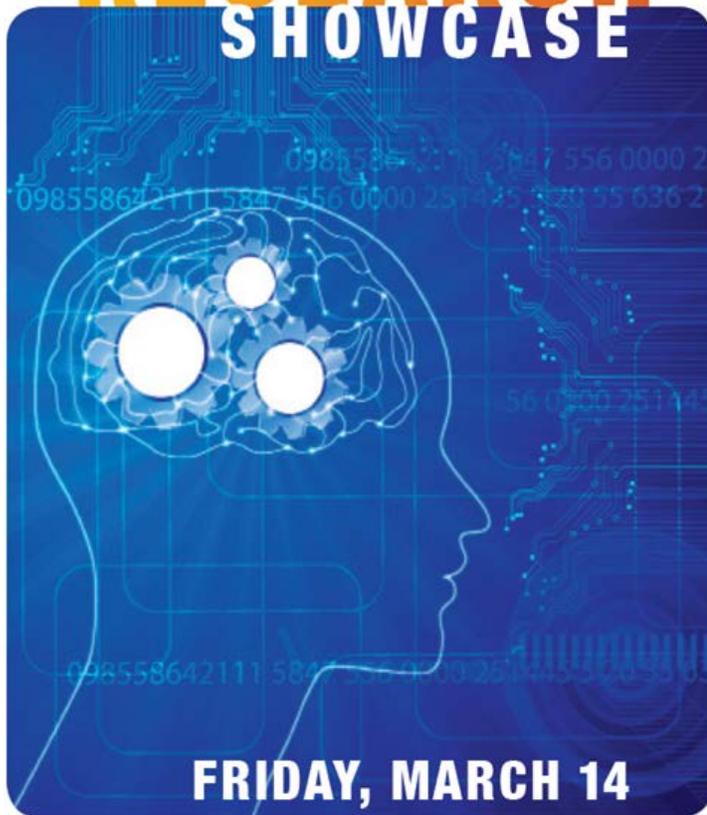


2014 RESEARCH SHOWCASE



Schedule of Events

12:00 – 1:30 p.m.

Poster Session 1, Demonstrations (131 LIS), and Pizza (East Foyer)

1:30 – 3:30 p.m.

Presentations (126 LIS)

3:30 – 5:00 p.m.

Poster Session 2, Demonstrations (131 LIS), and Reception (East Foyer)

Graduate School of Library & Information Science
501 East Daniel Street
Champaign, IL 61820

Overview of Sessions

12:00 – 1:30 p.m. • *Poster Session 1, Demonstrations (131 LIS), and Pizza (East Foyer)*

SERVICE ROLES, COMMUNITY ENGAGEMENT AND DIGITAL LITERACIES: A QUALITATIVE STUDY OF ILLINOIS PUBLIC LIBRARIES IN UNDERSERVED COMMUNITIES

Jeff Ginger

K-POP GENRES: A CROSS-CULTURAL EXPLORATION

Jin Ha Lee, Kahyun Choi, Xiao Hu, J. Stephen Downie

SUPPORTING ALL LEARNERS IN SCHOOL-WIDE COMPUTING: A CROSS-CASE QUALITATIVE ANALYSIS

Martin Wolske, Maya Israel, George Reese, Avigail Snir

INSECT BYTES: USING INFORMATICS TO SUPPORT DISEASE CONTROL

Ian Brooks

SENTINETs: USER CLASSIFICATION BASED ON SENTIMENT FOR SOCIAL CAUSES WITHIN A TWITTER NETWORK

Shubhanshu Mishra, Sneha Agarwal, Jinlong Guo, Kirstin Phelps, Johna Picco

“SHERLOCK HOLMES MEETS GODZILLA”: FILM DESCRIPTION THROUGH COMPARISON

Peter Organisciak, Michael Twidale

SUMMARIZING SCIENTIFIC TEXT WITH TYPED-DEPENDENCY GRAPHS

Henry A. Gabb

A THEORETICAL MODEL FOR TRUST IN AUTOMATED SYSTEMS

Masooda Bashir, Kevin Hoff

NERDFIGHTERIA’S LITERARY SALON: UNDERSTANDING THE SOCIAL NETWORKS IN AN ONLINE COMMUNITY OF READERS AND AUTHORS

Alaine Martaus

FROM INFORMATION TO INNOVATION: EXPLORING SCIENCE TEACHERS’ LEARNING AND INFORMATION PRACTICE FOR INNOVATION IN SCIENCE TEACHING

Wei Gao

1:30 – 3:30 p.m. • *Presentations (126 LIS)*

COMPUTATIONAL IMPACT ASSESSMENT OF SOCIAL JUSTICE DOCUMENTARIES

Jana Diesner, Jinseok Kim, Shubhanshu Mishra, Kiumars Soltani, Sean Wilner, Amirhossein Aleyasen

CLOSING THE APP GAP I: TABLETS IN SUMMER READING FOR YOUTH

Deborah Stevenson, Kate McDowell, Cass Mabbott

THE PRIMACY OF PRIVACY (A SURVEY OF KNOWLEDGE AND OPINIONS)

Masooda Bashir, Jay Kesan, Kevin Hoff, Carol Mullins

CYBERORGANIZING EVERYDAY HERITAGE IN AND AROUND PUBLIC LIBRARIES: AN EXPLORATORY STUDY IN ILLINOIS

Noah Lenstra

SOCIO-TECHNICAL DATA ANALYTICS

Catherine Blake

KNOWING WHAT KNOWLEDGE TO MANAGE: WHAT KNOWLEDGE IS WORTH DOCUMENTING, WHAT DOCUMENTS ARE WORTH KEEPING, AND WHEN IS DOCUMENTATION TOO MUCH?

Ingbert Schmidt

THE GEOGRAPHY OF CENSORSHIP: COMMUNITIES, CHALLENGERS, AND HARRY POTTER

Emily Knox

I'VE GOT A PHD IN COMPUTER SCIENCE: HOW COME I CAN'T GET THIS DRATTED GIZMO TO WORK?

Michael Twidale

3:30 – 5:00 p.m. • *Poster Session 2, Demonstrations (131 LIS), and Reception (East Foyer)*

AUTOMATICALLY SUMMARIZING MEDICAL LITERATURE

Ana Lucic, Catherine Blake

THE IMPACT OF GROUND OZONE ON ASTHMA: A CASE STUDY USING PROJECT INDICATOR IN CHAMPAIGN-URBANA

Elizabeth J Surbeck, Catherine Blake

INVESTIGATING WRITERS' ATTITUDES BY MINING A LARGE CORPUS OF BOOKS

Sayan Bhattacharyya

**USING COLLECTIONS AND WORKSETS IN LARGE-SCALE CORPORA: PRELIMINARY FINDINGS FROM THE
WORKSET CREATION FOR SCHOLARLY ANALYSIS PROJECT**

***Harriett Green, Katrina Fenlon, Megan Senseney, Sayan Bhattacharyya, Craig Willis,
Peter Organisciak, J. Stephen Downie, Tim Cole, Beth Plale***

SPECIALIZATION IN DATA CURATION: PRELIMINARY RESULTS FROM AN ALUMNI SURVEY, 2008-2012

***Cheryl A. Thompson, Megan Senseney, Karen S. Baker, Virgil E. Varvel, Carole L.
Palmer***

**AN EXPLORATION OF THE H1N1 OUTBREAK IN CHAMPAIGN-URBANA ELEMENTARY SCHOOLS DURING THE 2009-2010 INFLUENZA
SEASON**

Christopher Komisarz, Ian Brooks

USABILITY AND DESIGN ANALYSIS: GOOGLE HANGOUTS

Sean Gordon, Taylor Kirch, Adriel Flores-Pagnani, Michael Twidale

USING NAMED ENTITY RECOGNITION AS A CLASSIFICATION HEURISTIC

Andrea K. Thomer, Nicholas M. Weber

HOW WORKFLOW DOCUMENTATION FACILITATES CURATION PLANNING

Karen M. Wickett, Andrea K. Thomer, Karen S. Baker, Timothy DiLauro, Abigail E. Asangba

INVESTIGATING RESEARCH DATA DESCRIPTION IN SCIENTIFIC JOURNAL PUBLICATIONS

Tiffany C. Chao

OUTCOMES OF THE "DATA CURATION FOR GEOBIOLOGY AT YELLOWSTONE NATIONAL PARK" WORKSHOP

***Andrea K. Thomer, Carole L. Palmer, Bruce W. Fouke, Ann Rodman, G. Sayeed Choudhury, Karen S. Baker,
Abigail E. Asangba, Karen M. Wickett, Timothy DiLauro, Virgil Varvel***

Posters with Demonstrations (during Poster Sessions 1 and 2)

THE ILLINOIS DISTRIBUTED MUSEUM PROJECT: ENGINEERING AND TECHNOLOGY INNOVATIONS AT
THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Michael Twidale, Susan Frankenberg, Kelsey Heffren, Jamie Wittenberg

during poster session 1 only

MEASURES OF NOVELTY IN BIBLIOMETRIC MODELS

Vetle Torvik, Shubhanshu Mishra

PATCI - A CITATION MATCHER OPTIMIZED FOR NON-PATENT CITATIONS

Vetle Torvik, Miles Lincoln, Sneha Agarwal, Haoyan Cai

USEFUL APPLICATIONS FOR CREATIVE COLLABORATION ON LARGE SCALE TOUCH DEVICES

Christopher Nixon, Michael Twidale

Center Posters (Displayed during Poster Sessions 1 and 2)

CENTER FOR CHILDREN'S BOOKS (CCB)

Deborah Stevenson

CENTER FOR DIGITAL INCLUSION (CDI)

Jon Gant

CENTER FOR INFORMATICS RESEARCH IN SCIENCE AND SCHOLARSHIP (CIRSS)

Carole L. Palmer, Catherine Blake

HATHI TRUST RESEARCH CENTER (HTRC)

J. Stephen Downie

Presentation Abstracts

COMPUTATIONAL IMPACT ASSESSMENT OF SOCIAL JUSTICE DOCUMENTARIES

Jana Diesner, Jinseok Kim, Shubhanshu Mishra, Kiumars Soltani, Sean Wilner, Amirhossein Aleyasen

Besides telling a story, the goal with social justice documentaries is to motivate change in people's knowledge and/ or behavior. How can we know if a film has achieved these goals, and how early in the life cycle of a production can we answer this question? The need for reliable, efficient and systematic ways to evaluate the impact of such media products has been repeatedly pointed out by funders, practitioners and researchers. We report on how we have been addressing this need by developing, implementing, applying and evaluating a theoretically-grounded, computational solution for assessing the impact of social justice documentaries in a scalable, empirical and comprehensive fashion. Based on the assumption that documentaries are produced and watched as part of larger and continuously changing ecosystems that involve multiple stakeholders and information, we map, monitor and analyze social and semantic networks that represent these types of entities. We integrate techniques from natural language processing and network analysis for this purpose. We report on a) our theoretical framework and methodology and b) evaluating the performance and usability of our solution by bringing it into real-world application contexts where we have been collaborating with funders and producers to conduct impact assessments.

CLOSING THE APP GAP I: TABLETS IN SUMMER READING FOR YOUTH

Deborah Stevenson, Kate McDowell, Cass Mabbott

The Center for Children's Books (CCB) Closing the App Gap I planning grant examines the advantages of summer reading programming that includes tablet-based use of apps and e-books with primary-grade children. We're exploring the possibility that tablets can help defray the problem of summer reading loss, the setback of skills over the long summer vacation, a problem particularly evident in low income communities; we're also exploring the role of library tablet use in addressing the "app gap," the digital divide that leaves low income children with lesser access to tablet and handheld technologies. In this stage of the grant, we are working toward a pilot study at the Douglass Branch Library this summer; en route, we have developed a compilation of existing research and practice guidelines, articulated our selection standards for apps, and begun to design the programming that will form the basis of the pilot study. In this poster (or presentation), we will outline our findings to date and discuss the significant challenges we've discovered in library-based research on young people's practice.

THE PRIMACY OF PRIVACY (A SURVEY OF KNOWLEDGE AND OPINIONS)

Masooda Bashir, Jay Kesan, Kevin Hoff, Carol Mullins

We present the preliminary results of an extensive, two-part privacy survey centered around people's knowledge and opinions. Part one, which assesses knowledge using objective third-person scenarios, provides valuable information about what people know on a range of topics including cloud computing, cyber security, online behavioral advertising, the privacy of educational records, and the legal aspects of online privacy. Initial results expose several knowledge gaps and suggest discrepancies between demographic groups. In part two, we explore opinions on privacy issues ranging from government surveillance of private citizens to the data collection methods used by

online advertising companies. The preliminary results provide fascinating data about what people think, know, and expect when it comes to the availability of personal information on the Internet. Responses also offer insight into the relationship between knowledge, opinions, and online behavior. All together, our survey provides thought-provoking information going significantly beyond prior empirical work.

CYBERORGANIZING EVERYDAY HERITAGE IN AND AROUND PUBLIC LIBRARIES: AN EXPLORATORY STUDY IN ILLINOIS

Noah Lenstra

Trends from both within and without public librarianship create new opportunities for libraries to serve the public's interest in heritage information. Based on a workshop series in Illinois, this article presents exploratory findings on: a) existing public library heritage services and b) how cyberorganizing can enhance these services. Public libraries offer heterogeneous heritage services in local contexts composed of multiple actors and institutions. Improving this service area involves organizing within libraries, across local institutions, and among individuals with diverse interests. Part of this organizational work involves digital technologies, but people are the core element in organizing everyday heritage services.

SOCIO-TECHNICAL DATA ANALYTICS

Catherine Blake

For many years funding agencies both in the US and abroad have focused heavily on cyber-infrastructure which has eased the difficulty in both storage and computation of large quantities of data. The volume, velocity and variety tenants of the "Big Data" movement show a shift in emphasis from the underlying computational mechanisms to the data that these mechanisms must support. This talk will use a series of case studies from medicine and business to illustrate why human computation must be added to this agenda if the true potential of the information revolution is to be realized.

KNOWING WHAT KNOWLEDGE TO MANAGE: WHAT KNOWLEDGE IS WORTH DOCUMENTING, WHAT DOCUMENTS ARE WORTH KEEPING, AND WHEN IS DOCUMENTATION TOO MUCH?

Ingbert Schmidt

What kind of knowledge is important to document, and what kind of knowledge ends up being intellectual clutter? When is it important to keep documents, and when are they better discarded? Brown and Duguid's (1991) insightful synthesis of research on how organizational knowledge functions clearly shows how instruction manuals and prescribed practices can be simultaneously too restrictive and lacking in needed information, thus forcing organization members to find workarounds. Yet documentation of work practice clearly has some value to organizations given the continued research in the related fields of knowledge management, the learning organization, organizational learning, and organizational knowledge (Easterby-Smith & Lyles 2011). Studies in these fields typically focus on topics like the nature of knowledge, individual and group learning, social factors like culture and identity, communities of practice, teams, and organizational structures (Easterby-Smith & Lyles 2011). Other studies focus on technologies and systems for "managing" knowledge. But what about the content: the knowledge and the documents? This study examines how efforts at preserving how-to knowledge over multiple generations of participants had both successes and failures, and examines how decisions about what knowledge was preserved, how it

was preserved, and what documents were preserved, affected how they were understood and received by their intended audience. Findings include: efforts at creating step by step instructions and other recipe-like documents could be both empowering and destructive, depending on the kind of how-to knowledge the documents contained; the presentation of historical examples of activities were often read very differently by the audience than what the presenter intended; providing comprehensive documentation of complex information often proved counter productive for a variety of reasons.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities of practice: Toward a unified view of working, learning and innovation. *Organizational Science*, 2(1), 40–57.

Easterby-Smith, M., & Lyles, M. A. (2011). Handbook of Organizational Learning and Knowledge Management, chap. The Evolving Field of Organizational Learning and Knowledge Management. Wiley.

THE GEOGRAPHY OF CENSORSHIP: COMMUNITIES, CHALLENGERS, AND HARRY POTTER

Emily Knox

The Harry Potter series was one of the most censored books when it was first published. Through the use of the census and other publicly available data attempts to answer the following questions using the Harry Potter series as a case study: Is the perception of the pervasiveness of challenges accurate? Are there any commonalities among communities that experience challenges? Are some types of communities more prone to challenges to others? What are the characteristics that might unite these communities? The paper investigates the commonalities and differences among 23 communities that experienced challenges to Harry Potter during the years 1999-2007.

I'VE GOT A PHD IN COMPUTER SCIENCE: HOW COME I CAN'T GET THIS DRATTED GIZMO TO WORK?

Michael Twidale

All of us need to learn new computer applications, new features, new versions and upgrades. Many of us are learning new things every day about the technologies we use. When it works well, this is wonderful. We discover new, better ways of doing things, improving our workplace productivity, and finding better ways of getting things done in our personal and family lives. But it doesn't always work so well, does it? Often people seem rather afraid to admit their confusions, and we tend to blame ourselves, ascribing our confusion to stupidity, or lack of technical skill - rather than poor design of the technologies. So what needs fixing – us or them? Using a series of small autoethnographies and anecdotes of people with substantial technical credentials who yet also struggle with tech, we can understand some of the barriers to tech learning – and therefore what we should be doing to help everyone learn how to use technology in a more productive and less annoying way. One issue that emerges is the growing power of searching online for tech help – provided you know how to search efficiently and deal effectively with what you find. What are the implications of this kind of tech learning for what a school of information does, and how its graduates can help millions of people to be more innovative in their use of computational technologies?

Poster and Demonstration Abstracts

Poster Session 1

SERVICE ROLES, COMMUNITY ENGAGEMENT AND DIGITAL LITERACIES: A QUALITATIVE STUDY OF ILLINOIS PUBLIC LIBRARIES IN UNDERSERVED COMMUNITIES

Jeff Ginger

In this dissertation study I investigate the role of the public library in fostering digital literacy in underserved Illinois communities. Over the course of two years I collected data about people, policies, activities and infrastructure related to how individuals learn, comprehend and apply digital technologies with the library. By reflecting upon a combination of interviews, report data and a three month case study I reveal a compelling story of what digital literacy really means when it comes to public libraries, and what they're dealing with on a day-to-day basis. As an activist researcher in community informatics who is interested in rearranging power I contend that all of the libraries I visited can teach us something about the challenges and innovations we face in successfully connecting social roles to the emergent service roles of the public library. This is especially important because the libraries I visited include some of the voices and issues that often go unnoticed in other studies. My sample specifically targeted libraries that largely serve socially excluded populations that in many cases may be in the most need and also the best equipped to inform us about challenges and best practices. My ultimate goal is two-fold. First, by exploring why difficulties and successes occur in different and similar library settings I hope to inspire and facilitate problem-solving and idea-sharing between libraries. Consequently, this work includes presenting a cutting-edge model for a media maker space developed as a collaboration between The Urbana Free Library and Champaign-Urbana Community Fab Lab. Second, by considering and reconciling the complexity of digital literacies in these settings I advance a provisional guide for activity and policy development based on a connection between community engagement and nine dimensions of digital literacies, in hopes of helping practitioners in public librarianship to more effectively make sense of their approach. This model also has notable implications for library and information science education.

K-POP GENRES: A CROSS-CULTURAL EXPLORATION

Jin Ha Lee, Kahyun Choi, Xiao Hu, J. Stephen Downie

Current music genre research tends to focus heavily on classical and popular music from Western cultures. Few studies discuss the particular challenges and issues related to non-Western music. The objective of this study is to improve our understanding of how genres are used and perceived in different cultures. In particular, this study attempts to fill gaps in our understanding by examining K-pop music genres used in Korea and comparing them with genres used in North America. We provide background information on K-pop genres by analyzing 602 genre-related labels collected from eight major music distribution websites in Korea. In addition, we report upon a user study in which American and Korean users annotated genre information for 1894 K-pop songs in order to understand how their perceptions might differ or agree. The results show higher consistency among Korean users than American users demonstrated by the difference in Fleiss' Kappa values and proportion of agreed genre labels. Asymmetric disagreements between Americans and Koreans on

specific genres reveal some interesting differences in the perception of genres. Our findings provide some insights into challenges developers may face in creating global music services.

SUPPORTING ALL LEARNERS IN SCHOOL-WIDE COMPUTING: A CROSS-CASE QUALITATIVE ANALYSIS

Martin Wolske, Maya Israel, George Reese, Avigail Snir

National initiatives have raised awareness of the importance of developing innovative approaches to science, technology, engineering, and math (STEM) education. Research indicates the essential need to capture the interest of students in STEM before they graduate from middle school. This poster will highlight approaches to advance computational thinking currently underway within a small Midwestern elementary school to research the impact of computing at the elementary school level. We especially emphasize work with students at risk of academic failure. We take a multi-stakeholder community building approach to the development of computational thinking, tapping into the school librarian as engagement leader. Building from the assets brought to the initiative by school administrators, teachers, librarians, and technical staff, as well as students, parents, university, and community leaders, the approach creates a school-wide and community-deep community of learners to share innovative approaches and lessons learned. A multidisciplinary team from the University of Illinois, including the Center for Digital Inclusion (CDI), Office for Math, Science, and Technology Education (MSTE), and students from the Graduate School of Library and Information Science (GSLIS) have served as advocates, trainers, and lead evaluators in this elementary school-led initiative. Preliminary findings will be reported based on our four research questions that guided the cross-case qualitative study of this initiative: 1. How was programming integrated into instruction across different instructional contexts? 2. What barriers to implementation occurred during the adoption of programming into instruction? 3. How did teachers in different instructional contexts support struggling learners, including students with disabilities and those at risk for academic failure due to poverty? 4. How did students engage in programming within the different instructional contexts?

INSECT BYTES: USING INFORMATICS TO SUPPORT DISEASE CONTROL

Ian Brooks

The successful control of endemic diseases requires that decision makers have timely access to information, including clinical, public health, environmental, financial, economic, and satellite imagery data. Our goal is to develop a public health information system that can: (1) handle all these unique sources of information, (2) provide decision makers access to models for combining them, and (3) be scaled both globally and across diseases. This talk will describe the development of advanced information systems at GSLIS and the National Center for Supercomputing Applications, in particular, INDICATOR, a system for infectious disease outbreak detection, modeling, and response that is currently in production in Illinois. It will also discuss plans to extend INDICATOR to support control of endemic disease.

SENTINETES: USER CLASSIFICATION BASED ON SENTIMENT FOR SOCIAL CAUSES WITHIN A TWITTER NETWORK

Shubhanshu Mishra, Sneha Agarwal, Jinlong Guo, Kirstin Phelps, Johna Picco

Available sentiment classifiers typically describe statements as either positive or negative. While helpful for consumer products or marketing initiatives, this sort of binary classification is limiting for other types of sentiments, particularly those related to social causes. Our research contribution is the

creation of new orthogonal sentiment classifiers unique to social causes. This new classification helps capture a more nuanced sentiment along level of support (enthusiastic/passive) and the degree of enthusiasm (enthusiastic/passive) toward a cause. Twitter data is noisy and content specific, making it difficult for any topic-specific approach. However, our findings show that Enthusiastic and Supportive tweets were more densely present in tweets about social causes in Twitter. Our research takes a computational approach to address how social media data, with a better classification of sentiment analysis for social causes, can be maximized by individuals and agencies. With a more nuanced classifier, users within social networks more receptive to social causes can be more easily identified for collective action and advocacy.

“SHERLOCK HOLMES MEETS GODZILLA”: FILM DESCRIPTION THROUGH COMPARISON

Peter Organisciak, Michael Twidale

Describing things briefly, clearly and well is hard work. We know that – we study it and try to do it in many parts of LIS. Describing a movie is quite a challenge – especially if you are trying to do a good job in just a few words. Using a set of nearly 8 million Amazon user reviews of films [1] we find that some people are able to use a very terse and yet surprisingly effective way of describing some aspects of what makes the movie stand out - a qualified mixture of other films. In the world of mixtures, Daddy Day Care becomes “a cross between Mr. Mom and Kindergarten Cop”, Looper is “12 Monkeys meets The Terminator” and The Incredibles plays as “a cross between Toy Story, Superman, and Office Space.” These descriptions seem to get to the heart of the movie, in a way that many people who have seen the movie can agree with. They are clearly inspired by the popular culture view of the movie pitch – where an idea for a movie has to be described to busy executives as clearly and quickly as possible. For example, in reviews for the science fiction family film Super 8, we saw recurring patterns of films used to alternately describe the film style (e.g., Cloverfield, Jurassic Park), story themes (e.g., Aliens, Close Encounters) and character themes (e.g., Stand By Me, The Goonies, E.T.). We explore examples of this activity by Amazon user reviews to understand how, why and when they are effective and speculate on the potential of this approach to inform novel information organization and access.

[1] J. McAuley and J. Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. WWW, 2013.

SUMMARIZING SCIENTIFIC TEXT WITH TYPED-DEPENDENCY GRAPHS

Henry A. Gabb

American grade-schoolers are sometimes taught how to diagram sentences to illustrate correct grammar and the subtleties of sentence structure and good writing. It emphasizes the relationships among words rather than their order in a sentence. In theory, a sentence that is difficult to diagram is likely to be grammatically incorrect or incomprehensible. Modern typed-dependency (TD) parsers generate sentence graphs that are similar in spirit to grade school sentence diagrams. An algorithm will be presented that uses the subject-verb, verb-object, and various object relationships from TD graphs to retain the salient information from scientific text while achieving good compression ratios. This approach is compared to a similar method that uses part-of-speech tags instead of TDs. The inherent limitations of the TD-based approach and potential solutions will be also discussed. This summarization technique relies on the theory of scientific sublanguages developed by Zellig Harris, wherein a practitioner from a particular scientific discipline naturally orders words and supplies missing information based on knowledge of that discipline. Taken a step further, it might even be

possible to use the TD-based approach described above to automatically extract the grammatical rules of a particular scientific sublanguage from a corpus of articles from that discipline.

A THEORETICAL MODEL FOR TRUST IN AUTOMATED SYSTEMS

Masooda Bashir, Kevin Hoff

The concept of trust in automation has received a great deal of attention in recent years in response to modern society's ever-increasing usage of automated systems. Researchers have used a variety of different automated systems in unique experimental paradigms to identify factors that regulate the trust formation process. In this work-in-progress report, we propose a preliminary, theoretical model of factors that influence trust in automation. Our model utilizes three layers of analysis (dispositional trust, situational trust, and learned trust) to explain the variability of trust in a wide range of circumstances. We are in the process of verifying certain aspects of the model empirically, but our current framework provides a useful perspective for future investigations into the intricacies of trust in automated systems.

NERDFIGHTERIA'S LITERARY SALON: UNDERSTANDING THE SOCIAL NETWORKS IN AN ONLINE COMMUNITY OF READERS AND AUTHORS

Alaine Martaus

This poster presents a visual representation of the complex network of social and new media connections at the core of the Nerdfighter community. Understanding this network provides insights into how this online community connects readers and authors and creates new perspectives on the relationship between literature and new media. In January 2007, author John Green and his brother Hank established their Vlogbrothers YouTube channel, which quickly garnered a significant young adult fan-base and a related online community known as the Nerdfighters. At the same time, their videos often featured other popular young adult authors, who in turn linked their own online content to that of the Vlogbrothers. These social media-driven connections, both between the authors themselves and between the authors and their readers, resulted in a kind of online literary salon, in which authors engage in a conversation about books and reading in full view of, and with the sometime participation of, a young adult audience. The purpose of my research is to analyze this on-going conversation about the reading experience, exploring the way that it intersects with theories about the relationship between books and new media. To fully understand the function and development of the conversation itself, it is essential to recognize the social networks that connect those participating in it. This poster therefore presents a visual representation of the increasingly complex network of inter-referential connections at the core of the Nerdfighter community. By mapping these relationships across various social and new media environments, including YouTube, Twitter, Tumblr, and individual author blogs, the poster provides insights into a variety of important questions: how online communities of readers and authors are created and maintained, how important voices in the community are established and recognized, and how those voices present a collective ideal about the value of the reading experience.

FROM INFORMATION TO INNOVATION: EXPLORING SCIENCE TEACHERS' LEARNING AND INFORMATION PRACTICE FOR INNOVATION IN SCIENCE TEACHING

Wei Gao

Research in educational innovation has mostly focused on scaling up innovation and expediting the diffusion of innovations in the context of top-down approaches to educational change and school

reform. In contrast, there is a dearth of studies that focus on individual teachers as active agents of change. As innovation diffusion scholar Everett Rogers states, the decisions and events occurring previous to diffusion have a strong influence on the diffusion process, thus understanding the activities and decisions of individual science teachers in terms of how an innovation begins and how a new idea is implemented is essential to realizing the ultimate goal of diffusing successful innovations. Since teachers are ultimately responsible for transforming the science classroom, their experiences with and perceptions of instructional innovations need to be understood in order to support teachers' innovation efforts. Ignoring the experiences of teachers in the frontline of teaching misses the lessons that can be learned from their experiences and perspectives. This study aims to fill this gap by focusing on the experiences and perspectives of science teachers with regard to innovation in science teaching, particularly at the generation and implementation stages of the innovation development process. The research explores the processes in which science teachers seek and utilize information from myriad learning resources to initiate and implement innovative teaching modalities. The study uses qualitative in-depth interviews to examine science teachers' experiences with innovations in teaching. Online science teacher learning community, science teaching discussion forums, blogs, Internet videos and other online materials may be used as secondary data for the purpose of data triangulation. The study can contribute to research and scholarship in educational innovation by revealing the ways science teachers learn and implement new ideas as well as factors that contribute to success and failure when science teachers implement innovations in teaching.

Poster Session 2

AUTOMATICALLY SUMMARIZING MEDICAL LITERATURE

Ana Lucic, Catherine Blake

Scientific articles in medicine often report results in the form of a comparison, where a new intervention is compared with either a control group or another intervention (or both). Comparisons are particularly important for Comparative Effectiveness

Research (CER) which recently recommended that "CER should directly compare tests or active treatments - so called head-to-head comparisons - of viable clinical alternatives within the current stand or practice (which in some cases may be no intervention)" (Sox, Helfand et al. 2010). Although from a language processing perspective comparison sentences have been called "almost notorious for its syntactic complexity" (Bresnan, 1973), our group has been developing methods that automatically discern comparison from non-comparison sentences (Park & Blake, 2012). This work addresses the more difficult task of associating a particular role to noun phrases in a comparison sentence. Specifically our goal is to automatically identify the two entities (agent and object) that are being compared and the way in which those entities are compared (basis of comparison). We describe how this goal can be framed as a classification problem, the features used, and the preliminary results which show an average accuracy of 75%, 85% and 72% for the agent, object and basis of the comparison respectively using a Support Vector Machine (SVM) classifier on sentences that are less than 40 words long.

Bresnan, W. J. (1973). "Syntax of the Comparative Clause Construction in English." *Linguistic Inquiry* 4(3): 275-343.

Hoon Park, D, & Blake, C. (2012). "Identifying comparative sentences in full-text scientific articles." Association of Computational Linguistics, Workshop on Detecting Structure in Scholarly Discourse, July 12, Jeju South Korea.

Sox, H. C., M. Helfand, J. Grimshaw, K. Dickersin, P. L. M. Editors, D. Tovey, J. A. Knottnerus and P. Tugwell (2010). "Comparative effectiveness research: challenges for medical journals." PLoS Med 7(4): e1000269.

THE IMPACT OF GROUND OZONE ON ASTHMA: A CASE STUDY USING PROJECT INDICATOR IN CHAMPAIGN-URBANA

Elizabeth J Surbeck, Catherine Blake

The purpose of this analysis is to examine the correspondence between short-term ozone exposure and asthma-related hospital reports, adopting from the methods used in "Meta-analysis of the Association between Short-Term Exposure to Ambient Ozone and Respiratory Hospital Admissions" by Meng Ji, Daniel S. Cohan, and Michelle L. Bell. This meta-analysis was published in 2011 on studies that have connected asthma cases in hospitals and ozone measurements in air quality reports. Studies seem to suggest that there is a strong correlation between asthma and ozone level increases in the outdoor air. The point of duplicating their methods is to see if one could take real time collected data from a town area like Champaign-Urbana and prove or disprove the results and methods of this meta-analysis. The datasets to attempt fulfilling this purpose have been drawn from Project Indicator and the Environmental Protection Agency's (EPA) Air Quality System (AQS). The tools required to accomplish this analysis included R and MS Excel and methods for analysis included linear regression to understand the interaction of the quantitative data from the two main data variables. The results of the study suggested the possibility of a connection between asthma and ozone though additional factors such as weather patterns may provide a better defined answer to this analysis's concerns.

INVESTIGATING WRITERS' ATTITUDES BY MINING A LARGE CORPUS OF BOOKS

Sayan Bhattacharyya

Given a large corpus of digitized books, such as the HathiTrust corpus, it is useful for researchers in history or literary studies to have an efficient discovery mechanism that can identify which texts from the large collection of texts is likely to repay close study, in connection with their specific research question. Also, the research questions that interest researchers can often be envisioned as compositional, made up of multiple distinct parameters. Our work-in-progress approaches the issue in a scalable way by utilizing the compositionality of research questions to treat the problem as one of searching for collocations using a list-based approach. Our work-in-progress aims to be of use to researchers within the restrictive limitations on access on derived data likely to be imposed by copyright rules and their implications. Our use case involves investigating similarities and differences in attitudes of French language and English-language writers towards women's work in the colonized world. Our approach is to use co-occurrences of the elements that constitute the decomposable constituents of this investigative question after starting with available bibliographic metadata as an initial filter. We are also exploring methods of validation and tuning, as well as analytical comparison across defined segments of the dataset.

USING COLLECTIONS AND WORKSETS IN LARGE-SCALE CORPORA: PRELIMINARY FINDINGS FROM THE WORKSET CREATION FOR SCHOLARLY ANALYSIS PROJECT

Harriett Green, Katrina Fenlon, Megan Senseney, Sayan Bhattacharyya, Craig Willis, Peter Organisciak, J. Stephen Downie, Tim Cole, Beth Plale

Scholars from numerous disciplines rely on collections of texts to support research activities. On this diverse and interdisciplinary frontier of digital scholarship, libraries and information institutions must 1) prepare to support research using large collections of digitized texts and 2) understand the different methods of analysis being applied to the collections of digitized text across disciplines. The HathiTrust Research Center's Workset Creation for Scholarly Analysis (WCSA) project conducted a series of focus groups and interviews to analyze and understand the scholarly practices of researchers that use large-scale, digital text corpora. This poster presents preliminary findings from that study, which offer early insights into user requirements for scholarly research with textual corpora.

SPECIALIZATION IN DATA CURATION: PRELIMINARY RESULTS FROM AN ALUMNI SURVEY, 2008-2012

Cheryl A. Thompson, Megan Senseney, Karen S. Baker, Virgil E. Varvel, & Carole L. Palmer

Data curation is an emergent field requiring a workforce with new skills to meet digital data needs across a wide variety of domains. This study surveyed masters-level graduates with a Specialization in Data Curation from the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign. The survey assesses the value of the program to graduates, career placement, and continuing education needs. Preliminary findings and trends are reported to inform the evolution of data curation education in library and information science (LIS) programs.

AN EXPLORATION OF THE H1N1 OUTBREAK IN CHAMPAIGN-URBANA ELEMENTARY SCHOOLS DURING THE 2009-2010 INFLUENZA SEASON

Christopher Komisarz, Ian Brooks

The H1N1 influenza outbreak presented a unique opportunity to track infection, treatment, and vaccination rates in real time due to advancements in technology. Local health departments worldwide were able to adapt their response to best fit their community's needs. Geographic analysis played a large part in response and determining where there were areas of high infection rates or low vaccination coverage. In using this kind of technology there are numerous analyses that can be achieved. This project examines the absence data among elementary school students in Champaign and Urbana. What is unique about these school districts is that the composition of the schools differs by design with Urbana schools only taking students from a specific area or neighborhood and Champaign schools being non-selective. Analysis of these data showed that there is a difference both in time and severity between the two school districts. Urbana schools show a distinct time of increased absences over the course of October 2009, while Champaign schools show a peak in absences in mid-October. There were no significant correlations among absences and low-income composition, school population, or ethnic composition. This analysis can be used to better inform public schools and local health departments should another outbreak like this occur in the future.

USABILITY AND DESIGN ANALYSIS: GOOGLE HANGOUTS

Sean Gordon, Taylor Kirch, Adriel Flores-Pagnani, Michael Twidale

Information systems are only as good as their interface. If a user cannot efficiently and consistently access the information contained in the system without training, an interface redesign should be considered. When Google redesigned Google Talk into Hangouts, users were less able to utilize the tool for communication and collaboration. Our project focused on applying qualitative user-centered design techniques to address the shortcomings of Hangouts. We utilized evaluation tools such as heuristics for usability, cognitive walkthroughs and user testing to inform our rapid prototyping techniques. Corporations often favor quantitative over qualitative methods, such as eye tracking, which require expensive hardware and many test subjects before patterns emerge that can be analyzed. These expensive and time consuming quantitative methods yield little benefit if any benefit over the qualitative user-oriented design process highlighted in our usability and design analysis. Through these techniques we found that Google Hangouts could be made more usable for collaboration and communication. Our findings indicated that Hangouts would benefit from a universal design, a reduction in branding language, and an increase in the functionality of the settings and contacts list. Our redesign focuses on addressing these findings.

USING NAMED ENTITY RECOGNITION AS A CLASSIFICATION HEURISTIC

Andrea K. Thomer, Nicholas M. Weber

As part of an on-going research project, we are exploring the role of acknowledgment statements found in a large corpus of bioinformatics texts to better understand collaborations between the diverse peoples, technologies, and research tools that produce computational biological knowledge. In particular, we want to better understand how successful interdisciplinary collaborative arrangements distribute credit, how material resources are cited, and how computational and biological knowledge have subtly blended in this field over time. Past work has shown that the “acknowledgments” of a journal article can be especially helpful in shedding light on the often neglected, or invisible work of collaboration (Cronin, Shaw and Labarre, 2003; 2004), especially in domains dependant on expert methodological knowledge and instrument building (Salager-Meyer et al, 2010). However, much of this prior work has relied on time consuming, manual classification methods; here, we wish to utilize text mining tools to speed and scale up our research. In this poster, we describe a method of using of Named Entity Recognition as a heuristic tool for improve manual classification of acknowledgement statements. Our research questions include the following:

- With little to no customization, can NLP tools like the Stanford Named Entity Recognizer (Stanford NER) help us initially evaluate the quality of a corpus of acknowledgment statements? And, can they identify “entity rich” acknowledgments on which we should focus our initial analysis?
- How effective are general, out-of-the-box NLP tools at recognizing entities in a domain specific corpus (such as bioinformatics)?
- How can we best leverage tools that deliver quantitative results (e.g., number of entities per acknowledgment statement) to support or aide further qualitative enquiry?

We demonstrate how uncertainty in our initial text mining results were ‘ground-truthed’ using Natural Language Processing tools in a quick-and-dirty fashion. To verify this technique’s validity, we offer some initial results from our larger study. [Originally presented at iConference 2014]

HOW WORKFLOW DOCUMENTATION FACILITATES CURATION PLANNING

Karen M. Wickett, Andrea K. Thomer, Karen S. Baker, Timothy DiLauro, Abigail E.

Asangba

The description of the specific processes and artifacts that led to the creation of a data product provide a detailed picture of data provenance in the form of a workflow. The Site-Based Data Curation project, hosted by the Center for Informatics Research in Science and Scholarship at the University of Illinois, has been investigating how workflows can be used in developing curation processes and policies that move curation "upstream" in the research process. The team has documented an individual workflow for geobiology data collected during a single field trip to Yellowstone National Park. This specific workflow suggests a generalized three-part process for field data collection that comprises three distinct elements: a Planning Stage, a Fieldwork Stage, and a Processing and Analysis Stage. Beyond supplying an account of data provenance, the workflow has allowed the team to identify 1) points of intervention for curation processes and 2) data products that are likely candidates for sharing or deposit. Although these objects may be viewed by individual researchers as "intermediate" data products, discussions with geobiology researchers have suggested that with appropriate packaging and description they may serve as valuable observational data for other researchers. Curation interventions may include the introduction of regularized data formats during the planning process, data description procedures, the identification and use of established controlled vocabularies, and data quality and validation procedures. We propose a poster that shows the individual workflow and our generalization into a three-stage process. We plan to discuss with attendees how well the three-stage view applies to other types of field-based research, likely points of intervention, and what kinds of interventions are appropriate and feasible in the example workflow. [Originally presented at AGU 2013]

INVESTIGATING RESEARCH DATA DESCRIPTION IN SCIENTIFIC JOURNAL PUBLICATIONS

Tiffany C. Chao

Metadata is essential in enabling the sharing and reuse of scientific research data yet remains immensely time and resource intensive to create and manage. A key component of this metadata is description about how data were generated, encompassing information on the research design and processes employed. While direct engagement with the data producer is an established technique to obtaining metadata, the use of scientific journal publications to identify contextual details about the research processes used for the production of data is an alternative and potentially timesaving approach that can contribute to metadata production for data. This poster presents an exploratory study to understand what information contained in journal publications can be used to inform metadata description for research data. A sample of full-text articles from the Earth Sciences published between 2006-2012 was examined using the Data Curation Profile (<http://datacurationprofiles.org/>) as a framework to guide analysis. Preliminary findings indicate three practices that are consistently described across all journal articles: sampling procedures for gathering data (i.e., physical samples), processing physical samples, and conducting statistical analysis on the processed data. The description of each practice included not only the data type but also referenced techniques and processes applied to data. While the Data Curation Profile provided the initial guidance for analysis, a more enhanced framework can be developed in the identification of data production processes within journal articles. With the anticipated growth of digital data, these findings can potentially contribute to the development of a systematic approach for enhancing metadata in data curation systems and services and fostering data reuse.

OUTCOMES OF THE “DATA CURATION FOR GEOBIOLOGY AT YELLOWSTONE NATIONAL PARK” WORKSHOP

Andrea K. Thomer, Carole L. Palmer, Bruce W. Fouke, Ann Rodman, G. Sayeed Choudhury, Karen S. Baker, Abigail E. Asangba, Karen M. Wickett, Timothy DiLauro, Virgil Varvel

The continuing proliferation of geological and biological data generated at scientifically significant sites (such as hot springs, coral reefs, volcanic fields and other unique, data-rich locales) has created a clear need for the curation and active management of these data. However, there has been little exploration of what these curation processes and policies would entail. To that end, the Site-Based Data Curation (SBDC) project is developing a framework of guidelines and processes for the curation of research data generated at scientifically significant sites. This poster reports outcomes from a workshop held in April 2013 at Yellowstone National Park (YNP) to gather input from scientists and stakeholders. Workshop participants included nine researchers actively conducting geobiology research at YNP, and seven YNP representatives, including permitting staff and information professionals from the YNP research library and archive. Researchers came from a range of research areas -- geology, molecular and microbial biology, ecology, environmental engineering, and science education.

Through group discussions, breakout sessions and hands-on activities, we sought to generate policy recommendations and curation guidelines for the collection, representation, sharing and quality control of geobiological datasets. We report on key themes that emerged from workshop discussions, including:

- participants’ broad conceptions of the long-term usefulness, reusability and value of data.
- the benefits of aggregating site-specific data in general, and geobiological data in particular.
- the importance of capturing a dataset’s originating context, and the potential usefulness of photographs as a reliable and easy way of documenting context.
- researchers’ and resource managers’ overlapping priorities with regards to ‘big picture’ data collection and management in the long-term.

Overall, we found that workshop participants were enthusiastic and optimistic about future collaboration and development of community approaches to data sharing. Outcomes from the workshop are guiding next steps in the SBDC project.

Posters with Demonstrations (during Poster Sessions 1 and 2)

THE ILLINOIS DISTRIBUTED MUSEUM PROJECT: ENGINEERING AND TECHNOLOGY INNOVATIONS AT THE UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Michael B. Twidale, Susan Frankenberg, Kelsey Heffren, Jamie Wittenberg

****during poster session 1 only****

This project is focused on developing techniques to educate students, faculty, alumni, prospective students, and the general public on the historically significant inventions and developments by faculty and students of the College of Engineering at UIUC. At present there are many exhibits and artifacts on display about the engineering and technological innovations developed at UIUC. These are spread across the campus in locations including the Spurlock Museum, the Sousa Archives, and many individual department’s lobbies and hallways. In addition, many more artifacts and exhibits are hidden in labs, offices, and closets. Finally, there are many innovations that have occurred at UIUC

that have no physical artifacts or markers but are represented on the Internet or in books and documents. The ultimate goal of this project is to develop an Illinois Distributed Museum of Engineering and Technology Innovations using the latest web technology to give the user an experience as deep as being in a traditional physical museum even though the artifacts are distributed across multiple locations. Our vision is to showcase the place of the University of Illinois as a catalyst of innovation, bringing together different people and different ideas, and sparking new unexpected solutions. To our knowledge, nothing like this has been attempted on this scale. Thus we are working on the cutting edge of the field of museum informatics. We have created one functioning prototype, but we are constantly experimenting and refining our work with whatever new technology we find. Our hope is that the Illinois Distributed Museum will one day not only contain information on important people and ideas from all of the units on the Illinois campus, but showcase that information in new and exciting ways. You can check out our prototype at distributedmuseum.blogspot.com (keep in mind that it is still very much a work in progress). We also are experimenting with Omeka, and would be very happy to show our efforts to utilize that technology as well.

MEASURES OF NOVELTY IN BIBLIOMETRIC MODELS

Vetle Torvik, Shubhanshu Mishra

This research work is aimed at identifying novel topics and ideas in a given PubMed record. It will identify topics and ideas using which MeSH terms and MeSH pairs, respectively for a given PubMed record are novel, hot and passé during the year of publication of that record. As a part of this research work a tool called Novelity will be developed which will focus on identifying MeSH terms among the Organism, Disease and Drug MeSH category are novel, hot and passé for the given PubMed record ID. The tool will also produce useful visualizations which will be helpful in identifying the growth of MeSH terms present in a given PubMed record. The tool will train piece wise regression models for identifying the curve of a given MeSH term.

PATCI - A CITATION MATCHER OPTIMIZED FOR NON-PATENT CITATIONS

Vetle Torvik, Miles Lincoln, Sneha Agarwal, Haoyan Cai

Patents, the 'innovation' literature, are increasingly citing academic literature. It helps us to see how scientific research is driving new discoveries. Although relatively easier for the human brain, identification of different parts of a citation like, publication year, journal volume, page numbers, from just a string is not an easy task for the computer. Using them to find the best match for that citation is another problem. Given a non-patent citation of a US Patent, Patci looks up for the best match in bibliographical database. This match is found using a few hundred attributes, created by combination of one or more features of the citation, that have been identified as important using a trained logistic regression model. Patci is an important resource for linking industrial and academic literature and motivates further research on the same.

USEFUL APPLICATIONS FOR CREATIVE COLLABORATION ON LARGE SCALE TOUCH DEVICES

Christopher Nixon, Michael Twidale

I am exploring useful applications for creative collaboration on large scale touch devices, utilizing two recent technologies, one which uses the Microsoft Kinect to turn any flat, white surface into a touch screen, and another which consists of a flexible transparent plastic sheet, similar to those used for overhead projector transparencies. Both of these technologies have the advantage of being relatively

affordable compared to traditional touch screen technologies, and they also have their own novel benefits. The Kinect-based technology can detect both touches on the surface as well as gestures in three dimensional space in front of the surface. The plastic sheet can be bent to adhere to curved surfaces, such as a pillar, and is also transparent so can be attached to a window or computer monitor. By focusing on exploring the possible interactions these new technologies afford, particularly in the space of creative collaboration, I have created a prototype of an online whiteboard software which can be used with these as well as other touch sensitive technologies. I am using the design thinking process to iteratively design, build, test, and evaluate my prototype.

Center Posters (Displayed during Poster Sessions 1 and 2)

CENTER FOR CHILDREN’S BOOKS (CCB)

Deborah Stevenson

The Center for Children’s Books (CCB) is a crossroads for critical inquiry, professional training, and educational outreach related to youth-focused resources, literature, and librarianship. The Center’s mission is to facilitate the creation and dissemination of exemplary and progressive research and scholarship related to all aspects of children’s and young adult literature; media and resources for young (ages 0-18) audiences; and youth services librarianship.

CENTER FOR DIGITAL INCLUSION (CDI)

Jon Gant

The Center for Digital Inclusion (CDI) fosters inclusive and sustainable societies through research, teaching, and public engagement about information and communication technologies (ICT) and their impacts on communities, organizations, and governments. Digital inclusion encompasses not only access to the Internet but also the availability of hardware and software; relevant content and services; and training for the digital literacy skills required for effective use of ICT.

CENTER FOR INFORMATICS RESEARCH IN SCIENCE AND SCHOLARSHIP (CIRSS)

Carole L. Palmer, Catherine Blake

An overview of the activities of the GSLIS Center for Informatics Research in Science and Scholarship (CIRSS), including highlights of selected current projects. CIRSS conducts research on information problems that impact scientific and scholarly inquiry, with projects and activities that focus on how digital information can advance the work of scientists and scholars, the curation and analysis of research data, and the integration of information within and across disciplines and research communities. Within CIRSS, the Socio-technical Data Analytics (SoDA) Group design, develop, and evaluate new technologies in order to better understand the dynamic interplay between information, people and information systems.

Current projects include Site-Based Data Curation at Yellowstone National Park (SBDC); Developing a Model for Socio-Technical Data Analytics Education (SODA); Data Curation in Education Research Centers Program (DCERC); Digital Humanities Data Curation: NEH Institutes for Advanced Topics in the Digital Humanities (DHDC).

HATHI TRUST RESEARCH CENTER (HTRC)

J. Stephen Downie

The HTRC is a collaborative research center launched jointly by Indiana University and the University of Illinois, along with the HathiTrust Digital Library. The HTRC is developing computational research access to some 10 million volumes (3 billion pages) in the HathiTrust corpus. At present, the HTRC enables computational access to published works in the public domain, and the Center is developing a secure computational and data environment for conducting non-consumptive research against digitized works in copyright. Ultimately, the HTRC aims to allow scholars to fully utilize the contents of the HathiTrust corpus while preventing intellectual property misuse within the confines of current United State copyright law. Visit <http://www.hathitrust.org/htrc> for more information.