Kevin Crowston; Syracuse University
How can document-genre metadata improve information-access for large digital collections?

Participant Individuals:
CoPrincipal Investigator(s): Barbara Kwasnik
Graduate student(s): Joe Rubleske; You-Lee Chun

Partner Organizations:

Activities and findings:

Research and Education Activities:

Research to date on Phase 1:

The project began in January 2005, so here we report on the first 7 months of activity (Phase 1 of the project). The goal of Phase 1 is to determine the range of document genres and their uses found in our target population. Some tasks and situations seem likely to benefit more than others do from identification of document genre.

The original target audience identified was K-12 teachers. However, we have expanded our study to include additional groups in order to enable reproduction and comparison of our results. Our study to date has included journalists as well as teachers and we are in the process of identifying a third target population.

Our activities in the first six months include:

- Development and pilot testing of a data elicitation protocol. We continue to develop and revise our techniques based on our experiences. In particular we continue to look for ways of overcoming the difficulty participants experience in articulating precise genre descriptions and genre features.

- Think-aloud sessions with 18 individuals (to date: 15 journalists and 3 teachers) and preliminary analyses of these sessions. These approximately hour-long sessions involve participants who verbalize their thoughts while conducting web searches. Each think-aloud session is followed by a debriefing discussion in which participants give explanatory details about the web pages they visited.
  - Of the 15 journalists we have interviewed:
    - 11 work as reporters for the Syracuse Post-Standard newspaper;
    - 2 work as editors for the Syracuse Post-Standard; and
    - 3 work as editors for WTVH-TV in Syracuse, NY.
  - The 3 teachers we interviewed teach at Lincoln Middle School in...
Syracuse, NY.

- Recruitment of additional subjects. We have agreement from Manlius-Pebble Hill, a private K-12 school in Manlius, NY, and Edward Smith Elementary School, a public K-6 school in Syracuse, NY, and additional think-aloud sessions will be conducted in August.

- Development of an internal project web site serving as a repository of the sites visited by participants. Each page in this repository is organized by participant and consists of a large screenshot of the page, its URL, the task conducted by the participant while using the page, and the participant’s comments about the page. It is hoped that this repository will be a fruitful resource for identifying emerging, unanticipated themes. Eventually, the database of pages and genre meta-data will be made available to other researchers (e.g., researchers interested in automatic genre identification could use our corpus for training or validation).

- Development of a comprehensive literature review whose aim is to expand our evolving understanding of the complex nature of genre. In particular, from our work thus far, we have a deeper appreciation of the intricate role of people’s context in their recognition and use of genre information.

**Findings:**

Analyses of these data thus far have yielded:

- a preliminary set of codes for content analysis.

- some new considerations that might impact our future experiments in the study of how genre might be useful in searching the web. These include:
  - the role of the order in which pages are viewed
  - the role of the structure of web pages as “links in a chain” (suggesting genre sets)
  - first-time vs. habituated genre identification

Preliminary observations:

- In general, participants have expressed difficulty articulating a description of the “type” of certain pages, especially:
  - those comprising chiefly content;
  - those that do not include a large proportion of links;
  - those that are not homepages.

We continue to adjust our data-collection techniques to account for this difficulty and to develop techniques for eliciting difficult-to-articulate and implicit cues to genre identification.

- Despite the lack of vocabulary, it is becoming quite evident to us that participants think, consciously or not, in genre-oriented terms. For example, journalists, in particular, are quite skilled at rapidly assessing the utility of a page, a task that appears to be based heavily on their assessment of its 'look and feel.' We are learning that their verbal reports contain more clues than we originally thought, especially for what they tell us the page means to them as members of a certain community of practice.
There is clearly a need among journalists for search tools that distinguish in terms of form and function. For example, every journalist is concerned with 'legitimate' or 'authoritative' sources. This confirms our initial impetus for conducting this research, and it is our intent in this project to understand what these subtle cues/clues are.

As an interesting side note, we’ve observed that many practitioners, such as journalists, fall back on pre-digital vocabulary in describing digital documents – that is they invoke genre information pertaining to print documents in describing web pages.

Training and Development:

The project has provided two PhD students with hands on training in literature review, research design, data elicitation and qualitative data analysis.

Journal Publications:


Book(s) of other one-time publication(s):

Other Specific Products:

Internet Dissemination:

http://genres.syr.edu/

The site has information about the project being carried out with NSF support. Eventually it will have publications and data.

Contributions:

Contributions within Discipline:

- By organizing a mini-track on genre at the Hawaii International Conference on System Science (HICSS) and subsequent special issue of the journal Information, Technology and People, the PIs have increased the visibility of genre as an important research topic and perspective on digital information. As well, organizing the mini-track has helped us improve our understanding of the richness of the phenomenon.

- Development and refinement of a data elicitation technique for genre should help other researchers interested in studying genre.
Contributions to Other Disciplines:

- Genre has been studied extensively in several disciplines—rhetoric, management, linguistics, education—as well as our home discipline of information science. The literature review is one step in increasing the integration and mutual informing of genre concepts from a variety of reference disciplines.

Contributions to Education and Human Resources:

- The project has supported the education of two PhD students, who we expect to go on to take faculty positions.

Special Requirements for Annual Project Report:

Unobligated funds: less than 20 percent of current funds
Categories for which nothing is reported:
Participants: Partner organizations
Participants: Other Collaborators
Outreach Activities
Products: Book or other one-time publication
Products: Other Specific Product
Contributions to Resources for Science and Technology
Contributions Beyond Science and Engineering
Special Reporting Requirements
Animal, Human Subjects, Biohazards

We welcome comments on this system