Undergraduate Informatics Research Projects FA2021
List of Project Titles (project details are on following pages)

Project 1: **Design and Optimization of a Web Portal for Identifying Adults in Need of Exercise Support** (Dr. Sean Mullen, KIN & CHLH)

Project 2: **Asteroid Impact game development** (Dr. Jacob Fisher, College of Media)

Project 3: **VR Contemporary Dance Adventure** (Dr. John Toenjes, Dance)

Project 4: **How do researchers work together?** (Dr. Matthew Caesar, Computer Science)

Project 5: **SALT lab** (Dr. Yun Huang and Dr. Yang Wang, iSchool)

Project 6: **Enabling Computational Analysis of Fish Species Names and Biodiversity Data** (Dr. Milton Tan, Illinois Natural History Survey)

Project 7: **Automating Targeted Sequence Assembly and Extraction from Genomic Data for Phylogenetics** (Dr. Milton Tan, Illinois Natural History Survey)

Project 8: **Prediction of motor impairments using wearable accelerometer data in persons with Parkinson's disease** (Dr. Manuel Hernandez, KIN &CHLH)

Project 9: **Using novel satellite data to study environmental impact on agriculture productivity** (Dr. Kaiyu Guan, NRES)

Project 10: **Design Guidance for Behavioral Feedback Visualizations to Support Health Self-Management for Older Adults** (Qiong Nie & Dr. Wendy Rogers, KIN & CHLH)

Project 11: **Evaluation of Mobile Apps for Privacy** (Dr. Masooda Bashir & Tian Wang, iSchool)

Project 12: **My Newbery at 100: Young people's responses to literature through hands-on making and game design** (Dr. Kyungwon Koh, iSchool)

Project 13: **Advancing the Communication Platform of the Socio-Technical Risk Assessment** (Dr. Zahra Mohaghegh, NPRE)

Project 14: **Building digital health solutions for patients with chronic or progressive health conditions** (Dr. Jessie Chin, iSchool)

Project 15: **User experience and human-computer interaction (HCI) studies of the diverse patients** (Dr. Jessie Chin, iSchool)

Project 16: **Learning in the wild** (Dr. Jessie Chin, iSchool)

Project 17: **Building cognitive games to understand human behavior during information proliferation** (Dr. Jessie Chin, iSchool)

Project 18: **Developing computational models of plant health for an interactive classroom simulation** (Dr. Michael Tissenbaum, Curriculum & Instruction)

Project 19: **Envirotyping dashboard for agricultural field trials** (Dr. Nicolas Martin, CPSC)

Project 20: **Quantification of crop growing conditions in the US Midwest** (Dr. Nicolas Martin, CPSC)

Project 21: **Analysis of University Donations that lead to Named Chairs** (Dr. Michael Twidale, iSchool)

Project 22: **Developing a small proof of concept of integrated cultural heritage** (Dr. Michael Twidale, iSchool)

Project 23: **Analysis of Magazine Covers (ie Rolling Stone) through Time** (Dr. Michael Twidale, iSchool)

Project 24: **Annotating methods-related information in biomedical papers** (Dr. Jodi Schneider, iSchool)

Project 25: **Developing a protocol for constructing a relatedness matrix for a user-supplied bibliography** (Dr. Jodi Schneider, iSchool)

Project 26: **Network analysis and visualization of the exercise & depression research community** (Dr. Jodi Schneider, iSchool)

Project 27: **Politicization and polarization of COVID-19 information in the news** (Dr. Jodi

Project 28: **Writing a literature review about retracted papers** (Dr. Jodi Schneider, iSchool)

Project 29: **Using graph databases and network analysis to detect financial bias in research** (Dr. Jodi Schneider, iSchool)

Project 30: **Understanding how information moves between different expert communities: a team science case study of misunderstandings about airborne transmission of SARS-CoV-2** (Dr. Jodi Schneider, iSchool)

Project 31: **Analyzing citations to a fraudulent clinical trial by using OpenCitations data** (Dr. Jodi Schneider, iSchool)

Project 32: **The Center for Health Informatics (CHI): WHO/PAHO Anti-infodemic Virtual Center** (Dr. Ian Brooks, iSchool)

Project 33: **The Center for Health Informatics (CHI): Data Analytics** (Dr. Ian Brooks, iSchool)

Project 34: **The Center for Health Informatics (CHI): Social Media Analytics** (Dr. Ian Brooks, iSchool)

Project 35: **The Center for Health Informatics (CHI): Modeling of COVID-19 and other respiratory diseases** (Dr. Ian Brooks, iSchool)

Project 36: **Data science and network analysis for Humanitarian Assistance and Disaster Relief (HADR)** (Dr. Jana Diesner, iSchool)

Project 37: **Collaborative norms in computational social science** (Dr. Jana Diesner, iSchool)

Project 38: **Using Natural Language Processing and Human-Centered Data Science to Identify and Extract Impacts based on Grant-Funded Project Documents** (Dr. Jana Diesner, iSchool)

Project 39: **Words and Networks – validating, advancing and open-sourcing methods for the relational analysis of text data** (Dr. Jana Diesner, iSchool)

Project 40: **Responsible Use of Data – identifying, extracting, and interpreting information from data regulation**s (Dr. Jana Diesner, iSchool)

Project 41: **Exploring the co-evolution of ribosomal proteins and RNA with bipartite networks and phylogenomic reconstruction** (Dr. Gustavo Caetano-Anollés, CPSC)

Project 1: Design and Optimization of a Web Portal for Identifying Adults in Need of Exercise Support (Dr. Sean Mullen, KIN & CHLH)

We are seeking motivated students with interest in applying informatics to understand and improve exercise behavior interventions.

Background:
Approximately 50% of adults who begin an exercise program discontinue within 3-6 months. We have developed a highly accurate ($\geq$90%) predictive model that can differentiate exercise dropouts from adherers. Once deployed, we believe our "PASSFIT Profile" will be invaluable to the fitness industry, healthcare providers, and academic researchers. It will inform clinical practice by helping medical providers "flag" patients early who are likely to discontinue behavioral prescriptions. Our continued data collection will also facilitate the development of timely and individualized interventions for preventing dropout. The profile is based on ~10 variables that can be assessed online via self-report & a brief cognitive assessment. We still need to conduct further testing and tweaking classifications to ensure optimal responses to sampling variations in profile data. More details provided during interview process.

Estimated Project Needs:
- 2-3 students needed to work with PI and his team (including INFO PhD student mentor)
- Weekly (flexible) communication with PI and team with option of 2 or 3 credits (5-7 or 8-10 hours per week). Work may be completed remotely, in part. Occasional meetings and in-person reports will be required.
    - Web portal interface design in accordance with accessibility standards https://www.w3.org/TR/wcag-3.0/
    - Database development
    - Algorithmic optimization via AI would be helpful
    - Data visualization/feedback responding to user-data input

Prerequisite Skills:
- None required; any experience in web development (e.g., Wordpress), survey design (e.g., Qualtrics), database design (e.g., MySQL, MS Access), coding (R, html) is preferred.
- Experience with https://shiny.rstudio.com/gallery/ would be preferable.

Project 2: Asteroid Impact game development (Dr. Jacob Fisher, College of Media)

Asteroid Impact is an open source video game designed to investigate attention, cognitive control, and motivation processes, and how these processes are influenced by varying task parameters. Students who join this project will assist in the continued development of Asteroid Impact, and will help implement studies conducted using the platform. Main tasks will include code development and documentation in Unity, containerized deployment to a web server, and data management. Opportunities will be available to assist in data collection, data analysis, and other research activities if desired.

Skills required: Advanced familiarity with Unity, C#, Git, and GitHub. Ability to work independently and collaborate with an interdisciplinary team.
Contact: Dr. Jacob T. Fisher (jtfisher@illinois.edu)

Project 3: VR Contemporary Dance Adventure (Dr. John Toenjes, Dance)

We are creating a "contemporary dance adventure" within Virtual Reality that will be designed to include game play and participant interaction in what is normally a passive observer activity. This project has several goals: to explore ways to create and present theatrical dance within VR; to investigate how to incorporate participant interaction in dance both in a competitive and creative way; to increase participants' understanding of contemporary dances as an art form; to create a community of users within the app to grow the art form in general.

This effort requires many different skill sets, including plot writing, VR programming, 3D graphics, and motion capture. Student researchers in this project will work in a team to take this "Dance Adventure" from conception to reality.

Students interested in this project could choose to focus on the construction of the VR program in its various aspects such as how to capture motion data and translate it into avatar movement or user responses, or on spatial audio technologies, or network integration. Another area of research would be usability and user reaction studies.

With such a wide variety of needs this project can accommodate a wide variety of interests. Interested students can contact me to fashion a research project that will contribute to the development and/or study of this new art form.

Contact: Dr. John Toenjes (jtoenjes@illinois.edu)

Project 4: How do researchers work together? (Prof. Matthew Caesar, Computer Science)

The health of our research community depends on the ability of professionals to interact and form scientific collaborations. In this work we will analyze a large-scale dataset of research papers to understand how computer science researchers work together, who is working together, what is the nature of those collaborations. It is our hope this analysis will lead to insights that can help us inform professional societies and funding agencies to improve their practices, and improve inclusion of underrepresented groups in our community.

Prerequisite skills: good knowledge of programming (e.g., Python). Natural language process, machine learning, computer vision, or frontend/Javascript programming is a plus.

Contact: Dr. Matthew Caesar (caesar@illinois.edu)

Project 5: SALT lab (Prof. Yun Huang and Prof. Yang Wang, iSchool)

The iSchool SALT lab (https://salt.ischool.illinois.edu/index.html) conducts social computing systems research. Our work received support from Google, Facebook, IBM Research, Japan NTT Research Lab, NSF, IMLS, and ACL.

Our lab graduates land dream positions: some are faculty (e.g., at UMBC and Yale-NUS) and some work at top IT companies (e.g., Facebook, Google) and the Big Four (e.g., EY).

Currently, we are looking for passionate students to collaborate on several projects, including but not limited to the following:

1. Designing a coaching chatbot for college students' safety, career development, and wellbeing;
2. Exploring effective XR (Extended Reality) for hybrid learning;
3. Creating a social media reporting tool for community knowledge building; and
4. Designing tools to help improve online security and preserve user privacy.

Students who are interested in pursuing careers as qualitative or quantitative data analysts, web developers, mobile app developers, graphic designers, and professional writers are welcomed to apply

Contact: Prof. Yun Huang (yunhuang@illinois.edu, http://yunhuang.web.illinois.edu/) and Prof. Yang Wang (yvw@illinois.edu, http://yvw.web.illinois.edu/).

Project 6: Enabling Computational Analysis of Fish Species Names and Biodiversity Data (Dr. Milton Tan, Illinois Natural History Survey)

The study of species in biology depends on the scientific names assigned to species. Despite their importance, scientific names can be ambiguous because they can be updated over time as scientists update their classification. These changes can lead to confusion, and complicates the analysis of biodiversity data from databases linked to species names, which is everything from geographical data, to ecological data, to genetic data. With millions of species, managing and reconciling species names can be a challenge, but luckily existing databases and taxonomy reconciliation software exist for many groups of organisms. However, there are current limitations in obtaining and using up-to-date species classification data in fishes. The goal of this project is to develop tools for the downloading and processing of scientific name data from the online database Eschmeyer's Catalog of Fishes, enabling better study of the ecology and biology of fish species. This work integrates with other projects in the Fish Biodiversity Genomics Lab on studying the diversity of fishes, and the project can be continued or extended in the future.

There are no prerequisites, but experience in R or python will be helpful. The student will be trained in the relevant computing skills needed to accomplish the project, including basic coding skills, interaction with databases, text processing, and data management.

Project 7: Automating Targeted Sequence Assembly and Extraction from Genomic Data for Phylogenetics (Dr. Milton Tan, Illinois Natural History Survey)

Reconstructing the Tree of Life is a central question in biology to understand the evolution and relationships among species. The field of phylogenetics based on DNA sequence data has greatly advanced reconstructing the relationships across life, and continued progress can build upon prior work and databases of available sequence data. Such phylogenies can enable answering questions about when and how species arose and diversified over time and over geography, as well as how species have evolved their traits. As sequencing technologies have advanced, various different types of data stored on different databases are now used in phylogenetics, and effectively integrating these in an automated fashion would accelerate the reconstruction of the Tree of Life. The goal of this work is to build on earlier work in reconstructing the tree of life to develop a tool to integrate genomic data into a phylogenetics pipeline using targeted sequence assembly. This work integrate with other projects in the Fish Biodiversity Genomics Lab on studying the evolutionary relationships of fishes, and the project can be continued or extended in the future.

There are no prerequisites, but experience in Unix/Linux command-line interface and python will be helpful. The student will be trained in relevant computing skills needed for the project, including using the Unix/Linux command-line interface, interaction with databases, basic coding skills, data management, and working with remote computing.

Project 8: Prediction of motor impairments using wearable accelerometer data in persons with Parkinson's disease (Prof. Manuel Hernandez, KIN &CHLH)

The project goal will be to develop machine learning/artificial intelligence frameworks to predict motor function in persons with Parkinson's disease using wearable accelerometer data. Using python or MATLAB, we will first refine features used to quantify motor function from wearable sensors and integrate with traditional and deep learning models to improve the classification performance of motor dysfunction and tremor in persons with Parkinson's disease. You will be working with members of our team and colleagues in the nascent Center for Wearable Intelligent Technologies to build ML/AI frameworks and guide development of future wearable sensors that can provide feedback to users, which may positively impact health and wellness.

Required and preferred qualifications are provided, but interest and willingness to learn will be most crucial.

Required Qualifications: programming skill - python or MATLAB - R

Preferred Qualifications: - machine learning - experience with biomedical data, neuroscience, or multimodal data

A student interested in contributing to writing is also encouraged.

Project 9: Using novel satellite data to study environmental impact on agriculture productivity (Dr. Guan, NRES)

Dr. Kaiyu Guan's lab is conducting research on using novel satellite data from the NASA satellites to study environmental impact on global and U.S. agriculture productivity, in the platform of the most powerful supercomputer in scientific research (**Blue Waters**). We are looking for highly motivated and programming-savvy students to join the lab. The chosen students will be closely mentored by Dr. Guan, and will be working on issues including processing large satellite data, understand and implement remote sensing algorithms, and solve questions that are related to the global food production and food security.

Project 10: Design Guidance for Behavioral Feedback Visualizations to Support Health Self-Management for Older Adults (Qiong Nie & Dr. Wendy Rogers, KIN & CHLH)

Technologies (e.g., sensors, mobile apps, wearable devices) and built-in interactive displays can facilitate health self-management by collecting personal information and providing feedback visualizations to help people better understand and reflect on their behavior. Providing feedback via visualization can increase feedback effectiveness. It can help people see patterns and trends of their behavior, provide knowledge or expertise to make correct conclusions about their observations, motivate and maintain behavior change. In this research project, we aim to determine the best approach to design understandable, reflective and actionable behavioral feedback visualizations for older adults to support health behavior change. There are two phases in the study. Phase 1 is a systematic review study. We organize empirical evidences for the effect of behavioral feedback visualizations on health behavior change and develop an organizational framework representing the visual feedback--action process. In Phase 2, we focus on developing feedback visualizations for medication management for older adults. We will recruit older adults to assess their needs for feedback visualizations of medication taking, and interview pharmacies and nurses to understand requirements for developing the visualizations. We will also develop and assess the feedback visual protocols with the subject matter experts and older adults. Central to this process is developing engaging and easy to understand visualizations so that older adults can track their progress and be motivated to take medications.

We are looking for an undergraduate/graduate researcher to contribute to the behavioral feedback and visualization design process. You will register with Dr. Wendy Rogers in the

Human Factors and Aging Lab, work closely with Qiong Nie (PhD candidate) and interact with other graduate and undergraduates to work on this project.

Intern responsibilities. Your work will include conducting literature reviews, collaborating and supporting the design and development of visualization prototypes, conducting usability tests and taking meeting notes, etc.

Intern requirements. Our ideal researcher is motivated, detail-oriented, and responsible. You should be willing to learn specialized software and research methods (such as interface design tools, user testing, and qualitative and quantitative statistical analysis).

Project 11: Evaluation of Mobile Apps for Privacy (Dr. Masooda Bashir & Tian Wang iSchool)

Mobile app privacy has been a serious concern as the development of information technology. Not only the service providers may sell users' information collected to advertisers or use it for their own purposes, but also the popularity and market size for mobile apps make individual users ideal targets for cybercriminals who see games as a platform for stealing users' information or invading their privacy. The goal of this project is to evaluate different types of mobile apps and see how they protect users' privacy, or if any privacy violations exist when using such app.

Role: Help with data pre-processing/cleaning and data analytics (developing data visualizations).

Prerequisite skills: At least be familiar with Excel, and basic data analysis skills (knowledge of R, Python, or other data analytic tools).

Project 12: My Newbery at 100: Young people's responses to literature through hands-on making and game design (Dr. Kyungwon Koh, iSchool)

Bestowed each year by the American Library Association [ALA], the prestigious Newbery Medal recognizes "the author of the most distinguished contribution to American literature for children" (ALSC [Association for Library Service to Children], 2021). The Newbery celebrates its 100th anniversary this year, and the Center for Children's Books (CCB) is marking the occasion by launching an archive-building project, My Newbery at 100, that seeks to collaborate with young people to better understand the Medal and its influence. How do youth relate to the award, the awarded books, and the stories that matter to them?

As part of My Newbery at 100, the C-U Community Fab Lab and its community partners will facilitate and collect young people's responses to Newbery titles through hands-on making

activities with a range of technologies and creative game design approaches. We seek to investigate questions, such as:
- What are young people's responses to Newbery Medal books?
- How do youth use maker technologies and game design principles to respond to literature and express themselves?
- How do youth and adult mentors successfully collaborate in informal learning and making contexts?

We are also interested in capturing youth's responses to recent critiques of the award that center on issues of equity, inclusion, and representation.

Students who are interested in youth, literature, making, technology, game design, and participatory and community-based research are encouraged to apply. Participating student researchers will help design and facilitate the program and engage in research activities, including unobtrusive observation, documentation, and artifact analysis. Student researchers will develop their skills in understanding research processes, conducting literature reviews, data collection and analysis, and preparing research for publication and presentation. They are expected to participate in weekly meetings. No prerequisite skills are needed. Applicants must be willing to and enjoy working with young people from different cultures and backgrounds in a respectful way.

Project 13: Advancing the Communication Platform of the Socio-Technical Risk Assessment (Dr. Zahra Mohaghegh, NPRE)

The Socio-Technical Risk Analysis (SoTeRiA) Research Laboratory in the Department of Nuclear, Plasma and Radiological Engineering (NPRE) at the Grainger College of Engineering at University of Illinois Urbana-Champaign (http://soteria.npre.illinois.edu/) is welcoming Undergraduate Researchers in diverse areas of multidisciplinary risk analysis. Probabilistic Risk Assessment (PRA) is the central risk assessment technique used in SoTeRiA research. PRA is a systematic risk methodology and a key pillar of policy setting in the U.S. Nuclear Risk-Informed Regulatory Framework. Its applications have been extended to diverse high-consequence industries (e.g., space, oil and gas, healthcare, etc.). Undergraduates will have the opportunity to research under the supervision of Professor Zahra Mohaghegh and in collaboration with other Undergraduates, Masters and Ph.D. students in the SoTeRiA Laboratory.

The SoTeRiA Laboratory has been working on real-world risk analysis problems that are scientific, innovative, and practical. One of the key components of research at the SoTeRiA Laboratory is to properly communicate the balance between risks posed by energy technologies and the potential public benefits from the development and deployment of these technologies. This project focuses on advancing the current SoTeRiA risk platform to improve the communication among the risk community, the nuclear industry, and the public. Major challenges are understanding and communicating risk associated with different energy technologies to people outside engineering domains.

The student will support the development of articles and design of videos/websites that can highlight the criticality of risk analysis in a simple and effective way with multi-disciplinary groups of students and faculty. We are looking for students with experience in marketing and social media to capture students' perspectives on communicating safety risk. The student will gain knowledge on risk assessment approaches during meetings with other graduate/undergraduate students and senior members in SoTeRiA. The ideal candidate is a highly motivated student with a marketing background who is interested in exploring different communication approaches to create a positive impact on a diverse group of communities with different backgrounds. It is expected that the student has a strong writing background as well as experience in web-design.

Project 14: Building digital health solutions for patients with chronic or progressive health conditions (Dr. Jessie Chin, iSchool)

In this research project, we are developing digital health solutions for patients with chronic or progressive health conditions (including patients with spinal cord injuries, cancer, etc.) to support them to manage and detect their healthcare needs. We aim at combining data sciences, behavioral sciences theories, and artificial intelligence to deliver accessible and active care to diverse patients in our community.

We are looking for students who are interested in coding and building technological solutions for healthcare applications. Students who have passion in helping people with health needs and are willing to learn new things are especially welcome. Students will be working with a multidisciplinary team (with faculty and graduate students in other disciplines) and the actual patients to build solutions for patients manage their daily challenges. Students will be assisting in prototype design, mobile application development, chatbox development or browser plugin development.

- Necessary prerequisite skills: experience in either of the following: mobile application, chatbox, database-backed web interfaces, or browser plugins development.
- Open to all students.
- Please find more information about our team. https://jessiechinlab.ischool.illinois.edu/

Project 15: User experience and human-computer interaction (HCI) studies of the diverse patients (Dr. Jessie Chin, iSchool)

In this research project, we aim at understanding the day-to-day challenges of the diverse patients (including patients with spinal cord injuries, cancer, multiple sclerosis, etc.), and their experience with different kinds of information technologies.

We are looking for students who are interested in human-computer interaction (HCI) or user studies. Students who have passion in helping people with health needs and are comfortable with working with human participants are especially welcome. Students will be working with a multidisciplinary team (with faculty and graduate students in other disciplines) and will be assisting in study materials creation, study set-up, participants' scheduling and follow-up, and data collection.

- Necessary prerequisite skills: Comfortable communicating with people
- Open to all students.
- Please find more information about our team. https://jessiechinlab.ischool.illinois.edu/

Project 16: Learning in the wild (Dr. Jessie Chin, iSchool)

With the ubiquity of access to the knowledge in the world (e.g., search engines, massive open online course), people are taking more active roles to regulate their learning in the information environments. The project is to examine how people learn with or without the coupling with technologies, and how to develop interactive technologies to support adults to learn on their own.

We are looking for students who are interested in developing and conducting behavioral studies. Students will receive training in the research in cognitive sciences. Students will be assisting in different tasks, including developing behavioral experiments, programming experiments, study materials design and creation, and data collection in human subject studies.

- Necessary prerequisite skills: Interests in behavioral studies
- Preferred skills: software development, web or mobile application programming, experience in human subject studies
- Open to all students.
- Please find more information about our team.
  https://jessiechinlab.ischool.illinois.edu/

Project 17: Building cognitive games to understand human behavior during information proliferation (Dr. Jessie Chin, iSchool)

In the contemporary digital environment, people tend to adapt to information proliferation with the assistance/companion with technologies. Few studies have examined human cognition outside of the lab environments. To better capture human behavior in the naturalistic setting, we aim at developing cognitive games to simulate ecological human behavior during information

proliferation in situ. The study will have implications on designing better information environments to augment human cognition during information proliferation.

We are looking for students who have skills in software development, web application or mobile application development, and would like to learn how to translate these skills to understand human cognition and behavior in the naturalistic setting. Students will receive training in the research in cognitive sciences. Students will be assisting in developing behavioral games, and learning how to obtain meaningful data to gain insights on human behavior.

- Necessary prerequisite skills: software development, web or mobile application programming (or willing to learn how to program studies)
- Open to all students.
- Please find more information about our team. https://jessiechinlab.ischool.illinois.edu/

Project 18: Developing computational models of plant health for an interactive classroom simulation (Dr. Michael Tissenbaum, Curriculum & Instruction)

We are developing an interactive science simulation that enables middle school students to investigate a digital plant habitat and learn how different environmental and genetic factors affect plant health. As part of this simulation, we need to model how different factors (water, light, heat, soil composition) impact the health, growth, and fruit/vegetable production of the plants that students grow in the simulation.

Your job, as part of the research team, will be to help us develop these computational models based on data we've collected on a range of plants. You will be given a range of inputs for the plants' habitat, which you will use to calculate output variables, such as the plants' health as a given score (in collaboration with the research team). No prior experience with plants or agriculture are required, but they are a bonus!

The final integration of the computational model into the simulation will be handled by our development team, so no direct programming experience is required, but experience with game or system modeling software (e.g., machinations.io) is helpful. We're also willing to develop the models using other software if you have experience using those instead.

Project 19: Envirotyping dashboard for agricultural field trials (Dr. Nicolas Martin, CPSC)

Successful agricultural experimentation involves understanding the expected conditions in growing regions. With this purpose, agricultural experiments generate data from cultivars tested at multiple locations and years before making results available to farmers. These trialing networks are often denominated as multi-environment trials (MET). Research programs also are expected to focus on and region where farmers will grow their crops. These regions

experience a mixture of environmental conditions that can be encountered across multiple years within a defined geography. A robust crop model (APSIM) will be validated on phenotypic data from a field experiment and then deployed to simulate soybean yields across the US Midwest. A protocol has been developed to define growing stages and then characterize growing conditions by crop growing step.

Specific goals:
1.   Development of an R Shiny tool to characterize field trials based on their geographical information.
2.   Design visualizations to display complex results intuitively

Suggested Skillsets: R, Shiny

Project 20: Quantification of crop growing conditions in the US Midwest (Dr. Nicolas Martin, CPSC)

The first step in this process is building datasets to quantify environmental conditions in potential research sites and cropping areas in Malawi and other African countries and seek feedback from stakeholders. Then this characterization can be used to quantify growing conditions in TPE, and he MET for this breeding project. A comprehensive, quantitative understanding of Target Population Environments would benefit the business intelligence efforts through improved decision-making, specifically for evidence-based decisions such as trial placement within areas and valuation of phenotypic performance. We will use high-resolution maps to identify suitable regions for trial allocation based on the expected stability of top-yielding cultivars—geospatial tasks and model deployment at larger scales to broaden impact.

Specific goals:
1.   Quantification of the potential soybean crop footprint by cropland frequency.
2.   Identification of dominant soil types season conditions in US cropland areas.

Suggested Skillsets: SQL, R, GIS

Project 21: Analysis of University Donations that lead to Named Chairs (Dr. Michael Twidale, iSchool)

A colleague and I are looking at university donations by people that lead to named chairs. It is inspired by all the concern about the Sackler family's donations to universities and the problems that can cause when we learned more about opioid abuse. But of course it can applyto all kinds of problems that can arise when we look a bit more closely at how the money was made, or who the person was doing the giving. We want to create a database of named chairs: who gave the money, when, who the chair in is honor of, any stipulations on what it is for. We'd like to do this for all major universities in Illinois. The work involves looking up information about these chairs, press announcements etc. The information may require ingenuity to find.

Project 22: Developing a small proof of concept of integrated cultural heritage (Dr. Michael Twidale, iSchool)

As more archives are digitized, it becomes possible to interconnect information. Example: You take Great-grandma's scrapbook you found in the attic. You see a concert program she saved. A date with great grandpa. What more can you find about this event? who what where when, etc.? which databases can we link it to?

The project does not require tech development skills. We are not going to code it up yet. Instead we will generate examples by doing manual searches and compiling results as worked examples as a way to communicate the vision of the idea.

Project 23: Analysis of Magazine Covers (ie Rolling Stone) through Time (Dr. Michael Twidale, iSchool)

Certain magazines have great cultural significance. Rolling Stone was one. Who appeared on the cover of Rolling Stone was A Big Deal in revealing the cultural zeitgeist. So who did appear month by month over the decades? What does that tell us about evolving societal and cultural issues?

The project will be to construct a simple database of covers. Also to look at the ages of who appears. In the 1960s, it was always young people: Mick Jagger, Paul McCartney. More recently it has included much older people: Pope Francis, Bernie Sanders - and Mick Jagger.
How has age, gender and race of who is on the cover changed over time .
We have a rough dataset. But it needs further tidying up. Then we will develop various visualizations of the data using simple tools.

Project 24: Annotating methods-related information in biomedical papers (Dr. Jodi Schneider, iSchool)

Do you have a background in biomedical-related fields (biology, bioengineering, bioinformatics, chemistry, etc.)? Want to gain research experience in biomedical informatics and text mining? In this project, you will read biomedical papers then identify and annotate important research methods and materials. We will use your annotations to train machine learning classifiers. The goal of our work is to understand how research methods or materials can directly impact the validity of the findings reported in a paper. We focus on identifying citations that support research methods and materials, which we call "methods keystone citations". This work will help develop automatic classifiers in our larger project.

REQUIRED QUALIFICATIONS
- have taken higher-level college courses in biology or related fields
- experience reading academic papers in biomedicine

MORE INFORMATION
Previously published papers:
http://jodischneider.com/pubs/jcdl2020.pdf
http://jodischneider.com/pubs/scik2021.pdf

Mentors: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org
& Yuanxi Fu, PhD student
https://ischool.illinois.edu/people/yuanxi-fu

Project 25: Developing a protocol for constructing a relatedness matrix for a user-supplied bibliography (Dr. Jodi Schneider, iSchool)

Do you have excellent programming skills? Want to apply them to a real-world challenge of collecting, cleaning, transforming, and analyzing data? In this project, you will collect data from APIs of scholarly databases like Scopus and Web of Science. Then, you will write programs to clean the data and compute "relatedness" between any two items to form a "relatedness matrix", based on measures we supply. You will also experiment with a variety of "relatedness" measures we supply and compare their differences. Then you will develop a protocol for constructing the "relatedness matrix" for any given set of bibliographies. Ultimately this will contribute to tools which will assist policymakers and grant reviewers in their decision-making.

REQUIRED QUALIFICATIONS
Proficiency in a scripting language such as R or Python

PREFERRED QUALIFICATIONS
Interest in or experience using APIs
Familiarity with CSV file format
Interest in or familiarity with scholarly databases such as Scopus and Web of Science

MORE INFORMATION
Mentors: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org
& Yuanxi Fu, PhD student
https://ischool.illinois.edu/people/yuanxi-fu

Project 26: Network analysis and visualization of the exercise & depression research community (Dr. Jodi Schneider, iSchool)

Have strong programming skills? Interested in network analysis and visualization? In this project, you will analyze and visualize author affiliations and collaborations from an existing dataset of about 500 research papers. You will write computer programs using network science

packages (e.g., igraph, networkX). You will find and read existing scientific literature about collaboration and knowledge diffusion in research communities and author networks, to get new ideas to apply to your code and analysis. Your work will contribute to our understanding of how information travels in the exercise & depression research community, using our dataset, "Systematic Reviews and Meta-Analyses Concerning Unipolar Depression and Physical Activity Interventions Published Between 2013-2019". No subject knowledge is required; we will collaborate with a domain expert on interpreting visualizations and iteratively designing them.   REQUIRED QUALIFICATIONS Proficiency in a scripting language such as R or Python Familiarity with network analysis (such as from being enrolled in/having taken IS527 Network Analysis or similar)

PREFERRED QUALIFICATIONS
Interest in psychology, sociology, or mental health is a plus, but not required

MORE INFORMATION
Mentors: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org
& Dr. Caitlin Vitosky Clarke, Teaching Assistant Professor, Department of Kinesiology and

Project 27: Politicization and polarization of COVID-19 information in the news (Dr. Jodi Schneider, iSchool)

Are you a budding computational social scientist? This project investigates how public health emergencies are reported and to assess the polarization and politicization of the U.S. news coverage. You will be responsible for testing and improving search parameters, investigating contextual information such as media bias and media circulation, using text mining and data science (particularly topic modeling through the Cline Center's new DataCapsule interface), and close reading of sample texts. You will work closely with a student who has worked on the opioid crisis. Possible publication credit.

REQUIRED QUALIFICATIONS
Proficiency in R

PREFERRED QUALIFICATIONS
Interest in any of the following is preferred but not required: COVID-19, news, political science, communications, media studies
Familiarity with or interest in learning: R Studio server, command line interfaces, DataCapsule secure computing environments (such as through the HathiTrust Research Center), topic modeling
Strong English reading & writing skills

MORE INFORMATION
Project press release:
https://ischool.illinois.edu/news-events/news/2020/07/schneider-awarded-linowes-fellowship
Mentor: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org

Project 28: Writing a literature review about retracted papers (Dr. Jodi Schneider, iSchool)

Interested in what happens when science gets things wrong? We are completing a literature review about scientific research that has been retracted due to serious errors or misconduct. Read papers, extract information, and write the literature review.

In this project you will:
- read scientific papers in medical sciences, information sciences, science of science, and related fields
- extract and manually classify information about the goals, methods, and data used for each paper
- create tables to bring together and analyze this information
- summarize the data in writing for a literature review paper

Depending on your interests you might:
- contribute to updates of the systematic search process
- update the data-driven website (https://infoqualitylab.org/projects/risrs2020/bibliography/) using existing Python code

REQUIRED SKILLS
- Strong English reading and writing skills
- Detail oriented and responsible
- Comfortable with Excel or similar spreadsheet programs

PREFERRED SKILLS
- Familiar with Pivot tables in Excel
- Willing to learn specialized software and data, such as EPPI-Reviewer, BiBTeX, RIS, and Zotero
- Interested in learning how to run an existing Python program
- Some familiarity with basics of web publishing

FURTHER INFORMATION
Bibliography: https://infoqualitylab.org/projects/risrs2020/bibliography/

Mentors: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider

Project 29: Using graph databases and network analysis to detect financial bias in research (Dr. Jodi Schneider, iSchool)

The objective of this project is to create the WhoFundedIt Knowledge Graph, which will (1) visualize the networks of individuals, institutions, and funders contributing to research; and (2) identify possible markers of financial bias in research. Graph databases have been incredibly effective in uncovering financial ethics violations that are obscured by use of intermediary people and organizations. However, graph database approaches have not been widely applied to bibliometric data. The starting point for this project will be the semantically enhanced Microsoft Academic Knowledge Graph (Färber, 2019). Your goal will be to integrate funding sources such as The Crossref Funder Registry, ProPublica's Dollars for Profs, OpenPayments from the Centers for Medicare & Medicaid, and Disclosure UK.

REQUIRED SKILLS
- Strong technical skills
- Experience in or interest in graph databases such as neo4j

PREFERRED SKILLS
- Detail oriented and responsible
- Familiarity with semantic data such as learned in Information Modeling (IS 515, previously IS 561) or strong interest in learning related skills.
- Familiarity with bibliometrics data (Web of Science)
- Interest in research ethics or science of science

FURTHER INFORMATION
Färber, 2019 Publication and data: https://www.aifb.kit.edu/web/Inproceedings3775

Mentor: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org

Project 30: Understanding how information moves between different expert communities: a team science case study of misunderstandings about airborne transmission of SARS-CoV-2 (Dr. Jodi Schneider, iSchool)

Is COVID-19 spread through the air? It took months for public health officials and aerosol scientists to agree, in part because differences in terminology caused significant

misunderstandings between experts. The goal of this project is to understand how information moves between different expert communities and how to facilitate better communication.

REQUIRED SKILLS
- Excellent English reading and writing skills

PREFERRED SKILLS
- Interest in terminology, information organization, team science, or related areas
- Interest in social media mining and/or citation analysis
- Familiarity with bibliographic management software such as Zotero

FURTHER INFORMATION

Wired, They Say Coronavirus Isn't Airborne—but It's Definitely Borne by Air: https://www.wired.com/story/the-teeny-tiny-scientific-screwup-that-helped-covid-kill/
Wired, The 60-Year-Old Scientific Screwup That Helped Covid Kill: https://www.wired.com/story/they-say-coronavirus-isnt-airborne-but-its-definitely-borne-by-air/

Mentor: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org

Project 31: Analyzing citations to a fraudulent clinical trial by using OpenCitations data(Dr. Jodi Schneider, iSchool)

This project provides a good opportunity to lead a well-defined data science research project and write a publishable manuscript in 1-2 semesters. This is a replication study with a twist: You will use OpenCitations data to analyze citations to a single retracted paper that my lab has previously analyzed. You will collect open data from OpenCitations, using methods described in (Heibi & Peroni, 2021), and compare the data you retrieve to datasets deposited for my 2020 Scientometrics paper. (Schneider et al., 2020). This will test us whether open data is a good stand-in for proprietary data. Optionally, if time allows, you can use text mining and topic modeling to give a new characterization of the citations, as (Heibi & Peroni, 2021) describes. I have a number of ideas for follow-on projects (e.g., a larger citation network covering multiple citations; use of Neo4J; ingest of full-text papers from the Crosssref API).

MORE INFORMATION

Heibi I, Peroni S. A protocol to gather, characterize and analyze incoming citations of retracted articles. arXiv preprint arXiv:2106.01781. 2021 Jun 3. https://arxiv.org/abs/2106.01781

Schneider, J., Ye, D., Hill, A.M. Whitehorn, A.S.. Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. Scientometrics

125, 2877–2913 (2020). https://doi.org/10.1007/s11192-020-03631-1 (especially datasets from this paper)

Mentor: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org

Project 32: The Center for Health Informatics (CHI): WHO/PAHO Anti-infodemic Virtual Center (Dr. Ian Brooks, iSchool)

The Center for Health Informatics (CHI) is a multi-disciplinary unit and World Health Organization Collaborating Center that is working with WHO and the Mayo Clinic in the response to COVID-19.

Alongside the COVID-19 pandemic has come a storytelling crisis, an "infodemic" of false and misleading health information. The Anti-Infodemic Virtual Center is an international proof-of-concept project developed by the Pan American Health Organization (PAHO), in collaboration with the Center for Health Informatics.

We are looking for multiple students who can contribute to the development of infodemic outbreak detection algorithms. These algorithms will vary in complexity and require a range of skills from web crawling, report generation, machine learning, to artificial intelligence.

Project 33: The Center for Health Informatics (CHI): Data Analytics (Dr. Ian Brooks, iSchool)

The Center for Health Informatics (CHI) is a multi-disciplinary unit and World Health Organization Collaborating Center that is working with WHO and the Mayo Clinic in the response to COVID-19.

We are working with the Mayo clinic to investigate the relationship between social media and other public discourse and the COVID-19 prescribing practices of physicians. We are looking for a student who can contribute to the statistical analysis of the relationship.

Project 34: The Center for Health Informatics (CHI): Social Media Analytics (Dr. Ian Brooks, iSchool)

The Center for Health Informatics (CHI) is a multi-disciplinary unit and World Health Organization Collaborating Center that is working with WHO and the Mayo Clinic in the response to COVID-19.
CHI has dedicated access to the Brandwatch social media aggregator that provides the ability to search and download more than 1.4 trillion public posts from thousands of sites including the

complete Twitter stream stretching back to 2010. We are looking for a student who will work with our partners at WHO to demonstrate the value of social media to public health.

> *Preferred but not required skills: python, text machine learning.*
> *Spanish or Portuguese language skills would be helpful, but not necessary.*

We are also looking for a student who will continue our work with dermatologists and medical students to understand the social media chatter on the burden of skin diseases including psoriasis, acne, lupus, alopecia, and skin cancers.

Project 35: The Center for Health Informatics (CHI): Modeling of COVID-19 and other respiratory diseases (Dr. Ian Brooks, iSchool)

The Center for Health Informatics (CHI) is a multi-disciplinary unit and World Health Organization Collaborating Center that is working with WHO and the Mayo Clinic in the response to COVID-19.

Leveraging CHI dedicated access to the Brandwatch social media aggregator, we want to explore the existence of a relationship between reported Flu or COVID-19 cases and related social media activity. The goal is to build a predictive model for respiratory infections based on social media data. We are looking for a student who will work with other team members to analyze social media and CDC Flu net data to build the predictive model. Preferred but not required skills: Python, R, Predictive analytics

COVID-19 and the public health measures implemented to control its spread have influenced other respiratory infections like Flu. We want to investigate the existence of a disruption in the incidence of Flu cases during the 2019-2020 and 2020-2021 Flu seasons and explore whether this hypothetical change in the patterns of Flu cases can be leveraged to build an outbreak detection model for respiratory infections. We are also interested in analyzing if this proxy can be used to assess indirectly the effectiveness of the control measures implemented for COVID-19. We are looking for a student who will work with other team members to analyze CDC Flu net data and build the outbreak detection model. Preferred but not required skills: Python, R, Predictive analytics

Project 36: Data science and network analysis for Humanitarian Assistance and Disaster Relief (HADR) (Dr. Jana Diesner, iSchool)

Given the complex nature of disaster management and maritime response, which requires tremendous amounts of planning, resources, and coordination among agencies with different functions, the involved tasks cannot be handled solely by one organization, but rather a network of organizations. Evidence of response collaborations can be empirically observed through discussions about the crisis and actions in response to crisis via a number of platforms,

including social media, news, official situational reports, and expert blogs. Hence, this project focuses on the relational extraction task to capture collaborations between response entities (e.g. organizations, persons, geopolitical entities) from sizable corpora of text from multiple data sources (i.e. Twitter vs. governmental/non-governmental situational reports). The results from relation extraction will be converted to network representations for analysis.

Student researchers working with us will have an opportunity to work hands-on with text-based data from four major hurricane events that made landfall in continental U.S., namely 2016 Hurricane Matthew, 2017 Hurricane Harvey, 2017 Hurricane Irma, and 2018 Hurricane Michael. In particular, the student will be familiarized with the process of crisis management for natural disasters, and how data analytics play important roles in providing responders with necessary situational awareness information (e.g. infrastructure damage, missing/trapped persons, donation needs). The student researcher will then participate in the process of analyzing and sensemaking of the resulting networks based on the relations extracted and the entities mentioned (e.g. authority figures, organizations, cities, countries). At the end of the research project, the student researcher will be able to construct networks based on unstructured text data, as well as detecting meaning structures within the networks.
Preferred (but not required) qualifications: experience with content analysis, text analysis with Python (e.g. nltk, spaCy).

Project 37: Collaborative norms in computational social science (Dr. Jana Diesner, iSchool)

Research in computational social science brings together theories, data, methods, and evaluation approaches from different disciplines. Innovation in computational social science makes contributions to a variety of fields. In this project, we review prior work to identify areas of contributions. In this Science of Science project, students assist with reviewing and annotating prior work from a range of disciplines to develop a new interdisciplinary framework for collaboration in computational social science. Students interested in interdisciplinary research with a background in computational methods and social science, or at least either one of these areas, are welcome to join us in this effort.

Project 38: Using Natural Language Processing and Human-Centered Data Science to Identify and Extract Impacts based on Grant-Funded Project Documents (Dr. Jana Diesner, iSchool)

The goal of this research project is to leverage and advance methods from natural language processing and human-centered data science to identify and extract
diverse impacts from funded projects based on project reports, publication records, and surveys/interviews. With this ongoing work, we aim to answer two main research questions as follows:

1. What are the impacts of grant-funded projects on academia, society, and natural environment?
2. What do we need to consider when building a multi-method approach that includes text analysis and machine learning to extract impact information and to summarize and categorize lengthy documents?

The findings from this project will help us to better understand the objectives, outcomes, and impacts of projects from diverse domains (e.g., artificial intelligence, biodiversity conservation, automobility, linguistics, and music), and how data science and text mining can be used to aid this process. This project will also contribute to automated text summarization and categorization through designing novel frameworks to summarize and categorize lengthy documents.

The undergraduate student will work with graduate students to conduct this research, thereby enhancing their research skills and gaining hands-on experience in research. The work involves studying and annotating data, conducting literature reviews, designing and evaluating algorithms and models for information extraction, text summarization, and document categorization, and interpreting quantitative and qualitative results.

Preference and Requirements:

Prior exposure to at least one programming language is required (Python or R preferably). In addition, basic knowledge about natural language processing, network analysis, and machine learning is strongly preferred. It's worth noting that we do not expect high-level computer skills, however, we hope that the students are eager to learn and improve their skills throughout the project.

Project 39: Words and Networks – validating, advancing and open-sourcing methods for the relational analysis of text data (Dr. Jana Diesner, iSchool)

This project advances methodological research at the nexus of text analysis and network analysis. This work matters as considering the content of text data and meta-data for understanding social interactions enables us to understand the impact of language use on social networks and vice versa. Research on "Words and Networks" has led to eminent work on language change, collaborative work, recommender systems, semantic computing, relation extraction, and the diffusion and use of (mis)information offline and online.

In this project, students validate and advance methods, and help to make reliable solutions available in ConText. ConText (http://context.ischool.illinois.edu/) is an open-source tool for the integrated analysis of text data and network data. It is designed to enable researchers with any level of technical skills to analyze text and network jointly. In this project, students will have the opportunity to explore various natural language processing and network analysis libraries and algorithms, integrating them into the open-sourced application, and improving the design

of a user-friendly interface. Moreover, students will be able to contribute to the open source community and help researchers from all over the world to use this product.

Students' primary tasks for this project will be to (1) understand and validate relation extraction methods, (2) add and update features in the ConText application with the guidance of senior members in the lab, (3) test the functions in the application and provide feedback, (4) assist with using the tool of data analysis research, and (5) assist with creating documentation, training and outreach material, and a user manual for the application. This work also enables students to gain experience in working with larger tools and pipelines, learn about regulations and legal aspects of using data and open sourcing tech, and making academic research practically useful to newcomers and non-academic practitioners and researchers.

Preferred Requirements: Programming / coding skills (mainly Java, some Python) are required. Proficient communication skills (especially in writing) are preferred. Knowledge in natural language processing and network analysis is strongly preferred. But most importantly, if you are detail-oriented, interested in learning new things, and passionate about helping others, please join us!

Project 40: Responsible Use of Data – identifying, extracting, and interpreting information from data regulations (Dr. Jana Diesner, iSchool)

Researchers use, collect, analyze, and share data as part of their daily research activities. These data are governed by multiple sets of norms and regulations, including but not limited to, institutional norms, intellectual property law (e.g., copyright), terms of service, privacy and security laws, technical constraints, and personal ethics. For many researchers, it is technically feasible to access data from social media platforms, online communities, or publicly available websites; however, researchers might be unaware, uninformed, or unequipped with the knowledge and skills to implement rules applicable to the use of these data. To address this issue, the project aims to identify and extract relevant information from data regulations, transform the text into easy-to-interpret "labels", and increase researchers' awareness, knowledge, and skills on responsible use of data.

In this project, students will work with senior members in the lab to conduct literature reviews, collect regulations applicable to the data use (e.g., terms of service), identify and annotate relevant data rules from the terms of service, design and evaluate models that summarize and extract multiple types of data regulations, and analyze and interpret research results.

Preferred skills: Prior exposure to at least one programming language is required (Python or R preferably). In addition, basic knowledge about natural language processing and machine learning is strongly preferred. Proficient communication skills (especially in reading and writing) are preferred. Most importantly, we hope that the students are highly interested in responsible conduct of research, and eager to learn and improve their knowledge and skills throughout the project.

Project 41: Exploring the co-evolution of ribosomal proteins and RNA with bipartite networks and phylogenomic reconstruction (Dr. Gustavo Caetano-Anollés, CPSC)

Ribosomes are the central machinery of the cell that is responsible for the synthesis of proteins. Its origin and evolution however remains mysterious. Here we explore ribosomal evolution with chronologies that provide ages to ribosomal proteins (r-proteins) and ribosomal RNA (rRNA) substructures of the molecules. The project seeks to study how r-proteins contact rRNA substructures by building a time-series of bipartite networks of r-proteins and rRNA segments, and its two projections. The outcome of this project is to study how molecules co-evolve in the 3.8 billion year history of this molecular ensemble.

We plan to use networks software (e.g. PAJEK, iGraph), so knowledge of python can be helpful.