# Undergraduate Research Projects SP2021

## List of Project Titles (project details are on following pages)

Project 1: **Data Visualization on Controversies in Health (Exercise Science & Depression)** (Prof. Schneider, iSchool)

Project 2: **Citation and Text Mining for Literature Reviews** (Prof. Schneider, iSchool)

Project 3: **Annotating Papers for a Literature Review** (Prof. Schneider, iSchool)

Project 4: **Using graph databases and network analysis to detect financial bias in research** (Prof. Schneider, iSchool)

Project 5: **Making sense of evidence in meta-analytic studies on psychotherapy outcomes** (Prof. Schneider, iSchool)

Project 6: **Designing a Mobile Application for Older Adults to Improve Hypertension Medication Adherence (Qiong Nie, PhD candidate in Human Factors and Aging lab)**

Project 7: **Response to the COVID-19 Infodemic** (Prof. Brooks, iSchool)

Project 8: **Social Media Analytics** (Prof. Brooks, iSchool)

Project 9: **Information Systems for Health** (Prof. Brooks, iSchool)

Project 10: **Modeling of COVID-19 and other respiratory diseases** (Prof. Brooks, iSchool)

Project 11: **Designing virtual reality environments for students with disabilities** (Prof. Shackelford, ANTH)

Project 12: **Using Story-telling and Crowdsourcing to Identify People's Moral Perception of Information from Images versus Texts** (Prof. Diesner, iSchool)

Project 13: **Words and Networks – validating, advancing and open-sourcing methods for the relational analysis of text data** (Prof. Diesner, iSchool)

Project 14: **Humanitarian Assistance and Disaster Relief (HADR) Data Annotation & Analysis** (Prof. Diesner, iSchool)

Project 15: **Using Natural Language Processing to Extract Long-term Impact of Investments in International Biodiversity and Bio-conservation** (Prof. Diesner, iSchool)

Project 16: **Dynamics of phthalates over time** (Prof. Smith, VetMed)

Project 17: **Authoring tools for creating 360-degree video and VR scenarios** (Prof. Toenjes, DANC)

Project 18: **Young Researchers: Introducing Teens to Early Research Experiences in Public Libraries** (Prof. Magee, iSchool)

Project 19: **Examining Teen and Adult Conceptions of Success Across Youth Serving Organizations** (Prof. Magee, iSchool)

Project 20: **Health misinformation modeling** (Prof. Chin, iSchool)

Project 21: **Building digital health solutions for patients with progressive conditions** (Prof. Chin, iSchool)

Project 22: **Web Development: ClassTranscribe** (Prof. Huang, iSchool)

Project 23: **HTRC Code Repository and Python Library Audit** (Prof. Downie, iSchool)

Project 24: **Workset event timeline algorithm** (Prof. Downie, iSchool)

Project 25: **Visualizing Geographic Entities in the HathiTrust Digital Library** (Prof. Downie, iSchool)

Project 26: **Creating HathiTrust Use Cases and How-To Guides** (Prof. Downie, iSchool)

Project 27: **Sentiment Analysis Research and Tool Development** (Prof. Downie, iSchool)

Project 28: **Transforming HTRC Extracted Features Data to Topic Models** (Prof. Downie, iSchool)

Project 29: **Analysis of Page-level Language Data in the HathiTrust Extracted Features Dataset** (Prof. Downie, iSchool)

Project 30: **How much are language models confused by errors in novels?** (Prof. Underwood, iSchool)

Project 31: **Building Out an Open-Source Community for the Social Media Macroscope** (Prof. Joe Yun, ACCY)

Project 32: **The Role of Collaborative Data Science for Solving Healthcare Challenges** (Prof. Narang, GIES)

Project 33: **Voice Based Decision Making** (Monica Porteanu, PhD Candidate INFO)

## Project 1 (Prof. Jodi Schneider, iSchool)
## Data Visualizations on Controversies in Health
## (Exercise Science & Depression)

Interested in health information & data visualization?

Develop visualizations about controversies in health! This project will extend a previous paper with a new dataset, in exercise science and depression. You will prepare data and create visualizations similar to those in the paper "Visualizing Evidence-Based Disagreement Over Time: The Landscape of a Public Health Controversy 2002-2014" (http://jodischneider.com/pubs/asist2020.pdf)

WHAT YOU WOULD DO
- Read and analyze health-related papers (exercise science depression)
- Prepare data in tabular/text format for input into existing R code
- Extract study design characteristics
- Extract included studies
- Modify and run R code (depending on your interest and experience)

PREFERRED SKILLS
- Familiarity reading health studies (exercise science depression)
- Interest in data visualization
- Familiarity or interest in systematic reviews and meta-analyses
- Familiarity identifying health study designs such as randomized controlled trial, observational study

FURTHER INFORMATION
https://infoqualitylab.org
https://ischool.illinois.edu/people/jodi-schneider

## Project 2 (Prof. Jodi Schneider, iSchool)
## Citation and Text Mining for Literature Reviews

Help evaluate our in-house citation analysis/text mining tool (built in Python by previous informatics students). Test other science mapping and network visualization software.

In this project you will:
- test an existing Python tools (e.g.
https://github.com/janinaj/lit-review-search )
- further develop Python tools for citation network analysis and text mining
- test existing automatic clustering and classification tools using text mining, network analysis, machine learning, visualization, etc.
- apply tools to an ongoing scoping review

REQUIRED SKILLS
- Some experience with Python and APIs

PREFERRED SKILLS
- Detail oriented and responsible
- Experience making tables, charts and infographics
- Interest in science of science, network analysis, and/or text mining

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

### Project 3 (Prof. Jodi Schneider, iSchool)
**Annotating Scientific Papers for a Literature Review**

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 and manually annotate and extract information using a web-based review tool.

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

REQUIRED SKILLS
No experience necessary.

PREFERRED SKILLS
- Strong English reading and writing skills
- Detail oriented and responsible
- Willing to learn specialized software and data (such as EPPI-Reviewer, BiBTeX and Zotero)
- Interested in learning how to run an existing Python program
- Some familiarity with basics of web publishing

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

<div align="center">**Project 4 (Prof. Jodi Schneider, iSchool)**</div>

**Using graph databases and network analysis to detect financial bias in research**

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

PREFERRED SKILLS
- Detail oriented and responsible
- Experience with graph databases such as NEO4J
- 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
Mentor: Jodi Schneider, Assistant Professor, School of Information Sciences
https://ischool.illinois.edu/people/jodi-schneider
https://infoqualitylab.org

<div align="center">**Project 5 (Prof. Jodi Schneider, iSchool)**</div>

**Making sense of evidence in meta-analytic studies on psychotherapy outcomes**

This project will be relevant to students interested in mental health practice, research, or policy and in mental health knowledge representation and communication.

Research over the last 40 years indicates that psychotherapy is effective for psychological problems. For many people these days, going to see a psychotherapist is not all that unusual. Most clients, though, don't necessarily ask about the evidence of its effectiveness and rely on their own experience with a therapist to decide if the therapy is helpful or not. Most psychotherapists depend on their clinical training and experience with each client to figure out what is helpful and what is not. Researchers, on the other hand, are all about trying to discern

evidence in a systematic and generalizable way, and there is no shortage of studies—thousands, in fact! Meta-analysis is a relatively new tool that helps researchers pool the evidence from many individual studies in order to make stronger claims about degree of effectiveness. As powerful as they are, though, what do meta-analyses of psychotherapy studies actually tell us? What counts as evidence in these studies and what kinds of conclusions can clients, practitioners, policy makers, and even other researchers draw from them?

REQUIRED SKILLS:
None

PREFERRED SKILLS:
- Strong English reading skills
- Ability to think logically and step-wise through a problem
- Interest in health outcomes research
- Interest in psychotherapy, psychology, or mental health research
- Familiarity with meta-analyses, systematic reviews, and/or evidence-based practice

MENTORS
Prof. Jodi Schneider, School of Information Sciences (contact person for informatics students)
Prof. Lydia Khuri, Educational Psychology

MORE INFO
https://ischool.illinois.edu/people/jodi-schneider
https://education.illinois.edu/faculty/m-lydia-khuri
https://infoqualitylab.org


**Project 6 (Qiong Nie, PhD candidate in Human Factors and Aging lab, Applied Health Sciences)**
**Designing a Mobile Application for Older Adults to Improve Hypertension Medication Adherence**

Hypertension is prevalent in older people and blood pressure (BP) medications are effective in controlling BP and reducing adverse health outcomes. In this project, we developed the theory-based Multifaceted Prospective Memory Intervention (MPMI), which improved older adults' adherence to BP medications. We further supported this intervention by developing the Medication Education, Decision Support, Reminding, and Monitoring System (MEDSReM) mobile application, which changes medication taking from an effortful process dependent on executive functions and cognitive processes that decline with age, to cue-driven associative processes that are mostly preserved with age. MEDSReM-2 will capitalize on technological advances to integrate additional functionalities into MEDSReM, including electronic BP monitoring and providing feedback about the relationship between medication adherence and BP. Central to this process is developing engaging and easy to understand visualizations so that older adults can track their progress.

We are looking for a researcher to contribute to the behavioral feedback and visualization design process. You will work closely with Dr. Daniel Morrow, interact with graduate students 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).

## Projects 7-10 (Prof. 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 in the response to COVID-19. We are also working with the WHO regional office to develop information systems for health throughout the Americas, and to apply advanced analytical techniques to large health datasets. We have opportunities for undergraduate research assistants to work on several projects in 2021:

### Project 7: Response to the COVID-19 Infodemic
The COVID-19 pandemic has been accompanied by a parallel infodemic – an overwhelming flood of information related to the pandemic that can be good, confusing, misleading, or just wrong. We are looking for a student who will work with the WHO and CDC infodemic team to support their efforts. One possible area of collaboration is to develop and implement tools for measuring the infodemic.

### Project 8: Social Media Analytics
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 9: Information Systems for Health
The IS4H project is a major focus of the Pan American Health Organization, the regional office of the WHO for the Americas.

*"The project seeks to support countries in defining strategies, policies, standards for interoperable and interconnected systems, and best practices in health data management for improving decision-making and well-being."*

We are looking for a student who will contribute to the development of "knowledge capsules" and COVID-19 factsheets. These are one-page summaries of important IS4H topics, such as big data, artificial intelligence, or social media, intended to give public health officials a brief summary of the topic and links to further information.

## Project 10: Modeling of COVID-19 and other respiratory diseases

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 11 (Prof. Shackelford, ANTH)

**Designing virtual reality environments for students with disabilities**
Our team has previously designed, created and delivered a college-level course in archaeology using immersive, interactive virtual reality (VR). This archaeology course – like VR games and applications more broadly – cannot accommodate students with mobility challenges or other physical disabilities. The ubiquity of video games and the increasing affordability of consumer VR systems have brought issues of accessibility and the need to support gamers and other VR users with physical disabilities to the forefront.
Prior to the adoption of immersive VR, technology has required minimal physical interaction to access (i.e. typing, clicking, swiping). Immersive VR increases the physical component of the interaction so that it may require full-body involvement. These increased interactions create immediate barriers for many users with alternative accessibility needs. While VR has the potential to make learning more accessible, the hardware and software technologies need to be accessible and compatible with assistive technologies used by those with disabilities.
The current NSF-supported project builds on prior work to evaluate and design a framework of hardware, software and instructional solutions to enable wheelchair users to fully participate in educational VR activities. We invite applications from students who are interested in virtual reality and its use in education as well as students interested in usability testing and human-

centered design. We have multiple projects ongoing and so there is flexibility as to the type of project based on the student's interests. But, we are specifically interested in students for the following tasks:

- Designing and distributing student surveys to wheelchair-using students about their gaming and VR experience
- Interviewing participants about VR experiences
- Designing and testing user interface technologies
- Applying universal design for learning rubric to current VR experiences
- Programming VR functionalities or environments using Unreal Engine

Contact information:
Laura Shackelford, Dept. of Anthropology
llshacke@illinois.edu
217-721-8805

## Project 12 (Prof. Diesner, iSchool)
## Using Story-telling and Crowdsourcing to Identify People's Moral Perception of Information from Images versus Texts

**Project Description:**
Using a combination of story-telling and crowdsourcing, we aim to analyze individuals' moral perception of information based on multi-modal data. With this work, we aim to develop a computational model to (1) analyze moral values in user-generated texts, (2) measure if morality in images correlates with morality in texts, and (3) examine the similarities and differences between the perceived information.
For this purpose, we selected used gold standard, annotated data and prepared an online questionnaire to collect user-generated text data and self-reported data.

We are currently collecting the user-generated stories using the Amazon Mechanical Turk platform. The student will work with a graduate student to analyze the data. Through this collaboration, the student will have the opportunity to enhance their research skills and gain hands-on experience in research. The work involves studying and annotating data, conducting literature reviews, running statistical analysis, implementing algorithms and models for analyzing texts, and interpreting quantitative and qualitative results.

**Preferred Requirements:**
Prior exposure to at least one programming language is required (Python preferably). In addition, basic knowledge about statistical analysis, natural language processing, 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 13 (Prof. Diesner, iSchool)**
**Words and Networks – validating, advancing and open-sourcing**
**methods for the relational analysis of text data**

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, you 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, you will be able to contribute to the open source community and help researchers from all over the world to use this product.

Your 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 you 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 preferred. But most importantly, if you are detail-oriented, interested in learning new things, and passionate about helping others, please join us!

**Project 14 (Prof. Diesner, iSchool)**
**Humanitarian Assistance and Disaster Relief (HADR) Data Annotation & Analysis**

**Project Description:** In times of crisis, substantial amounts of information about the crisis are shared and discussed on a number of platforms, including social media, news, official situational reports, and expert blogs. This project focuses on reliably extracting relevant information from sizable corpora of text from multiple data sources. We define relevant information as both about (1) the needs of the affected and at-risk populations as well as (2) the entities involved in the response operations of a particular crisis. Student researchers working with us will have an opportunity to work hands-on with data from multiple disaster events in the past, namely earthquakes and hurricanes. 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 an annotation task to identify and label the named entities mentioned in official situational reports. The completion of the annotation process enables both the student researcher and the collaborators to understand the types of entities involved (e.g. authority figures, organizations, cities, countries) in response operations and the extent to which these entities work together. The student will then be involved in using the annotated data in machine learning pipelines and for algorithmic evaluation tasks.

Preferred (but not required) qualifications: experience with content analysis, text analysis with Python (e.g. nltk, spaCy).


## Project 15 (Prof. Diesner, iSchool)
## Using Natural Language Processing to Extract Long-term Impact of Investments in International Biodiversity and Bio-conservation.

**Project Description:**
The goal of this research project is to leverage and advance methods from natural language processing (mainly information extraction) and human-centered data science to identify the impact (i.e., objectives or outcomes) of grant-funded investments in global conservation based on project documents, research papers, and surveys. With this ongoing work, we aim to answer the following questions:

1. What are the long-term impacts of grant funding on peoples' environments?
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 lengthy documents?
3. How does impact based on bibliometric versus alternative assessment methods compare?

The findings from this project will help us to better understand the objectives and outcomes of projects, and how data science and text mining can be used to aid this process. This project will also contribute to automated text summarization through designing a novel framework to summarize lengthy bio-conservation documents.

The 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 and text summarization, and interpreting quantitative and qualitative results.

**Preferred Requirements:**
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. 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 16 (Prof. Smith VetMed)
### Dynamics of phthalates over time

Phthalates are chemicals ubiquitous in modern life, found in plastics, personal care products, and medical equipment. They are known to be endocrine-disrupting, blocking proper hormone function. While they have an extremely short half-life in the body, near constant exposure results in continuous presence, which can be measured in the urine. Many studies have looked at the impacts of these chemicals, but most have considered only snapshot values. The Midlife Women's Health Study measured a variety of phthalates in the urine of more than 700 women once a year for up to 7 years, as they progressed into menopause. We are curious about the stability and variability of these concentrations over time. The student would work with these data to describe the dynamics in different groups of women and to visualize these dynamics in novel and compelling ways. If time allows, analysis of how the dynamics affect menopausal symptoms would be considered.

Skills needed: statistics, R

### Project 17 (John Toenjes, Dance)
### Authoring tools for creating 360-degree video and VR scenarios

DANC 456 Choreographic Lab course in Spring Semester 2021 is going to be approaching social media issues through interactive dance. Using a game structure, we will experiment with how to engage audiences in self-reflection through participating in a game within a performance of contemporary dance. We will most likely be collaborating with a class at Florida International University in Miami, and possibly other remote institutions. This will require employing technologies such as 360-degree video, VR, projections, peer-to-peer Internet collaborations, and a cellphone app. I need someone to help with learning authoring tools for creating 360-degree video and VR scenarios, and in creating app programs on cellphones using a custom utility for audience interaction within the dance. Research into how to link personal VR with public view of the VR-users headset will be needed, as well as into how to live stream this view to remote locations in Miami and Irvine, California. Research might also be required of how to gather tracking information from the headset to be used as data to determine game flow and progression and user choices. At the end of the semester I would like the student to help write an academic paper about this semester's experiments for submission to various publications and conferences. Skills and aptitudes desired (but it's not necessary to know all of) are ability to think creatively, an interest in games, ability to learn high-level authoring tools (such as Max/MSP, Unity, Isadora, Touch Designer), ability to research how to apply technologies to artistic problems, computer programming expertise (any language), editing video in Adobe Premiere and still images in Adobe Photoshop, good writing skills.
Contact Professor John Toenjes, jtoenjes@illinois.edu, 217-244-9190

## Project 18 (Prof. Magee, iSchool)
## Young Researchers: Introducing Teens to Early Research Experiences in Public Libraries

This project is focused on connecting teens to the original research process through topics that matter to them. We're working with five partner libraries (each in a different state) to train local teens on how to design and conduct research, including study design, data collection and analysis, and presentation/publication. At this stage of the project, we're providing support to each library in their work with teens and guiding the development and implementation of research studies in each location. This spring we'll be conducting interviews, surveys, and focus groups with library staff and teens at each location. You'll contribute to the project with literature searching, transcription, developing community profiles (using Census and other data), contributing to designing interviews/surveys/focus groups, observing data collection, and observing and contributing to qualitative data analysis, with the opportunity to contribute to writing and presentations about that analysis.

Prerequisites: None. Preferred interest in youth, technology practices, public libraries, and/or qualitative research.
Open to all students (undergrad and grad).

## Project 19 (Prof Magee, iSchool)
## Examining Teen and Adult Conceptions of Success Across Youth Serving Organizations

This project is focused on understanding how adults working in youth serving organizations (like libraries, 4-H, after-school clubs, etc.) describe their successful work, and comparing and contrasting their perceptions to how the teens they serve describe their experiences. We are also interested in whether and how nondominant teens' (Black, Indigenous, People of Color and LGBTQIA+ youth) experiences are distinct from other youth, and how we can build connections across community organizations to improve positive outcomes for nondominant youth. We will be working with six organizations in one mid-sized Illinois community to conduct interviews and focus groups around these topics, and to plan and implement an event that connects the organizations to build better youth services. You'll provide support to the project through literature searching, developing a community profile (with Census and other data), contributing to designing interviews and focus groups, and observing and contributing to data analysis, with the opportunity to contribute to writing and presentations about that analysis.

Prerequisites: None. Preferred interest or experience with youth serving organizations like libraries, 4-H, after-school clubs, etc.; preferred interest in justice and equity.
Open to all students (undergrad and grad).

## Project 20.  (Prof Chin, iSchool)
## Health misinformation modeling

While health misinformation is prevalent, studies showed that people are vulnerable to identify the low-quality health information and eliminate the consequences of misbeliefs. The project is to understand how adults across the lifespan process and interact with health misinformation through both social media analysis and behavioral experiments.

We are looking for students who are interested in data analytics, text analysis and behavioral studies. Students will be working with a multidisciplinary team (including graduate students and faculty in other disciplines) in this project. Students will be assisting in different tasks, including data preprocessing, annotation, study materials creation, data collection and analysis.

· Necessary prerequisite skills: none.

· Preferred skills: experience in Python and R for data analysis, natural language processing, machine learning.

· Open to all students. · Please find more information about our team. https://jessiechinlab.ischool.illinois.edu

## Project 21.  (Prof Chin, iSchool)
### Building digital health solutions for patients with progressive conditions

In this research project, we are developing digital health solutions for patients with progressive conditions (including patients with spinal cord injuries, strokes, cancer, etc.) to support them to manage and detect their healthcare needs. We aim at delivering accessible and active care to diverse patients, especially for patients in the underserved community or rural area.

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 to help 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 22 (Dr. Huang, iSchool)
### Web Development: ClassTranscribe

**ClassTranscribe** (classtranscribe.illinois.edu), an open source video learning platform is looking for students to help with both Frontend and server-side web development. The ideal contributor would be available for 2 at least semesters, because it is expected that it will take one semester to get up to speed with the technologies and coding standards. **Frontend** development is based on **react.js** and Redux. It provides a sophisticated accessible user interface. The FrontEnd project is available on Github at https://github.com/classtranscribe/FrontEnd/)

**Server-side development** is based on asynchronous C# together with Entity Framework to provide a json-based WebAPI and background taskengine. It uses several technologies including Docker, RabbitMQ, Postgres, and Elastic Search. WebAPI developers will typically use Microsoft's free Code IDE and free Visual Code Community edition editors to create, test & debug code. The project also has a tiny amount of Python code for some simple integration tasks (e.g. downloading videos) but C# development is the main skills needed to succeed in this project. The source code of the API & taskengine is here (https://github.com/classtranscribe/WebAPI ).

**Can I get started without C#/react skills?** If you do not yet have the skills, e.g. no C# experience (backend) or no react.js experience (frontend) but are eager to join the project then one method of demonstrating competency is to work with the development team and address a simple bug or feature already listed on GitHub. We have plenty to work on!

**Copyright & Attribution.** Please note, all code contributed to the project will be copyrighted by the University of Illinois.  Code development will be attributed from git commit history; so we can write letters of recommendation that highlight your contributions and impact to this project.

Please contact **Lawrence Angrave** (CS)**, Yun Huang** (yunhuang@illinois.edu at iSchool)**, Jon Tubbs** (Business), or **Rob Kooper** (NCSA) for more information.

## Projects 23 - 29   (Prof. Downie, iSchool, HathiTrust Research Center)

The HathiTrust Research Center (HTRC) enables computational analysis of the HathiTrust corpus. It is a collaborative research center launched jointly by Indiana University and the University of Illinois, along with HathiTrust, to help meet the technical challenges researchers face when dealing with massive amounts of digital text. It develops cutting-edge software tools and cyberinfrastructure to enable advanced computational access to the growing digital record of human knowledge.
Leveraging data storage and computational infrastructure at Indiana University and the University of Illinois at Urbana-Champaign, the HTRC builds tools and services for scholars to perform research using data from the HathiTrust Digital Library. The Center is breaking new ground in the areas of text mining and non-consumptive research, allowing scholars to fully utilize HathiTrust content while preventing intellectual property misuse within the confines of current U.S. copyright law.

HTRC Analytics: https://analytics.hathitrust.org/
HathiTrust: https://hathitrust.org/

## Project 23: HTRC Code Repository and Python Library Audit

Synopsis of project: Student will audit HTRC's public code base in GitHub and make recommendations for standard repository structure, documentation and licensing based on current and emerging best practices in open source software development, including analogous organizations or projects. Additionally, student will choose one piece of user-facing HTRC

Python software to use as an exemplar library for future HTRC software releases, building out robust documentation and deploying it as a Python package (https://pypi.org/help/#basics) on pypi.org for easy install.

Project outcomes
- A document detailing a recommended approach to repository directory structures, licensing and documentation, grounded in relevant other use cases and current best practices reflected in literature.
- (Optional) A model repository following the above guidelines.
- An exemplar, documented Python library released in pypi.org as a Python package.

Skills/expertise required
- Experience coding with Python
- Familiarity with GitHub
- Interest in open source software development

## Project 24: Workset event timeline algorithm

In this project the goal is to create an algorithm and tool that allows researchers to identify, retrieve, and visualize dated events occurring in a given HathiTrust workset. Visualizing the timeline of events portrayed in a collection in chronological order is a great way to follow along and understand the key milestones or cause-and-effect relationships in the narrative. For an example of what a timeline of events looks like, see http://www.simile-widgets.org/timeline. The HTRC can wrap/package the algorithm and include it in the set of available (pre-canned) algorithms provided by the Analytics Gateway for the benefit of the entire HT/HTRC community. Researchers can use the generated timeline visualizations to better understand the geographic flow of events portrayed in the text.

Project outcomes
An algorithm/library that takes as input a workset (set of HathiTrust volume IDs) and produces a web-friendly interactive timeline visualization of the dated events mentioned in the text.

Skills/expertise required
- programming experience
- NLP experience (identifying/extracting date entities from text)
- web development experience

## Project 25: Visualizing Geographic Entities in the HathiTrust Digital Library

This project proposes the creation of an algorithm/library/tool that can be used to identify, extract, and visualize on a map geographic entities mentioned in a collection of texts. Oftentimes texts make references to locations such as cities, countries, or points of interest to localize actions or events that are described in the text. Having the ability to identify those mentions, disambiguate them when necessary (Paris the capital of France, or Paris a city in Illinois), and plot them on a map, provides a visual aid that helps in better understanding the geographic

component of the ideas presented in the text. This project will develop an algorithm or tool to identify and visualize mentioned locations in texts from the HathiTrust Digital Library.

Project outcomes
An algorithm/codebase/library that, when given as input a HathiTrust workset (list of volume IDs), produces as output a web-friendly visualization of the location references from the text with their surrounding context.

Skills/expertise required
- programming experience
- NLP experience (identifying/extracting location entities from text)
- web development experience

## Project 26: Creating HathiTrust Use Cases and How-To Guides

The HathiTrust Research Center provides its users with a substantial collection of research examples, use cases, and how-to guides (see here, for example) that has grown organically over its nearly 10 years of existence.  This project will first review, update (as needed), and organize these existing user support materials using tagging, faceting, or other information-organizational methods.  The project also includes the opportunity to create new user-facing materials in any of the genres currently used: how-to guides, use-case descriptions, or lesson modules (e.g., as Python or R notebooks).
HTRC's wide variety of analytical tools have a long history and a set of user-facing examples and how-to guides that have grown organically over many years. Getting them organized will help make these tools, and all of HTRC's products and services, more accessible to users with a broad range of experience and knowledge.

Project outcomes
- High-level overview of current user-facing how-to guides, use cases, and tutorials
- Creation and implementation of a tagging or faceting scheme to organize these materials
- Creation of new user-facing how-to and tutorial content

Skills/expertise required
- Interest in and some knowledge of Digital Humanities research methods
- Willingness to learn HTRC tools and methods
- Interest in creating user documentation

## Project 27: Sentiment Analysis Research and Tool Development

Sentiment analysis is a technique that detects polarity within text (for example, positive or negative opinions) at a particular level (for example, the sentiment of an entire document, a chapter, a paragraph, or even a sentence). This project proposes the creation of an algorithm/library/tool that can be used to determine the emotional tone (sentiment) behind words and produce a visualization of the evolution of the degree of positive or negative sentiments expressed in a text as the story unfolds. There are several approaches to performing sentiment

analysis on text - from machine learning to rule-based techniques - and many resources on the internet describe ways in which this can be achieved. A "lit review" of the most popular approaches would be beneficial before deciding on which approach to use to implement this algorithm for HTRC.

Project outcomes
- An executive summary of the insights gained from the literature review of sentiment analysis approaches and techniques.
- An algorithm/codebase/library that, when given as input a workset (list of volume IDs), produces as output a web-friendly visualization of the evolution of the level of positive or negative sentiments expressed in the text of each of the volumes in the workset.

Skills/expertise required
- programming experience
- experience or strong interest in natural language processing (NLP)
- web development

### Project 28: Transforming HTRC Extracted Features Data to Topic Models

The HathiTrust Research Center (HTRC) Extracted Features (EF) Dataset (https://analytics.hathitrust.org/datasets) is a dataset derived from the 17+ million volumes in the HathiTrust Digital Library. This dataset is fully non-consumptive (https://www.hathitrust.org/htrc_ncup), which allows for it to be distributed and without restrictions on its use under the doctrine of Fair Use. However, as a heavily structured, bag-of-words (https://en.wikipedia.org/wiki/Bag-of-words_model) dataset, it can pose data processing challenges for some users who wish to use it for certain text and data mining purposes that the data itself supports, but the structure complicates. This project will seek to generate an easy-to-use and well-documented Python script that a user can run to process the EF Dataset into a form that is suitable for topic modeling.

Project outcomes
- A documented, reasonably easy-to-use script to transform HTRC Extracted Features data into a format suitable for topic modeling
- An example use case of deploying the script to generate topic models on a set of HTRC EF data

Skills/expertise required
- Experience coding with Python
- Knowledge of or strong interest in topic modeling

**Project 29: Analysis of Page-level Language Data in the HathiTrust Extracted Features Dataset**

The HTRC Extracted Features dataset contains algorithmically-identified language data for each page of the 17+ million volumes in the dataset (that's a total of almost 6 billion pages). This project aims to analyze, summarize, and visualize these page-level language tags in ways that provide useful high-level views of the distribution of languages across the dataset. The analysis would also look at how well the volume-level, metadata-supplied language tag aligns with the predominant page-level identified language and report any misalignment; additionally, we aim to identify and report potential dataset anomalies where, for example, it is found that a volume features a large number of languages, or that consecutive pages in a volume have varying language tags, etc.

Project outcomes
The expected deliverable of this project is a report and accompanying dataset including the following:
• graphs and statistical summaries providing useful overviews of the dataset
• reports of misalignments between volume-level metadata language tag and predominant algorithmically-identified volume language
• reports of potential anomalies found in the dataset

Skills/expertise required
• programming experience (processing JSON files and extracting data)
• data analysis / statistics (compute summaries, identify outliers, …)
• data visualization (create graphs, charts, and reports)
• interest in multilingual analysis

**Project 30 (Prof Underwood, iSchool)**
**How much are language models confused by errors in novels?**

You have a great idea for a research project on the history of children's literature. You gather a few thousand works of fiction, and start to feed them into a neural language model, then notice that the optically-transcribed texts look like this:

"Dorothy thankcd the little mice warinly for saving hcr frºcnds ft - >= *"

Transcription errors are common in digital libraries. How much do these errors distort researchers' conclusions? Are some methods more robust than others? We're trying to answer these questions. We have room for one or two people to join the lab and help us run experiments.

Experience with Python needed; a little exposure to machine learning would be helpful, but is not required.

**Project 31 (Prof. Joe Yun, ACCY)**
**Building Out an Open-Source Community for the Social Media Macroscope**

The Social Media Macroscope (www.socialmediamacroscope.com) is an open-source project created here at the U of I in which we give access to researchers, students, and general users without coding and computational backgrounds the ability to analyze social media data within a fully open-source environment. Nothing within the environment is black-box and we have gained users from over one hundred research institutions around the world. The problem is that we are getting a bit too big for our team to handle on our own, thus we need to create a broader community around the project to provide mutual user support, more code-builders into our GitHub, and general awareness about the project to more people that could benefit from it.

We are looking for students that can analyze the user community that we already have, think about how we can build out awareness of the project, and help select platforms that we can use to provide forums, support, and overall community growth. No coding or data science background is needed for this project, but students have the ability to learn a great deal about open-source software in general as well as data science applied to social media data.

**Project 32 (Profs. Unnati Narang and Mehmet Eren Ahsen, Gies College of Business)**
**The Role of Collaborative Data Science for Solving Healthcare Challenges**

We are evaluating the effect of a crowdsourcing platform on solving healthcare challenges. The platform allows participants to register their own pre-existing teams, or form new teams, submit their code and algorithms, engage in online forum discussions, and ultimately be recognized for winning entries on publicly visible leaderboards and/or through monetary incentives. The proposed research will examine the impact of this collaborative (open) data science platform on innovation, particularly in developing successful innovative solutions to healthcare challenges. The research project will involve collecting data from this platform, summarizing and analyzing it. Based on the results from the proposed study, we will also design future experiments.

Skills preferred but not required are some statistical background, knowledge of excel, R or python.

**Contact details:**
Eren Mehmet Ahsen: ahsen@illinois.edu
Unnati Narang: unnati@illinois.edu

**Project 33 (Monica Porteanu, PhD Candidate in Informatics. Advisor: Professor Kevin Hamilton, New Media & Dean, Fine & Applied Arts)**
**Voice Based Decision Making**

**Context:** Using voice to engage with our devices has become increasingly popular. At the same time, voice is a key identifier and biometric, while emergent aggregated data sets and algorithmic emotional and affective interpretations promote voice-based decision-making. However, our knowledge of our voice journey, access to its print, and its interpretation across devices, platforms, clouds, and algorithms is limited.

**Hypothesis**: AI-mediated voice-based decision-making and the assumptions the technology makes about a user are subjective as our voices operating in today's complex technology ecosystem have new effects that we are not fully aware of.

**Research topic:** Triggering user awareness of their voice journey through technology and its affective interpretations as it relates to voice-based decision-making that impacts user's contextual identity and privacy

**Looking for** curious and self-starter critical thinker(s), eager to learn and interested in multi-media research and app experimentation. **Desired skills**: prior exposure to at least one programming language (e.g., Python, R), API platforms (e.g., Azure, Google), voice app interpretation and development, affective computing, rapid prototyping, and basic knowledge of machine learning.