

▼ Introduction to Genres in Communication and Digital Documents Minitrack

Kevin Crowston
Syracuse University
School of Information Studies
crowston@syr.edu

Carina Ihlström Eriksson
Halmstad University, Sweden
carina.ihlstrom_eriksson
@ide.hh.se

Tero Päiväranta
University of Agder, Norway
Dept. of Information Systems
tero.paivarinta@uia.no

Rhetoricians since Aristotle have attempted to classify communications into categories or “genres” with similar form, topic or purpose. Given a socially recognized need to communicate, individuals will typically express similar social motives, themes and topics in a communication with similar physical and linguistic characteristics, that is, they will communicate in a recognized genre.

The Web provides a particularly interesting setting in which to study the use and development of digital genres and genre repertoires since it provides a way for diverse groups to interact and to evolve new genres of communication. In some cases, a genre may act as a type of boundary object, providing a common point of contact between different groups but this mixing may instead lead to genre confusion, meaning that there is a practical need to understand the way genres enable communication. As well, a document’s genre is a piece of meta-data that may be quite useful for information access systems, but since genre is socially constructed, it is challenging to use in an automated way.

The three papers for this year’s minitrack all focus on automatic genre recognition and classification issues. In the first paper, “Using Visual Features for Fine-Grained Genre Classification of Web Pages”, Levering, Cutler and Yu explore the idea of improving classification of fine-grained genres of e-commerce web sites by utilizing visual features of HTML documents. Use of visual features in addition to the traditional approach of extracting textual features of documents is suggested for improving fine-grained classification of genres.

Kim and Ross investigate the correlation between visual, stylistic and topical content features and genre classes in the paper “Examining Variations of Prominent Features in Genre Classification”. Their results provide evidence that genre classification should be regarded as a multi-dimensional task

consisting of several classification tasks, which involve a varying distribution of feature type strengths as distinguishing factors of classification. In practice, they propose using an array of varying strengths of feature types to express classes of document genres for their automatic classification.

The final paper, “An Examination of Genre Attributes for Web Page Classification” by Lei Dong, Carolyn Watters, Jack Duffy and Michael Shepherd, describes a set of experiments to examine the effect of various attributes on the automatic identification of the genre of web pages. Four different genres are used in the data set, namely, FAQ, News, E-Shopping and Personal Home Pages. The effects of the number of features used to represent the web pages (5, 20, or 100) as well as the types of attributes (content, form, and functionality), singly and in various combinations are examined. The results indicate that fewer features produce better precision but more features produce better recall, and that attributes in combinations will always perform better than single attributes.

The articles in this year’s minitrack focus on one specific stream of genre-oriented research. We believe that this development illustrates the obvious usefulness of genre for information access. Past mini-tracks have included as well papers that use the genre lens to analyze and theorize on the co-evolving phenomena of organizational communication and media in general. The absence of this perspective this year suggests perhaps that it has become more challenging to come up with new and appealing theoretical insight into the field of communication genres in general. Time will tell whether this is a temporary phenomenon in the long history of the mini-track, or whether also the other genre-oriented research streams will continue to contribute in the years to come.

We thank all authors who submitted papers and the reviewers for their contributions to the minitrack.