V-CSIT2023: Feb. 24-25

“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)

• Digital Libraries
  – 5S, Services, Scenarios
  – Building (extensible)

• Piloting: IMLS, CS5604 PBL instances

• Summary
Acknowledgements

• NDLTD, and Worldwide ETD Initiative

• Mentors (Licklider, Kessler, Salton); IMLS, NSF, . . .

• Virginia Tech, CS, Digital Library Research Laboratory (DLRL)

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


1. Sampanna Yashwant Kahu, Figure extraction from scanned electronic theses and dissertations, 2020, [http://hdl.handle.net/10919/100113](http://hdl.handle.net/10919/100113)

2. Palakh Mignonette Jude, Increasing Accessibility of Electronic Theses and Dissertations (ETDs) Through Chapter-level Classification, 2020, [http://hdl.handle.net/10919/99294](http://hdl.handle.net/10919/99294)

3. Sung Hee Park, Discipline-Independent Text Information Extraction from Heterogeneous Styled References Using Knowledge from the Web, 2013, [http://hdl.handle.net/10919/52860](http://hdl.handle.net/10919/52860)


ETD-related Class Projects

1. Kaushal, Kulendra Kumar; Kulkarni, Rutwik; Sumant, Aarohi; Wang, Chaoran; Yuan, Chenhan; Yuan, Liling. Collection Management of Electronic Theses and Dissertations (CME) CS5604 Fall 2019 (Virginia Tech, 2019-12-23); http://hdl.handle.net/10919/96484

2. Aromando, John; Banerjee, Bipasha; Ingram, William A.; Jude, Palakh; Kahu, Sampanna. Classification and extraction of information from ETD documents (Virginia Tech, 2020-01-30); http://hdl.handle.net/10919/96645

3. Alotaibi, Fatimah; Abdelrahman, Eman. Otrouha: Automatic Classification of Arabic ETDs (Virginia Tech, 2020-01-23); http://hdl.handle.net/10919/96571

4. Ma, Yufeng; Jiang, Tingting; Shrestha, Chandani. ETDseer Concept Paper (Virginia Tech, 2017-05-03); http://hdl.handle.net/10919/77868
ETD-related Summarization Class Projects


- Fall 2018 CS4984/5984 (Big Data Text Summarization) projects by teams 10, 16, 17: http://hdl.handle.net/10919/86418, http://hdl.handle.net/10919/86406, http://hdl.handle.net/10919/86420
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.
New Journal: 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)
search.ndltd.org

Global ETD Search

Search the 6,357,361 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
Digital Libraries: Content
5S Layers

- Societies
- Scenarios
- Spaces
- Structures
- Streams
While the 5S framework may be used to describe many types of information systems, and is likely to have even broader utility and appeal, we focus here on digital libraries. Our view of digital libraries is broad, so further generalization should be straightforward.
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
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)
# Supporting Services across the Lifecycle

<table>
<thead>
<tr>
<th>Infrastructure Services</th>
<th>Add Value</th>
<th>Information Satisfaction Services</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repository-Building</strong></td>
<td><strong>Add</strong></td>
<td><strong>Add</strong></td>
</tr>
<tr>
<td>Creational</td>
<td><strong>Value</strong></td>
<td></td>
</tr>
<tr>
<td>Acquiring</td>
<td>Annotating</td>
<td>Browsing</td>
</tr>
<tr>
<td>Cataloging</td>
<td>Classifying</td>
<td>Collaborating</td>
</tr>
<tr>
<td>Crawling (focused)</td>
<td>Clustering</td>
<td>Customizing</td>
</tr>
<tr>
<td>Describing</td>
<td>Evaluating</td>
<td>Filtering</td>
</tr>
<tr>
<td>Digitizing</td>
<td>Extracting</td>
<td>Providing access</td>
</tr>
<tr>
<td>Federating</td>
<td>Indexing</td>
<td>Recommending</td>
</tr>
<tr>
<td>Harvesting</td>
<td>Measuring</td>
<td>Requesting</td>
</tr>
<tr>
<td>Purchasing</td>
<td>Publicizing</td>
<td>Searching</td>
</tr>
<tr>
<td>Submitting</td>
<td>Rating</td>
<td>Visualizing</td>
</tr>
</tbody>
</table>

| Preservational           |                                       |
| -------------------------|                                       |
| Conserving               | Annotating                            |
| Converting               | Classifying                           |
| Copying/Replicating      | Clustering                            |
| Emulating                | Evaluating                            |
| Renewing                 | Extracting                            |
| Translating (format)     | Indexing                              |
|                          | Measuring                             |
|                          | Publicizing                           |
|                          | Rating                                |
|                          | Reviewing (peer)                      |
|                          | Surveying                             |
|                          | Translating                           |
|                          | (language)                            |

**Add Value**

- Annotating
- Classifying
- Clustering
- Evaluating
- Extracting
- Indexing
- Measuring
- Publicizing
- Rating
- Reviewing (peer)
- Surveying
- Translating
- (language)
## 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

Service x  Service y  Service z

Catalog of system req. descriptions
Case  UX ARTIFACTS

<table>
<thead>
<tr>
<th>Case1</th>
<th>-----</th>
<th>----</th>
<th>----</th>
</tr>
</thead>
<tbody>
<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>

ARTIFACT MANAGER

UXR

Conduct UX Research

SME

Capture User Requirements

Query

Results

PRODUCE

UXR
Developer
Analyst

GRAPH INTERFACE

WORKFLOW MANAGER
SERVICES REGISTRY

CS1 - Information System
CS2 - Information System

Case 1
Case 2
Case 3

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

User

G1

G2

Gn

Goals

Workflow 1

Workflow 2

Workflow n

System Solution

Service 1A

Service 1B

Service 1M

Service 2A

Service 2B

Service 2M

Service nA

Service nB

Service nM
Workflow definition process

- Collaboration between users (SMEs), UX researchers, and developers
- Step 1: Extract goals
- Step 2: Identify tasks
  - Breakdown of tasks determines workflow steps
- Step 3: Model workflows
  - Identify functions/services to support each task
- Step 4: Represent goal-workflow knowledge graph

Methodology to map user needs to workflows
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)
CASE STUDY: CS5604 (Information Retrieval)

Problem-based learning:
Build novel intelligent system

Approach: co-designed architecture -> CI/CD
https://vtechworks.lib.vt.edu/handle/10919/19081
Inference is accomplished via the best performing model trained by Samppanna and others.
microscope observations of live bundles, and studies of kinocilium height (Fontilla 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 spread, 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.

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 E1 axis. Tapering at the base of stereocilia was

**Cropped images**
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

Extracted Text
CHAPTER 1: INTRODUCTION AND BACKGROUND

Balb body on th vestibular e the brain wi the larvae p semicircular detect orient of the head inside these stimuli int more detail

The Semi

The detect an acceleration
Rec epithelium l above the a up into the acts on and This mech system acce

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.

Accurate knowledge of the geometry of a bundle is crucial. Cilia diameters, numbers of and locations of cilia, and cilia heights all have significant effects on bundles stiffness, as elaborated on in chapter 3. Although not discussed in detail, even factors such as stereocilia base tapering, and tip link diameters can noticeably influence stiffness.

Certainly, modeling a bundle as a simple row or column neglects a significant amount of information and can give incorrect results.

Equally important in accurate modeling are the material properties, such as elastic moduli and shear moduli. Of particular importance is the tip-link elastic modulus, which is important both in affecting overall bundle stiffness, as well as influencing the behavior of the ionized ion-gated channels.

All of these factors are of extreme importance just in static response of bundles! The complexities of dynamic response are surely even more challenging and dependent on (or 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 stereocilial/kinesin 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. Lumped parameter models and simple 2-D row models are not sufficient. They
CS5604 ETD Team: Classification

Extracted Text

Dublin Core XML

Classifier

Subject: ["Biomedical Engineering"]

Labels for ETD
CS5604 Fall 2022 SMEs

- 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
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
Summary

• Acknowledgments

• NDLTD (Networked Digital Library of Theses and Dissertations: ndltd.org=theses.org)

• Digital Libraries
  – 5S, Services, Scenarios
  – Building (extensible)

• Piloting: IMLS, CS5604 PBL instances

• Summary
Questions?
Discussion?

Thank You!

fox@vt.edu
fox@ndltd.org
j-etd@ndltd.org