Interim Report

Grant Number: HD-51433-11

Title of Project: Rose (Research-oriented Social Environment): Bibliographical
                                     Knowledge as Social Knowledge

Name of Project Directors: Alan Liu (Director); Rama Hoetzlein & Rita Raley (Co-directors)

Name of Grantee Institution: University of California, Santa Barbara

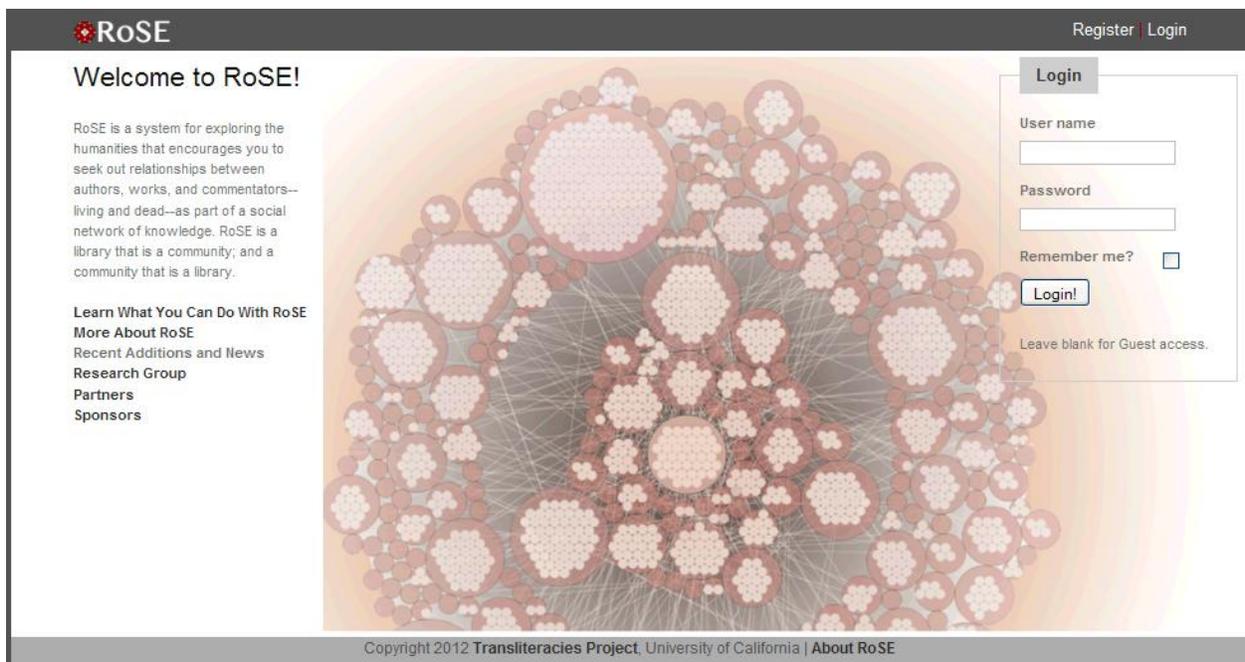Date Report Submitted: March 18, 2012

Figure 1: Front Page of RoSE

Narrative Description

Summary:

The RoSE project team has met weekly since the start of the NEH grant in September 2011 for general developer meetings (involving the project director and two co-directors, four graduate-student programmer developers, and four graduate-student humanities developers). The programmer developers have also met for coding sessions each week. RoSE has made timely progress on its stated grant tasks as follows:

1. Data-mine and import metadata from SNAC to add to the base information in RoSE

At the beginning of the NEH grant period, RoSE had processed and imported metadata about people and documents from YAGO and Project Gutenberg to provide users with a foundation of loosely linked humanities bibliographical networks on which to build thicker descriptions and relationships. YAGO provided data for 7,557 people and 11,395 documents, with keywords offering opportunities for clustering studies and also "influence" identifiers to establish relations between people. Project Gutenberg provided data for 11,964 people and 34,077 documents, but with no keywords or information about relationships.

After a conference call and correspondence with Daniel Pitti (director of the Social Networks & Archival Contexts project) and other SNAC developers, RoSE worked with 125,000 individual XML files provided by SNAC to harvest information useful for RoSE. This was essentially a twofold process of parsing SNAC entries into selected component name, date, document, and keyword metadata and also removing parsing errors to produce a clean, relatively compact set of XML files compatible with RoSE metadata format. We concentrated on the approximately 74% of the above-mentioned 125,000 SNAC files that are entries for people, and concentrated in addition on the 3% of these SNAC files representing highly connected people. The latest results of our selective harvesting from SNAC is data for 3,412 people and 28,036 documents. Currently, we are continuing work with processing the SNAC data to improve its usefulness for RoSE. In particular, we are identifying principles of selection (since SNAC has so much data) and methods for filtering, parsing, and cleaning up the information to better match RoSE's purposes (especially to select and process data that allows users to see connectivity between people and documents).

During development, we are moving data imported from SNAC into and out of the online production version of RoSE as we continue to experiment (sometimes SNAC data is not included in the online production version of RoSE). We expect to have selected SNAC data in the system in time for our use-scenarios studies beginning at the end of April 2012.

2. Improve visualizations in RoSE; add more visualization types (e.g., social network graphs, geographic maps)

Visualization (together with user-interface improvement) has been one of the most intensive areas of development in RoSE during the NEH grant, though the direction of visualization development has somewhat altered due to innovating important new functions in RoSE. Besides continuing to work on interactive social-network style, radial, and packed-radial style visualizations of relations between people and documents, we decided that a crucial need was for end-users to collect and arrange their findings. We thus designed "history," "collection," and "storyboard" functions whose database

functions are represented and manipulated in part through visualizations. "History" tracks the items a user locates or moves between during a session in RoSE; "collection" allows a user to build a shareable collection of resources; and "storyboard" allows a user to drag items from a history or collection and arrange them narratively on a whiteboard to facilitate the presentation of findings or the brainstorming of interpretive arguments. This development work is proceeding rapidly, with an expected finish date of approximately the end of April 2012 (in time for the project's use scenario studies).
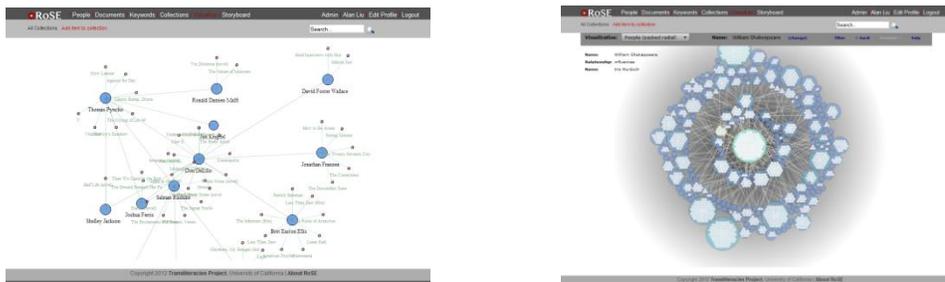
Figures 2 & 3: Examples of current interactive visualizations in RoSE

3. Use Scenario Studies/Evaluation.

RoSE's planned use-scenario studies are scheduled for the latter portion of Spring quarter at UCSB (beginning at the end of April 2012). Progress to date has occurred in three areas. First, we have concentrated on improving the user-interface and visualization functions of RoSE to facilitate using the system. This work is proceeding along a tightly-planned timeline of programmer tasks. Secondly, we have submitted a Human Subjects protocol (now under review) for the use-scenario studies that involve collecting user response and feedback to RoSE. Third, we are developing materials and plans to support the three particular use-scenario studies: (a) study of the way RoSE functions for a student course assignment (in Prof. Rita Raley's English 146, "Distracted Reading"); (b) study of use of the system to support a student research showcase event (a poster show); and (c) study of the use of the system to support an individual graduate student's dissertation research. Materials being prepared include how-to recipes and examples of using RoSE; plans for the specific student assignment in Prof. Raley's course; and information pages on the RoSE site.

In addition to the above tasks, we have migrated the RoSE system to a production server in a virtualized environment in the UCSB English Department's infrastructure (as well as created multiple offline development servers running locally on the developers' personal computers).

Problems and Challenges

A key challenge in the project's progress so far has been recruiting and retaining graduate-student programmer-developers from the UCSB Media Arts & Technology program due to unpredictable changes in these students' other funding sources. One key visualization programmer was awarded a graduate fellowship for spring quarter 2012 that proscribes her from other work (we have just recruited a replacement). Another programmer has had to drop out of the project due to the workload demands of his other funding sources. So far, we have been successful in maintaining the necessary level of developer staffing for the project. We also note that the very fact that we have this problem reflects on the success of RoSE as an interdisciplinary collaboration between the humanities and other programs

whose students are much in demand for project work. Our collaboration with MAT graduate students has enabled a wider range of intellectual and technical experiments than would otherwise be possible.