

Reproduced and emergent genres of communication on the World-Wide Web

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

Marie Williams
Web Architects
Syracuse, NY
mfw@vcomm.net

Abstract

The World-Wide Web is growing quickly and being applied to many new types of communications. As a basis for studying organizational communications, Yates and Orlikowski [1, 2] proposed using genres. They defined genres as, “typified communicative actions characterized by similar substance and form and taken in response to recurrent situations” [1, p. 299]. They further suggested that communications in a new media will show both reproduction or adaptation of existing communicative genres as well as the emergence of new genres. We studied this phenomena on the World-Wide Web by examining randomly selected Web pages (100 in one sample and 1000 in a second) and categorizing the type of genre represented. Perhaps most interestingly, we saw examples of genres being adapted to take advantage of the linking and interactivity of the new medium, such as solicitations for help and genealogies. We suggest that Web site designers consider the genres that are appropriate for their situation and attempt to reuse familiar genres.

1. Introduction

The World-Wide Web (or the Web) is an Internet client-server communication system for retrieving and displaying multi-media hypertext documents [3]. Documents are identified by their address, called a Uniform Resource Locator or URL. The Web's main advantage over earlier Internet systems is its merger of retrieval and display tools, its capacity for handling formatted text, embedded graphics and other media and point-and-click links to other documents (hence the name). As well, many browsers are capable of seamlessly retrieving information using older protocols (e.g., FTP, Gopher, Usenet News) and automatically launching other applications to display diverse Internet data types (e.g., sound, animation).

Over the past several years, the Web has clearly become the most popular Internet application. For example, Bray [4] reported that 223,851 servers at 89,271 sites provided 11,366,121 unique URLs in November 1995; Woodruff et al. [5] collected 2.6 million Web documents. These numbers continue to

grow—current size estimates are in excess of 50 million pages—as individuals, businesses and other organizations rush to establish Web presences.

What is less clear is how the Web can or will be used by these diverse groups. Many organizations see the Web primarily as a cheap means of publishing information and are simply moving existing documents to the new medium (so-called “brochure-ware”). However, a few are experimenting with its capabilities to communicate and interact in novel ways, creating new “genres” of communication [6-8]. For example, Andersen Consulting is experimenting with a Web-based agent, Bargain Finder, to comparison shop for music CDs. Simulations, access to customer information databases and interactions with experts are all possible in this new medium.

Rather than working on the technology, we were interested in studying the Web as a social phenomenon, focusing in particular on the variety of communicative practices it supports. The purpose of our study was to begin to describe the range of genres of communication in use on the Web. Communicative genre is defined generally as accepted types of communication sharing common form, content or purpose, such as an inquiry, letter, memo or meeting. Note that genre is not simply the medium of communication—a memo genre may be realized on paper or in an electronic mail message (two different media), while the electronic mail medium may be used to deliver memos and inquiries (two different genres). However, medium does influence which genres are accepted. We were particularly interested in studying how the adoption of a new communication medium, the Web, might be leading to adaptation of existing genres and the emergence of new ones.

2. Theoretical background

Rhetoricians since Aristotle have attempted to classify communications into categories or “genres” with similar form, topic or purpose. Numerous definitions of genre have been debated in that community [e.g., 6, 7, 8]. Other groups have also struggled with the notion of document types; for example, information scientists have included rules for document types in SGML [9].

More recently, Yates and Orlikowski [1, 2] proposed using genre as a basis for studying communications in organizations. They defined genres as, “a distinctive type of communicative action, characterized by a socially recognized communicative purpose and common aspects of form” [2, p. 543]. In other words, given a socially recognized need to communicate [i.e., a purpose, 10], individuals will typically express similar social motives, themes and topics in a communication with similar physical and linguistic characteristics (i.e., form), that is, they will communicate in a recognized genre. Some genres are defined primarily in terms of purpose, such as a proposal or inquiry; others in terms of the physical form, such as a booklet or brochure. However, most genres imply a combination of purpose and form, such as a newsletter, which communicates “the news of the day”, including multiple short articles and is distributed periodically to subscribers or members of an organization.

This paper, for example, is an example of the social science paper genre, commonly used when communicating scientific results in a community of social scientists to advance the state of knowledge in a field, as well as the authors’ careers. It has a form familiar to a social scientist: a title, authors and affiliations, sections for introduction, theory, method, data and discussion, citations, a bibliography, etc. Other common genres include letters and memos, project team meetings and TV sitcoms, all immediately recognizable by their typical purpose and characteristic form.

The notion of genre can be extended in several directions. First, as the examples illustrate, genres form a hierarchy. The social science paper is a special case of a more general research paper genre, which in turn is a type of paper. Other types of research papers include computer science implementation papers, and biology papers. These genres share some similarities, such as a title and bibliography, but differ in other particulars, such as the expected section headings, types of arguments, etc. Similarly, TV sitcoms are a special case of TV shows in general, project team meetings of meetings, and so on. Rather than argue about the proper level of analysis for a genre, we believe it is most useful to follow Yates and Orlikowski and consider genres at any of these different levels.

Second, multiple genres may be linked or embedded to form a more complex pattern of communication. As Orlikowski and Yates pointed out, some communications use multiple genres simultaneously, such as a proposal embedded in a memo. As well, multiple communications may be performed in a recognizable pattern, what Bazerman [11] called a genre system. Examples include the sequence of examination and cross examination in a trial, or the cycle of article submission to a journal or conference, reviews, final acceptance or rejection letters and publication. Features of a genre may enable their use in a genre system: for example, page numbers in a

technical paper make it possible to cite concepts or quotations from the paper, thus binding the paper into the literature.

Finally, Orlikowski and Yates [2, p. 546–547] introduced the notion of a genre repertoire, that is, the set of genres in use within a community. They noted that different communities use different genres in their communication, and use common genres with different frequencies. These differences provide one source of insight into the communicative (and other) practices of the community. For example, a community of social scientists and computer scientists can be distinguished by the frequency of use of different paper genres, as well as the paucity of computer programs and program documentation created in the former, reflecting different modes of research.

Genres are useful because they make communications more easily recognizable and understandable by recipients. Because we drew on the social science paper genre, for example, another social scientist can more quickly determine the purpose and content of our communication and begin to evaluate its contribution. On the other hand, a genre may be unfamiliar or hard to understand for someone outside of the community. In fact, recognition of a particular genre is one sign of membership in a particular community. Freedman and Medway [12, p. 14] suggest that incomprehensible genres may even be used deliberately to defend positions of privilege.

2.1. Genre change

Drawing on Giddens’ [13] structuration theory, Orlikowski and Yates [2, p. 545] argued that, “People produce, reproduce and change genres through a process of structuring”. As members of the community draw on their knowledge of a genre repertoire to communicate, they reinforce the use of these genres, making them more appropriate or legitimate for use in the given situation. For example, by creating an order entry Web page that draws on the genre of an order form, a designer reinforces the appropriateness of the order form genre for this type of communication, making its use in future situations more likely. In other words, the set of genres in use (i.e., the genre repertoire) is both a product of and a shaper of the communicative practices of a community.

Orlikowski and Yates [2, p. 547] suggested that in a new situation individuals will typically draw on their existing genre repertoires, reproducing genres they have experienced as members of other communities. For example, traditional genres such as the book or academic article have moved intact to the Web. These reproductions may be immediately accepted or there may be a transition period during which the limits of the genre are renegotiated. For example, the electronically distributed journal article is still in transition. It is being used, but this adapted genre is not yet completely accepted or considered legitimate for all purposes (e.g.,

as evidence for a tenure case) by the academic community as a whole.

However, people are also free to modify a genre and communicate in a way that invokes only some of the expected aspects of a form. If these changes become repeatedly used, they too may become accepted and used together with or instead of existing genres, thus extending or altering the genre repertoire. For example, the journal article will likely change as it moves on to the Web to take advantage of the possibilities of linking or embedding information; the eventual form may bear only passing resemblance to the self-contained 20–25 page articles of today. As well, modifications of genres that are parts of genre systems may require corresponding changes to the rest of the system. For example, changes in citation habits will be necessary before page numbers can be dropped from the technical paper genre. Such interdependencies between genres will tend to slow the adoption of a new genre. Of course, too novel a communication may also be rejected by the community, in which case no genre will be established.

Because the definition of genre relies on social acceptance, it is impossible to define the exact point at which a new genre emerges from the old one. Acceptance may take many years. However, after some period of coexistence, the new combination of form and purpose may become generally recognized and named as a separate genre, as, for example, the FAQ (Frequently Asked Questions) has emerged as a distinct genre on the Usenet (an AltaVista search returns at least 30,000 Web pages with FAQ in their title). As well, genres may be accepted in different communities at different rates. The emergence of new genres would be one sign of the formation of a new community with new communicative practices.

2.2. Why study genres on the Web?

The Web provides a particularly interesting setting in which to study the use and development of genres and genre repertoires. First, the capabilities of the new media seem likely to result in the development of new genres of communication. Furthermore, the rapid development of this media suggests a high level of experimentation with potential genres. Bearman [14, pp. 160–161], for example, notes the rapid evolution in what he refers to as “forms of material” in electronic media in general.

Second, because the majority of Web sites are public, many examples of Web communication are easily available. Furthermore, because there is no central management of the Internet or the Web, there is no explicit management or enforcement of genres of communication, as might happen in the introduction of a communication system in a corporate environment [15]. Instead, individual Web site developers individually choose how to present their information, drawing on their understanding as members of a community, what Orlikowski et al. [15] called implicit structuring (in this

case, from the point of view of the Web page developer rather than the recipient of the communication).

Finally, there are many communities meeting on the Web, bringing experiences with different genres and using the Web for many different purposes. The Web is sometimes used for direct communication where someone with a Web server “delivers” a document to members of a known community by giving them a URL. For example, some academics use the Web to communicate with colleagues by publishing their own papers, and with students by publishing syllabi and assignments. Another example of communication within a predictable community is computer companies announcing new products, publishing catalogs, and providing troubleshooting tips online. Since their customers by definition have the computer necessary for Web access, computer companies have been early and heavily into Web site development in expectation of directly reaching their customers.

However, in many other cases the audience is unpredictable. Unlike the Usenet or electronic mail groups, there are no clear separation of communities into different channels of communication. Therefore, the resulting genre repertoire will likely be the result of interactions among them. In some cases, a genre may act as a type of boundary object [16], providing a common point of contact between different groups [12]. In others, this mixing may lead to genre confusion, meaning that there is a practical need to understand the way genres enable communication. For example, organizations have used the Web to publish information such as product brochures, annual reports, country, state, and city home pages, government agency press releases, etc. These organizations tend to use existing genres when putting information on the Web. However, a person happening to reach a document on one of their Web sites has a good chance of being outside the community in which that genre evolved. As a result the content may be confusing and the communicative purpose lost.

3. Method

To document the range of genres currently in use on the Web, we sampled and classified randomly selected Web pages. We chose the individual Web page as the unit of analysis because we felt that sampling Web sites (e.g., picking randomly from Yahoo’s list of sources) would have resulted in mostly “cover pages” for Web content. Instead, we sampled pages without regard to where they appeared in a site. As a result, our sample included parts of documents as well as whole documents, allowing us to see adaptations of existing genres in different parts of an electronic document. However, this procedure means that we were more likely to choose pages from sites with many pages and from the interior of a site, since there are many interior pages and typically only one top page (Bray [4] found that the

majority of pages are pointed to only by other pages at the same site).

3.1. Sample

A pilot sample of 100 pages was generated by clicking on the “Anything Goes” link under the “Surprises” category of the AltaVista search engine put up by Digital Equipment Corporation (<http://altavista.digital.com>). For this sample, language was restricted to English so that the authors would be able to determine the purpose of Web pages that did not follow traditional forms. One-hundred and twenty-five hits were required to obtain 100 useable Web pages. The 25 unusable pages included 12 non-English, 7 “Error 404: URL not found”, 2 “Error 403: client not authorized to get this URL”, 2 unidentifiable binary or other files, 1 Usenet newsgroup archive and 1 listserv archive. The pages were examined between January 15 and 28, 1996; some links may already be obsolete.

AltaVista’s “Anything Goes” link selects pages by generating a random number n , selecting the n^{th} word in the inverted file [i.e., a listing of all the words in all the documents, 17, p. 82–89] and then recalling the document corresponding to that word (Monier, 1996, personal communication). Because this approach samples in proportion to the number of words in the document, long documents were over-represented in our pilot sample. To avoid this size bias, a second sample was created by uniformly sampling 1000 URLs from the URL database created by the developers of AltaVista. As with the pilot sample, these 1000 URLs include pages in numerous languages (at least 9 so far), as well as 313 obsolete URLs and 35 URLs where the server did not respond. As the coding of the larger sample is still in progress, this paper will focus primarily on results from the pilot sample. However, data from coding completed to date seems to conform with these initial results.

It is worth noting that even this larger sample is imperfect for identifying innovative genres. A random sample of pages tends to return more leaf nodes, simply because there are more leaf nodes than “top-level” pages. More heavily trafficked top-level pages might show more innovation in use of genres, which would not be reflected in our sample. While we feel our sample is appropriate to catalog the current range of genres in use, we will discuss the possibilities of a non-random sampling method in the conclusion.

3.2. Coding

In their study of genres in electronic mail, Orlikowski and Yates [2] coded the purpose of each message as well as specific features such as the presence of embedded messages, subheadings or lists. They then defined genres

in terms of combinations of these features. Such an approach was necessary to inductively identify specific genres. As well, their study needed such precision because their messages were mostly quite similar, the differences between genres were not blatant and because they wanted to reliably classify hundreds of messages. In our case, however, we had only a few examples of many different genres and the differences between most were pronounced, obviating the need for such precision.

Therefore, genres were identified based on the authors’ experience with the Web and with other forms of communication. The determination of the genre of each Web page was done by each author separately. The authors then discussed the classification of each page and either agreed on a genre or that the page was unclassifiable. Because genres form a hierarchy, as discussed above, we attempted to code at the most precise level, noting that some pages are also examples of other less specific genres. (In contrast, Orlikowski and Yates [2] worked exclusively at the most precise level.) Interestingly, in coding the second sample we found many cases where it was possible to tentatively assign a genre to a page written in a language neither of the authors could read, underlining the importance of form in defining certain genres.

4. Data

The URLs and titles of the 100 Web pages in the pilot sample are shown in the Appendix. The pages had diverse origins, as shown in Table 1. Twelve countries were represented in the sample, despite our restriction to English language pages. About equal numbers were from educational and commercial sites, while a smaller number were from government sites. The second sample includes pages from at least 40 countries, as shown in Table 1.

To determine the effect of the sampling bias, we calculated the number of words in the text of each page using the UNIX `wc` command. Pages ranged from 9 to 111,586 words, with an average of 5900 words and a median of 2000 (as might be expected, the distribution was quite skewed). Our pages are clearly much longer than average—the average size of a Web page is about 6500 bytes or 1050 words [4]. We have not yet determined the average page length in the second sample, although we expect it to be smaller.

About half (53) of the pages had hyper-text links in the body of the document, 24 had links on the page but outside the body (e.g., for navigation or e-mail), while 23 had no linking at all. Links are one of the defining characteristics of the Web, so we were somewhat surprised by the low use. However, Bray [4] also found that just under 75% of pages included a link.

Table 1. Top-level domains of sample sites.

Domain	Pilot	Second
edu	34	223
com	32	260
gov	6	28
org	5	34
net	3	58
ca (Canada)	3	45
us (USA)	3	20
nl (Netherlands)	3	9
se (Sweden)	2	23
au (Australia)	2	16
uk (United Kingdom)	1	34
de (Germany)	1	33
ch (Switzerland)	1	8
ie (Ireland)	1	5
za (South Africa)	1	2
in (India)	1	0
pl (Poland)	1	0
jp (Japan)		46
fr (France)		13
fi (Finland)		9
it (Italy)		8
no (Norway)		8
be (Belgium)		7
dk (Denmark)		6
mil		6

Domain	Pilot	Second
at (Austria)		5
es (Spain)		5
hk (Hong Kong)		5
hu (Hungary)		4
tw (Taiwan)		4
br (Brazil)		3
cz (Czech Republic)		3
is (Iceland)		3
kr (South Korea)		3
sg (Singapore)		3
cl (Chile)		2
my (Malaysia)		2
co (Colombia)		1
gr (Greece)		1
hr (Croatia)		1
il (Israel)		1
mx (Mexico)		1
nz (New Zealand)		1
pt (Portugal)		1
ru (Russian Federation)		1
si (Slovenia)		1
sk (Slovakia)		1
tr (Turkey)		1
IP address		11

The genres we assigned to each page are also given in the Appendix. We found a surprisingly wide range of genres. Table 2 lists the genres we found along with a short definition of each one, drawn in many cases from the Oxford English Dictionary (OED). Table 3 shows the number of pages which used familiar genres, new genres, a mix of new and familiar genres and which we could not classify. In most cases, we agreed on the genre represented, but for 7 we encountered some difficulties, which are discussed below.

5. Discussion

In our survey we found examples of the reproduction and adaptation of genres. As well, we found components of genre systems. In this section, we will briefly discuss these findings.

5.1. Genres on the Web

Most of the pages (80) we studied more or less faithfully reproduced genres or combinations of genres familiar in traditional media, such as the book (#4), FAQ (#98), meeting minutes (#43), box scores (#31) and script (#48). In some cases, we recognized the purpose and form of the page, although we were at a loss for a convenient term. For example, one page described an organization (#19), while several described residential communities or cities (#15, 68) in stereotyped ways. These pages may represent genres that are common in a community of which we are not members (e.g., public relations, real estate). Other pages represented types of communication that are stereotyped, but not usually named, such as someone displaying and describing photographs of family members (#29).

Table 2. Characteristics of identified genres.

Genre	Characteristics of form and purpose
Book	Long work (one or more volumes) on any topic, often divided into chapters, with a table of contents and index
Report	A formal statement of the results of an investigation or of any matter on which definite information is required, made by some person or body instructed or required to do so (OED)
Newsletter	Medium length work, including multiple articles, titled, issued periodically (e.g., with date or volume), communicates “the news of the day”, distributed periodically to subscribers or members of an organization
Essay	A composition of moderate length on any particular subject (OED)
Pamphlet	A small treatise occupying fewer pages or sheets than would make a book, issued as a separate work unbound, on some subject or question of current or temporary interest, personal, social, political, ecclesiastical, or controversial, on which the writer desires to appeal to the public (OED)
Article	A literary composition forming materially part of a journal, magazine, encyclopædia, or other collection, but treating a specific topic distinctly and independently (OED)
News wire article	An article prepared for a news wire, with a dateline and news content
Column	A special feature, esp. one of a regular series of articles or reports (OED)
Memorial	An article by which the memory of a person, thing, or event is preserved
Concert review	Retrospective description of a concert
Product reviews	Description and evaluation of products to advise potential purchasers
Ratings	Numeric evaluations of products or services
Submission instructions	Instructions for preparation of material to be submitted for consideration for publication
Table of contents	A summary of the matters contained in a book, in the order in which they occur, usually placed at the beginning of the book (OED)
Index	An alphabetical list, placed (usually) at the end of a book, of the names, subjects, etc. occurring in it, with indication of the places in which they occur (OED)
Discography	A catalogue raisonné of gramophone records; a list of the recordings of a single composer or performer (OED)
Filmography	A list of the films of a particular director, producer, actor, etc., or of those dealing with any particular theme (OED)
Regulation or rule	A rule prescribed for the management of some matter, or for the regulating of conduct; a governing precept or direction; a standing rule (OED)
Product information	A description of the features and/or benefits of a product, written by manufacturer for potential purchasers
Government program description	A description of the features of government program, written by agency for potential beneficiaries
Testimonial	A letter of recommendation of a person or thing (OED)
University course listing	A list of numbers, titles and description of course offered in some department or departments, possibly including information such as prerequisites, instructor, meeting time and place
Problem set	A collection of problems assigned for a course
Faculty information	Short descriptions of individual faculty members
Vitae	A brief account of one's career (OED)
Publications list	A list of publications by members of some group

Genre	Characteristics of form and purpose
List of research projects	A list of research projects underway in some group
Directory	A book containing one or more alphabetical lists of the inhabitants of any locality, with their addresses and occupations; also a similar compilation dealing with the members of a particular profession, trade, or association (OED)
Library acquisitions list	List of books acquired during a particular time period
Order form	A fill-in form for obtaining the information needed to arrange payment and shipment of some product.
Meeting minutes	The record of the proceedings at a meeting of an assembly, corporate body, society, company, committee, or the like (OED)
Box score	A record of players and plays in a game
Chronicle	A detailed and continuous register of events in order of time; a historical record, esp. one in which the facts are narrated without philosophic treatment, or any attempt at literary style (OED)
Script	The typescript of a cinema or television film; the text of a broadcast announcement, talk, play, or other material (OED)
Political party platform	A public declaration of the principles and policy on which a political party proposes to stand (OED)
Genealogy	An account of one's descent from an ancestor or ancestors, by enumeration of the intermediate persons; a pedigree (OED)
Demographic data	Data on the characteristics or composition of a population
Guide	A book of information on places or objects of interest in a locality, city, building, etc. (OED)
Archive item	Description of item stored in a collection of historical documents
FAQ	Edited collection of questions and answers on some topic; often labelled as such
Users' manual	Instructions on the use of a product
Computer documentation	Instructions on the use of a computer program or of a computer programming system
Source code	A computer program as written by the programmer.
File directory listing	A list of files in a computer directory
E-mail directory listing	A list of links to e-mail messages
Hot list	A list of Web sites not controlled by the list's author, often organized by topic
Home page	A web page presenting personal or organizational information or the page at the hierarchical top of a website presenting this information.
Server statistics	Records of accesses to a Web server

A few pages (11) appeared to be novel genres, yet ones that are already well accepted by the general Web community. These included the hotlist, home page and Web server statistics. Yates and Orlikowski [1] suggested that these new genres are most likely derived from earlier genres that might have seemed appropriate to the situation

We defined a hotlist as a series of links to material not controlled by the page developer (as opposed to the table of contents of a document or the directory of a site's content), often on a related set of topics. In our sample, we found hotlists on topics such as multimedia (#35), music (#41) and anime (#90). Such lists are the

online equivalent of a jotted list of discoveries or the hotlist menus maintained by some browsers. These lists were especially useful before the development of good resource discovery tools, such as Yahoo, AltaVista or Lycos. They appear now to be developed as a way to express an interest or to add value to a Web site (such as a commercial site) as much as for personal navigation. The hotlist form may also have drawn on earlier forms of posted lists of useful FTP sites or bulletin boards.

An especially easily identified and commonly accepted genre is the home page (defined in Table 2). In our sample, we found several (e.g., #80, #75), containing personal or professional information. Interestingly, one

of these (#71) stated that it had been developed to meet the expectation of IRC users that everyone have one as a form of introduction.

The antecedents of the home page genre are unclear. One hypothesis suggested by JoAnne Yates (Yates, 1996, personal communication) is that they are adaptations of the .plan files maintained on some UNIX machines. (A .plan file is created by the owner of an account and printed whenever another user “fingers” or requests the status of the account. It typically includes contact information and a brief description of the owner’s interests.) There are also organizationally-created home pages, which seem to be adaptations of entries from a university’s faculty profiles book. As well, many organizations maintain home pages as overviews to the organization, their site or some collection of information. For example, one of our pages (#81) presented information about the Greater Eau Claire Area.

Table 3. Count of genres found.

Familiar genres	80
Named	73
Didn’t know name	4
Unnamed	3
New, but accepted genres	11
Hotlists	4
Home pages	3
Web server statistics	4
Unknown genres	9
Unknown purpose	3
Mixed features	6

Finally, 4 of the pages we encountered reported Web server statistics (#17, #27, #69, #86). These we felt were an example of the potentials for genre confusion. The statistics are reports of interest primarily to the managers and sponsors of the Web site; however, it turns out to be technically easy to provide them via the Web, which makes them available to everyone, even though most are unlikely to find them useful or even understandable.

The most interesting genres we found showed signs of adaptation, taking advantage of the capabilities of the media. For example, we found two pieces of genealogies (#59, #65) which used linking to display and navigate the usually unwieldy amount of data in a family tree. Two pages (#34 and #62) were both examples of what we called a letter column, for want of a better label, where readers could write in to request help or information from other readers. However one page (#62) was simply a transfer of this concept to the Web, where readers wrote or e-mailed to an editor, who then included the letters in a regular column. The second (#34) was an

automated version, where readers could use their Web browsers to read and post questions and replies, thus speeding the process by making use of the increased interactivity of the media. Similarly, one newsletter (#100) included a survey implemented as a Web form which could be filled out online, again speeding up the flow of information.

5.2. Embedded genres

We encountered several examples of genres embedded in other genres. For example, one page (#16) was at one level an entry in an archive, but included a letter, within which were stories for a folklore collection. Each wrapping (e.g., an archive entry around a letter around a set of stories) created a new genre without completely losing the characteristics of the previous instantiation. This phenomenon is not new, as Orlikowski and Yates [2] point out, but we believe that it is much more relevant to the Web because of the ease of reusing text in electronic form. For example, wrapping seemed particularly common in an e-mail digest. The surface genre is a simple listing of e-mail messages, but when these are selected on a particular topic, e.g., discussions or reviews of a type of product (#88, #24) they create a review or FAQ. In some cases, this selection might even be done automatically.

Through the use of linking, a single document can serve multiple purposes, typically something plus an index. For example, page #76 is in one way a filmography, listing all the films of Laurel and Hardy; however, it also has links to catalog descriptions of the films that can be ordered from the company, thus creating an index or catalog. Similarly, one page was a list of ratings for zines (#46), but also included links to longer reviews and to the zines themselves.

5.3. Parts of documents

Finally, we found numerous pages that were parts of longer documents, e.g., part of an index or a chapter in longer document (#83, 9, 56). In many cases these pages were still recognizable as a distinctive genre, although the purpose was sometimes hard to determine.

5.4. Unclassified pages

As mentioned earlier, we had difficulty assigning genres to a number of the pages, most often when we agreed there was a genre, but simply did not know the name. In 2 other cases (#32, #72), we could not determine the purpose of the communication, making the assignment of a genre problematic. However, we believe that some of these cases may be examples of genres in the process of adaptation to the Web. At this early stage, they are usually seen as variants of an accepted genre, missing some features and possibly adding others. For example, the selected collections of e-mail (#24, #89)

were not considered proper e-mail archives because they were incomplete and not sorted by time or author, but they were not quite FAQs either, because they were not edited into a coherent document. If their use continues, then they may eventually become independent genres. Others, such as the single entry from a table of contents (#2), may represent dead-ends, experiments with linking that do not become common accepted usage.

6. Conclusions

We argue that genres provided a useful theoretical tool for analyzing uses of the Web. In general, the concept of genre seemed to be easily applied to the Web pages we studied, with some caveats. Genres distinguished primarily by difference in physical form (e.g., a brochure vs. a booklet vs. a flyer) are not very useful. Instead, we focus more on the purpose of the information, e.g., product description, services, etc. We believe it would be useful to characterize more precisely the common purposes for which information is distributed and how these are reflected in different genres.

We found numerous examples of genres being reproduced on or adapted to the new media as well as a few examples of new genres, such as the hotlist and home page. There are already some puzzles, like the origins and antecedents of the home page. Therefore, we hope to study the Web over a longer period to better document the processes by which genres are being adapted and new genres emerging. We expect to find further examples of this process in our larger sample.

We believe that the size of the genre repertoire is a reflection of the many different communities on the Web and their varied uses of the medium. While our sample included a surprisingly large variety of genres, the limitations of the sampling technique makes it impossible to draw any firm conclusions on the relative frequency of the use of different genres. (For example, our sample included 4 sets of Web server statistics, but we believe this result is due to their length rather than the frequency of this genre.) Our larger sample was created without the size bias and should be more useful in examining this point.

Some of our pages seemed to be parts of genre systems. As people start to build more interactive Web applications, e.g., to support internal and external processes rather than simply broadcasting information, defining such genre systems will be more important. As a result, we think it might be useful to define the form of a hyper-document (and thus the overall genre) by the pattern of links it exhibits. For example, a hotlist is a linear list of links, all of which go to other sites; an online book is (usually) a sequential list of chapters, each linked to the next, perhaps with a table of contents pointing to each chapter; and a hyper-document has a pattern of densely inter-linked pages. Similarly, a glossary could be identified by the links from throughout

a set of documents to regular points on a separate set of pages. Thus, the genre of a hyper-document might be determined in part by examining how its component parts are linked together.

Perhaps our biggest surprise was just how mundane our sample was. The 100 pages in our pilot sample did not include anything particularly radical, such as a remote controlled robot (an example of a user interface genre) or video camera (perhaps a TV monitor genre). The absence of such pages may be due to the bias in our sample, which over-selected long text-only documents, combined with the general rarity of such applications. As well, some of the most interesting pages may be hidden from the search engines used to create the database from which we drew our samples. Therefore, we believe that it would be interesting to conduct a follow-up study using a theoretical sample of "interesting" pages. It would be quite difficult to create a reliable sample of genres this way, but reliability would not be a key concern in a study describing innovative uses of the Web, their antecedents or the process of adaptation.

6.1 Implications for Web site design

Finally, we believe that our research has some implications for the practice of Web site design. First, designers may want to draw on accepted genre where they are appropriate for their purpose. One HTML design book takes this approach, offering samples of home pages for individuals and large and small organizations, brochures, surveys and hotlists [18]. Most do not, focusing on the mechanics of formatting a page to the exclusion of the communicative intent. As well, designers should be aware of users' expectations of a genre. For example, we are accustomed to tables of contents and indices that list the entire contents of a book. It can be very disconcerting, therefore, to encounter site home pages that appears to be tables of contents, or a searches that appear to be indices, but which are incomplete or are actually hotlists. Of course, the difficulty of meeting a reader's expectations is compounded by the diverse audience of the Web. Even so, on too many pages we found it hard to determine the communicative intent, if any.

We believe that it is useful to establish the genre of even a single Web page. In a physical document, pagination is determined by the physical dimensions of the book and so is not usually meaningful. However, there is no particular limit to the length of a page on the Web, so division into pages should reflect the actual structure of the communication. Furthermore, with the growing use of indexing systems such as Lycos or AltaVista, it is not uncommon for a user to start reading in the middle of a document as we did, again suggesting that the purpose and form of even a single page should be evident. However, we noted that numerous sites lacked navigational aids to help a reader figure out

where they were in a longer document or the purpose of that communication.

On the other hand, the technology of the Web enables novel applications, such as a shift from static documents to “live” data. For example, Yan et al. [19] describe how patterns of user access can be used to suggest which information should be viewed next. Designers should be free to modify or reject genres when it is necessary to take advantage of the technology. In creating novel applications, however, designers must be aware that new genres are often misunderstood or resisted. Therefore, more attention is needed to clearly define the community in which the communication makes sense and to identify already accepted genres that can serve as a basis for evolution. We believe that such explicit attention to genre will speed the wider acceptance of newly emerging genres of communication unique to the Web.

7. References

- [1] J. Yates and W. J. Orlikowski, “Genres of organizational communication: A structurational approach to studying communications and media,” *Academy of Management Review*, vol. 17, pp. 299–326, 1992.
- [2] W. J. Orlikowski and J. Yates, “Genre repertoire: The structuring of communicative practices in organizations,” *Administrative Sciences Quarterly*, vol. 33, pp. 541–574, 1994.
- [3] T. Berners-Lee, R. Cailliau, A. Luotonen, H. F. Nielsen, and A. Secret, “The World-Wide Web,” *Communications of the ACM*, vol. 37, pp. 76–82, 1994.
- [4] T. Bray, “Measuring the Web,” presented at Fifth International World Wide Web Conference, Paris, France, 1996.
- [5] A. Woodruff, P. M. Aoki, E. Brewer, P. Gauthier, and L. A. Rowe, “An investigation of documents from the World Wide Web,” presented at Fifth International World Wide Web Conference, Paris, France, 1996.
- [6] K. K. Campbell and K. H. Jamieson, “Form and Genre: Shaping rhetorical action,” . Fall Church, VA: Speech Communication Association, 1978.
- [7] J. Harrell and W. A. Linkugel, “On rhetorical genre: An organizing perspective,” *Philosophy and Rhetoric*, vol. 11, 1978.
- [8] C. R. Miller, “Genre as social action,” *Quarterly Journal of Speech*, vol. 70, pp. 151–167, 1984.
- [9] C. Dollar, “Electronic objects circa 2001: Problems or opportunities?... Yes,” presented at Conference on Electronic Records in the New Millennium, Vancouver, BC, 1994.
- [10] L. F. Bitzer, “The rhetorical situation,” *Philosophy and Rhetoric*, vol. 1, pp. 1–14, 1968.
- [11] C. Bazerman, “Systems of genres and the enactment of social intentions,” in *Genre and the New Rhetoric*, A. Freedman and P. Medway, Eds. London: Taylor and Francis, 1995.
- [12] A. Freedman and P. Medway, “Locating genre studies: Antecedents and prospects,” in *Genre and the New Rhetoric*, A. Freedman and P. Medway, Eds. London: Taylor and Francis, 1994, pp. 1–22.
- [13] A. Giddens, *The Constitution of Society: Outline of the Theory of Structuration*. Berkeley: University of California, 1984.
- [14] D. Bearman, *Electronic Evidence: Strategies for Managing Records in Contemporary Organizations*. Pittsburgh: Archives and Museum Informatics, 1994.
- [15] W. J. Orlikowski, J. Yates, K. Okamura, and M. Fujimoto, “Shaping electronic communication: The metastructuring of technology in the context of use,” *Organization Science*, vol. 6, pp. 423–444, 1995.
- [16] S. L. Star and J. R. Griesemer, “Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39,” in *Social Studies of Science*, vol. 19. Newbury Park: Sage, 1989, pp. 387–420.
- [17] J. Rowley, *Organizing Knowledge*, 2nd ed. Brookfield, VT: Ashgate, 1992.
- [18] L. Arronson, *HTML Manual of Style*. Emeryville, CA: ZD Press, 1994.
- [19] T. W. Yan, M. Jacobsen, H. Garcia-Molina, and U. Dayal, “From user access patterns to dynamic hypertext linking,” presented at Fifth International World Wide Web Conference, Paris, France, 1996.

8. Acknowledgments

We would like to thank Louis Monier of Digital Equipment Corporation’s Palo Alto lab for his assistance in understanding AltaVista and for providing the second sample of URLs and JoAnne Yates for her advice on this paper.

9. Appendix

Due to space limitations, the Appendix could not be included in this version of this document. It is available on the Web at:

<http://istweb.syr.edu/Projects/Faculty/Crowston.html>