

## Genres of Digital Documents

Barbara H. Kwaśnik and Kevin Crowston  
*Syracuse University School of Information Studies*  
*bkwasknik@syr.edu, crowston@syr.edu*

There has recently been an upsurge of interest in document genres—the fusion of form and function that makes a document instantly recognizable to its community of users. The utility of genre identification is now accepted as a premise, but many interesting and challenging questions remain. The papers assembled here create a truly interdisciplinary body of work, drawing on literary and communication theories, library and information science, natural language processing, writing, and machine learning, among others. One of the most challenging aspects of research in genre is to create a rigorous yet flexible way of representing document genres that will reflect their form and properties, as well as their communicative role as embedded in a particular community of discourse.

Crowston and Kwaśnik approach this dilemma by suggesting a faceted approach to representation. This approach allows for an expressive and hospitable scheme that sees genre as fundamentally multidimensional. Schmid-Isler and Oehninger draw on the Media Reference Model (MRM) as a rich representational framework. The MRM takes into account several views: the infrastructure view, the service view, the interaction view of processes, and the community view. The authors recognize the importance of representing genre in its context, that is, in the human endeavors in which a given genre is invoked. Their paper offers a first step in organizing genre features, which may in turn, lead to a typology of genres that will take into account the most important situational aspects. Lee and Myaeng also address the issue of representation, but from the perspective of automatic identification. They build on previous work and explore the link between the genre of a document and its subject. One of the interesting aspects of their paper is their identification of a set of linguistic features, such as parts of speech and punctuation, that might help automatically identify and cluster genres and subjects.

Another approach is to study genres from within, as it were, by conducting intensive investigations of particular genres. If we understand a given document genre deeply—how the genre functions and is recognized—then perhaps we can extend such understanding to other genres. At the very least we gain experience with the

research methods and can build up a toolkit of approaches for studying all genres. Herring, Bonus, Scheidt and Wright make a contribution to methodology by conducting a content analysis of Weblogs (blogs)—frequently modified web pages. Their analysis yields a useful vocabulary of criteria (structural components, links, page characteristics, temporal aspects, and author characteristics) by which other researchers might proceed with genre description. As an interesting result, the authors found that their data did not support the commonly held assumptions about blogs, resulting in a, perhaps, more valid description of the context and communicative purposes of blogs than could an anecdotal approach. Marlow reports on his study of aspects of technical trouble tickets—communications about technical problems, such as network outages. He employs an analysis that invokes lexical, mechanical, and grammatical clues, as well as situational factors that influence a person's communicative choices. This study addresses not only the specific structure and characteristics of trouble tickets, but also the *communicative style* of this genre—yet another conceptual framework for studying documents in use, rather than documents as artefacts. Ihlström and Åkesson address a troublesome aspect of genres: their ability to undergo evolutionary change, which is particularly true of document genres that have migrated from print to electronic formats. The authors studied one such genre, investigating the stages of evolution of Swedish online newspapers. Interestingly, they find that not all online newspapers have evolved to the same level. This suggests that in studying migrated genres it is necessary to describe not only the changes from one format to another, but also the placement of the more recent format on a continuum of change.

Identifying, describing, and utilizing document genres is a complex and challenging task. These papers contribute both general models and approaches and specific techniques for further research. We wish to thank the authors of the papers selected for this session, and also the authors of those many excellent contributions that we were not able to include. As well, we thank the reviewers for their extensive, constructive, yet gentle comments.