The Journal of Electronic Theses and Dissertations

Volume 1 Article 5

2021

Building and Using Digital Libraries for ETDs

Edward A. Fox
Virginia Tech, fox@vt.edu

Building and Using Digital Libraries for Electronic Theses and Dissertations

Edward A. Fox

Department of Computer Science

Virginia Tech

Blacksburg

VA 24061 USA

ORCiD: 0000-0003-1447-6870

Email: fox@vt.edu

Abstract

Despite the high value of electronic theses and dissertations (ETDs), the global collection has seen limited use. To extend such use, a new approach to building digital libraries (DLs) is needed. Fortunately, in recent decades a vast amount of "gray literature" has become available through a diverse set of institutional repositories, as well as regional and national libraries and archives. ETDs are among the most numerous of the works in those collections, and are often freely available in keeping with the open access movement, but such access is limited by the services of supporting information systems. As is explained through a set of scenarios, ETDs can better meet the needs of diverse stakeholders if customer discovery methods are used to identify personas and user roles, as well as their goals and tasks. Fortunately, DLs, with a rich collection of services, as well as newer, more advanced ones, can be organized so that those services, and expanded workflows building on them, can be adapted to meet personalized goals, as well as traditional ones, such as discovery and exploration.

Index terms: digital preservation, distributed information systems, electronic publishing, image classification, indexing, information retrieval, information systems, libraries, natural language processing, open access, optical character recognition software, recommender systems, search engines, software engineering, taxonomy, text processing, text mining, user-centered design, user experience

Relevant conference themes: institutional repositories and platforms, open scholarship, open access, digital archives, technical processes

Building and Using Digital Libraries for Electronic Theses and Dissertations

I. Introduction

Electronic theses and dissertations (ETDs) are important academic research artifacts that mark significant milestones in graduate students' careers. Their life cycle includes research, authoring, review, approval, archiving, dissemination, and long-term preservation, which involve the university: students, faculty, administration (research division and graduate school), library, and archives (Halbert, 2012). As the closely allied field of digital libraries (DLs) has developed alongside the shift from paper theses/dissertations to ETDs, this paper explores the connection between DLs and ETDs, and aims to guide the improvement of DLs that support ETDs, so they can better aid each stakeholder community involved across that life cycle.

The remainder of the paper is organized as follows. Section II highlights the important role of ETDs. Section III presents an overview of the history of work with ETDs. Subsection III-A summarizes broader work with "gray literature", while the next subsections describe open access and the related Open Access Initiative (OAI) (Lagoze, Van de Sompel, & Warner, 1998–2017), as well as more general work on repositories. Section IV provides pointers for user experience (UX) activities that help ensure there is appropriate support for each ETD stakeholder user community. In particular, Subsection IV-A describes customer discovery efforts associated with UX research (Interaction Design Foundation, 2020). Reflecting the needs and wishes of such users, Section V suggests some possible broader uses of ETDs, regarding the realization of which, Section VI applies the context of work with DLs, including Subsection VI-A on services and Subsection VI-B on related research advances. Section VII concludes this paper with a vision of how future DLs can be applied to ETDs to further improve and broaden their use.

II. ETD Importance

Millions of ETDs are available worldwide, covering every field and discipline, with works (or at least abstracts) available in scores of languages. This body of literature is unique in its breadth and scope, with origins in Germany in the 1700s, and is likely to continue as long as graduate education exists (Barton, 2005). While studying ETDs can be particularly valuable to students aiming to learn about related studies, and later to model their good qualities when they prepare their own ETDs, they also have a broader value. A shortlist of helpful content as well as the advantages of these ETDs includes algorithms and code snippets; comprehensive nature, reflected in the length, detail, and inclusion of negative as well as positive results; datasets and descriptions thereof; definitions, glossaries, and terminology lists; details of experimentation; details of methodologies; discussion of open problems and possible future work; documentation of studies approved by institutional review boards; extensive bibliographies along with well-organized specialized literature reviews; figures and visualizations that clarify discoveries; illustrations and discussions of art, architecture, and archaeology works; lemmas, proofs, and theorems; lists and descriptions of research questions along with findings that address them; motivation for undertaking a particular line of research; conceptualization of research hypotheses along with results regarding their falsification; supplements with almost any type of multimedia, hypertext, or hypermedia content; and tables of results.

III. History and Current Situation

Work on ETDs emerged from a November 1987 workshop hosted by UMI in Ann Arbor, Michigan, U.S.A. aiming to connect dissertations with the SGML-oriented Electronic Manuscript Project (Mehringer, 1986). The NDLTD (NDLTD, 2020) began as a national effort in the United States receiving funding from the U.S. Department of Education, among others (Fox et al., 1996). In 1997, it broadened into an international consortium, with the same acronym (Fox et al., 1997). The team at Virginia Tech, and others involved in the NDLTD traveled extensively to broaden support (Fox, McMillan, & Eaton, 1999). With a clear mission (Suleman et al., 2001a), the NDLTD served graduate student authors (Phanouriou, Kipp, Sornil, Mather, & Fox, 1999) and undertook an initial set of services and research (Suleman et al., 2001b). It called for changes that would benefit universities (Fox et al., 2002). A key contribution was the construction of a Union Archive (Suleman, 2020b; Suleman & Fox, 2002b), which supports a global search system (Suleman, 2020a). As part of its role as an agent of change, the NDLTD holds an annual international conference; ETD 2020 is the 23rd in the series.

A. Gray Literature

Work involving ETDs can be viewed in a broader context. Communication, an essential human activity, can be categorized according to multiple schemes, such as informal vs. formal, synchronous vs. asynchronous, in-person vs. remote, or popular vs. scholarly. Scholarly communication is best known through the work of publishers, but author/institutional preferences and economic considerations have led to an enormous body of useful, free, "gray literature" (Farace, 1992–2020). This label is often attached to theses and dissertations. Another thread of work in the gray literature entails technical reports (TRs), which used to be quite popular in computer science (CS) departments. The shift from paper to electronic distribution of TRs worked especially well in the CS environment (French, Fox, Maly, & Selman, 1994, 1995) and spread widely to other disciplines. For example, arXiv has reports and other scholarly articles not only in CS but also from at least seven other fields (arXiv, 2020). With over 1.7 million works as of October 2020, this freely accessible archive, despite not necessarily benefiting from peer reviews, does benefit from the fact that many authors submit their works immediately after first documenting new research and can make revisions based on early feedback from the community, leading to revised submissions to arXiv and/or submissions to workshops, conferences, or journals.

B. Open Access, Repositories, and OAI

One advantage of most gray literature is that it promotes open access (Suber, 2012), which is particularly appealing to researchers, students, and academic institutions, especially given the escalating costs of purchases and subscriptions to publisher services.

Many scholars are vocal advocates of open access both to ETDs and other genres, including Peter Murray-Rust (Wikipedia, 2020), who occasionally attends ETD-related events. Others, such as Stevan Harnad, have advocated extending open access to journals (Harnad, 2015). A recent work has highlighted not only the theory but also the history and current practice of open access (Pinfield, Wakeling, Bawden, & Robinson, 2020).

C. Repositories

Such access is often supported by software systems such as DSpace (DuraSpace, 2020). Instances or variants of DSpace, which include Virginia Tech's VTechWorks system, often support a local institutional repository, sometimes called an open repository. However, not much attention has been given to the application of the latest information retrieval methods to such repositories, which is one of the motivations for our current research.

A seminal contribution to the evolution of repositories and DLs is the Kahn–Wilensky framework (Kahn & Wilensky, 1995), which formalized terms such as "repository"; "digital object"; "handle" (i.e., an identifier for a digital object, now more commonly called digital object identifier or DOI), along with a supporting "handle server"; and "repository access protocol." It states that some digital objects are elemental, while others are composite, as they have other digital objects as constituents. Digital objects may have associated metadata that can be processed separately as metadata objects.

In connection with the open-access movement, and building on the Kahn–Wilensky framework, an initiative to automate the open sharing of metadata was conceived, which became known as the Open Archives Initiative (OAI) (Lagoze et al., 1998–2017; Suleman

& Fox, 2002a). This encouraged the construction and use of repositories, each of which supported a repository access protocol called the OAI Protocol for Metadata Harvesting (OAI-PMH). The NDLTD was an early supporter and adopter of OAI-PMH, which allows its Union Archive (Suleman, 2020b; Suleman & Fox, 2002b) to collect metadata about ETDs from universities and national libraries, as well as other aggregators worldwide.

IV. Personas and User Roles

To address the needs of users in various ETD stakeholder communities, it is essential to identify user types (i.e., personas), describe their characteristics, and ensure they are engaged in a suitable UX (Hartson & Pyla, 2012). Besides documenting each persona, the UX process clarifies their roles, goals, subgoals, tasks, and subtasks. The aim is for them to accomplish their goals and fulfill their roles. Many DL users are familiar with role-related login screens that have them indicate whether they are an author, reviewer, program committee member, editor, or a publisher employee. It remains to be seen whether these roles will be extended to include library- and archive-related functions, such as that of a curator, cataloger, data scientist, teacher, student, and collaborator.

A. Customer Discovery

In the world of businesses and start-ups, the UX process is often preceded or aided by the customer discovery process, preferably executed by interviewing at least 100 people in the relevant ecosystem. This is more effective than surveys or focus groups at the early stages of devising an appropriate aid for users. It is best accomplished by one person (preferably a CEO or a senior person with product/service responsibility) who is often accompanied by another who takes notes, engaging in a directed but open-ended interview, usually 20–60 minutes long (Constable, 2014). The interview should lead to a description of the work, activities, and tasks carried out by users, as well as other "jobs to be done" to better fulfill their role (Wunker, Wattman, & Farber, 2016). Proceeding this way makes it more likely that the resulting products and services address users' actual needs and desires

(Ulwick, 2005). The assumption is that it is better to test early (and, if necessary, fail softly, after minimal investment of time and resources) any hypotheses regarding personas, jobs, needs, wants, goals, and tasks. Accordingly, these should be verified through discussions first and then using prototypes and pilot testing (Constable, 2018). As work progresses, with the aim of delighting users, it is essential that UX considerations, as well as design-thinking, guide change and innovation processes (Brown, 2019).

V. Possibilities

Scenario-based design advocates (Carroll, 1999) asserted that a representative set of scenarios, ideally obtained through customer discovery or similar efforts, may offer a clearer understanding of possibilities associated with building and using DLs for ETDs. In this regard, the following may help:

1. Researcher P has trained Researcher B in a new approach to building DLs. This involves B interviewing subject-matter experts (SMEs) familiar with ETDs and their use. The SMEs explain their ETD-related jobs, which leads to B listing the associated tasks. B asks for more details about the reasons for those tasks, identifies SME goals and subgoals, and validates the lists of tasks and goals with multiple SMEs, developing a set of the most frequent tasks and goals with the strongest support. Next, B finds a senior DL software developer, D, and meets with D along with one of the more articulate and engaged SMEs, A, to go over the lists of tasks and goals. B, D, and A then meet and consider the lists as well as the existing services in the current DL (that is used to support ETD work) and develop a list of additional required services, along with a mapping of the goals and tasks, to the service workflows. After a review by the DL product manager, several artifacts resulting from following P's methodology are built into a new version of the ETD DL.

2. Student S has volunteered to test the new DL. S recently decided to work toward her Ph.D. and is looking for a good problem to solve. S uses a DL service called "find

open problem.” After explaining some of her interests and her background, S receives a list of problems that others have identified in their ETDs as open problems or planned future work. S selects five problems of interest and asks the DL to extend the list with the dates of the source ETDs along with a one-page summary of the motivation given earlier in those ETDs. After reviewing the five entries, S narrows them down to two, fetches the related ETDs, studies them, and then meets with her advisor to construct a research plan.

3. Student L has identified a problem for his M.S. thesis and wants to learn more about prior research so he can prepare a bibliography and a literature review chapter. L prefers the Overleaf service to prepare documents using the LaTeX approach. L uses the new DL to describe his problem, and in turn is given a BibTeX file that can be loaded into Overleaf. After reviewing the entries, L picks a set of 10 that seem particularly interesting. L asks the DL to check if these are open access or if there are open access–related works that are frequently co-cited with the entries in the list. L then starts his literature review with the identified open-access works, which fortunately include a recent ETD that discusses one aspect of the identified problem.

4. Student W has completed a study of literature relevant to her doctoral research. Her bibliography is in good shape, full of details that could save time for those interested in extending her work. She is preparing a chapter to serve as a literature review, which she hopes to extend later into a survey article. She asks the new DL for assistance. It then clusters the works in her bibliography, and after some interaction, the DL and W produce a skeletal version of the chapter, with sections and subsections, along with cite entries in each subsection to the works in the bibliography that fit. The DL also returns excerpts from the literature review chapters in completed ETDs that correspond to each cite so she can see what others have said about which work, which she will then cite and explain. W is delighted

with all of the help provided with this important chapter and is confident that it will be useful for others working in this specialty area.

5. Faculty member F is running an advanced graduate seminar course in his area of specialization. He asks the DL to help generate a reading list of ETDs that students could read and discuss in the seminar. He also asks the DL to prepare a supplemental bibliography of highly cited open-access publications that are included in the bibliographies of at least two of the ETDs being covered.

6. Senior faculty member G is on the examining committee of a graduate student planning a final defense. G has just received a draft of the related ETD. She submits it to the DL and asks for related ETDs that were released in the previous year. She finds that the draft ETD cites none of these works and immediately contacts the student involved to ensure that the final defense and related ETD will clarify how the work being defended reflects advancements beyond those recent works.

7. Associate head of University U's Research Division R has been asked to review the dossier of a faculty member, M, who is up for promotion. She takes the list of graduate students for whom M has served as co-advisor and gives it to the new DL. As she has other resources to help her assess M's impact in his profession, R focuses on gauging the influence of the ETDs of students who were advised by M. Much to her delight, R finds hundreds of downloads from U's institutional repository of the ETDs of those students.

8. Librarian L is assessing the use of the local institutional repository's ETD collection. She asks the new DL to mine the download logs. She is delighted to find that the ETDs are used not only by local graduate students but also by many in the faculty and a large number of undergraduate students, along with a significant number of external users from highly rated and diverse universities worldwiide.

VI. Digital Libraries

Digital libraries (DLs) emerged in part through a series of workshops in the early 1990s (Fox, 1993). Many overviews have been provided regarding early efforts and conceptualizations (Fox, Akscyn, Furuta, & Leggett, 1995; Fox & Marchionini, 1998, 2001; Fox & Urs, 2002). As in other CS fields, such as programming languages and database management, where theory has revolutionized technology, theory can also provide a foundation for DL innovation (Candela, Castelli, Fox, & Ioannidis, 2010). Precise definitions of the five key aspects of DL—streams, structures, spaces, scenarios, and societies (5S)—provide a firm theoretical basis (Fox, Goncalves, & Shen, 2012; Gonçalves, Fox, Watson, & Kipp, 2004). The 5S framework facilitates work on key DL issues such as evaluation and integration (Shen, Gonçalves, & Fox, 2013) and can also guide the implementation of various technologies (Fox & da Silva Torres, 2014) and support different applications (Fox & Leidig, 2014).

A. Services

A full-function DL provides an integrated suite of services, which, in their most elaborate structure, reflect a formal ontology (Gonçalves, Fox, & Watson, 2008). Figure 1 provides one such organization of those services, with commonly considered user-facing services on the right and the other columns covering the infrastructure building processes. Those, in turn, include some for building the repository and others for adding value. Supporting the long view of curation and the information life-cycle, the second column lists preservation services.

Each of these services is ETD related. Most users are concerned with those that help them satisfy their information needs, starting with searching and continuing with browsing (Shen & Fox, 2009). From a theoretical and practical perspective, these, along with visualization (Yang et al., 2013), are all part of the broader activity of exploration (Shen, Vemuri, Fan, da Silva Torres, & Fox, 2006). Even broader is the need to support

Figure 1

DL Service/Activity Taxonomy

Infrastructure Services			**Information Satisfaction Services**
Repository Building		***Adding Value***	
Creational	**Preservational**		
Acquiring	Archiving	Annotating	Browsing
Cataloging	Conserving	Classifying	Collaborating
Crawling (focused)	Converting	Cleaning	Comparing
Describing	Copying	Clustering	Customizing
Digitizing	Emulating	Evaluating	Deduplicating
Federating	Migrating	Extracting	Downloading
Harvesting	Renewing	Indexing	Exploring
OCRing	Replicating	Parsing	Filtering
Purchasing	Standardizing	Publicizing	Providing access
Submitting	Translating (format)	Rating	Recommending
		Reviewing (peer)	Requesting
		Summarizing	Searching
		Surveying	Sharing
		Translating (language)	Visualizing

knowledge discovery with ETDs (R. Richardson, Srinivasan, & Fox, 2008).

One way to facilitate knowledge discovery is to employ information extraction (IE). This is often coupled with disambiguation methods applied to extracted data, such as author names (Godoi et al., 2013). IE can be applied to many problems and can therefore yield multiple benefits, the simplest of which is metadata enhancement. Machine-learning methods, such as SVM, have been helpful in this regard (Han et al., 2003). Another application is to mine phrases that can be added to keyword lists and assist with multilingual knowledge discovery (R. Richardson & Fox, 2007). Even more difficult is the extraction of metadata from scanned works, which, through leveraging optical character recognition (OCR) techniques, must still deal with substantial noise and often leads to different results from user- or cataloger-supplied metadata, sometimes because of errors but other times providing additional information (Choudhury, Wu, Ingram, & Fox, 2020).

While those familiar with e-commerce services often rely on recommender services, they are only available in a limited number of DLs (Perugini, Gonçalves, & Fox, 2004). Building on a long tradition of information retrieval that includes computing document–document similarity, such as during clustering, recommendations can be based on similarities between ETD pairs. This can be enhanced through an analysis of records of the relations between users and documents and among users (Akbar, Shaffer, Fan, & Fox, 2014).

B. Advances

In addition to the abovementioned typical DL services, advanced DLs can benefit from both additional services and new ways to integrate and apply them. We can build on decades of related work with distributed expert-based information systems (Belkin et al., 1987). DLs can be tailored to communities and use new approaches to intelligent systems (J. P. Leidig & Fox, 2014). Services can log activities to implement metrics (Gonçalves et al., 2003). These can lead to DL quality models (Gonçalves, Moreira, Fox, & Watson, 2007). Testing and evaluation can be based on automated quality assessment (Moreira, Gonçalves, Laender, & Fox, 2009). In addition, periodic formative and summative usability studies can lead to improved UX (Kengeri, Seals, Harley, Reddy, & Fox, 1999).

Regarding collection building, many institutions that have been gathering "born-digital" ETDs only from the time they instituted an ETD submission requirement are now paying the penalty of the delay in instituting that policy by having to spend time and labor to scan older ETDs. While this clearly helps the global research community by expanding the worldwide collection of ETDs, it requires additional effort to make those works accessible. Fortunately, improvements in OCR technologies help in those efforts, which in turn enable new research, such as the extraction of figures based on their visual characteristics (Kahu, 2020).

The Kahn–Wilensky framework (Kahn & Wilensky, 1995) indicates that finding additional useful ETD elements is possible when an ETD is viewed as a composite digital object. Identifying an ETD's constituent elements, such as its figures, then becomes another way to add digital objects and value to an ETD collection. While that can proceed according to visual characteristics, it can also benefit from text analytics.

Accordingly, another area of advancement builds on work with natural language processing (NLP) (Bird, Klein, & Loper, 2009). NLP has rapidly improved in recent years because of progress in deep learning (Torfi, Shirvani, Keneshloo, Tavaf, & Fox, 2020). As Python is often used for NLP as well as deep learning, it is a natural environment for those wanting to learn more about deep learning (Robinson, Whelan, Burton, Bolon, & Ellis, 2020).

Another related approach is the use of deep learning to facilitate text classification (Zhang, Zhao, & LeCun, 2015). Based on a fairly large sample, a machine-learning study of ETD classification targeting Library of Congress categories (Srinivasan & Fox, 2016) aided categorization regarding science, technology, engineering, and/or mathematics (STEM) (Srinivasan & Fox, 2016). Good results were achieved using machine-learning methods, with better ones coming from deep-learning techniques when building classifiers for ETDs targeting the upper levels of the ProQuest subject categories (Jude, 2020). However, many such methods require large amounts of training data to build accurate and reliable classifiers (Matykiewicz & Pestian, 2012). Thus, it is not surprising that preliminary work on classifying a small collection of Arabic ETDs identified the need for more data (Abdelrahman, Alotaibi, Fox, & Balci, 2020). Such scarcity in available training data for supervised machine-learning can be partly remedied through active learning (Settles, 2009) and/or semi-supervised learning (Zhu, 2005) so that even with minimal training data and reduced human-in-the-loop involvement, good classifiers can be produced.

Another development concerns citation parsing. ETDs tend to include large bibliographies and literature reviews. While references can be found in footnotes, chapter reference lists, or an integrated bibliography, a combined list can be generated. However, because formats and styles involved vary widely across thousands of options, parsing a full set of references often requires some machine learning (Park, Ehrich, & Fox, 2012). Further improvements are being explored using deep learning (Prasad, Kaur, & Kan, 2018).

Given that most ETDs lack chapter-level synopses, summarization is another service of interest. Early work leveraged IE techniques and methods to generate concept maps as summaries at both the ETD and chapter levels (W. R. Richardson, 2007). This also facilitated discovery across languages, as it is easier to apply machine translation to short text strings in a concept map than to long text abstracts. Summarization research has matured since, especially with regard to news stories but also regarding tweets, webpages, and collections thereof (Li, 2020; Li, Geissinger, Ingram, & Fox, 2020). However, our work with long documents has highlighted the need for sophisticated tailored pipelines, including aspect or topic classification, in summarizing such documents (Chakravarty, Chekuri, Mehrotra, & Fox, 2020).

VII. Future Digital Libraries

Supported by the U.S. Institute for Museum and Library Services, a team at Virginia Tech and Old Dominion University is engaged in research to build a next-generation DL for ETDs (Ingram, 2020). We plan to apply many of the aforementioned ideas.

Our early efforts have benefited from the work of multiple graduate class team term projects: one in a course on information retrieval (Kaushal et al., 2019) and two in a class on DLs (Alotaibi & Abdelrahman, 2019; Aromando, Banerjee, Ingram, Jude, & Kahu, 2019). These have explored segmenting documents into chapters, indexing metadata, indexing full text, parsing citations, extracting figures, classifying English works, and categorizing Arabic works.

As this project involves developing prototypes and conducting pilot testing, it is imperative that those processes be well integrated and quick. Accordingly, we will avoid the pitfalls of conventional information system construction (Kim, Behr, & Spafford, 2018). The latest methods to integrate development and operations (DevOps) will be used (Kim, Debois, Willis, & Humble, 2016).

Going beyond that, we aim to ensure that future DLs will be extensible and built on the latest practices from UX research and software engineering (Hartson & Pyla, 2012). They will leverage the movement toward low-code development as well as DevOps and continuous integration/continuous deployment (CI/CD).

That can be further enhanced by strengthening formal foundations and working on systematic DL generation (J. Leidig & Fox, 2013). A key point here would be using the abovementioned services and advances, which can be described in a declarative specification, so that suitable DLs can be generated (Gonçalves & Fox, 2002). This has been repeatedly prototyped, including the context of helping rapidly configure a suitable DSpace instance (Gorton, Fan, & Fox, 2007). It can work alongside any suitable set of components that act like building blocks (Suleman, Fox, Kelapure, Krowne, & Luo, 2003) to support desired user scenarios. As user goals and tasks have been considered when building those components, the generated system can support user needs (Kelapure, Gonçalves, & Fox, 2003).

While these methods help create and tailor DLs, a step beyond involves DLs becoming extensible systems. One approach (Fox & Chandrasekar, 2021) involves connecting the services dynamically into workflows (Liew et al., 2016), which can be coupled with knowledge graphs (Ehrlinger & Wöß, 2016). In one study, after developing different DL services, students analyzed project documents to identify user goals and tasks, representing them in a knowledge graph so they could be connected to services (and workflows built atop those services) (Meno & Vincent, 2020).

As such research progresses, findings must be implemented, and should be shared in library-related educational settings, such as through additions to curricular materials related to DLs (Oh, Yang, Pomerantz, Wildemuth, & Fox, 2016).

VIII. Acknowledgments

We are grateful to the IMLS for their support through grant LG-37-19-0078-19. We also thank the Virginia Tech Department of Computer Science, its Digital Library Research Laboratory, and University Libraries, as well as everyone who has helped us with the efforts described herein; the works of many of them are listed in the references.

References

Abdelrahman, E., Alotaibi, F., Fox, E. A., & Balci, O. (2020). Otrouha: A corpus of Arabic ETDs and a framework for automatic subject classification. In *Proc. Electronic Theses and Dissertations, ETD 2020: Discovery, Data, and Dates, virtual conference, Nov. 16-18, 2020, UAE, 8 pages.*

Akbar, M., Shaffer, C. A., Fan, W., & Fox, E. A. (2014). Recommendation based on deduced social networks in an educational digital library. In *IEEE/ACM Joint Conference on Digital Libraries, JCDL 2014, London, United Kingdom, September 8-12, 2014* (pp. 29–38). Retrieved from `http://dx.doi.org/10.1109/JCDL.2014.6970147` doi: 10.1109/JCDL.2014.6970147

Alotaibi, F., & Abdelrahman, E. (2019). *Otrouha: Automatic Classification of Arabic ETDs.* (Virginia Tech CS6604 team term project, December, Blacksburg, VA, http://hdl.handle.net/10919/96571)

Aromando, J., Banerjee, B., Ingram, W. A., Jude, P., & Kahu, S. (2019). *Classification and extraction of information from ETD documents.* (Virginia Tech CS6604 team term project, December, Blacksburg, VA, http://hdl.handle.net/10919/96645)

arXiv. (2020). *arXiv.org.* Retrieved from `https://arxiv.org/` (a free distribution service and an open-access archive for 1,771,036 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics)

Barton, M. D. (2005). *Dissertations: past, present, and future* (Doctoral dissertation, FL: University of South Florida). Retrieved from `https://scholarcommons.usf.edu/etd/2777`

Belkin, N., Borgman, C., Brooks, H., Bylander, T., Croft, W., Daniels, P., . . . Walker, D. (1987). Distributed expert-based information systems: An interdisciplinary approach. *Information Processing & Management*, *23*(5), 395–409.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python* (1st ed.). O'Reilly Media, Inc.

Brown, T. (2019). *Change by design, revised and updated: How design thinking transforms organizations and inspires innovation.* Harper Business. (ISBN 978–0062856623)

Candela, L., Castelli, D., Fox, E. A., & Ioannidis, Y. E. (2010). On digital library foundations. *International Journal on Digital Libraries*, *11*(1), 37–39. Retrieved from http://dx.doi.org/10.1007/s00799-011-0063-y doi: 10.1007/s00799-011-0063-y

Carroll, J. M. (1999). Five reasons for scenario-based design. In *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, Jan. 5–8, 1999, Maui, HI, USA. HICSS-32. Abstracts and CD-ROM of Full Papers* (Vol. Track3). (11 pages, DOI 10.1109/HICSS.1999.772890)

Chakravarty, S., Chekuri, S., Mehrotra, M., & Fox, E. A. (2020). *Aspect classification for legal depositions.* Retrieved from https://arxiv.org/abs/2009.04485

Choudhury, M. H., Wu, J., Ingram, W. A., & Fox, E. A. (2020). A heuristic baseline method for metadata extraction from scanned electronic theses and dissertations. In *Proc. JCDL 2020, Aug. 1-5, Virtual Event, China, poster, 2 pages, Honorable Mention for Best Poster/Demo.* Retrieved from https://doi.org/10.1145/3383583.3398590

Constable, G. (2014). *Talking to Humans: Success starts with understanding your customers.* Author. (ISBN 978–0990800927)

Constable, G. (2018). *Testing with Humans: How to use experiments to drive faster, more informed decision making.* Author. (ISBN 978–0990800934)

DuraSpace. (2020). *DSpace: The software of choice for academic, non-profit & commercial organizations building open digital repositories.* Retrieved from https://duraspace.org/dspace/

Ehrlinger, L., & Wöß, W. (2016). Towards a definition of knowledge graphs. *SEMANTiCS*

(Posters, Demos, SuCCESS), *48*, 1–4.

Farace, D. (1992–2020). *GreyNet International website.* Retrieved from http://www.greynet.org/ (Research, Publication, Open Access, Education, and Public Awareness to Grey Literature)

Fox, E. A. (1993). *Source Book on Digital Libraries.* Retrieved from http://fox.cs.vt.edu/DigitalLibrary/DLSB.pdf (prepared for and sponsored by the National Science Foundation, covering a series of NSF Invitational Workshops and Related Information, Dec. 6, 1993)

Fox, E. A., Akscyn, R. M., Furuta, R., & Leggett, J. J. (1995). Digital libraries - introduction to the special section. *Communications of the ACM*, *38*(4), 22–28. Retrieved from http://doi.acm.org/10.1145/205323.205325 doi: 10.1145/205323.205325

Fox, E. A., & Chandrasekar, P. (2021). How should one explore the digital library of the future? *Data and Information Management*, *5*.

Fox, E. A., & da Silva Torres, R. (2014). *Digital library technologies: Complex objects, annotation, ontologies, classification, extraction, and security.* Morgan & Claypool Publishers. Retrieved from http://dx.doi.org/10.2200/S00566ED1V01Y201401ICR033 doi: 10.2200/S00566ED1V01Y201401ICR033

Fox, E. A., Eaton, J. L., McMillan, G., Kipp, N. A., Mather, P., McGonigle, T., … DeVane, B. (1997). Networked Digital Library of Theses and Dissertations: an international effort unlocking university resources. *D-Lib Magazine*, *3*(9). Retrieved from http://www.dlib.org/dlib/september97/theses/09fox.html doi: 10.1045/september97-fox

Fox, E. A., Eaton, J. L., McMillan, G., Kipp, N. A., Weiss, L., Arce, E., & Guyer, S. A. (1996). National Digital Library of Theses and Dissertations: A scalable and sustainable approach to unlock university resources. *D-Lib Magazine*, *2*(9). Retrieved from http://www.dlib.org/dlib/september96/theses/09fox.html doi: 10.1045/september96-fox

Fox, E. A., Gonçalves, M. A., McMillan, G., Eaton, J. L., Atkins, A., & Kipp, N. A. (2002). The Networked Digital Library of Theses and Dissertations: changes in the university community. *Journal of Computing in Higher Education*, *13*(2), 102–124. Retrieved from http://dx.doi.org/10.1007/BF02940968 doi: 10.1007/BF02940968

Fox, E. A., Goncalves, M. A., & Shen, R. (2012). *Theoretical foundations for digital libraries: The 5S (societies, scenarios, spaces, structures, streams) approach.* Morgan & Claypool Publishers, San Francisco. Retrieved from http://dx.doi.org/10.2200/S00434ED1V01Y201207ICR022 doi: 10.2200/S00434ED1V01Y201207ICR022

Fox, E. A., & Leidig, J. P. (2014). *Digital libraries applications: CBIR, education, social networks, eScience/simulation, and GIS.* Morgan & Claypool Publishers, San Francisco. Retrieved from http://dx.doi.org/10.2200/S00565ED1V01Y201401ICR032 doi: 10.2200/S00565ED1V01Y201401ICR032

Fox, E. A., & Marchionini, G. (1998). Toward a worldwide digital library - introduction. *Communications of the ACM*, *41*(4), 28–32. Retrieved from http://doi.acm.org/10.1145/273035.273043 doi: 10.1145/273035.273043

Fox, E. A., & Marchionini, G. (2001). Digital libraries: Introduction. *Communications of the ACM*, *44*(5), 30–32. Retrieved from http://doi.acm.org/10.1145/374308.374329 doi: 10.1145/374308.374329

Fox, E. A., McMillan, G., & Eaton, J. L. (1999). The evolving genre of Electronic Theses and Dissertations. In *32nd Annual Hawaii International Conference on System Sciences (HICSS-32), January 5-8, 1999, Maui, Hawaii, USA.* Retrieved from http://dx.doi.org/10.1109/HICSS.1999.772646 doi: 10.1109/HICSS.1999.772646

Fox, E. A., & Urs, S. R. (2002). Digital libraries. *ARIST*, *36*(1), 502–589. Retrieved from `http://dx.doi.org/10.1002/aris.1440360113` doi: 10.1002/aris.1440360113

French, J. C., Fox, E. A., Maly, K., & Selman, A. L. (1994). Wide area technical report service - technical reports online. *ACM SIGACT News*, *25*(3), 124–127. Retrieved from `http://doi.acm.org/10.1145/193820.193849` doi: 10.1145/193820.193849

French, J. C., Fox, E. A., Maly, K., & Selman, A. L. (1995). Wide area technical report service: technical reports online. *Communications of the ACM*, *38*(4), 45. Retrieved from `http://doi.acm.org/10.1145/205323.205330` doi: 10.1145/205323.205330

Godoi, T. A., da Silva Torres, R., Carvalho, A. M. B. R., Gonçalves, M. A., Ferreira, A. A., Fan, W., & Fox, E. A. (2013). A relevance feedback approach for the author name disambiguation problem. In *13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, Indianapolis, IN, USA, July 22 - 26, 2013* (pp. 209–218). Retrieved from `http://doi.acm.org/10.1145/2467696.2467709` doi: 10.1145/2467696.2467709

Gonçalves, M. A., Fox, E. A., & Watson, L. T. (2008). Towards a digital library theory: a formal digital library ontology. *International Journal on Digital Libraries*, *8*(2), 91–114. Retrieved from `http://dx.doi.org/10.1007/s00799-008-0033-1` doi: 10.1007/s00799-008-0033-1

Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. A. (2004). Streams, structures, spaces, scenarios, societies (5S): A formal model for digital libraries. *ACM Trans. Inf. Syst.*, *22*(2), 270–312. Retrieved from `http://doi.acm.org/10.1145/984321.984325` doi: 10.1145/984321.984325

Gonçalves, M. A., Moreira, B. L., Fox, E. A., & Watson, L. T. (2007). What is a good digital library? - A quality model for digital libraries. *Information processing & management*, *43*(5), 1416–1437. Retrieved from `http://dx.doi.org/10.1016/j.ipm.2006.11.010` doi: 10.1016/j.ipm.2006.11.010

Gonçalves, M. A., Panchanathan, G., Ravindranathan, U., Krowne, A., Fox, E. A., Jagodzinski, F., & Cassel, L. N. (2003). The XML log standard for digital libraries: analysis, evolution, and deployment. In *ACM/IEEE 2003 Joint Conference on Digital Libraries (JCDL 2003), 27-31 May 2003, Houston, Texas, USA, Proceedings* (pp. 312–314). Retrieved from http://doi.ieeecomputersociety.org/10.1109/JCDL.2003.10010 doi: 10.1109/JCDL.2003.10010

Gonçalves, M. A., & Fox, E. A. (2002). 5SL: A language for declarative specification and generation of digital libraries. In *Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '02, Portland, Oregon, USA, June 14-18, 2002* (p. 263–272). New York, NY, USA: Association for Computing Machinery. Retrieved from http://doi.acm.org/10.1145/544220.544276 doi: 10.1145/544220.544276

Gorton, D., Fan, W., & Fox, E. A. (2007). Specification and generation of digital libraries into DSpace using the 5S framework. In *Research and Advanced Technology for Digital Libraries, 11th European Conference, ECDL 2007, Budapest, Hungary, September 16-21, 2007, Proceedings* (pp. 567–569). Retrieved from http://dx.doi.org/10.1007/978-3-540-74851-9_71 doi: 10.1007/978-3-540-74851-9_71

Halbert, M. (2012). *The Life Cycle of Electronic Theses and Dissertations, keynote presentation of 47 slides.* Retrieved from (https://digital.library.unt.edu/ark:/67531/metadc78271/

Han, H., Giles, C. L., Manavoglu, E., Zha, H., Zhang, Z., & Fox, E. A. (2003). Automatic document metadata extraction using support vector machines. In *ACM/IEEE 2003 Joint Conference on Digital Libraries (JCDL 2003), 27-31 May 2003, Houston, Texas, USA, Proceedings* (pp. 37–48). Retrieved from http://dx.doi.org/10.1109/JCDL.2003.1204842 doi: 10.1109/JCDL.2003.1204842

Harnad, S. (2015). Open Access: What, Where, When, How and Why. In *Ethics, Science,* *Technology, and Engineering: An International Resource, eds. J. Britt Holbrook & Carl Mitcham, (2nd of Encyclopedia of Science, Technology, and Ethics, Farmington Hills MI: MacMillan Reference).* Retrieved from https://eprints.soton.ac.uk/361704/1/ESTEarticle-OA-Harnad.pdf

Hartson, R., & Pyla, P. (2012). *The UX book: Process and guidelines for ensuring a quality user experience* (1st ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Ingram, W. (2020). *Opening Books and the National Corpus of Graduate Research.* Retrieved 2020-04-04, from https://dlib.vt.edu/content/dlib_vt_edu/en/Research/etd.html (Team on IMLS grant LG-37-19-0078-19, Virginia Tech, Blacksburg, VA)

Interaction Design Foundation. (2020). *What is UX research?* Retrieved from https://www.interaction-design.org/literature/topics/ux-research (Interaction Design Foundation, accessed January 15, 2020)

Jude, P. M. (2020). *Increasing accessibility of Electronic Theses and Dissertations (ETDs) through chapter-level classification* (Master's thesis, Virginia Tech, Department of Computer Science, Blacksburg, VA 24061 USA). Retrieved from http://hdl.handle.net/10919/99294

Kahn, R., & Wilensky, R. (1995). *A framework for distributed digital object services.* Retrieved from http://www.cnri.reston.va.us/k-w.html (cnri.dlib/tn95-01)

Kahu, S. Y. (2020). *Figure extraction from scanned electronic theses and dissertations* (Master's thesis, Virginia Tech, Department of Electrical and Computer Engineering, Blacksburg, VA 24061 USA). Retrieved from http://hdl.handle.net/10919/100113

Kaushal, K. K., Kulkarni, R., Sumant, A., Wang, C., Yuan, C., & Yuan, L. (2019). *Collection Management of Electronic Theses and Dissertations (CME) CS5604 Fall 2019.* (Virginia Tech CS5604 team term project, December, Blacksburg, VA, http://hdl.handle.net/10919/96484)

Kelapure, R., Gonçalves, M. A., & Fox, E. A. (2003). Scenario-based generation of digital library services. In *Research and Advanced Technology for Digital Libraries, 7th European Conference, ECDL 2003, Trondheim, Norway, August 17-22, 2003, Proceedings* (pp. 263–275). Retrieved from http://dx.doi.org/10.1007/978-3-540-45175-4_25 doi: 10.1007/978-3-540-45175-4_25

Kengeri, R., Seals, C. D., Harley, H. D., Reddy, H. P., & Fox, E. A. (1999). Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. *International Journal on Digital Libraries*, *2*(2-3), 157–169. Retrieved from http://dx.doi.org/10.1007/s007990050044 doi: 10.1007/s007990050044

Kim, G., Behr, K., & Spafford, G. (2018). *The Phoenix project, 5th anniversary edition: A novel about IT, DevOps, and helping your business win.* IT Revolution Press. (ISBN 9781942788294)

Kim, G., Debois, P., Willis, J., & Humble, J. (2016). *The DevOps Handbook: How to create world-class agility, reliability, and security in technology organizations.* IT Revolution Press. (ISBN 978–1942788003)

Lagoze, C., Van de Sompel, H., & Warner, S. (1998–2017). *Open Archives Initiative website.* Retrieved from https://www.openarchives.org/ (Standards for Web Content Interoperability, including ResourceSync, Protocol for Metadata Harvesting (OAI-PMH), and Object Reuse and Exchange (OAI-ORE))

Leidig, J., & Fox, E. A. (2013). Formal foundations for systematic digital library generation. In *13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, Indianapolis, IN, USA, July 22 - 26, 2013* (pp. 391–392). Retrieved from http://doi.acm.org/10.1145/2467696.2467774 doi: 10.1145/2467696.2467774

Leidig, J. P., & Fox, E. A. (2014). Intelligent digital libraries and tailored services. *Journal of Intelligent Information Systems*, *43*(3), 463–480. Retrieved from http://dx.doi.org/10.1007/s10844-014-0342-3 doi: 10.1007/s10844-014-0342-3

Li, L. (2020). *Event-related collections understanding and services* (Doctoral dissertation, Virginia Tech). Retrieved from http://hdl.handle.net/10919/97365

Li, L., Geissinger, J., Ingram, W. A., & Fox, E. A. (2020). Teaching natural language processing through big data text summarization with problem-based learning. *Data and Information Management*, *4*(1), 18–43. Retrieved from https://content.sciendo.com/downloadpdf/journals/dim/4/1/article-p18.xml doi: 10.2478/dim-2020-0003

Liew, C. S., Atkinson, M. P., Galea, M., Ang, T. F., Martin, P., & Hemert, J. I. V. (2016, December). Scientific workflows: Moving across paradigms. *ACM Computing Surveys*, *49*(4). doi: 10.1145/3012429

Matykiewicz, P., & Pestian, J. (2012). Effect of small sample size on text categorization with support vector machines. In *BioNLP@HLT-NAACL* (p. 193–201). Proceedings of the 2012 workshop on biomedical natural language processing. Association for Computational Linguistics.

Mehringer, A. (1986). The Association of American Publishers' Electronic Manuscript Project: a status report. PubMed: 3947774. *Bulletin of the Medical Library Association*, *74*(1), 27–30. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC227768/

Meno, E., & Vincent, K. (2020). *Twitter-based knowledge graph for researchers.* (Virginia Tech CS4624 team term project, May, Blacksburg, VA, http://hdl.handle.net/10919/98239)

Moreira, B. L., Gonçalves, M. A., Laender, A. H. F., & Fox, E. A. (2009). Automatic evaluation of digital libraries with 5SQual. *Journal of Informetrics*, *3*(2), 102–123. Retrieved from http://dx.doi.org/10.1016/j.joi.2008.12.003 doi: 10.1016/j.joi.2008.12.003

NDLTD. (2020). *Networked Digital Library of Theses and Dissertations.* Retrieved from http://ndltd.org/

Oh, S., Yang, S., Pomerantz, J., Wildemuth, B. M., & Fox, E. A. (2016). Results of a digital library curriculum field test. *International Journal on Digital Libraries*, *17*(4), 273–286. Retrieved from http://dx.doi.org/10.1007/s00799-015-0151-5 doi: 10.1007/s00799-015-0151-5

Park, S. H., Ehrich, R. W., & Fox, E. A. (2012). A hybrid two-stage approach for discipline-independent canonical representation extraction from references. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '12, Washington, DC, USA, June 10-14, 2012* (pp. 285–294). Retrieved from http://doi.acm.org/10.1145/2232817.2232871 doi: 10.1145/2232817.2232871

Perugini, S., Gonçalves, M. A., & Fox, E. A. (2004). Recommender systems research: A connection-centric survey. *Journal of Intelligent Information Systems*, *23*(2), 107–143. Retrieved from http://dx.doi.org/10.1023/B:JIIS.0000039532.05533.99 doi: 10.1023/B:JIIS.0000039532.05533.99

Phanouriou, C., Kipp, N. A., Sornil, O., Mather, P., & Fox, E. A. (1999). A digital library for authors: recent progress of the Networked Digital Library of Theses and Dissertations. In *Proceedings of the fourth ACM conference on Digital Libraries, August 11-14, 1999, Berkeley, CA, USA* (pp. 20–27). Retrieved from http://doi.acm.org/10.1145/313238.313244 doi: 10.1145/313238.313244

Pinfield, S., Wakeling, S., Bawden, D., & Robinson, L. (2020). *Open Access in Theory and Practice: The theory-practice relationship and openness*. Routledge. (ISBN 978–0367227852)

Prasad, A., Kaur, M., & Kan, M.-Y. (2018, Nov 01). Neural ParsCit: a deep learning-based reference string parser. *International Journal on Digital Libraries*, *19*(4), 323-337. Retrieved from https://doi.org/10.1007/s00799-018-0242-1 doi: 10.1007/s00799-018-0242-1

Richardson, R., & Fox, E. A. (2007). Using bilingual ETD collections to mine phrase translations. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2007, Vancouver, BC, Canada, June 18-23, 2007, Proceedings* (pp. 352–353). Retrieved from http://doi.acm.org/10.1145/1255175.1255244 doi: 10.1145/1255175.1255244

Richardson, R., Srinivasan, V., & Fox, E. A. (2008). Knowledge discovery in digital libraries of electronic theses and dissertations: an NDLTD case study. *International Journal on Digital Libraries*, *9*(2), 163–171. Retrieved from http://dx.doi.org/10.1007/s00799-008-0046-9 doi: 10.1007/s00799-008-0046-9

Richardson, W. R. (2007). *Using concept maps as a tool for cross-language relevance determination* (Doctoral dissertation, Virginia Tech). Retrieved from http://hdl.handle.net/10919/28191

Robinson, M., Whelan, A., Burton, S., Bolon, B., & Ellis, H. (2020). *Deep learning course.* Retrieved from http://hdl.handle.net/10919/98244 (open-source course with Python and PyTorch on deep learning topics hosted on GitHub in the Machine Learning Mindset repository built with the guidance of VT CS4624 course client Amirsina Torfi)

Settles, B. (2009). *Active learning literature survey.* Retrieved from http://digital.library.wisc.edu/1793/60660 (University of Wisconsin-Madison Department of Computer Sciences TR1648)

Shen, R., & Fox, E. A. (2009). Browsing in digital libraries. In *Encyclopedia of Database Systems* (pp. 268–273). Retrieved from http://dx.doi.org/10.1007/978-0-387-39940-9_40 doi: 10.1007/978-0-387-39940-9_40

Shen, R., Gonçalves, M. A., & Fox, E. A. (2013). *Key issues regarding digital libraries:*

Evaluation and integration. Morgan & Claypool Publishers. Retrieved from `http://dx.doi.org/10.2200/S00474ED1V01Y201301ICR026` (ISBN 9781608459124) doi: 10.2200/S00474ED1V01Y201301ICR026

Shen, R., Vemuri, N. S., Fan, W., da Silva Torres, R., & Fox, E. A. (2006). Exploring digital libraries: integrating browsing, searching, and visualization. In *ACM/IEEE Joint Conference on Digital Libraries, JCDL 2006, Chapel Hill, NC, USA, June 11-15, 2006, Proceedings* (pp. 1–10). Retrieved from `http://doi.acm.org/10.1145/1141753.1141755` doi: 10.1145/1141753.1141755

Srinivasan, V., & Fox, E. A. (2016). Progress toward automated ETD cataloging. In *Proc. 19th International Symposium on Electronic Theses and Dissertations, ETD 2016 "Data and Dissertations", 11-13 July 2016, Lille (France), in Session 3.1 on 7/13 "ETD metadata and cataloging".* Retrieved from `http://etd2016.sciencesconf.org/92334`

Suber, P. (2012). *Open access.* MIT Press. Retrieved from `http://mitpress.mit.edu/sites/default/files/titles/content/9780262517638_Open_Access_PDF_Version.pdf` (ISBN 9780262517638)

Suleman, H. (2020a). *Global ETD Search.* Retrieved from `http://search.ndltd.org/`

Suleman, H. (2020b). *NDLTD Union Archive.* Retrieved from `http://union.ndltd.org/portal/`

Suleman, H., Atkins, A., Gonçalves, M. A., France, R. K., Fox, E. A., Chachra, V., ... Young, J. (2001a). Networked Digital Library of Theses and Dissertations: bridging the gaps for global access - part 1: mission and progress. *D-Lib Magazine*, *7*(9). Retrieved from `http://dx.doi.org/10.1045/september2001-suleman-pt1` doi: 10.1045/september2001-suleman-pt1

Suleman, H., Atkins, A., Gonçalves, M. A., France, R. K., Fox, E. A., Chachra, V., ... Young, J. (2001b). Networked Digital Library of Theses and Dissertations: bridging the gaps for global access - part 2: services and research. *D-Lib Magazine*, *7*(9).

Retrieved from http://dx.doi.org/10.1045/september2001-suleman-pt2 doi: 10.1045/september2001-suleman-pt2

Suleman, H., & Fox, E. A. (2002a). The Open Archives Initiative: realizing simple and effective digital library interoperability. *Journal of Library Automation*, *35*(1/2), 125–145.

Suleman, H., & Fox, E. A. (2002b). Towards universal accessibility of ETDs: building the NDLTD Union Archive. In *Proceedings of ETD'2002, Provo, UT: Brigham Young University, May 30 - June 1, 2002.*

Suleman, H., Fox, E. A., Kelapure, R., Krowne, A., & Luo, M. (2003). Building digital libraries from simple building blocks. *Online Information Review*, *27*(5), 301–310. Retrieved from http://dx.doi.org/10.1108/14684520310502252 doi: 10.1108/14684520310502252

Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2020). *Natural language processing advancements by deep learning: A survey.* Retrieved from https://arxiv.org/abs/2003.01200 (arXiv:2003.01200)

Ulwick, A. (2005). *What Customers Want: Using outcome-driven innovation to create breakthrough products and services, 1st edition.* McGraw-Hill. (ISBN 978–0071408677)

Wikipedia. (2020). *Peter Murray-Rust biography page.* Retrieved from https://en.wikipedia.org/wiki/Peter_Murray-Rust

Wunker, S., Wattman, J., & Farber, D. (2016). *Jobs to be Done: A roadmap for customer-centered innovation.* AMACOM. (ISBN 978–0814438039)

Yang, S., Chung, H., Lin, X., Lee, S., Chen, L., Wood, A., … Fox, E. A. (2013). PhaseVis1: What, when, where, and who in visualizing the four phases of emergency management through the lens of social media. In *10th Proceedings of the International Conference on Information Systems for Crisis Response and Management, Baden-Baden, Germany, May 12-15, 2013.* Retrieved from http://idl.iscram.org/files/yang/2013/1122_Yang_etal2013.pdf

Zhang, X., Zhao, J. J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *NIPS, Advances in neural information processing system* (pp. 649–657).

Zhu, X. (2005). *Semi-supervised learning literature survey.* Retrieved from https://minds.wisconsin.edu/handle/1793/60444 (University of Wisconsin-Madison Department of Computer Sciences TR1530)