td>Repository</td>
<td>Completeness</td>
</tr>
<tr>
<td></td>
<td>Consistency</td>
</tr>
<tr>
<td>Services</td>
<td>Composability</td>
</tr>
<tr>
<td></td>
<td>Efficiency</td>
</tr>
<tr>
<td></td>
<td>Effectiveness</td>
</tr>
<tr>
<td></td>
<td>Extensibility</td>
</tr>
<tr>
<td></td>
<td>Reusability</td>
</tr>
<tr>
<td></td>
<td>Reliability</td>
</tr>
</tbody>
</table>
Scenarios of Future Use / Building DLs

1. UX: Customer discovery: subject-matter experts

2. UX: Validated list of:
   - Jobs-to-be-done, tasks, sub-tasks, goals, sub-goals

3. Personas
   1. Curators
   2. Experimenters
   3. Researchers (students, faculty, …)

4. DL software developer: knowledge graph mapping:
   - Goals, Sub-goals, Tasks, Sub-tasks
   - Workflows of services: Existing, Desired

5. Operations (Docker, Airflow; DevOps with CI/CD)

(Doctoral work of Prashant Chandrasekar)
Prashant Chandrasekar’s DL Architecture

Digital Library of SME System Req. Descriptions

Catalog of system req. descriptions
Case  UX ARTIFACTS

<table>
<thead>
<tr>
<th>Case1</th>
<th>Service x</th>
<th>Service y</th>
<th>Service z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-- -----</td>
<td>---- -----</td>
<td>---- ----</td>
</tr>
<tr>
<td>Case2</td>
<td></td>
<td>--- ----</td>
<td>---- ----</td>
</tr>
<tr>
<td>Case3</td>
<td></td>
<td>--- ---</td>
<td>---- ----</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UXR

Capture User Requirements

SME

Conduct UX Research

Results

Query

UXR

PRODUCE

UXR

Developer

Analyst

WORKFLOW MANAGER

SERVICES REGISTRY

GRAPH INTERFACE

CS1 - Information System

CS2 - Information System

http://hdl.handle.net/10919/103815, https://doi.org/10.2478/dim-2021-0003
Workflow-defining Goal Decomposition

User -> G1 -> Workflow 1 -> Service 1A
 -> G2 -> Workflow 2 -> Service 1B
 -> ... -> Gn -> Workflow n -> Service 1M

Goals -> System Solution

Opening Graduate Research
IMLS; 2019-2023; PI: William Ingram

- Activities
  - Collecting: 500,000+ from USA
    - Large universities, HBCUs, HSIs + Arabic corpus
  - Analyzing: parsing / detecting (texts, images)
  - Extracting: tables, figures, equations, references…
  - Scanned ETDs -> improved metadata
  - Classification, Topic Modeling -> Browsing
  - Segmenting: chapters -> Chapter summaries

- Results: New methods & technologies, pilot system (search, browse, recommend, viz)
CS5604 SMEs

• Aman Ahuja: topic modeling, object detection/document parsing (http://hdl.handle.net/10919/115817)
• Bipasha Banerjee, Sara Ahmadi: segmentation, language models, transformers, classification, summarization
• Prashant Chadrasekar, Dhanush Dinesh: integration, workflows, extensibility, DevOps
• Satvik Chekuri: search, recommendation
• Sung Hee Park, Bill Ingram: database, files
CS5604 Fall 2023 Teams
Inference is accomplished via the best performing model trained by Samppana and others.
microscope observations of live bundles, and studies of kinocilium height (Forstilla and Peterson, 2000), were used to define heights of stereocilia and the kinocilium. The height data was obtained from various bundles that were different from, but similar to, the original bundle. In this manner a realistic representation of a bundle was assembled. The computer-generated graphic for each bundle in Figure 2.2 is based on the model input into broad, and shows the deformed state of the bundle. Although it may not be clear from Figure 2.2, cells 1, 2, 4, and 5 are “loose-packed”, and cells 3 and 6 are “tight-packed”, as defined in Chapter 1.

![Figure 2.2: Six utricle cells – electron micrograph and 3-D rendering](image)

Obviously, many approximations were made in modeling the cell bundles. Stereocilia diameters and spacing were approximated as constant throughout a given bundle. Perfect hexagonal layouts do not exist in biological bundles, but they are much easier to model. Cilia heights were based on similar bundles, and were approximated so as to linearly decrease in height along the E-I axis. Tapering at the base of stereocilia was

![Cropped images](image)
CHAPTER 1: INTRODUCTION AND BACKGROUND

Balls body on th vestibular e the brain wi

The labyrinth p semicirculad detect ori

The semicircular canals detect orientation of the head inside the skull. Stimuli from these structures

CHAPTER 2: METHODS AND MATERIALS

CHAPTER 3: THREE-DIMENSIONAL BUNDLE MECHANICS

CHAPTER 4: ION GATES

CHAPTER 5: CONCLUSIONS AND FUTURE WORK

If one were to try and sum up the conclusions obtained from this research into one statement, perhaps the best summary would be to say that bundles are mechanically complex, and all details are important in accurately modeling them.

Equally important in accurate modeling are the material properties, such as elastic moduli and shear moduli. Of particular importance is the tan modulus, which is important both in affecting overall bundle stiffness, as well as influencing the behavior of the theoretical ion gate.

All of these factors are of extreme importance in static response of bundles! The complexity of dynamic response is surely even more challenging and dependent on these (and other) factors.

The implications of these conclusions are three-fold. First, and unsurprisingly, better information about bundles is needed to improve modeling efforts. The material properties of tip and lateral links need to be known more precisely. Unfortunately, it is currently impossible to measure these properties directly; testing values in a model is presently the best possible way to determine these values. Geometric properties of individual bundles being modeled need to be measured more exactly. The details are important; rough estimates are insufficient. The importance of the stereocilia/kinematically height ratio suggests that accurate height data is particularly crucial, but cilia diameters, taper ratios, and other values are also vital. Second, modeling needs to be as precise as possible. Lump parameter models and simple 2-D row models are not sufficient. They
Analysis of Vestibular Hair Cell Bundle Mechanics Using Finite Element Modeling

Joseph Allan Silber

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Engineering Mechanics

J. Wallace Grant, Chair
Ellengene H. Peterson
John R. Cotton

November 18, 2002
Blacksburg, Virginia

Keywords: Vestibular System, Hair Cell, Finite Element
Copyright 2002, Joseph A. Silber

Joseph Allan Silber

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Engineering Mechanics

J. Wallace Grant, Chair
Ellengene H. Peterson
John R. Cotton

November 18, 2002
Blacksburg, Virginia

Keywords: Vestibular System, Hair Cell, Finite Element
Copyright 2002, Joseph A. Silber
Example: ETD-Topics (Architecture)

Aman Ahuja, William A. Ingram, Chenyu Mao, Chongyu He, Jianchi Wei and Edward A. Fox. Analyzing and Navigating ETDs Using Topic Models. ETD 2022 conference, Novi Sad, Serbia, September 7-9, 2022
Your To Do List

• Use, and Encourage Others to Use:
  http://search.ndltd.org/

• Support NDLTD(.org) and Worldwide ETD Requirements

• Use Digital Libraries and Repositories

• Improve Information Systems using 5S & AI

• Apply PBL in Courses (or Help as a SME)
Questions?
Discussion?

Thank You!

fox@vt.edu
fox@ndltd.org
j-etd@ndltd.org
https://fox.cs.vt.edu/cv.htm