



The iSchool at Illinois

2015 RESEARCH SHOWCASE



FRIDAY, APRIL 3

Schedule of Events

12:00 – 1:30 p.m.

***Poster Session 1 (131 LIS) and
Pizza (East Foyer)***

1:30 – 3:30 p.m.

Presentations (126 LIS)

3:30 – 5:00 p.m.

***Poster Session 2 (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 (131 LIS) and Pizza (East Foyer)*

PATRON PRIVACY IN JEOPARDY: A PILOT ANALYSIS OF THE PRIVACY POLICIES OF DIGITAL CONTENT VENDORS IN AMERICAN PUBLIC LIBRARIES

April Lambert, Michelle Parker, Masooda Bashir

ON THE NEED FOR COMBAT LIBRARIANS: A CASE STUDY OF KNOWLEDGE MANAGEMENT AT THE END OF THE IRAQ WAR

Caryn Anderson

PARAMETER TUNING: EXPOSING THE GAP BETWEEN DATA CURATION AND EFFECTIVE DATA ANALYTICS

Catherine Blake, Henry Gabb

CLOSING THE APP GAP: FROM PLAN TO PROJECT I

Deborah Stevenson, Kate McDowell, Cass Mabbott

A NEW PERSPECTIVE OF STUDYING USERS' ATTITUDES TOWARD INFORMATION PRIVACY AND CULTURE IN GLOBAL CONTEXT: AN EXPLORATIVE STUDY

Hsiao-Ying Huang, Kim Naples, Masooda Bashir

AUTOMATING THE IDENTIFICATION OF CRITICAL COMMUNITY HEALTH INDICATORS AND THEIR PRESENTATION TO THE PUBLIC

Robert Sarwark, Julia Bello-Bravo, Barry R. Pittendrigh, Tom Szyrka, Ian S. Brooks

INCREMENTAL SENTIMENT PREDICTION BASED ON HUMAN IN THE LOOP LEARNING

Jana Diesner, Shubhanshu Mishra, Liang Tao, Elizabeth Surbeck, Jason Byrne

EXPERT-GUIDED CONTRASTIVE OPINION SUMMARIZATION FOR CONTROVERSIAL ISSUES

Jinlong Guo

IMPACT OF ENTITY RESOLUTION ON NETWORK PROPERTIES

Jinseok Kim, Jana Diesner

DIGITAL INNOVATION LEADERSHIP PROGRAM (DILP)

Jon Gant

INFORMATION INTEGRATION: A CASE STUDY OF AIR QUALITY IN CHICAGO AND ST. LOUIS

Jooho Lee, Catherine Blake

COLLECTIVE LEADERSHIP FOR COLLABORATIVE ACTION: A NEW LEADERSHIP FRAMEWORK FOR LIBRARY ENGAGEMENT AROUND DIGITAL LITERACY EDUCATION

Kirstin Phelps

THE IMPACT OF MUSIC GENRE ON MUSIC SIMILARITY: A QUALITATIVE EXPLORATION

Yun Hao, Kahyun Choi, J. Stephen Downie, Ning Chen

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

WHO ARE THEY, AND IF SO, HOW MANY? PROPAGATION OF ENTITY DISAMBIGUATION ERRORS TO NETWORK ANALYSIS RESULTS

Jana Diesner (presenter), Craig Evans, Jinseok Kim

MECHANICAL LITERACIES OF THE INDUSTRIAL REVOLUTION

Elizabeth Hoiem

THE GSLIS CARNEGIE SCHOLARS

Nicole A. Cooke

PORNOGRAPHY, BOMB-BUILDING, AND GOOD INTENTIONS: WHAT WOULD IT TAKE FOR AN INTERNET FILTER TO WORK?

Emily Lawrence (presenter), Richard Fry

ADVANCED TOOLS AND TECHNIQUES FOR LOGIC-BASED KNOWLEDGE REPRESENTATION, PROCESS DOCUMENTATION, AND DATA CURATION

Bertram Ludäscher

#DIVERSE CHILDREN'S LITERATURE: EXAMINING SOCIAL MEDIA'S ROLE

Melissa Hayes, DeAnza Williams

THE HATHITRUST+BOOKWORM PROJECT: EXPLORING CULTURAL AND LITERARY TRENDS IN MILLIONS OF SCANNED BOOKS

Peter Organisciak (presenter), Loretta Auvil, Benjamin Schmidt, Sayan Bhattacharyya, Mari Colleen Fallaw, Muhammad Saad Shamim, Yun Hao, Robert McDonald, Erez Aiden Lieberman, J. Stephen Downie

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

THE IMPOSITION OF EMPATHY: THE PEDAGOGY OF EMPATHIC DESIGN

Andrea Gannon, Michael Twidale

COLORADIO: A PRODUCT OF INTERDISCIPLINARY COLLABORATION IN AN ISCHOOL ENVIRONMENT

Christopher Nixon, Kinyetta Nance, Berenice Vargas, Helen Zhou

CONCEPTUALIZING WORKSETS FOR NON-CONSUMPTIVE RESEARCH

Jacob Jett, Chris Maden, Colleen Fallaw, Megan Senseney, J. Stephen Downie

EDUCATE, EMPOWER, ENGAGE: THE CENTER FOR DIGITAL INCLUSION (CDI)

Jon Gant, Kirstin Phelps, Martin Wolske

TOPIC MODELING USERS' INTERPRETATIONS OF SONGS TO INFORM SUBJECT ACCESS IN MUSIC DIGITAL LIBRARIES

Kahyun Choi, Jin Ha Lee, Craig Willis, J. Stephen Downie

LIS ONLINE EDUCATION: COMPARING LEEP WITH OTHER PROGRAMS

Kate McDowell, Wei Gao, Caroline Nappo, Mikki Smith

MINING THE POACHERS: QUANTITATIVE ANALYSIS OF FANFICTION WORKS ACROSS TWO ARCHIVES

Elizabeth Wickes

DIGITAL LITERACY FOR ALL LEARNERS

Martin Wolske, Kim Naples, Hailley Fargo, Sarah Butt, Travis Faust

DIGITAL HUMANITIES DATA CURATION INSTITUTES: CHALLENGES AND PRELIMINARY FINDINGS

Megan Senseney, Trevor Munoz, Julia Flanders, Ali Fenlon

TARGET FUNCTION ASSESSMENT FOR IMPACT GOALS OF ISSUE FOCUSED PUBLIC MEDIA

Rezvaneh Rezapour, Jana Diesner

EXTRACTING FEATURES FROM TEXT FOR NON-CONSUMPTIVE READING WITH THE HATHITRUST RESEARCH CENTER

Sayan Bhattacharyya, Boris Capitanu, Ted Underwood, Peter Organisciak, Loretta Auvil, Colleen Fallaw, J. Stephen Downie

EXPLORING CULTURAL DIFFERENCES IN LANGUAGE USAGE: THE CASE OF NEGATION

Sveta Stoytcheva, Dov Cohen, Catherine Blake

Posters with Demonstrations (during Poster Sessions 1 and 2)

THE DIGITAL INNOVATION LEADERSHIP PROGRAM (DILP)

Lisa Bievenue, Jeff Ginger, Kinyetta Nance, Jarai Carter

MAPS - APPLYING LIS TO CULTURAL HERITAGE

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)

Bertram Ludäscher, Catherine Blake

HATHI TRUST RESEARCH CENTER (HTRC)

J. Stephen Downie

Presentation Abstracts

WHO ARE THEY, AND IF SO, HOW MANY? PROPAGATION OF ENTITY DISAMBIGUATION ERRORS TO NETWORK ANALYSIS RESULTS

Jana Diesner (presenter), Craig Evans, Jinseok Kim

Electronic archives of communication and collaboration data are a common source for constructing social network data. This approach to acquiring network data is efficient, but bears the risk of entity disambiguation errors, which come in two forms: First, failing to consolidate instances of the same entities that were recorded with different surface forms. Second, erroneously merging entities with identical surface forms that truly represent different individuals. How much do these errors matter for network analysis results and any conclusions drawn? We provide an answer to this question based on a series of controlled experiment, where we compare the results from using 1) convenient, raw, off-the shelf data or disambiguation approaches to 2) moderately refined solutions to 3) the best disambiguation solutions we could obtain or produce. For this purpose, we use two largescale, over-time data sets that are primarily subject to merging (email data) and splitting (coauthorship data). Our findings suggest that analysis results in terms of common network metrics, topologies and key players as well as derived implications can change strongly depending on disambiguation resolution efforts. This means any observed changes in the networks we construct are independent from any underlying social processes; an attribution error that might be made if not controlling for these effects. We observe that for email data, failing to merge alternative address references to the same individual can let organizational communication networks appear less cohesive, integrated and small than they really are. This might lead to a falsely perceived need for more communication. For co-authorship data, failing to keep apart distinct individuals who happen to have the same name can make scientific collaboration networks seem to be more integrated, prolific and cohesive than they really are, and individual researchers look more productive and diversified. This might erroneously reduce the perceived need for supporting collaborations, e.g., in terms of research funding.

MECHANICAL LITERACIES OF THE INDUSTRIAL REVOLUTION

Elizabeth Hoiem

My presentation considers how perceptions of early literacy and children's books shifted during Britain's first industrial revolution (1780-1850), when educators conceptualized learning and reading in materialist terms. As one periodical explains in 1826, schools must provide students with "sensible signs." Whereas "the old plan of instruction was to teach by words and books," cultivating the

“memory,” in the “new system, words and books are mere instruments towards the elucidation of things themselves, or of their representatives, submitted to your senses.” Called “object learning” or “the education of things,” these pedagogical practices used a child’s physical awareness of her body and her observations of her immediate environment as the foundation for learning to read. Many children’s authors and book publishers helped to publish books and ephemera used for these reading practices, including moveable books, literacy games, model machines, mechanical toys, and other sense-engaging books. The investment of early experiential educators in object lessons, which they considered an indispensable accompaniment to book learning, suggests that children of the industrial age experienced texts and objects in nurseries and classrooms as interdependent tools. I suggest that reading, during the early industrial revolution, was closely tied to other manual literacies, which were equally associated with social mobility and class hierarchy. Object learning claimed to empower children by placing them above the elements composing their environment, which they control through what I call “mechanical literacy”—that is, by learning the dependable laws governing how things are sensed, manipulated, created, purchased, manufactured, and exchanged. Just as today we discuss “digital literacies” or “information literacy,” parents in the late-eighteenth and nineteenth centuries were anxious for their children to acquire these “mechanical literacies.” Some of the ways mechanical literacies were taught continue in today’s K-12 curriculum and have experienced a revival through such movements as library Maker Spaces.

THE GSLIS CARNEGIE SCHOLARS

Nicole A. Cooke

During the years of 1970-1972, under the supervision, advocacy, and mentorship of former faculty member Terry Crowley, GSLIS recruited and graduated two diverse cohorts of students, the Carnegie Scholars. Composed of 28 students from underrepresented backgrounds, the scholars completed graduate work in the Masters of Science program and successfully entered the field through a grant funded by the Carnegie Foundation of New York. Many of the scholars went on to be leaders in the library profession. Not since that time has GSLIS had so many diverse candidates enrolled at one time. This research consists of two phases: 1) compiling the archival documents that exist about this groundbreaking group of students,* and 2) interviewing as many of the Carnegie Scholars as possible. It is hoped that the stories and experiences of the scholars will help inform the narrative about the recruitment, retention, and mentorship of underrepresented library and information science students at the local and national levels. Learning from the successes and failures of this program will enable GSLIS to strategize ways to improve its educational programs and plan for the future, successful recruitment and retention of diverse students. These are lessons that will be applicable to the field at large. *This presentation will focus on the 1st phase of the study, the archival research.

PORNOGRAPHY, BOMB-BUILDING, AND GOOD INTENTIONS: WHAT WOULD IT TAKE FOR AN INTERNET FILTER TO WORK?

Emily Lawrence (presenter), Richard Fry

Librarians often object to Internet filters on the grounds that filters are prone to overblocking and underblocking. This argument implies that a significant problem with contemporary filters is that they are insufficiently fine-grained. In this paper, we posit that present-day filters will always be conceptually capable of failure, regardless of how granular their content analysis becomes. This is because, we argue, objections to content are best understood as objections to problematic interactions between content and particular knowers. We import the concept of the *situated knower* from feminist epistemology to capture the heterogenous, socially embedded nature of patrons, about whom we cannot make blunt generalizations for filtering purposes. A successful filter would need information about these differently situated patrons, the content they seek, and the interactions between the two. We conclude that a genuinely successful Internet filter would therefore need to be both mind-reading and fortune-telling.

ADVANCED TOOLS AND TECHNIQUES FOR LOGIC-BASED KNOWLEDGE REPRESENTATION, PROCESS DOCUMENTATION, AND DATA CURATION

Bertram Ludäscher

Euler is a toolkit to solve taxonomy and classification alignment problems by employing logic-based reasoning engines, e.g., based on first-order logic, answer set programming, and other reasoners. Euler can detect and diagnose inconsistent inputs and then suggest repairs, i.e., minimal changes to the user-provided input articulations that eliminate the logical inconsistency. YesWorkflow (YW) is a set of tools that aim to provide users of scripting languages with many of the benefits of scientific workflow systems. YW requires neither the use of a workflow engine nor the overhead of adapting code to run effectively in such a system. Instead it enables scientists to annotate existing scripts with comments that reveal the computational modules and dataflows otherwise implicit in these scripts. YesWorkflow tools extract and analyze these comments, represent the scripts in terms of entities based on the typical scientific workflow model, and provide graphical renderings of this workflow-like view of the scripts. In the Kurator project, we are developing a toolkit that allows biodiversity data collection managers to easily program data curation pipelines, consisting of services and functions for data quality control and (semi-)automated repair. We are developing software components (“actors”) to examine data collections and perform checks, e.g., on scientific names, name authorship, collecting date, collector name (recorded by), georeference, locality, and phenological state (where applicable). Kurator allows to “clean data with data”: In addition to checking the internal consistency of records, we can employ external resources to spot quality issues and suggest repairs. Human curators remain in control: Kurator tools keep track of processing history and data lineage (computational provenance) to show original records, alternative forms and the respective sources, thus allowing human curators to make informed decisions about which suggested repairs and flagged records require action.

#DIVERSE CHILDREN'S LITERATURE: EXAMINING SOCIAL MEDIA'S ROLE

Melissa Hayes, DeAnza Williams

The Cooperative Children’s Book Center has been tracking diversity in children’s books since 1994. However, little was known about their research outside of the field until an infographic of their findings on the diversity gap in children’s literature was posted on publisher Lee and Low's blog and spread across social media platforms in 2013. Beginning in the spring of 2014, the goal of the #WeNeedDiverseBooks campaign was to spread awareness about the lack of diversity in children’s literature. In light of the #WeNeedDiverseBooks campaign, we are interested in investigating the role that social media plays in creating awareness for social justice campaigns. Our presentation will examine the background, key points, and the current responses to the campaign on the social media platform of Twitter (@DiverseBooks) from the spring of 2014 to the present. Our key question is: How has Twitter been used by #WeNeedDiverseBooks to communicate the organization’s message? We will use content analysis to examine tweets sent by @DiverseBooks (We Need Diverse Books) to communicate with and engage supporters. From this research, we hope to gain understanding about how Twitter is being used by social justice campaigns.

THE HATHITRUST+BOOKWORM PROJECT: EXPLORING CULTURAL AND LITERARY TRENDS IN MILLIONS OF SCANNED BOOKS

Peter Organisciak (presenter), Loretta Auvil, Benjamin Schmidt, Sayan Bhattacharyya, Mari Colleen Fallaw, Muhammad Saad Shamim, Yun Hao, Robert McDonald, Erez Aiden Lieberman, J. Stephen Downie

The Hathitrust Digital Library contains over 14 million works scanned from cultural heritage and library collections around the world. A collection spanning centuries at such a massive scale is valuable for tracking cultural and literary themes through their prevalence in published work. However, large scales also introduce unwieldiness, where technical hurdles challenge casual exploration and burden scholars

with additional skill requirements. To address some of the complexities of large-scale analysis, the Hathitrust+Bookworm (HT+BW) project enables real-time analysis of the Hathitrust Digital Library through faceted data visualization. HT+BW makes it possible to “slice and dice” the corpus data in real time: one can compare different terms or topics (e.g. the usage of 'Istanbul' versus 'Constantinople' in English publications), over time, and across custom metadata or content-based subsets of the data (e.g. comparing themes between British and American poets). HT+BW is improving the accessibility of the HathiTrust corpus while enabling the sorts of complex, multifaceted research questions that scholars may ask. At the same time, this project is contributing toward a tool, Bookworm, which can be implemented by other institutions for their own collections. In our presentation, we will present these benefits, introduce the challenges involved in developing the project, and demonstrate examples of interesting questions communicated through bookworm.

Poster and Demonstration Abstracts

Poster Session 1

PATRON PRIVACY IN JEOPARDY: A PILOT ANALYSIS OF THE PRIVACY POLICIES OF DIGITAL CONTENT VENDORS IN AMERICAN PUBLIC LIBRARIES

April Lambert, Michelle Parker, Masooda Bashir

While the library profession has long defended readers’ privacy, a public library patron’s personal information is no longer solely in the hands of intrepid librarians determined to defend intellectual freedom. Libraries use vendors to provide a large portion of their digital content. These vendors gain access to extensive personal information about patrons. Libraries must now work with content providers to negotiate privacy protections for their patrons that are in accordance with the American Library Association’s Code of Ethics. This poster presents the pilot results of a content analysis of the privacy policies of five of the top digital content vendors of American public libraries. This analysis examines whether these privacy policies meet the privacy standards of the library community, meet other industry standards, and are accessible and understandable to public library patrons. Policies were assessed using a code book consisting of thirty-five questions related to (1) user consent and access, (2) vendor collection of personally identifying information, (3) vendor reasons for sharing information, (4) vendor efforts to enforce their privacy policies, and (5) vendor security efforts and data storage. Based on this pilot research, we conclude that while vendors are largely meeting the Fair Information Practices standards of American industry, the policies fail to meet the heightened standards of the library community. This project is in the process of being expanded to include the privacy policies of nearly forty major public library digital content vendors to assess whether these initial findings hold true throughout the industry. If they do, this research can provide support for librarians trying to identify privacy issues they must negotiate with their vendors

ON THE NEED FOR COMBAT LIBRARIANS: A CASE STUDY OF KNOWLEDGE MANAGEMENT AT THE END OF THE IRAQ WAR

Caryn Anderson

Despite operating in complex, dynamic, and frequently changing environments, organizations have a vested interest in making sure that knowledge created today is available for use tomorrow. This paper describes a case study of knowledge management (KM) at the end of the Iraq war, where the breadth and depth of KM was extreme. The “Iraq Knowledge Management (KM) Transition,” a joint Department of State (DOS) and Department of Defense (DOD) effort, ensured that the intelligence, relationships, capacity building, and reconstruction developed by the United States Government (USG) over eight years in Iraq could continue to be leveraged by the U.S. diplomatic mission after the departure of the

U.S. military in December 2011. The effort involved over 100 USG offices and agencies to transition ~500 information sources and ~40 networked software applications. KM activities are analyzed using Anderson's KM and Kling's social informatics frameworks. The results reveal that many of the challenges mirror the socio-technical interdependencies identified in other organizational settings and suggests that lessons learned from this extreme case can be used to inform KM practices of information professionals in non-combat environments.

PARAMETER TUNING: EXPOSING THE GAP BETWEEN DATA CURATION AND EFFECTIVE DATA ANALYTICS

Catherine Blake, Henry Gabb

The "big data" movement promises to deliver better decisions in all aspects of our lives from business to science, health, and government by using computational techniques to identify patterns from large historical collections of data. Although a unified view from curation to analysis has been proposed, current research appears to have polarized into two separate groups: those curating large datasets and those developing computational methods to identify patterns in large datasets. Our initial project goal was to automatically identify results (also called findings or claims) from empirical studies. The problem was framed as a binary classification task, where the classifier was trained to discern a result from a non-result sentence. The case study presented here demonstrates the enormous impact that parameter tuning can have on the resulting accuracy, precision, and recall of a computational model that is generated from data. It also illustrates the vastness of the parameter space that must be searched in order to produce optimal models and curated in order to reproduce previous experiments and avoid redundant experiments. This highlights the need for research that focuses on the gap between collection and analytics if we are to realize the potential of big data. This work was published in the 2014 ASIS&T Proceedings.

CLOSING THE APP GAP: FROM PLAN TO PROJECT I

Deborah Stevenson, Kate McDowell, Cass Mabbott

This poster reports on our IMLS-funded Closing the App Gap planning grant, which examines the use of tablet computers, apps, and e-books in public libraries for primary grades children in underserved communities. As digital media grows in importance, the resource disparity between affluent families and lower-income families becomes even more pronounced. In this project, we hoped to combine the public library's traditional summer reading strengths with a technology-based approach to bring new tools in the fight against summer reading loss, enhance technological literacy, and mitigate the effect of the digital divide on children in lower-income families. We partnered with the Douglass Branch Library in Urbana, IL, which serves a low-income, predominantly minority population, in a pilot program during summer reading.

A NEW PERSPECTIVE OF STUDYING USERS' ATTITUDES TOWARD INFORMATION PRIVACY AND CULTURE IN GLOBAL CONTEXT: AN EXPLORATIVE STUDY

Hsiao-Ying Huang, Kim Naples, Masooda Bashir

The intersection of attitudes about information privacy and culture has been studied by academia in different fields. However, the lack of consistency in situations where privacy matters might be the reason why cultural perspectives on information privacy are difficult for researchers to study. To rebuild the connection between the theory and practice in information privacy and culture, our research is dedicated to explore new approaches to investigate users' perspectives and the employment of informational norms in global contexts according to the current social issues, including privacy rights of consumers, employees, and citizens. We created two different studies that attempted to draw conclusions on how culture influences how people think about privacy. Both studies concluded that people will respond to privacy situations differently in different context, in support of Nissenbaum's Contextual Integrity Framework. The first study (a short answer and multiple-choice

survey) asked participants how they felt about certain entities having access to their information. It also sought to explore whether participants felt stronger about a national identity or a global identity and how that identification influenced how they felt about sharing their information. We suggested a new perspective to investigate information privacy and culture and further proposed a new contextual information privacy index (CIPI) that can be an instrument for policy makers to legislate the relevant online regulations on information privacy. The second study (a long-form vignette survey) concluded that people's values influence how they think about privacy and these values do not always correlate to their nationality. We concluded that nationality was not a predictor on how people would respond to a privacy situation. There is further exploration to be done into what culture is made up of and what role values play in that make-up. The paper resulting from our research was an attempt to practically apply Helen Nissenbaum's theories. Based on the findings, we suggest future studies adopt normalized measurements to explore informational norms in different contexts.

AUTOMATING THE IDENTIFICATION OF CRITICAL COMMUNITY HEALTH INDICATORS AND THEIR PRESENTATION TO THE PUBLIC

Robert Sarwark, Julia Bello-Bravo, Barry R. Pittendrigh, Tom Szyrka, Ian S. Brooks

Every local health district in Illinois is required to conduct a community health assessment and develop a community health plan every five years. This project provides tools to help with both aspects of this requirement. 1. The community assessments rely heavily on data about health indicators provided by the state department of health. Making sense of this data is challenging for most local health districts and burdensome for small rural ones. Working with the state department of health we have developed a prototype algorithm that automates this process by analyzing 93 health indicators to establish a ranked list of those of greatest importance to a health district. 2. Once the local health district has identified its priorities for the next five year cycle it is important for them to explain their plan and the underlying reasons to the public. To support this education step we have developed prototypes of a template driven presentation targeted at a 7th grade level of data understanding. These presentations are automatically produced based on the results obtained from the algorithm and for each identified health indicator describe the indicator, why it is important for health, and the reason it has been identified as a priority for the community. The graphical elements of the presentations are designed to be languageless, enabling easy adaptation to different languages through voiceover narration, and take advantage of public health animations developed by the University of Illinois Scientific Animations Without Borders project.

INCREMENTAL SENTIMENT PREDICTION BASED ON HUMAN IN THE LOOP LEARNING

Jana Diesner, Shubhanshu Mishra, Liang Tao, Elizabeth Surbeck, Jason Byrne

Predicting various types of sentiment for textual data sources is an intensely studied problem in linguistics and computing. Despite progress with computational solutions beyond deterministic look-up dictionaries, practitioners often use dictionary-based, off the shelf tools in order to classify consumer products as being perceived in a positive, negative or neutral way. Advanced scientific solutions have moved far beyond this approach; providing probabilistic solutions that consider a variety of lexical, syntactic and surface form features. How useful are such scientifically rigorously built and tested solutions for practitioners? We provide an answer to this question in three ways: First, by comparing the results from a solution we built via supervised learning to the predictions from a main commercial benchmark tool. For this, we were provided with about ~25K of unique tweets that were hand-tagged by subject matters experts from a large corporation from the food industry sector. Our solution builds upon existing approaches for sentiment classification of twitter data, including suitable parts of speech tagging. We achieve about 80% accuracy (F1-score). Second, by having practitioners use our solution to classify their consumer product data and provide feedback on usability, accuracy beyond precision and recall, and scalability. Third, by refining our prediction model based on incremental learning, where the same subject matter experts inspect and if applicable relabel prediction outcomes; resulting in gradually improved and changing models. This incremental, human

in the loop learning also accounts for the dynamic nature of trends and patterns in language use on social media. To support this process, we have designed visual analytics routines that are aimed to support end-users in efficient data inspection, error analysis and close readings. Overall, this comprehensive and collaborative approach to sentiment analysis and evaluation contributes to making scientific research available for industry scale applications.

EXPERT-GUIDED CONTRASTIVE OPINION SUMMARIZATION FOR CONTROVERSIAL ISSUES

Jinlong Guo

This paper presents a new model for the task of contrastive opinion summarization (COS) particularly for controversial issues. Traditional COS methods, which mainly rely on sentence similarity measures are not sufficient for a complex controversial issue. We therefore propose an Expert-Guided Contrastive Opinion Summarization (ECOS) model. Compared to previous methods, our model can (1) integrate expert opinions with ordinary opinions from social media and (2) better align the contrastive arguments under the guidance of expert prior opinion. We create a new data set about a complex social issue with “sufficient” controversy and experimental results on this data show that the proposed model are effective for (1) producing better arguments summary in understanding a controversial issue and (2) generating contrastive sentence pairs.

IMPACT OF ENTITY RESOLUTION ON NETWORK PROPERTIES

Jinseok Kim, Jana Diesner

This presentation shows how the choice and parameterization of certain data preprocessing methods for disambiguating author names affects our understanding of the structure and evolution of co-publication networks. Two case studies were conducted: (1) Thirty years of publication records from 125 Information Systems journals from DBLP (110,000 papers), and (2) more than 50 years of domestic journal publication records of scholars in Korea (494,000 papers). Author names in the datasets were pre-processed via algorithmic disambiguation. We applied the commonly used all-initials and first-initial based disambiguation methods to the data, generated over-time networks with a yearly resolution, and calculated standard network metrics on these graphs. Our results show that initial-based methods underestimated the number of unique authors, average distance, and clustering coefficient, while overestimating the number of edges, average degree, and ratios of the largest components. These self-reinforcing growth and shrinkage mechanisms amplify over time. This can lead to false findings even about fundamental network characteristics such as topology and the inference of underlying social processes. It can also cause erroneous predictions of future network evolution and suggest unjustified policies, interventions and funding decisions. The findings of the study suggest that scholars need to be more attentive to data quality such as the ratio of ambiguous names and their impact on data analysis.

DIGITAL INNOVATION LEADERSHIP PROGRAM (DILP)

Jon Gant

The Digital Innovation and Leadership Program (DILP) is a project funded through the Illinois Extension, engaging a rich set of stakeholders across campus, extension, and community. DILP focuses on building local community capacity around essential 21st century skills such as data analytics, digital manufacturing, and digital media production. However, the way DILP orients this process is through the unique combination of collective leadership and critical interpretative sociotechnical frameworks. A critical interpretative sociotechnical (CIS) framework advances agency and challenges a number of assumptions embedded within the dominant narrative regarding technology, technology expertise, and technology’s role in society (Rhinesmith and Wolske, 2014). The CIS framework brings together several rich traditions such as sociotechnical systems, science and technology studies, and interpretive and critical meta-theoretical frameworks. It has been developed over the past fifteen years as part of

pedagogical research into a social justice approach to LIS technology education. Collective leadership is defined as an informal, distributed process where leadership is shared among various team members rather than centralized within a single individual (Pearce & Conger, 2003). Such a framework recognizes the dynamic nature and unique challenges of collaborative partnerships among various stakeholders while providing ways to approach the social dynamics of processes with attention towards inclusive, self-reflective leadership practice. Through DILP, the opportunity to bring together these two frameworks, along with the rich tradition of Extension, allows us to uniquely address local issues around digital manufacturing, digital media production, and data analytics. The program brings together skill development in areas such as computational thinking and demystifying technology and combines it with digital innovation resources like FabLabs and community resources such as public libraries to further the ultimate goal of helping communities build stronger community. The presentation will provide an overview of the CIS and collective leadership frameworks underpinning the program. This presentation is part of an overall group of similarly themed projects, submitted to the poster presentation session, which are funded through the Center for Digital Inclusion.

INFORMATION INTEGRATION: A CASE STUDY OF AIR QUALITY IN CHICAGO AND ST. LOUIS

Jooho Lee, Catherine Blake

Government agencies are increasingly making raw data available to citizens, but merely having access to data is not sufficient to realize the potential of “big data.” Answering questions in science, business, and public policy requires data integration which is challenging when data from different sources are used for different reasons. This project provides a detailed case study of how to integrate public data to understand the relationship between demographic factors and air quality. Demographic factors from US Census and American Community Survey were collected for two major cities (Chicago and St. Louis) and then integrated with air quality from the US Environmental Protection Agency (US EPA). Results show that air quality has improved in both cities between 2000 and 2012. Determining correlations between ethnicity, education, level of income and air quality warrant further exploration.

COLLECTIVE LEADERSHIP FOR COLLABORATIVE ACTION: A NEW LEADERSHIP FRAMEWORK FOR LIBRARY ENGAGEMENT AROUND DIGITAL LITERACY EDUCATION

Kirstin Phelps

The environment within which public libraries are embedded is changing. Infrastructure allowing for increased connectivity to the web, as well as opportunities to support new technologies, i.e. makerspaces, have raised questions on how libraries can best respond to the social and technical goals of their communities. Reflected in these questions is a need to explore how libraries can develop effective community partnerships around such opportunities; for example, through the provision of digital literacy education. The Digital Innovation Leadership Program (DILP) provides a framework for developing capacity within libraries to form such partnerships, focused on community identified development goals. DILP aims for innovative approaches to empower and engage libraries to be successful in dealing with the relationships among socio-technical systems. These issues cannot be fully addressed with traditional models of leadership where a single organization is in charge, or where a single individual is responsible for providing essential services. DILP recognizes this by focusing on a process of collective leadership, which is defined as an informal, distributed process where the responsibility of leadership is shared among team members [1]. This poster will present DILP and its collective leadership process as a way to re-envision the frameworks used within public libraries for community engagement. Outcomes include capacity building and the fostering of sustainable community partnerships responsive to the changing goals of the context in which they are embedded. Additionally, the components of the DILP toolkit will be outlined, which includes collective leadership skills, stakeholder alignment, digital literacy preparedness, and technology needs assessment.

References: [1] Contractor, N. S., DeChurch, L. A., Carson, J., Carter, D. R., & Keegan, B. (2012). The topology of collective leadership. *The Leadership Quarterly*, 23(6), 994-1011

THE IMPACT OF MUSIC GENRE ON MUSIC SIMILARITY: A QUALITATIVE EXPLORATION

Yun Hao, Kahyun Choi, J. Stephen Downie, Ning Chen

That to what extent music similarity can be explained or predicted by music genre is a big topic in the field of music information retrieval. According to Oxford dictionary, “genre” is “a category of artistic composition, as in music or literature, characterized by similarities in form, style, or subject matter.” There is no doubt that people tend to consider two pieces of music similar if the two music works are of the same genre. But is genre the most distinguishable feature of music? The answer may differ among people, and this is the reason why music similarity remains intractable – people determine music similarity using their own standard. The Audio Music Similarity and Retrieval task (AMS) of Music Information Retrieval Evaluation eXchange (MIREX) provides a post hoc framework of evaluating music similarity and retrieval algorithms by volunteers who compare queries and returned results for similarity. With the recent nine years’ results of AMS available, we have similarity scores for 8049 songs of which genre information is also available. Similarity scores come in the format of both categorical (i.e., not similar, somewhat similar, very similar) and fine scores (scores on a continuous scale of 0 – 10). Our current work aims to statistically measure how representative music genre is of music similarity using this dataset. Assuming that music similarity is mostly attributed to genre, we are conducting a 10-fold music similarity prediction experiment by converting the pairwise similarity fine scores into genre similarity scores and using genre similarity information to predict categorical similarity scores. Our findings provide some insights into chances to improve music services such as automatic playlist generation systems and music recommendation.

Poster Session 2

THE IMPOSITION OF EMPATHY: THE PEDAGOGY OF EMPATHIC DESIGN

Andrea Gannon, Michael Twidale

In my independent study with Dr. Twidale this semester, I am researching the broader implications of using empathy to design technology without limiting it to a specific context, such as the medical field, and instead thinking about its uses in a variety of areas. We are examining theoretical applications of empathy, particularly the way in which past research is being used to shape future project ideas because this is still a rather novel and untested concept. As we are coming from the information science perspective with a usability/user experience slant, we are looking at the scope of teaching and applying empathic design as well as the cost and feasibility of relying on empathic data that does not always fit the quantifiable metrics to which programming directors and product developers are accustomed. In particular, I am focusing on the role that empathy plays in information pedagogy and how one can teach the effective communication of user issues. Empathic design is useful in the creation of tools that are more in tune with their targeted audiences; however, it can also lead to products which create empathy through the process of interacting with those items or applications themselves. Some of the questions that have emerged include asking if one can then use empathy to design products that instill the values of empathy in others and what this could potentially look like. I believe it pertains significantly to the realms of library and information science, business architecture, educational scenarios, and several other product and application development contexts. My poster would explore the possibility of how, by using empathy, not only can one design better solutions that address users’ needs with heightened understanding but also potentially create solutions which teach empathy themselves. Empathy is essential for the creation of successful and sustainable designs, and it is a lens that we often use without realizing it.

COLORADIO: A PRODUCT OF INTERDISCIPLINARY COLLABORATION IN AN ISCHOOL ENVIRONMENT

Christopher Nixon, Kinyetta Nance, Berenice Vargas, Helen Zhou

In the fall of 2014, three GSLIS students and one Industrial Design student collaborated on a project to design and build a prototype tangible user interface. By drawing on our diverse set of experiences and skills, we were able to go from nothing to finished prototype and documentation in only four days. Through this experience we were able to understand and value the benefits of design thinking methodologies in interdisciplinary work. This presentation will be an overview of our process and insights into how to conduct more, better interdisciplinary work from an LIS perspective, in particular focusing on how to use design thinking, and an iterative process to address user experience challenges.

CONCEPTUALIZING WORKSETS FOR NON-CONSUMPTIVE RESEARCH

Jacob Jett, Chris Maden, Colleen Fallaw, Megan Senseney, J. Stephen Downie

The HathiTrust (HT) digital library comprises 4.5 billion pages (composing 12.9 million volumes). The HathiTrust Research Center (HTRC) – a unique collaboration between University of Illinois and Indiana University – is developing tools to connect scholars to this large and diverse corpus. This poster discusses HTRC’s activities surrounding the discovery, formation and optimization of useful analytic subsets of the HT corpus (i.e., workset creation and use). As a part of this development we are prototyping a RDFbased triple-store designed to record and serialize metadata describing worksets and the bibliographic entities that are collected within them. At the heart of this work is the construction of a formal conceptual model that captures sufficient descriptive information about worksets, including provenance, curatorial intent, and other useful metadata, so that digital humanities scholars can more easily select, group, and cite their research data collections based upon HT and external corpora. The prototype’s data model is being designed to be extensible and fit well within the Linked Open Data community. Note that this is a poster which will have been presented at iConference the week before.

EDUCATE, EMPOWER, ENGAGE: THE CENTER FOR DIGITAL INCLUSION (CDI)

Jon Gant, Kirstin Phelps, Martin Wolske

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. This poster will provide an overview of the goals and mission of CDI, as well as insight into current and previous projects supported by the Center to empower, engage, and educate individuals and communities around ICTs. This poster is part of an overall group of similarly themed projects which are funded through the Center for Digital Inclusion.

TOPIC MODELING USERS' INTERPRETATIONS OF SONGS TO INFORM SUBJECT ACCESS IN MUSIC DIGITAL LIBRARIES

Kahyun Choi, Jin Ha Lee, Craig Willis, J. Stephen Downie

The assignment of subject metadata to music is useful for organizing and accessing digital music collections. Since manual subject annotation of large-scale music collections is labor-intensive, automatic methods are preferred. Topic modeling algorithms can be used to automatically identify latent topics from appropriate text sources. Candidate text sources such as song lyrics are often too poetic, resulting in lower-quality topics. Users' interpretations of song lyrics provide an alternative source. In this paper, we propose an automatic topic discovery system from web-mined user-generated interpretations of songs to provide subject access to a music digital library. We also propose and evaluate filtering techniques to identify high-quality topics. In our experiments, we use 24,436 popular songs that exist in both the Million Song Dataset and songmeanings.com. Topic models are generated

using Latent Dirichlet Allocation (LDA). To evaluate the coherence of learned topics, we calculate the Normalized Pointwise Mutual Information (NPMI) of the top ten words in each topic based on occurrences in Wikipedia. Finally, we evaluate the resulting topics using a subset of 422 songs that have been manually assigned to six subjects. Using this system, 71% of the manually assigned subjects were correctly identified. These results demonstrate that topic modeling of song interpretations is a promising method for subject metadata enrichment in music digital libraries. It also has implications for affording similar access to collections of poetry and fiction.

LIS ONLINE EDUCATION: COMPARING LEEP WITH OTHER PROGRAMS

Kate McDowell, Wei Gao, Caroline Nappo, Mikki Smith

The field of library and information science (LIS) has long been at the forefront of online education, and the LEEP program was among the first, in 1996, to push the boundaries of pedagogical opportunities in online learning. This poster examines the current status of online education in library and information science by looking at the different approaches taken by top LIS schools in the US. To this end, various aspects of online education were examined across selected LIS schools, including curriculum, residency requirements, delivery mode, and financial options. This poster visually summarizes findings from 13 LIS schools selected based on the most recent (2013) US News and World Report Ranking of Best Graduate Schools in Library and Information Studies. Comparing Illinois to other schools, only Syracuse also has a similar initial week-long residency, and only Florida also has synchronous courses. Despite the recent elimination of required residency for most LEEP courses (effective summer 2015), our combination of initial residency and synchronous courses is unique among these schools. Because we are one of only nine programs that offer options for online Master's degrees, these features make us quite distinctive as we maintain our high standard of instruction and support for our students.

MINING THE POACHERS: QUANTITATIVE ANALYSIS OF FANFICTION WORKS ACROSS TWO ARCHIVES

Elizabeth Wickes

Fanfiction, or stories written based on an existing franchise or canon, was first researched in depth by Jenkins (1992) with his publication "Textual Poachers." Research has been completed into the culture and practices of these writing communities since then, but not much quantitative analysis has been completed on the works themselves. This project focuses on aggregating and providing basic descriptive analytics on works held within two archives: Fanfiction.net and Archive of Our Own. The purpose is to provide aggregate descriptive information on written fan works to inspire future qualitative research questions. Metadata for 2.5 million works from Fanfiction.net was compiled, with Archive of Our Own currently in progress. The results suggest that not all fanfiction stories are created equal. This is a fact implicitly understood by fans and researchers, but not measured or described across entire archives. Differences have been found with word length across source medias, franchises, and archives. This presentation will summarize the basic descriptive statistics for stories from these two domains, and highlight the significant differences between archives, medias, and franchises. Methods for data collection and processing workflows will also be discussed.

DIGITAL LITERACY FOR ALL LEARNERS

Martin Wolske, Kim Naples, Hailley Fargo, Sarah Butt, Travis Faust

The Digital Literacy for All Learners project within the Center for Digital Inclusion is partnering with the CU Community Fab Lab to develop and share models for digital literacy that empower citizens to affect social change. Guided by the Critical Interpretive Sociotechnical (CIS) framework that seeks to foster knowledge power and challenge exclusionary forces, our participatory leadership process seeks to advance community agency and self-efficacy at each point of the information, knowledge, action, power cycle. The three core foundations of our demystifying technology pedagogy encourage

participants to progress from passive use of technical artifacts to co-creation of innovations-in-use by community, in community, for community: 1) Computational thinking & inquiry-based learning to advance critical thinking; 2) The critical interpretive sociotechnical framework to advance a critical perspective regarding the relationship between the social & the technical; and 3) A revolution of values to advance a humanizing & people-centered approach grounded in popular education. Highlights from the work of four research assistants from the Graduate School of Library and Information Science will include: the growth of a two-year old Teen Open Lab program at The Urbana Free Library; an 18-month old community-led, cross campus project at Kenwood elementary in Champaign; expansion of programming at long-term GSLIS partners, the Urbana Neighborhood Connections Center and Tap-In Leadership academy; and new teen-focused Maker/Fab Lab programming at the Champaign Public Library. This poster is part of an overall group of similarly themed projects which are funded through the Center for Digital Inclusion.

DIGITAL HUMANITIES DATA CURATION INSTITUTES: CHALLENGES AND PRELIMINARY FINDINGS

Megan Senseney, Trevor Munoz, Julia Flanders, Ali Fenlon

The growth of digital humanities research has made the curation of DH research data a priority for humanities scholars and their institutions. Data curation “addresses the challenges of maintaining digital information produced in the course of research in a manner that preserves its meaning and usefulness as potential input for future research.” More fully integrating data curation into digital research involves fluency with topics such as publication and information sharing practices, descriptive standards, metadata formats, and the technical characteristics of digital data. This poster presents lessons learned from a series of workshops on digital humanities data curation conducted from June 2013 to April 2014 with funding from the National Endowment for the Humanities Institutes for Advanced Topics in the Digital Humanities program.

TARGET FUNCTION ASSESSMENT FOR IMPACT GOALS OF ISSUE FOCUSED PUBLIC MEDIA

Rezvaneh Rezapour, Jana Diesner

Philanthropic foundations as well as scientific organizations have recently started to rethink how to measure the impact of the work they are funding in a meaningful, robust and scalable fashion. In collaboration with funders and producers of issue focused public media products, we have been leveraging data analytics, more specifically a combination of text mining and network analysis, to develop a theoretically grounded solution to this problem. How well does our approach meet the needs of funders? In this presentation, we identify which of the main impact goals as defined and used in the social change community can be captured with our approach, illustrate our solution with an empirical case study, and compare our findings to those one would obtain by using traditional impact assessment methods and metrics. We find that our solution complements and enhances the findings and interpretations resulting from using state of art solutions used in the given application domain, especially when applying data mining techniques to natural language text data, such as representations of public awareness, dialogue and engagement around various issues in their cultural contexts.

EXTRACTING FEATURES FROM TEXT FOR NON-CONSUMPTIVE READING WITH THE HATHITRUST RESEARCH CENTER

Sayan Bhattacharyya, Boris Capitanu, Ted Underwood, Peter Organisciak, Loretta Auvil, Colleen Fallaw, J. Stephen Downie

We introduce the Extracted Features (EF) dataset, consisting of textual features currently derived from a preliminary collection of over 250,000 volumes (likely to grow soon into 4 million volumes) of digitized text — a subset of the holdings of the HathiTrust Digital Library. The EF dataset constitutes a means for researchers to interact with the text corpus within the constraints of copyright restrictions on material in the collection. These restrictions necessitate providing facilities for “non-consumptive” reading of the collection (distant reading without “consuming” the works as happens in any act of direct reading).

A digital library cannot provide copies of the copyrighted work directly to researchers for algorithmic analyses, as non-consumption on the part of the reader cannot be ensured. One solution to this problem that the HathiTrust Research Center (HTRC) has undertaken is to provide researchers with features from the text instead. The EF dataset is an outcome of this initiative. Features are quantified properties of the page text. Per-page features are packaged in the JSON lightweight data interchange format, in one file per volume, with the relevant page section (that is, header, footer and body) identified for each page. Occurrence counts of part-of-speech-tagged words per page-section constitute one of several features provided through the EF dataset. Several benefits follow from this translation of the text data into features. Pre-packaged extraction saves users' effort and time: much pre-analysis, such as tokenization, tagging, de-hyphenation and content extraction, is already performed for the user. Consistent pre-processing also diminishes the risk of potential methodological inconsistencies between different studies on the data, making it easier to compare and reproduce such studies. Our presentation will describe and discuss the current EF dataset as well as the HTRC Feature Reader tool, which provides a facility for interactively reading and representing the EF data.

EXPLORING CULTURAL DIFFERENCES IN LANGUAGE USAGE: THE CASE OF NEGATION

Sveta Stoytcheva, Dov Cohen, Catherine Blake

Prior research suggests that speakers of Asian languages are more likely to use negation than English speakers. Our goal in this work is to explore this theory using empirical data from news stories. Specifically, we used natural language processing to compare negation usage in two newspapers: the New York Times and Xinhua News (English Edition). Overall, negation represents 0.55% of typed dependencies in the New York Times (versus 0.18% in Xinhua News). Additionally, 9.28% of sentences and 86.56% of articles in the New York Times contain one or more instances of negation (compared to 3.33% of sentences and 24.94% of articles in Xinhua News). In contrast to the prevalent theory, negation is approximately three times more common in the New York Times than in Xinhua News (English Edition).

Posters with Demonstrations (during Poster Sessions 1 and 2)

THE DIGITAL INNOVATION LEADERSHIP PROGRAM (DILP)

Lisa Bievenue, Jeff Ginger, Kinyetta Nance, Jarai Carter

The Digital Innovation Leadership Program (DILP) is a collaborative project, funded through Illinois Extension, focused on building capacity within local communities around digital literacy needs. DILP engages a diverse range of stakeholders from campus, within extension, and within local communities to address social and economic needs around essential 21st Century skills for economic and social development. DILP is formed from a unique combination of collective leadership and critical interpretative socio-technical frameworks that focus on empowering individuals to address local needs, and encourage social change. DILP represents a new model for engagement between LIS, Extension, and the community. The partnership leverages the strengths of each partner to address concerns of digital literacy through STEM focused education in digital manufacturing, digital media production and digital analytics. The model for providing this training is a live learning lab environment comprised of mobile digital design space - mobile Fab Labs. In addition, DILP connects subject matter experts in fields such as computational thinking, creativity, news literacy, empathetic design, and more, to create unique digital literacy programming resources. Access to subject matter experts and the associated curriculum will be provided to communities through a learning management system (LMS). The LMS is part of an overall toolkit which will be provided to communities, in order to empower the creation of localized programming around topics of relevance. The toolkit includes resources for conducting community needs assessments, user studies on youth media use, collective leadership skills, as well as equipment and planning resources for creating community Fab Labs. The toolkit reflects the research-grounded resources available to communities to identify, implement, and sustain local

community needs in digital literacy and applications. This poster will provide an overview of DILP and its associated parts. This poster is part of an overall group of similarly themed projects which are funded through the Center for Digital Inclusion.

MAPS - APPLYING LIS TO CULTURAL HERITAGE

Christopher Nixon, Michael Twidale

Maps is a tourist map with an integrated RFID chip, enabling customized, personalized information to be displayed at kiosks around a city, site, or museum. Not only does the Map provide useful information to the tourist, such as recommendations on places to visit, it provides useful data to the city or tourism organization, such as patterns of sites tourists visit. Public kiosks are another key component of Maps, as they provide the check-in points for tourists. Once home, travelers can review their trip on their smartphones or computers via a web interface. Maps was developed as a project in Professor Michael Twidale's Museum Informatics class, on how to use information technology to address a need in a cultural heritage context. The demo will consist of a working prototype demonstration, where people can use a Map and "travel" to different locations in Hong Kong, get recommendations of places to go, and review their trip on their smartphone.

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)

Bertram Ludäscher, 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.