

Utilizing Information “Shape” as an Interface Metaphor Based on Genre

Elaine G. Toms
School of Library and Information Studies
Dalhousie University
etoms@is.dal.ca

D. Grant Campbell
Faculty of Information and Media Studies
University of Western Ontario
gcampbel@julian.uwo.ca

ABSTRACT

Documents have conventions which have evolved within discourse communities and which facilitate document use. These conventions are represented in a document by visual cues that define a shape and serve as an interface metaphor in a user's interaction with a digital document. In this paper we report on the results of two studies, one of which examined the impact of medium on genre recognition and one which examined the impact of discourse community. Results indicate that document structure plays a role in document recognition and that the composition of a document's structure delineates a 'shape' with metaphoric qualities.

INTRODUCTION

Communities of academic, business, or professional groups in particular fields have developed and continue to develop conventions of presenting information. Some conventions such as those represented by the lowly office memo are shared by multiple groups while others like chemical structures and the university calendar are highly specialized with unique purposes. These conventions are represented in documents by features such as columns, headers, lists and so on which give the document a distinctive visual appearance. Documents fall into categories, or genre, based on common characteristics, purposes and conventions. In the digital document universe, genre can be used as a powerful enabling concept that will enhance interaction with and navigation through these documents.

We suggest that a digital document typically has a configuration of visual features that characterize it as a particular genre, and that the relationship between visual appearance and semantic content is relevant as an instrument both of document presentation and of document retrieval. From the user's point of view, the visual appearance of different genre can function as an interface metaphor, in which the user of a computer system, when provided with a visual metaphor such as a desktop or house, loads a set of expectations into working memory (Carroll and Thomas 1982). The user, therefore, by seeing and recognizing these visual cues, identifies the document as an instance of a particular genre class, and loads a set of expectations at the very outset of the reading experience. Furthermore, when formatting features for instances of a document type are reasonably consistent, they provide the basis of a new mode of automatic document recognition, filtering and retrieval.

To capture a user's interpretation of a genre for computer processing, either of the text or of retrieval programs that filter by genre, certain questions need to be answered. Above all, we need to determine whether or not a digital genre possesses a set of parsimonious attributes that appear in visual form. If so, are these visual cues sufficiently stable, either within or across specific discourse communities, to enable them to function as interface metaphors? This paper, then, presents the partial results to date of a project which is studying document types in terms of their visual characteristics. We compare the results of two studies, which analyzed user interpretation of the visual features of a select group of document types used in the academic environment.

BACKGROUND

Genre

Critics of literature and art traditionally classify works according to genre. Since the 1950s, genre theorists have recognized that "the study of genres is based on analogies of form" (Frye 95): works of a particular genre have common structural characteristics, even when their topics and overt meanings differ. These literary genres are not static categories; rather, they are systems which evolve with each new instance of a class. Concepts of genre provide the critic with "an interpretive and evaluative frame for a work of art" (Swales 37). For authors and readers, they provide "a communication system, for the use of writers in writing, and readers and critics in reading and interpreting" (Fowler 1982, 286). Business has adapted this concept of genre as a system for communication. In an organizational context, a genre is "a typified communicative action invoked in response to a recurrent situation" (Yates & Orlikowski 1992, 301). Genre, therefore, arises from regularities in business communication: its characteristics are not predefined, but emerge with each additional instance of a "typical" business communication.

Genre and Digital Documents

With the rise of digital documents, genre has acquired a new relevance. With many digital documents, the semantic content is more dynamic than that of printed documents; it is subject to change in unpredictable ways, and the new text supplants and effaces the old. Genre has emerged as an alternative source of social and discursive stability: "Genre(s), akin to social structure(s), provide rules and resources to be deployed in the construction of a text by actors within a specific set of social relations (a community) and sharing similar social practices" (Yates & Sumner 1997, 5). Furthermore, digital environments, particularly the World Wide Web, have created new genres and transformed old ones. Books, minutes and scripts have migrated to the Web, existing alongside new genres such as the hotlist, the home page, and Web Server Statistics (Crowston & Williams 1997). Numerous studies have charted the emergence of new genres such as the personal home page (de Saint-Georges 1998; Roberts 1998), online conversations (Erickson 1997), and the travel genre (Panko & Panko 1998). Others have studied the transformations of traditional genres in the online environment: the limerick (Erickson 1999), advertising (Fortanet, Palmer & Posteguillo 1998), the newspaper (Morin 1998; Watters & Shepherd 1997), dissertations (Fox, McMillan & Eaton 1999), and accounting source records (Tallberg 1997).

In addition to these studies, researchers are searching for theories that would explain how genres are transformed when transferred onto the Web. Studies of linking properties in digital documents suggest that certain kinds of hypertext linking effect a genre transformation, by enhancing navigation within a document or linking it to other information sources (Crowston & Williams 1999). Others describe the migration of a genre into the digital environment as a process of transformation, beginning with simple replication of the print source in digital form, succeeded by various transformations which enhance digital functionality until it emerges as a “cybergenre” in its own right (Shepherd & Watters 1998, 98). These various studies confirm that digital text, even at its most unstable and unpredictable, has underlying regularities of characteristics, structure and purpose, and that these regularities significantly enhance the reader’s ability to use text efficiently.

Interface Metaphor and Information Shape

Researchers have begun to study ways of exploiting these regularities of conventions to facilitate user-document interaction in digital environments. Two different but complementary approaches have been taken. Toms & Campbell (1998) have suggested that digital genre communicate to the user about the document and function as an interface metaphor while Vaughan and Dillon (1998, 561) have applied the principles of cognitive psychology, particularly schema theory and mental models.

Toms and Campbell (1998, 1999) suggest that genre, as manifested in the visual layout of a document, provides an organizing metaphor. In this case, the visual cues enable the user to load the appropriate framework of expectations into working memory facilitating document recognition. Recognition however comes when the user connects the visual shape of the document with the document’s purpose and function.

Vaughan and Dillon (1998), on the other hand, suggest that humans organize their knowledge about the world by forming mental representations in hierarchically organized structures, which are built from the bottom up by repeated exposure to regularly-occurring phenomena. This approach implies that a crucial relationship exists between physical space and semantic space. The appearance of a document – the configuration of its visual parts as they appear on a screen – is a distinct from, but related to, the document’s semantic meaning. The connection between form and meaning implied in the concept of shape enables users not just to locate desired information, but to “assimilate information into their own knowledge structures, an activity that extends processing beyond layout (the classic physical navigation concern) to meaning” (Dillon & Vaughan 1997).

In our research, then, we are assessing the visual aspects of digital genres, to measure their significance in terms of document type recognition. We believe that the visual cues that delineate shape heighten the metaphor qualities of document genre. If genre is to be an effective means of facilitating such recognition, we must understand how users recognize document types in terms of their visual characteristics.

STUDY #1

Overview

In this initial study, we tested proof of concept. A complete and full description of the methods used for this study can be found in Toms and Campbell (1999). Four questions guided the study:

1. Can users identify a document when only the document's visual structure is presented?
2. Can users more efficiently identify a document when only the visual structure is presented?
3. What is the most parsimonious characteristic that uniquely identifies a document type?
4. Are the characteristics that are present in paper documents transferrable to digital documents?

Methods

Participants

Fifteen participants were recruited, representative of the academic community and experienced in Web navigation and use.

Materials

Six types of documents familiar to the academic community were used: journal article, dictionary, course reading list, memo, meeting minutes and course calendar. Representative documents of each type were obtained in both print and digital form. Both print and digital documents were converted into two surrogates (See Figure 1):

Form: This version exhibited all of the physical characteristics of the original document, but the text was masked. Each alphabetical character was represented by an "x" (upper or lower depending on the original text) and each digit by a "9." Thus the visual shape remained intact, but all semantic meaning was removed from the document.

Content: In this case, a document was reduced to its semantic content. The document was represented as one single paragraph of information presented in regular text (i.e., 12 pt. Times Roman). None of its physical shape was retained.

Variables

Using these surrogates, three variables were measured:

- **document type** (article, memo, etc.)
- **genre element** (form or content)
- **medium** (paper or digital).

Procedure

Participants identified eight print and eight digital documents, four representing form and four representing content. Their answers were recorded, as was the time taken to identify the document. The numerical data was analyzed using descriptive statistics and repeated-measures analysis-of-variance. Participants were also asked to identify the distinguishing characteristics that identified the document type, and the most discriminating feature of the document. These responses were transcribed and categorized by the researchers, but were not quantitatively analyzed.

Results and Discussion

Overall, participants recognized 63% of the documents, and recognized significantly more by content than by form. On average, they took 20 seconds to identify the document, and there was no significant difference between the time taken to recognize by form and that taken to recognize by content. More surprising, we found that there was no interaction of genre, media and document type; no differences were perceived between print and digital documents.

Participants explained which features of the document led them to identify it as a particular type, and which feature was the most discriminating. In general, we found that when identifying documents with only content (content version) and no format, participants largely tended to read or scan in sequence until a key word that identified the document was found: "Abstract" in the case of the journal article, "course offerings" for the course calendar, and "To:" and "From:" in the memo. They also tended to look for repeated occurrences of words or numbers: personal pronouns in the memo, for instance, or course numbers in the course calendar. Also, they discerned visual cues, even in the unformatted blocks of text: parentheses in the reading list, or the red link in the electronic journal article. Sometimes, even the block itself was interpreted visually, as an e-mail message or an abstract.

When looking at the formatted documents with no content, the participants showed three primary tendencies:

1. They identified format features in terms of a purpose or semantic content which they associated with that feature. An italicized passage, for instance, was a title; numbers following a colon were page numbers.
2. They responded to combinations of features, such as the alternating bolded and non-bolded sections of the dictionary, or the contrast between headings and short paragraphs in the course calendar.
3. Participants frequently took a bird's eye view and responded to an overall visual impression: noting the unequal lengths of the paragraphs, or the "structure of chunks" in the dictionary.

In essence, structural features in the documents provided meaningful clues about the document type. But we were unable to isolate in participants' responses the most parsimonious set of attributes that make a document a course calendar and not a memo. Clearly, a document's visual structure had metaphoric qualities for participants; noteworthy was the way participants saw not just isolated attributes, but sets of features that appeared to have shape and profile.

From this study, we had proof of concept: users can identify documents by visual cues alone and have difficulties identifying documents when no structure is present. In addition, we discovered that documents with the same structure are equally recognizable in print or in digital form; participants did not miss the three-dimensional cues that are present in print documents. More interestingly, they discerned delineated patterns that can be perceived as shape.

STUDY #2

Overview

In the second study, we discarded medium as a contributing factor to document recognition, but added another element: discourse community. In the pilot study, the participant group represented one community: academia. To contrast its response, we added as a control group members of the general public. We also increased the number of participants to 72, a number more likely to lead to statistically significant results. We increased the number of documents to 24, although only six are reported in this comparative analysis, and added as a control the original document to contrast with the form and content versions used in the first study. Thus, in the second study we anticipated that:

- Participants would, to some extent, interpret the semantic cues in terms of visual cues and vice versa.
- Participants would respond to combinations of visual cues and repetitions that were seen from an overall perspective.
- Genre recognition would be most successful in cases where the visual cues were closely associated with one specific genre. We expected, for example, that the memo, with its "To/from/subject/date" headers, would be easy to identify.
- Participants from different discourse communities have deferring degrees of success in recognition documents outside their respective communities

A fuller description of the methods for this study can be found in Toms, Campbell and Blades (1999). We retained three of our research questions from the original study. The research questions guiding this study, therefore, were:

1. Can users identify a document type when only the form is presented?
2. Can users more efficiently identify a document type when only the form is presented?
3. What are the most parsimonious characteristics that uniquely identify a document type?
4. Is document type recognition affected by membership within a discourse community?

Methods

Participants

Seventy-two (44 females and 28 males) participants were recruited from the Halifax area. Ages ranged from teens to senior citizens, but half were under 30. All had been to university and about two-thirds had graduated. Eighty per cent used computers for more than 5 years. All had used the WWW, and more than half for more than one year. About 50% continued to use the WWW at least once per day.

Thirty-six participants currently worked in the academic environment as faculty, staff or students while the remaining 36 were members of the general public. There were no significant differences in the constitution of the two groups except in frequency of use of the WWW, age and education. As anticipated, the members of the general public had more education, tended to be older and tended to use the WWW less frequently. All participants were paid \$20.00 for participating in the study.

Materials

The types of documents in the test collection duplicated those used in the pilot study. All documents came from electronic sources, either created for Web distribution or created in electronic form for later paper distribution. Each document was subjected to a verification procedure, in which the structural elements of each document were compared against a set of ten examples. The documents finally chosen, therefore, were good representatives of that document type. Six of the twenty-four documents closely resembled those used for the earlier study: the non-annotated bibliography, course calendar, dictionary, journal article, memo and meeting minutes. As with the previous study, each document was converted into two surrogates, one based on form and the other on content. Unlike the pilot study, a control set that contained the original documents was also used.

Variables

Two of the variables were the same as the previous study. The third was new:

Type of document: Bibliography, Call for papers, etc.

Genre element: Form vs. Content vs Original

Membership in discourse community: Belonging or not belonging to academia.

Procedures

Each participant examined each document only once, and in only one version. Participants examined two sets of documents: eight documents from the control set, followed by sixteen documents from the two surrogates. The first set was deemed necessary to prime participants about the task. The documents were randomized within each set and each set was randomly assigned to each participant. The participants viewed the documents using Netscape on a Pentium computer with a 15-inch monitor, and were instructed to use only the back button and the scroll bar. One research assistant worked with the participants, while comments were recorded on an audio recorder. On viewing each document, participants were asked the following questions:

- What type of document do you think this is?
- What feature of the document was most helpful in making your decision?
- Were there any other features that assisted you in making your choice?
- Were there any features which puzzled you, or seemed inconsistent with your choice?

As before, the quantitative data was analyzed using descriptive statistics and repeated-measures analysis-of-variance. Two persons not associated with the study reviewed the names of the documents provided by participants and determined the accuracy of each ($\kappa = .82$). In addition, a preliminary analysis of the attributes was completed by the two researchers.

Results

Genre Recognition

Each participant examined only one version of the document, but examined all documents. On average participants correctly identified 62% of the documents in the form and content versions.

By Discourse Community

The members of the general public were able to identify as many documents as the academic community ($F(1,70)=.05, ns$) as illustrated in Table 1. In fact there were no statistically significant differences on any factor or interaction. Thus this variable is removed from analyses.

Group	Genre Element		
	Full	Form	Content
Academic	1.94	.81	1.6
Public	1.89	.75	1.8
Average	1.9	.78	1.7

Table 1. Average Number of Documents identified by Group and Genre Element

By Genre Element

Not surprisingly, there were differences among the document versions ($F(2, 140)=27.92, p<.0001$). On average, participants correctly identified the 97% of the full version of the documents, 39% of the form version and 85% the content version. Most of the documents in the control set were correctly identified as were those showing Content only. Less than half the documents in the Form version were identified by participants.

By Document Type

Participants reacted differently to the document types ($F(5,414)=7.78, p<.0001$) as illustrated in Figure 2. Additionally there is an interaction between document type and genre element ($F(10,414)=8.03, p<.0001$). Not surprisingly, most people recognized the document when its full version was visible, but not all. Documents for which no shape was evident did not fare as well. The dictionary in particular was recognized by only half when the structural cues were not present; most were confused by the content of the dictionary. When semantic content was absent from documents, about half could identify the journal article, dictionary, and the memo. Our raters viewed memo, letter and electronic mail message as being equivalent responses which accounts for the accuracy of memo in both the form and content versions.

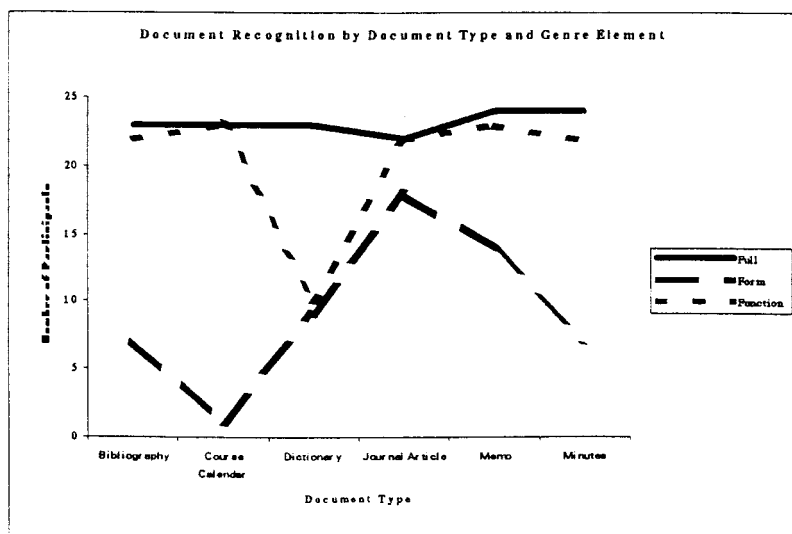


Figure 2. Number of Participants who Recognized Documents by Type and Genre Element

Discriminating Characteristics of Document Genres

Semantic Cues

When viewing the content-only documents, the participants behaved much like the participants in Study 1. In most cases, they read sequentially until they encountered key words which enabled them to identify the type. In particular, titled documents were identifiable by self-describing words, such as “bibliography,” “minutes,” and “to” and “from.” Participants also noticed repetitive patterns throughout the text as a whole. One participant noticed that the dictionary had the same kind of information repeated; another noticed the prevalence of course numbers in the course calendar, while still another participant found that even the block-like nature of the journal article was suggestive of a magazine layout.

Visual Cues

As expected, participants frequently responded to the overall look of the document, and made observations on that basis: one participant noted that the course calendar, as a whole, looked like “a reference list;” the memo was noted for being “concise” and “not very long,” while others noted that the non-annotated bibliography was “highly organized” and “formal.”

In addition, participants were sensitive to combinations and repetitions of visual features. The journal article attracted attention for the obvious relationship between the top line and the two lines beneath it; the sections at the top, which were centered, appeared to form a unit apart from the rest of the text, which was left-justified. Once again, participants noted the dictionary’s tendency to alternate between bolded headings and short passages of non-bolded text. They also noted, in both the minutes and the course calendar, the relationship between single lines beginning with a number and the short paragraphs that separated them.

When we studied the responses to various sections of the documents, however, our expectations were confounded. As we had expected, participants tended to congregate around specific visual features of the documents: the columns in the dictionary, the lists in the bibliography and the meeting minutes, the centered headings in the journal article and the numbered headings in both minutes and course calendar. However, as Table 2 indicates, the semantic interpretations of these features varied widely, even with features which are traditionally associated closely with one genre. The headers at the top of the memo, for instance, which contain the “to/from” information, were interpreted as descriptions, introductions, authors, locations, or publication dates. The lines below the title of the journal article, giving the authors’ names, were interpreted as subtitles, authors, or vital information.

Document	Feature	Interpretation
Non-annotated bibliography	Bulleted list	Point form; pieces of information; different works
	Repeated headings	Minor headings; small headings; headings of different topics; subtitles; chapter headings; subjects
	Information in parentheses at end of items in list	Dating and brackets; brackets and dates; index reference number; publication info; brackets at end of each sentence; all end in dates
Course calendar	First line	Title; name; display at top could be someone's name; name of the person
	Repeated numbered lines throughout document	Course titles; titles
	Text following the numbered lines	Text; description of titles; short little abstract; reference to that title; content; abstract; different bunches of paragraphs
	Numbers at beginning of numbered lines	Course numbers; dates
Dictionary	Columns	
	Bolded text	Headings; bolded entries; bold headings; titles
	Word in upper corner	Keyword; text notation at top; test identifier
Journal article	Top line	Title
	Two lines beneath top line	Authors; vital information; subtitle.
	Centered paragraph	Abstract
Memo	4 short lines below horizontal line	To/from/subject/date; introduction; description; key information; authors; locations; list of things;
	Words at top	Title; heading; name; letterhead
	Numbers in header	Date
	Paragraph form	
Minutes	List	Table of contents
	Numbered lines repeated throughout document	Subheadings; numbers and text; categories
	Title	Department title

Table 2. Interpretation of the Visual Features of Format-Only Documents

DISCUSSION

The results of these studies indicate that visual appearance plays an important role in document recognition. While participants easily identified documents in which form and content were both expressed, when either was removed from the formula, participants had trouble making an accurate identification; both exist as complementary factors contributing to the document as a whole and serving to identify its overall purpose and function.

In addition, whether the document was represented in a digital or print form made no difference to the recognition process. Participants were just as likely to recognize a document in print as in digital form. Similarly, discourse community had no affect. Much of the literature on digital genre assumes that discourse communities serve this function: by belonging to a particular group of users which have their own conventions, the user can easily recognize features that are defined by those conventions (e.g., Yates & Sumner 1997). But in our study, users outside the academic community identified documents with as much success as those inside it. The documents in this study were used in academic communities, although it can be argued that some like the memo are in general use. However, others like the course calendar and bibliography tend to be artifacts of academia. This factor needs further study with different communities to be conclusive.

In the process of attempting to identify a document, participants exhibited two strategies of great importance. First, they did not see the visual features of a document as isolated elements. Rather they grouped them together in terms of sequence, relationship and hierarchy. In so doing, they deconstructed the document into a series of linked attributes that are inter-related in a tree-like structure. Together these attributes define the notion of document shape, assist users in identifying document genre, and serve as an interface metaphor.

Second, when participants interpreted these attributes, they made no clean distinctions between visual appearance and semantic content. Even when studying the document shape, emptied of semantic content, as in Figure 3, participants perceived the features of that shape in terms of two facets which parallel the elements of a genre:

- A *form* or set of stylistic features which determine its physical appearance: text that is emphasized, positioned horizontally, of a particular length, or a distinctive size, and
- *Content* which the user readily associates with it: title, author, abstract, to/from header, etc.

Our data suggests that users need to match the stylistic features with the content in order to define the purpose of a document and ultimately to identify the document type. Without the visual shape, even the most self-evident words can be taken out of context and misinterpreted. And when users try to make the association on the visual cues alone, two factors complicate the process:

- A single stylistic feature can have multiple purposes in different documents: a bulleted list can be part of a bibliography, but it can also be an agenda or attendees list for meeting minutes.
- Content can be expressed in different styles; dictionaries don't *need* to be in columns; nor does an abstract need to be centered.

Reading the words is not sufficient; examining the form is not sufficient. The two are melded together at a moment in time when the user identifies the pattern.

Readers, then, recognize documents as hierarchical structures of linked attributes, which have visual and semantic facets which are interdependent. These findings are of great importance at this time when rich markup languages such as XML and XSL are about to come into common use on the World Wide Web. If, in fact, readers respond to tree structures, then SGML and XML, both of which conceive documents as hierarchical structures, open the door to a closer match between text design and user cognition. Document can be designed in ways that match the structure of their genre, and formatted with stylesheets that are tailored to their generic conventions.

However, our findings also remind us that formatting is more than mere “window-dressing.” While SGML and XML conventionally separate form and content for computer processing purposes, we believe that the separation is an artificial construct. Labelling and tagging the semantic parts of a document are crucial steps forward, but variations in the formatting of those parts affect document recognition. Participants for example linked a **big bold heading** with a journal title. Had the title been expressed as a little italic heading, we doubt that participants would have recognized it as a title. Each attribute of a document has a particular content and a *limited* range of stylistic features. The advent of XML as a web authoring tool has not banished formatting considerations; it has made a richer combination of content and form possible.

CONCLUSIONS AND FUTURE DIRECTIONS

Our data suggests that other factors are present, which need to be explored further. We have collected valuable evidence of how users classify documents, and how they perceive relationships between document types. Documents can be grouped into families based on similar purposes; a memo, a letter and an e-mail message, for instance, can be considered correspondence. But when do users make a clear distinction among those types? Documents can also be associated on the basis of shared visual cues: bibliographies and minutes and resumes all share lists; course reading lists and bibliographies share bibliographic citations. What is the range of values that those cues may have before they are misinterpreted? If we can understand how these relationships and associations work, these could be exploited to facilitate document recognition and ultimately used to create and retrieve digital documents.

In our studies to date, users have been exposed to documents in a very random way, much like the interaction one would expect in using the WWW. We are continuing our work by examining the opposite task much like one would expect to find in a work environment. How much does is document structure contribute to the recognition process when users are primed? When asked to find a journal articles or a dictionary, is the recognition process easier than it was for the participants of these initial studies?

All of this future work, however, rests upon what we have discovered in these initial studies. We have discovered that document “shape” plays a much more complex role than the literature on shape and digital genre had led us to believe. Shape is delineated by a set(s) of attributes that are rendered on the screen in a particular configuration. When that configuration matches a user’s prior exposure to a document type, recognition takes place in much the same way that an interface metaphor illuminates a process. Document recognition, therefore, depends largely

on document structure in its fullest sense: a hierarchical configuration of attributes which have both visual and semantic meaning, both of which are determined by the conventions of a particular genre. In this way, document genre recognition plays a crucial role in facilitating user-document interaction. And with the rise of more sophisticated and powerful markup languages, we have the power to exploit this process of recognition in new and exciting ways.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of OCLC Inc. which funded this project through its Research Grant Program. They also wish to thank the School of Library and Information Studies, Dalhousie University, for the use of facilities, and research assistants, Ruth Blades who administered the second study, Shelley Gullikson, Bill Jeffries, Alison Learning and Gillian Byrne.

REFERENCES

- Carroll, J.M. & Thomas, J.C. (1982). Metaphor and the cognitive representation of computing systems. *IEEE transaction so Systems, Man and Cybernetics* SMC-12: 107-16.
- Crowston, K. & Williams, M. (1999). The effects of linking on genres of Web documents. In *Proceedings of the thirty-second Hawaii International Conference on System Sciences: Digital Documents Track*. Los Alamitos: IEEE Computer Society.
- Crowston, K. & Williams, M. (1997). Reproduced and emergent genres of communication on the World Wide Web. In *Proceedings of the thirtieth Hawaii International Conference on System Sciences. Volume VI: Digital Documents Track*. Los Alamitos: IEEE Computer Society, xx-xx.
- de Saint-Georges, I. (1998). Click [here](#) if you want to know who I am. Deixis in personal homepages. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 68-77.
- Dillon, A. & Vaughan, M. (1997). It's the journey and the destination: shape and the emergent property of genre in evaluating digital documents. *New review of multimedia and hypermedia* 13, 91-106.
- Erickson, T. (1997). Social interaction on the Net: virtual community as participatory genre. In *Proceedings of the thirtieth Hawaii International Conference on System Sciences. Volume VI: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 13-21.
- Erickson, T. (1999). Rhyme and punishment: the creation and enforcement of conventions in an on-line participatory limerick genre. In *Proceedings of the thirty-second Hawaii International Conference on System Sciences. Volume ??: Digital Documents Track*. Los Alamitos: IEEE Computer Society.

- Fortanet, I., Palmer, J.C. & Posteguillo, S. (1998). Netvertising: content-based subgeneric variations in a digital genre. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 87-96.
- Fowler, A. (1982). *Kinds of literature: an introduction to the theory of genres and modes*. Cambridge: Harvard University Press.
- Fox, E., McMillan, G. & Eaton, J.L. (1999). The evolving genre of electronic theses and dissertations. In *Proceedings of the thirty-second Hawaii International Conference on System Sciences: Digital Documents Track*. Los Alamitos: IEEE Computer Society.
- Frye, N. (1957). *Anatomy of criticism: four essays*. Princeton: Princeton University Press.
- Morin, J. (1998). HyperNews, a hypermedia electronic newspaper environment based on agents. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 58-67.
- Panko, R. & Panko, D. (1998). Where do you want to fly today? A user interface travel genre based on flight simulators. The evolution of cybergenres. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 110-118.
- Roberts, G.F. (1998). The home page as genre: a narrative approach. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 78-86.
- Shepherd, M. & Watters, C. (1998). The evolution of cybergenres. In *Proceedings of the thirty-first Hawaii International Conference on System Sciences. Volume II: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 97-109.
- Swales, J.M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tallberg, A. (1997). Towards paperless accounting source records: a transaction cost approach. In *Proceedings of the thirtieth Hawaii International Conference on System Sciences. Volume VI: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 60-67.
- Toms, E.G., Campbell, D.G. & Blades, R. (1999). Does genre define the shape of information? The role of form and function in user interaction with digital documents. *Proceedings of the 62nd annual meeting of the American Society for Information Science*. ASIS, forthcoming.
- Toms, E. & Campbell, D.G. (1999). Genre as interface metaphor: exploiting form and function in digital documents. In *Proceedings of the thirty-second Hawaii International Conference on System Sciences: Digital Documents Track*. Los Alamitos: IEEE Computer Society.
- Toms, E.G. & Campbell, D.G. (1998). Proposal: Genre as interface metaphor: exploiting form and function in digital environments. OCLC Library and Information Science Research Grant Program.

Vaughan, M.W. & Dillon, A. (1998). The role of genre in shaping our understanding of digital documents. *Information access in the global economy: proceedings of the 61st annual meeting of the American Society for Information Science*. ASIS, 559-566.

Watters, C.R. & Shepherd, M.A. (1997). The digital broadsheet: an evolving genre. In *Proceedings of the thirtieth Hawaii International Conference on System Sciences. Volume VI: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 22-29.

Yates, J. & Orlikowski, W. (1992). Genres of organizational communication: a structurational approach to studying communication and media. *Academy of Management Review*, 17, 299-326.

Yates, S.J. & Sumner, T.R. (1997). Digital genres and the new burden of fixity. In *Proceedings of the thirtieth Hawaii International Conference on System Sciences. Volume VI: Digital Documents Track*. Los Alamitos: IEEE Computer Society, 3-12..

Figure 1: Three Different Versions of a Sample Memo

Figure 3: Single Feature Used in Different Document Types

NON-ANNOTATED BIBLIOGRAPHY

XXXXXXXX XXXXX XXXXXXXXXXXXXXXX
 XXXX XXXXX XXXXX: 9/9/99
 XXXXXXXX XXXXXXX XXXX: 99/99/99 99:99:99
 xx X. X. XXXXX, XXXXXXX XXXXXXXXXXXX: XXXX://xx9.XXXXXX.XXX/-XXXXXXXX/XXX.XXXX
 XXXXXXX XXXX XXXXXXX, XXXXXXX, X X 99999-9999

LIST:

Form Features:
Sequence of single
or double lines

**Assumed content
or purpose:**
List of books
List of attendees

MEETING MINUTES

- 9. XXXXXX XX XXXXXXXX XXXXXXXX
- 9. XXXXXX XXXX XXXXXXXXXXXX XX XXXXXXXX XXXXXXXX
- 9. XXXXXXXXXXXX XX XXXXXXXX XXXXXXXX
- 9. XXXXXXXXXXXX XXXX XXXXXXXXXXXX XX XXXXXXXX XXXXXXXX
- 9. XXXXXX XX XXXXXXXXXXXXXXXX XX XXXXXXXX XXXXXXXX

9. XXXXXX XX XXXXXXXX XXXXXXXX

- X. X. Xxxx, Xxxx-Xxxxxxxx Xxxxxx (Xxxxxxxx, Xxx Xxxx, 9999).
- X. X. Xxxxx xxx X. Xxxxxxx, Xxxxxxxx Xxxxx Xxxxx Xxxxxxxx (Xxxxxxxx, Xxx Xxxx, 9999).
- X. X. Xxxxxxxx xxx X. X. Xxxxxxx (Xxx.), Xxx Xxxxx Xxxxxxxx (Xxxxxxxx, Xxx Xxxx, 9999).
- X. X. Xxxxx (Xx.), Xxxx Xxxxx Xxx Xxxxxx (Xxxxxxxx-Xxxxxx, Xxxxxx, 9999).
- X. X. Xxxxx (Xx.), Xxxxxxxx Xxxxx xx Xxx Xxxxx (XXXX, Xxxxxxxx, Xx, 9999).
- X. X. Xxxxx (Xx.), Xxxxxx Xxxx Xxxxxxxx (Xxxxxx Xxxxx, Xxx

```

XXXXXXXX XX XXXXXXXX XXXXXXXXXXX XXXXXXXXXXXX
XXXXXXXX XXX XXXXXXXXXXXXX XXX
XXXX XXXXXXXXXXXXX XXXXXXXX
XXXXXXXX XX XXXXXXXX. XXXX 99, 9999

```

~~Xxxxx XXXXXXXX xx XXXXXXXX xx XXXX XXXXXXXXXXXX
XXXXXXXXXX XXXXX, XXXXXXXX, X.X.~~

Xxxx 99, 9999

Xxxxxxxx: 99

Xxxx	XXXXXXXXXX
Xxxxx, Xxxxxx	- XXXXXXX XXXXXXXX xx XXXXXXXXX
Xxxxx, Xxxx	- Xxxxx XXXXXXXX Xxxx XXXXXXXXX
XXXXXXXXX, XXXXXXX	- XXXXX XXXXXXXX Xxxx XXXXXXXXX
Xxxx, XXXXX	- XXXXX XXXXXXXX xx Xx Xxxx xx XXXXXXX
Xxxxxxx, XXXXXXXXX	- XX Xxxxx XXX
XXXXXXXXX, Xxx	- XXXXXXXX-XXXXXX XXXXXXXXXX
XXXXXXXXXX XXXXXXX	- XXXXXXXXXX xx XXXXXX, XXXXXXX
Xxxxx, XXXXX	- XXXXXXXXXX XXXXXXXXXX, XXXXXXX (XX)
XXXXXXXXXXXXX, XXXXXXX	- XXXXXX XXXXXXXXXX
Xxx, XXXXXXX	- XXXXXXXX XXXXXXXXXX
XXXXXX, Xxxx	- XXXXXX XXXXX XXXXXXXXXX
XXXXXXXXX, Xxx	- Xx XXXXXXXX XXXXXXXXXX
Xxxxx, XXXXXXX	- XXXXXX XXXXXXXXXX
XXXXXX, Xxx	- XXXXXXX XXXXXXXXXX xx XXXXXXXXXX,
XXXXXXXXX.	