

The School of Information Sciences

For more than 100 years, the top-rated University of Pittsburgh School of Information Sciences (SIS) has been educating leaders in the information professions through its graduate and undergraduate programs. SIS graduates work with "all things information" including books, artifacts, digital documents, databases, routers, and networks. They solve problems in the boardroom, operating room, and classroom. They develop strong analytical and problem solving skills with a keen understanding of effective uses of technology needed to access and manage information, all rooted in the human experience. SIS is comprised of three program areas that all work together to develop unique solutions to information and technology challenges: Information Science & Technology, Telecommunications & Networking, and Library & Information Science.

SIS is a founding member of the "iSchool consortium," which is a collective of information schools formally convened in 2005 and committed to advancing the field of information sciences. The iSchool consortium now numbers 65+ member institutions globally.

Pitt is a National Center of Academic Excellence in Information Assurance/Cybersecurity Research, and is one of only 20 institutions in the country with five Committee on National Security Systems certifications. SIS is ranked sixth in the nation for IT Security, ranked seventh out of 400+ international programs for cybersecurity education, and has a Top 10 American Library Association-accredited MLIS program.

The Department of Computer Science

The University of Pittsburgh hosts one of the oldest CS Departments in the world with research and teaching excellence since 1966. The CS Department covers many areas of research/teaching: AI, algorithms, databases, software engineering, systems and networks, data visualization/imaging, security/privacy, embedded systems, and natural language processing. Substantial research funding is provided by the National Science Foundation, the National Institutes of Health, and other federal agencies.

CS at Pitt is ranked 25th out of 473 ranked programs (19th with respect to gender diversity) in the most recent PhDs.org evaluation, and 34th out of 128 in the most recent (2011) National Research Council (NRC) evaluation, jumping up nine positions since the 1995 NRC rankings



University of Pittsburgh

School of Information Sciences

2015 Board of Visitors Poster Session

October 12, 2015

1. “Exploring the Effects of Open Social Student Model Beyond Social Comparison” by Julio Guerra, PhD Student (IS)

Presenter: Julio Guerra; Advisor: Peter Brusilovsky

In our journey exploring the effects of Open Student Model (OSM) on students working with programming problems and examples, we have incorporated the idea of social visualizations to extend OSM to Open Social Student Modeling (OSSM). Although comparison features in OSSM, where a student can compare herself to the group or individual peers, have shown to increase students’ work, we now shift our attention to other effects. The goal is to explore the OSSM effects beyond comparison, particularly metacognitive support, and we propose a representation of the OSSM towards these lines.

2. “Tweeting Questions in Academic Conferences: Seeking or Promoting Information?” by Xidao Wen, PhD Candidate (IS)

Presenter: Xidao Wen; Advisor: Peter Brusilovsky and Yu-Ru Lin

The fast growth of social media has reshaped the traditional way of human interaction and information seeking behavior, which draws research attention on characterizing the new information seeking paradigm. However, results from previous studies might not be well grounded under certain social settings. In this paper, we leverage machine learning techniques to identify different types of question tweets within academic communities as an example of one particular social context. By studying over 160 thousand tweets posted by 30 academic communities, we discovered a different landscape of information-seeking behaviors, where less tweets are regarded as question tweets, and more real information-seeking tweets are observed. We also found that users respond differently to different types of question tweets. We believe our study would be beneficial for understanding the information seeking behaviors in social media.

3. “Geo-Social-RBAC: A Location-Based Socially Aware Access Control Framework” by Nathalie Baracaldo, PhD Candidate (IS)

Presenter: Nathalie Baracaldo; Advisors: James Joshi

The ubiquity of low-cost GPS-enabled mobile devices and the proliferation of online social networks have enabled the collection of rich geo-social information that includes the whereabouts of the users and their social connections. This information can be used to provide a rich set of access control policies that ensure that resources are utilized securely. Existing literature focuses on providing access control systems that control the access solely based on either the location of the users or their social connections. In this poster, we argue that a number of real-world applications demand an access control model that effectively captures both the geographic as well as the social dimensions of the users in a given location. We propose, Geo-social-RBAC, a new role based access control model that allows the inclusion of geo-social constraints as part of the access control policy. Our model, besides capturing the locations of a user requesting access and her social connections, includes geo-social cardinality constraints that dictate how many people related by a particular social relation need to be present in the required locations at the time of an access. The model also allows specification of geo-social and location trace constraints that may be used to dictate if an access needs to be granted or denied.

4. “System Dreams: Public Libraries and the Politics of Regional Library Service in Allegheny County, 1935-1989” by Michael M. Widdersheim, PhD Student (LIS)

Presenter: Shelia Corral; Advisor: Kip Currier

Large-scale coordination of infrastructure is a politically contentious issue in Pittsburgh due to the region’s corrugated topography and fractured governmental landscape. In the early 1990s, however, public libraries in the region coalesced into a countywide unit, providing exemplary social services at scale. The purpose of this study is to understand the historical trajectory of this amalgamation. Primary sources were collected from Heinz History Center Detre Library and Archives, Oliver Special Collections Room at Carnegie Library of Pittsburgh (CLP), and the University of Pittsburgh Archives Service Center. Data was then analyzed according to a public sphere theory proposed by Habermas (1996). In this model, communicative power flows to legislative bodies where it is transformed into administrative power through law, leading to system-level developments. In this study, communicative power and its resulting effects during five periods of activity are identified and explained. This study contributes to a political understanding of infrastructural development.

5. “Accessibility Cyberscapes and the Digital Divide” by Jessica Benner, PhD Student (LIS)

Presenter: Jessica Benner; Advisor: Jung Sun Oh and Hassan Karimi

This poster discusses early findings on a study of user generated accessibility information. We utilize data from OpenStreetMap that include tags related to wheelchair accessibility and compare the distribution of these data to other explanatory data collected in the American Community Survey. We perform a linear regression and find that the population of people with disabilities in an area had a stronger relationship with the frequency of data points than household income or the urban status. Future work includes analyzing the accessibility data within the context of other data found in OpenStreetMap, and the use of additional measures associated with the digital divide such as level of education or the number of available access points.

6. “Development of the EDDA Study Design Terminology to Enhance Retrieval of Clinical and Bibliographic Records in Dispersed Repositories” by Ashleigh Faith, PhD Student (LIS) (with Eugene Tseytlin & Tanja Bekhuis, Department of Biomedical Informatics at Pitt)

Presenter: Ashleigh Faith; Advisor: Alison Langmead

Medical terminology varies across disciplines and reflects linguistic differences in communities of clinicians, researchers, and indexers. Inconsistency of terms for the same concepts impedes interoperable metadata and retrieval of information, such as clinical reports and scientific articles in various repositories. To facilitate information retrieval and, more recently, data sharing, the medical community maintains an assortment of terminologies, thesauri, and ontologies. Valuable resources include the US National Library of Medicine Medical Subject Headings (MeSH), Elsevier Life Science thesaurus (Emtree), and the National Cancer Institute Thesaurus (NCIt). It is increasingly important to identify medical investigations by their design features. The EDDA Group at the University of Pittsburgh is developing a terminology of designs. Among the resources analyzed, inconsistent entry points, semantic labels, synonyms, and definitions are common. The EDDA Study Design Terminology is freely available in the NCBO BioPortal (<http://purl.bioontology.org/ontology/EDDA>). A primary goal is to improve identification and retrieval of electronic records describing studies in dispersed data warehouses or electronic repositories.

7. “Understanding Health Information Intent via Crowdsourcing: Challenges and Opportunities” by Di Lu, PhD Candidate (IS) and Wei Jeng, PhD Student (LIS)

Presenter: Di Lu & Wei Jeng; Advisors: Yu-Ru Lin, Rosta Farzan

Social Q&A sites have been emerging as a platform for people to seek information and social supports around health topics. Identifying users' information needs from the questions can significantly help social Q&A sites serve their users better. Prior research had attempted to understand askers' intentions and implicit needs by classifying hidden intent from questions, while the non-trivial categorization was only able to be conducted with a limited size of data. In this study, we aim to develop a scalable categorization method that can categorize the askers' intent in a large set of health-related questions via crowdsourcing. We conducted a preliminary experiment on Amazon Mechanical Turk to evaluate our categorization method. Our results suggests both challenges and opportunities for understanding health information intent via crowdsourcing.

8. “Wireless Network Virtualization: Opportunities for Spectrum Sharing in the 3.5GHz Band” by Marcella Gomez, PhD Candidate (IS with concentration in Tele)

Presenter: Marcella Gomez; Advisor: Martin Weiss

In this research work we aim at evaluating the opportunities that Wireless Network Virtualization can bring for Spectrum Sharing. Our study is framed within the regulatory guidelines presented by the Federal Communications Commission (FCC) for shared operations in the 3.5 GHz band.

Through its three-tiered sharing proposal, this regulatory approach envisions the creation of an “innovative band”, which could result in an efficient sharing arrangement due to the access and usage possibilities it creates for a wider range of users. In this light, by pairing this framework with a novel technical approach, such as Wireless Network Virtualization, we worked toward assessing the opportunities that could be found, not only from a regulatory and technical perspective, but also taking into account the economic factors that influence the viability of such a sharing scenario. As a result, this work lays out a comprehensive ground for further exploration and development of virtualized networks that would provide significant opportunities for enabling and enhancing current spectrum sharing proposals. In addition, it leaves an open door for adapting social considerations, which became evident and essential as we developed our analysis.

9. “The Spine Concept: A new approach to provide high resilience and differentiated services” by Abdulaziz Alashaikh, PhD Candidate (IS with concentration in Tele)

Presenter: Abdulaziz Alashaikh; Advisor: David Tipper

The reliance of critical services (e.g., financial transactions, emergency calls, smart grid communications, etc.) on communications networks are expected to grow in the near future. It is also expected that these services would require high end-to-end availability that might be higher than what basic protection schemes would offer. Thus providing extremely high availability levels for critical services has become a major concern for network providers. In addition, communications networks typically serve many customers with different requirements in terms of quality of service as well as availability. Hence supporting differentiated services is needed to control cost and meet customers' requirements. Our proposal "The Spine" gives a solution to both problems. The basic idea is to embed a highly available subnetwork in the physical layer and use this subnetwork as a basis to differentiate between traffic classes.

10. “Scalable Processing of Aggregate Continuous Queries in a Distributed Environment” by Anatoli Shein (Department of Computer Science)

Presenter: Anatoli Shein; Advisors: Panos Chrysanthis; Alexandros Labrinidis

Data Stream Management Systems (DSMSs) performing on-line analytics rely on the efficient execution of large numbers of Aggregate Continuous Queries (ACQs). In this paper, we study the problem of generating high quality execution plans of ACQs in DSMSs deployed on multi-node (multi-core and multi-processor) distributed environments. Towards this goal, we classify optimizers based on how they partition the workload among computing nodes and on their usage of the concept of Weavability, which is utilized by the state-of-the-art WeaveShare optimizer to selectively combine ACQs and produce low cost execution plans for single-node environments. For each category, we propose an optimizer, which either adopts an existing strategy or develops a new one for assigning and grouping ACQs to computing nodes. We implement and experimentally compare all of our proposed optimizers in terms of (1) keeping the total cost of the ACQs execution plan low and (2) balancing the load among the computing nodes. Our extensive experimental evaluation shows that our newly developed Weave-Group to Nodes (WG_tn) and Weave-Group Inserted (WG_i) optimizers produce plans of significantly higher quality than the rest of the optimizers. WG_tn minimizes the total cost, making it more suitable from a client perspective, and WG_i achieves load balancing, making it more suitable from a system perspective.

11. “Annotation & Classification of Argumentative Writing Revisions” by Fan Zhang (Department of Computer Science)

Presenter: Fan Zhang; Advisor: Diane Litman

This paper explores the annotation and classification of students' revision behaviors in argumentative writing. A sentence-level revision schema is proposed to capture why and how students make revisions. Based on the proposed schema, a small corpus of student essays and revisions was annotated. Studies show that manual annotation is reliable with the schema and the annotated information helpful for revision analysis. Furthermore, features and methods are explored for the automatic classification of revisions. Intrinsic evaluations demonstrate promising performance in high-level revision classification (surface vs. text-based). Extrinsic evaluations demonstrate that our method for automatic revision classification can be used to predict a writer's improvement.

12. “Collaborative Logistic Regression for User Behavior Prediction” by Zitao Liu (Department of Computer Science)

Presenter: Zitao Liu; Advisor: Milos Hauskrecht

In this poster presentation, we develop a collaborative logistic regression (CLR) model for large scale user recommendations for each Tmall shops, which is the largest business-to-consumer online retail store in China. Instead of solely leveraging raw user behavior features (user click count, number of items put in the cart, number of favorite items, and number of purchases), we borrow the idea of collaborative filtering and develop collaborative features (first order features) based on shop-to-shop similarities. We test our CLR model on a 42-million-user, 153-thousand-shop, 10-billion-record dataset which gives us a 3% increase in terms of recall.