Digging into Human Rights Violations: Anaphora Resolution and the Emergent Witness (DHRV)
Team Overview

Ben Miller, Georgia State University, US team lead, PI. (miller@gsu.edu)
Lu Xiao, Western University, Canadian team lead. (lxiao24@uwo.ca)

Overview of goals and objectives of your DiD project
US Team, Information Extraction and Data VIsualization: Contribute to analysis of human rights violations reports with a tool for extracting, relating, and enriching sparse and imprecise information about locations, entities, and times; develop or adapt methods in information extraction, NLP, and data visualization to help narrate the stories of victims and perpetrators of human rights violations; expand work on cross-document coreference; develop corpus level visualizations that highlight time at location and entity movement across locations; develop big data research capacity and network of practitioners and scholars working in data and text mining within a human rights domain.

Canadian Team, User-centered approach in Developing Visual Analytic Tools that Combine Artificial Intelligence and Visualization Techniques: Apply a user-centered development approach in the processes of design software programs that support qualitative analysis of large-scale human rights violation data. Our approach manifests through several aspects: facilitating the requirements analysis of the software programs; design, rapid prototyping, and formative evaluation of user interfaces of the US team's Martus system based software program; and developing a tool for qualitative analysis of collections of audio/video recordings and transcripts through participatory design.

Challenges and lessons learned from international collaboration across disciplines and domains
International: Many scheduling issues such as semesters, budget years, grant periods, student hiring periods, and holidays occur at different times; geographic distance between the two teams makes in-person meeting rare and expensive.
Disciplines: expectations for methods vary. Both the requirements of human rights researchers and our own experiences of working in this project show that researchers of different disciplines have their own respective practices of conducting research activities. The lack of common ground in research practices makes it challenging to communicate and collaborate across disciplines in this international team. Moreover, the team did not have collaboration history prior to the grant, adding more challenges to the development of team identity.

Digital humanities, social sciences and computational based research methods in the context of big data projects
US Team: Many of our methods are core to DH work, including text analytics, information extraction, machine learning based clustering, text mining, topic modeling, geospatial data visualization, and data visualization. Our project followed a pipeline from primary source documentation through to researcher-based analysis of extracted material.

Canadian Team: In our user-centered design approach, we've conducted tool survey to understand the existing tools that have been in use by human rights researchers; interviewed and administered online questionnaires to human rights researchers about the data they deal with and their data analysis practices, and their expectations and concerns regarding software programs that facilitate their data analysis processes; and applied participatory design methodology in developing a visual analytic tool that combines text mining and visualization techniques.

Indicators of success
1. The prototype of the algorithms for information extraction, NLP, and data vis are developed and produce meaningful content, valid relations, and useful visual analytics tools

2. Human rights researchers express interest in using the software program prototype for their analysis

3. Research lines begun for DHRV lead to new work

4. Partner institutions have increased capacity for and interest in the study of big data phenomena

5. Project's techniques and perspective on HR violations reports are adopted by practitioners

Measuring impact

1. *The project's work has been well-received by the various stakeholders' disciplines.* For example, data visualization techniques developed for DHRV have been presented at IEEE VISSOFT, a special session of the 29th IEEE Int. Conference on Software Maintenance, HASTAC, DH2013, and IEEE Big Data 2013, and a panel at the WSSF that includes our partner organization, archivist, and CS and HCI researchers.

2. As with traditional scholarship, *the quality of publications and presentations* produced by a project

3. *The project's success of transitioning participating graduate students into permanent academic positions.* As DH is a newly manifested field, one measure of lasting impact is the career opportunities of graduate students from various fields who commit significant time to DH projects. In two years, an MA, PhD, and Postdoctoral researcher have secured their next positions in part because of their participation on DHRV.

4. The *sustained partnership with human rights research organizations* that extends this research work after the completion of the project

5. Offer *understanding of the human rights researchers' data analysis practices*

6. Provide guidelines on design of visual analytic tools for human rights researchers

Knowledge dissemination mechanism and tools

Conference presentations, scholarly publications, availability of the user interface design and the developed software program to potential users, meetings with human rights organizations, public scholarship

Importance of working with libraries, archives and data repositories

US Team: Data meant the transcribed or OCR'd free text narratives of witnesses to human rights violations. These records have frequently been collected for four purposes: truth and reconciliation proceedings, prosecution, awareness documentation, or as a byproduct of bureaucracy. In each of these cases, various agencies act as custodians for the collections. Without a strong relationship with a custodian of HR records, no NDA, however binding, would be sufficient. Librarians have traditionally been the interface between academic researchers and data vendors, and the local curators of data at the academic institution. Although DHRV did not go that route, it may have been wise to. Alternatively, an archivist consulted with the project throughout, and her advice helped guide our data management plan, our NDAs, and our perspective on data handling.

Canadian Team: Our researchers have dealt with the audio/video interview recordings that are publicly available. However, we acknowledge that the archives have leveled access to the interviews depending on the level of the interviewees' consent to data release.

Capacity building and training (students and highly qualified personnel)

Institutional: Push IT and university infrastructure to accommodate the hardware and configurations necessary for humanities big data; work with various departments to explore options for crediting non-traditional academic labor.

Students and Personnel: Involve Masters and Doctoral students of different disciplines in the process; conduct regular meetings with the students; meet with advisors to help align project contributions to project needs; facilitate participation by students in skills workshops (i.e., fund enrollment in a workshop on corpus linguistics); and raise awareness of directors of graduate study in relevant departments of new sources of interesting projects and visibility (i.e., an NSF funded information extraction project emerging from an English department that presents at CS conferences).