Syracuse University
SURFACE

The School of Information Studies Faculty Scholarship

School of Information Studies (iSchool)

1999

The Effects of Linking on Genres of Web Documents

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

Marie Williams
Web Architects

Follow this and additional works at: http://surface.syr.edu/istpub

Part of the Library and Information Science Commons

Recommended Citation
Crowston, Kevin and Williams, Marie, "The Effects of Linking on Genres of Web Documents" (1999). Former Departments, Centers, Institutes and Projects. Paper 3. http://surface.syr.edu/ischool_other/3

This Article is brought to you for free and open access by the School of Information Studies (iSchool) at SURFACE. It has been accepted for inclusion in The School of Information Studies Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

The Effects of Linking on Genres of Web Documents

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

Marie Williams
Web Architechs
williams@web-arch.com

Abstract

Documents on the Web can be composed of multiple Web pages, suggesting the need to consider how linking between pages affects a document's form. We illustrate this point by considering patterns of linking in a common genre of document, the Frequently Asked Questions file or FAQ. In a sample of 70 FAQs, we found four patterns of linking: no links, links within the page, links to pages on the same host and links to other hosts. We suggest that links that tie together document pieces simply recreate the already accepted FAQ genre, but links that provide navigation within the document or that link to other information sources begin to extend and adapt the FAQ genre to the needs and capabilities of the Web.

1.Introduction

The World-Wide Web (or the Web) is an Internet client-server communication system for retrieving and displaying multi-media hypertext documents [1]. 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).

The purpose of our study was to describe how genres of communication might evolve given the capability of the Web to link between pages. Communicative genre is generally defined as an accepted type 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 the possible form of documents—an email memo has a somewhat different form than a paper memo. In this paper, we will examine in particular how Web linking affects form, possibly creating new genres.

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., 2, 3, 4]. More recently, Yates and Orlikowski [5, 6] 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" [6, p. 543]. In other words, given a socially recognized need to communicate [i.e., a purpose, 7], 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", includes multiple short articles and is distributed periodically to subscribers or members of an organization.

Genres on the Web

Crowston and Williams [8] analyzed Web pages to see how established genres were being adapted to this new medium. They examined 100 randomly selected Web pages to categorize the genre represented. They found many pages that recreated genres familiar from traditional media and a few adapting to take advantage of the linking and interactivity of the new medium. As well, a few novel genres were seen emerging to fit the unique communicative needs of the Web audience. Crowston and Williams [8] analyzed the genre of single Web pages. However, examining only single page is limited, because many documents, especially long ones, span multiple

Web pages (e.g., to reduce download time when a reader wants only a portion of the document). Therefore, a discussion of genre of Web documents should consider how division of documents into pages affects document form.

Genres of multi-page documents

In the case of multi-page documents on the Web, the form of the hyper-document (and thus its overall genre) is defined in part by the pattern of links it exhibits. The form of the document is reflected in the pattern of linking between the pages because links govern the pattern of access to the information. For example, a document designer can force sequential access to information by linking only between a page and the page following it. This technique is used on the Web to present a single narrative rather than allowing viewers to move around at random (e.g., links only between sequential pages of a photo album of a trip to prevent the photos from being viewed out of order). Many technical manuals moved to the Web have hierarchical links (up, down, previous, next) that reflect the structure of the document, divided into chapters, sections, subsections, etc. A glossary might have a dense and random pattern of links within the glossary for cross-references, and to the glossary from a referring document.

In this paper, we argue that only linking that affects the purpose of the document changes the genre of the document; merely dividing a document into pages does not, any more than routine repagination affects the genre of a paper document. Stated alternately, we argue that there is a class of changes to form, namely those related to pagination rather purpose, that do not affect genre.

Of course, determining the purpose of a document is difficult without knowing more about the intended uses and users. However, we can at least put ourselves in the shoes of a potential user and examine the suitability of a given document for the traditional purpose of the genre it appears to represent.

For example, omitting links may make the document unsuitable for its usual purpose, which would change the genre. For instance, a reference manual with only sequential links from one chapter to the other would, we argue, no longer be a reference manual since it would be nearly impossible to lookup random pieces of information, a key purpose of a manual. Contrariwise, adding links may enable the document to serve new purposes, thus creating a new genre. For example, an index to a novel might provide random access to each scene where a particular character appears, creating something more than a traditional novel.

3.Method

In this section we will review how we went about collecting and analyzing data about linking in FAQ documents. We will first describe the sample of pages we created, then how link data were extracted from those pages, and finally, how those link data were analyzed.

A sample of pages

To support our thesis, we decided to study documents of an established genre that had been moved to the Web to see how linking was used and what effect the linking had on the document genre. To start, we chose to study Frequently Asked Question documents, or FAQs. An FAQ is an edited collection of questions and answers on some topic. Their origins are unclear, although they are quite popular on Usenet, a distributed world-wide computer conferencing system. Usenet is organized into a hierarchy of “newsgroups” on a diversity of topics, including computer systems, social issues, hobbies and current events. Users create messages and post them to a particular newsgroup or newsgroups, where they can be read and replied to by anyone who chooses to subscribe to that newsgroup. Many newsgroups have FAQs that are maintained and periodically posted to the list to document the commonly sought collected wisdom of the newsgroup and as a starting point for new members of the group.

We decided to start our study with FAQs because:

- they are common (an AltaVista search indicates approximately 170,000 Web pages with FAQ or “Frequently asked questions” in their title);
- they are easy to find and identify (FAQs are often labelled as such and have a distinctive question and answer format); and
- because they have been converted to HTML in a variety of ways, thus making them an appropriate initial focus for our study.

Since the purpose of the study was to find thought-provoking examples of the use of links, rather than drawing statistically significant conclusions about the population of Web pages, we did not attempt to create a representative random sample of FAQs. Instead, candidate FAQs were found by searching in the Yahoo directory of Web sites. We chose to use Yahoo because we wanted a broad range of items that had been classified as FAQs.

A search on Yahoo for “FAQ” returned 2863 sites (clearly a small subset of the universe of FAQs). For this paper, we analyzed the first 100 FAQs found in this search, which yielded a set of 95 distinct URLs. We then

filtered this set to eliminate out-of-date URLs or pages that were not actually FAQs. We first eliminated cases where the server or document was not found (404s) or the document was empty, leaving 82 pages. Six of these pages indicated that the FAQ itself had moved, in which case we used the new URL and page in the analysis.

We then examined each page to eliminate those that were not FAQs. In some cases, the page catalogued by Yahoo was of some other genre. We considered a document an FAQ if the document called itself an FAQ (e.g., in the title) and included some information in the form of questions and answers. A few pages were marginal, since they included other information not as questions and answer (e.g., in the form of an announcement). If the page wasn't an FAQ but included a link to an FAQ, we used the linked page instead. For example, we found 3 title pages that named the document and provided a link to the content. A couple of sites offered a choice between an HTML and a text-only FAQ, in which case we included the HTML version. Eight URLs used frames, which present multiple pages simultaneously (although not all of the eight were FAQs). We observed that frames were typically used to display a list of the sections of an FAQ together with a particular question and answer, so we analyzed the frame that included the list of sections. A total of 27 pages were eliminated from, and 15 new pages added to the sample, giving a sample of 70 FAQs.

The URLs and titles of the documents are shown in the appendix. The FAQs in the sample were on a diversity of topics, though the bulk were about religion, sexuality, rock groups or programming languages. The pages came from 6 different countries, as shown in Table 1, although the majority were from the United States, reflecting Yahoo's origins. Again, this sample is not necessarily representative of all FAQs on the Web, since it reflects the biases of Yahoo's creators and of our sampling from Yahoo. However, it is adequate to suggest a range of possible approaches to adapting FAQs to the Web.

Table 1. Origins of Web sites in sample.

Origin	Count
US Commercial	26
US Educational	15
Network	10
Non-Profit	7
United Kingdom	5
Denmark	2
Canada	2
US Government	1
Finland	1
Italy	1
Grand Total	70

Extracting link data

Having located a sample of FAQs, we next examined each top-level page to see how links were used. We identified and counted the links on each page using a parser written in Perl, with a CPAN HTML library module to do the actual parsing of the HTML code. We wanted to distinguish links that connected different parts of the same FAQ document from other links that might have been used on the page. Of course, HTML does not type links in any way, so it is impossible to be sure about the purpose of a link, and thus, about the boundaries of the document. For this analysis, we used the host name in the URL as a proxy for authorship of the linked pages and assumed that pages on the same host were part of the same document.

By comparing the URL of the page and the destination of the link, the parser differentiated between links:

1. within the page,
2. to other URLs with the same host name (part of the same document), and
3. to URLs with different host names (other documents).

We originally considered counting in the third category links to pages with the same host name but in different directories (e.g., a link from a page back to the top level of the Web server). Grouping pages this way would have treated pages in different directories as part of a different document, which would reflect the common practice of storing all pages for a particular Web site in a common directory. However, this grouping did not change our results dramatically, so we elected to use the simpler definitions for our analysis.

Link usage

The count of links found in each document is shown in Table 2. The total number of links (i.e., the number of "<a href=…>"tags) is shown in the column labelled "Links". The next three columns, labelled "Same file", "Same host" and "Different host", correspond to the three different types of links discussed above. The means, standard deviations and ranges of these variables are shown in Table 2; again these counts are highly skewed. The final column of Table 2 gives the count of zeros, that is, how many pages had none of that kind of link.

Table 2. Descriptive statistics for link counts.

Variable	Mean	Std Dev	Minimum	Maximum	Zero
Links	40.09	52.03	0	315	7
Same file	15.57	42.33	0	293	40
Same host	13.59	20.53	0	134	16
Different host	10.90	17.17	0	87	26

Our first observation is that most pages use links. Only 7 have none at all. The average number of links on a page is about 40, although the distribution of counts is highly skewed, as shown in Figure 1. Examination of the 7 pages with no links at all revealed that they are text Usenet FAQs, moved to the Web (e.g., to provide broader access) but without any other modification. In other words, these pages are examples of the "classic" FAQ genre.

We next examined the kinds of links used in the other pages. Just under half of the pages (30 out of 70) use the first kind of link, links within the same file. These links are commonly used to provide a table of contents for a longer document. For example, an FAQ document could have a list of the questions at the start of the document with links to the answers later in the document.

All pages with links included the second kind of link, links to URLs with the same host name. Such links would be used to connect the pieces if the document were split into several pages.

Finally, about two-thirds of documents used the third kind of link, links to other hosts. Such links are not inherently part of the FAQ genre, but if the questions concern access to information, URLs of other sites are frequently part of the answers. A hypertext version of a page converts these URLs into clickable links.

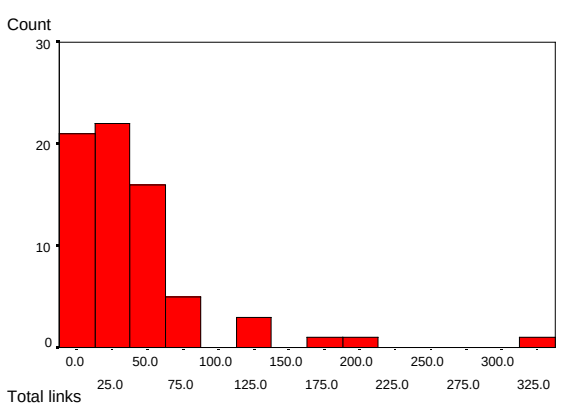

Figure 2. Histogram of total number of links.

Analysis of the link data

To determine the patterns of linking, we performed a hierarchical cluster analysis on the link counts of the pages. Because pages were of different lengths and had different number of links, we based the analysis on the proportions of each type of link (i.e., the counts of the three types of links divided by the total number of links on each page). For pages with no links we set all three proportions to zero.

The analysis was performed using SPSS's hierarchical cluster command, an agglomerative hierarchical method, using between-groups linkages (i.e., average cluster distances) and squared Euclidean distances. Average linkage was used because it is not as sensitive to poorly separated clusters. For comparability of the resulting coefficients, we normalized the distances to range from 0 to 1.

Hierarchical cluster analysis

The dendogram from the hierarchical cluster analysis of the pages is shown in Figure 2 on the next page. The final steps of the analysis are shown in Table 3. Each line in Table 3 represents one step in the clustering, as two clusters are merged together to create a larger cluster. The coefficient indicates how far apart the two merged

Table 3. Intermediate results of hierarchical clustering analysis.

Stage	Coefficient	Number of clusters
64	0.133717	6
65	0.136673	5
66	0.162637	4
67	0.301108	3
68	0.466048	2
69	0.589289	1

Note: Results from the first 63 stages have been omitted from this table.

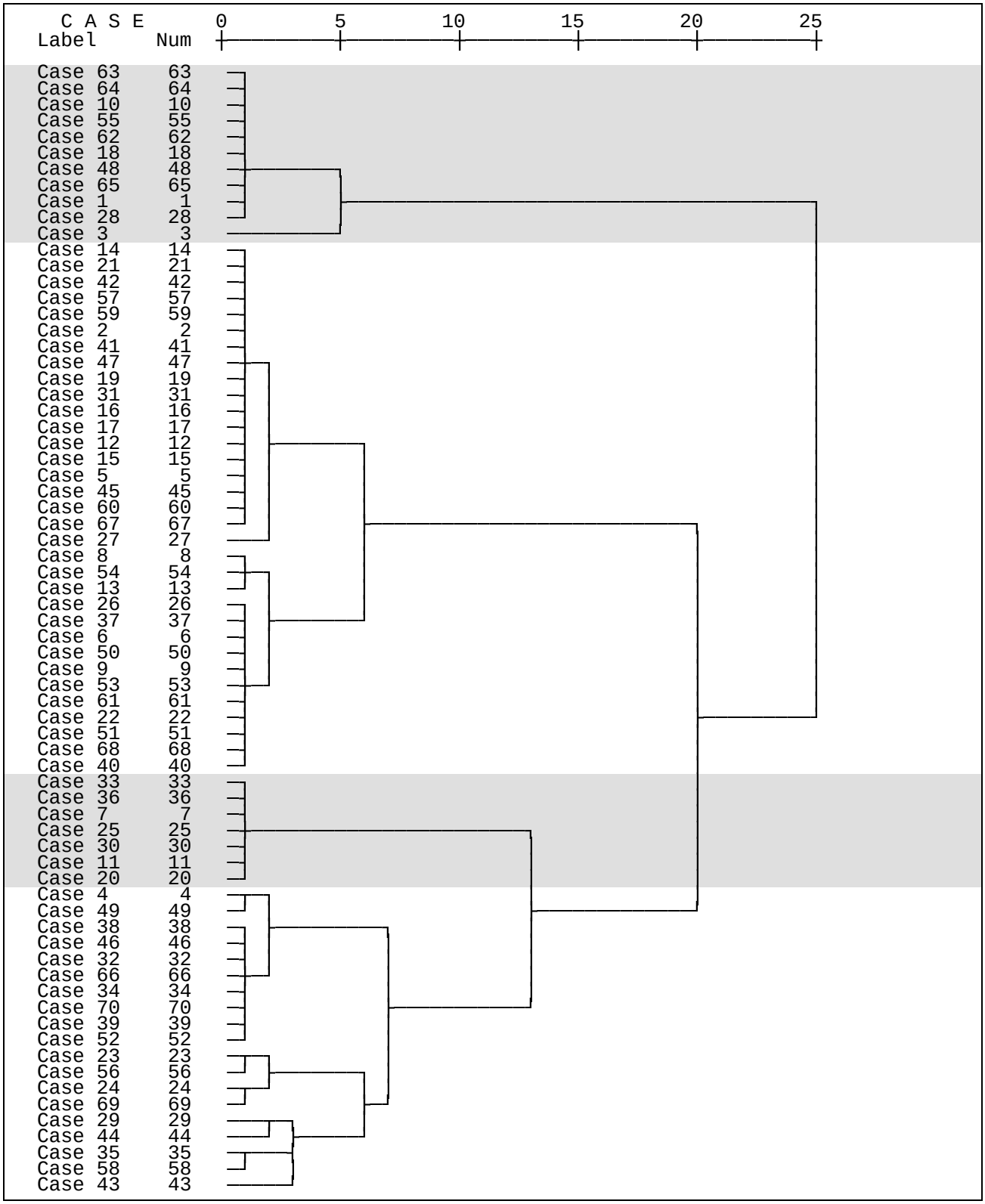

Figure 3. Average-linkage dendogram for distances between pages. Shaded bars indicate clusters.

clusters were, calculated as the average distance between the points in the two clusters. Based on the jump in the coefficient, we elected to analyze the four intermediate clusters (i.e., the results at stage 66). The four clusters are shown by the shaded bars in Figure 2.

4.Discussion

To interpret these clusters we examined the average link counts of the pages in each. We developed the following interpretations for the clusters:

1. No links at all (7 pages);
2. Links primarily on the same page (19 pages);
3. Links primarily to URLs with the same host name (33 pages); and
4. Links primarily to URLs with different host (11 pages).

Figure 3 shows a matrix scatter plot of the three link proportions versus each other. The colour of the dots indicates how the corresponding page was clustered by the hierarchical cluster analysis. The four clusters will be discussed in turn.

As we said above, the pages in the first cluster appear to be examples of the "classic" FAQ genre as it originally developed on Usenet. These documents have simply been made accessible via the Web (in addition to Usenet), but without any adaptation to the new medium.

The pages in the second cluster use linking to provide navigation within the document, reducing the need to scan the FAQ from beginning to end to find the answer to a question. These pages represent an extension to the traditional FAQ document to fit the needs of the Web. On the Usenet, reading the FAQ is considered good practice before contributing to a group, because it avoids the need to ask already discussed questions, wasting members' time. This use of the FAQ reflects its role in forming and maintaining the newsgroup's social structure.

However, on the Web, there is typically not a group to join, so the primary purpose of an FAQ seems instead to be to answer specific questions about a topic. Links within the document from the list of questions helps serve this purpose by facilitating access to specific questions. We contend that if the document still serves primarily as a repository of questions and answers that have come up in a group, it remains an FAQ. However, it is unclear to us at what point such navigational aids allow a sufficiently new purpose to be served, giving rise to a novel genre. Clearly, systematic investigation of this matter will require data beyond simply examination of the pages themselves.

The pages in the third cluster use the linking of the Web to divide a long document into smaller, more convenient pieces. When we looked more closely at these pages, we saw two patterns of division of FAQs into pages. The first pattern again simply recreates the

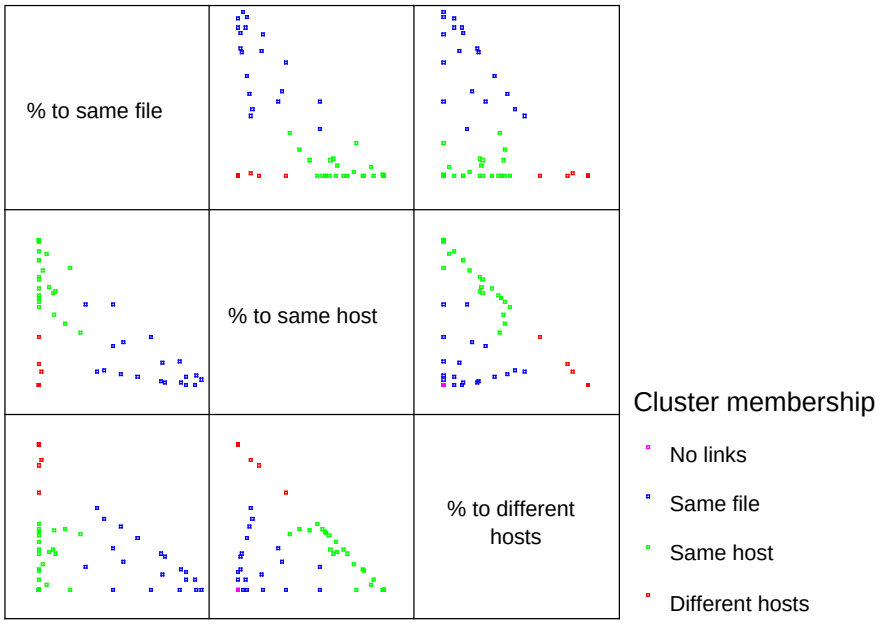

Figure 3. Scatter plot of proportion of different types of links, coded by cluster membership.

appearance of Usenet FAQs. Because Usenet postings have a maximum size, large FAQs have to be divided for distribution (just as a paper document must be split into page-sized units for printing). To ensure that these pieces appear near one another in the display of a newsreader, it is usual to give the pieces nearly identical names, such as “Rolling Stones FAQ [1/4]”, “Rolling Stones FAQ [2/4]”, etc. Several of the pages we examined simply recreated this display, listing and linking to pieces of an FAQ but providing no information about the contents of the pieces.

A more adapted version of an FAQ divides the document into pages on particular topics. For example, one FAQ offered links labelled, “Section A - What is anarchism?”, “Section B - Why do anarchists oppose the current system?”, etc. Similarly, several sites used frames to simultaneously display navigational information and the contents of the FAQ, making it more convenient for the user to navigate through a long FAQ.

In both cases, the form of the document has been changed somewhat by conversion to multiple pages. However, the first version does not facilitate searching for particular topics, while the second does. The difference between these two forms parallels the difference between the first two clusters of pages we just discussed. In other words, while both uses of links take advantage of the capabilities of the Web, the first simply recreate the existing FAQ genre, while the second begin to extend the genre to meet the needs of the Web.

The pages in the fourth cluster use links primarily to provide access to other sites. These pages appear to have been converted from FAQs without division into pages. However, where reference was made to an outside source, the URL has been converted to a clickable link. In this case, the basic form of the document has not changed, but the purpose has been extended. Instead of simply providing answers, these documents now also provide an organized set of references to a broader set of material. In this way, the documents have been transformed from simple FAQs to a hybrid of FAQ and hotlist (defined as a series of links to material not controlled by the page developer, on a related set of topics).

We also examined the results with five intermediate clusters (i.e., the hierarchical clustering at stage 66). In this case, the first four clusters have the same interpretations as above, while the fifth cluster includes pages with intermediate proportions of the three types of links, as shown in Figure 4. (These pages were merged with the same file cluster in the four-cluster result). Since these pages make use of all three kinds of links, these FAQs show the greatest signs of adaptation to the capabilities of the Web.

5.Conclusions

In summary, we noted that documents on the Web are sometimes composed of multiple Web pages, suggesting the need to consider how linking affects a document’s

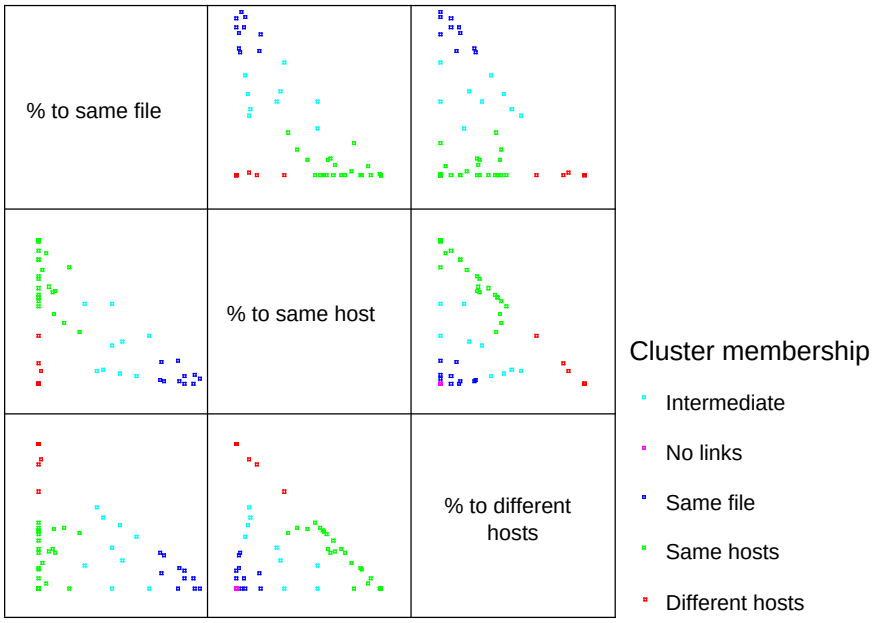

Figure 4. Scatter plot of proportion of different types of links, coded by cluster membership, with 5 clusters.

form. We illustrated this point by considering patterns of linking in a common genre of document, the Frequently Asked Questions file or FAQ. Our analysis revealed four clusters of link usage, namely, 1) no links, 2) links primarily within the file, 3) links primarily to other pages on the same host and 4) links primarily to other pages on other hosts.

We argued that only linking that affects the purpose of the document changes the genre of the document; merely dividing a document into pages does not, any more than routine repagination affects the genre of a paper document. More specifically, we suggested that links that tie together document pieces (i.e., cluster 3, links to other pages on the same host) simply recreate the already-accepted FAQ genre. On the other hand, links that provide navigation within the document (cluster 2) or that link to other information sources (cluster 4) begin to extend and adapt the FAQ genre to the needs and capabilities of the Web. However, it remains uncertain at what point such adaptations can be said to have led to the establishment of a new genre. Further studies of these novel genres will likely require data from the user community as well as from the documents they create.

To test our principle, we hope in the future to perform a similar analysis with documents of other genres. For example, because of their length, manuals and reports are often broken up into multiple pages when they are moved to the Web and might therefore be good candidates for our study. It would also be interesting to examine novel genres, such as shopping cart systems, to determine if it is meaningful to assign a genre to such Web-based interactive systems.

References

[1] 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.

[2] K. K. Campbell and K. H. Jamieson, "Form and Genre: Shaping rhetorical action," . Fall Church, VA: Speech Communication Association, 1978.

[3] J. Harrell and W. A. Linkugel, "On rhetorical genre: An organizing perspective," *Philosophy and Rhetoric*, vol. 11, 1978.

[4] C. R. Miller, "Genre as social action," *Quarterly Journal of Speech*, vol. 70, pp. 151–167, 1984.

[5] 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.

[6] W. J. Orlikowski and J. Yates, "Genre repertoire: The structuring of communicative practices in organizations," *Administrative Sciences Quarterly*, vol. 33, pp. 541–574, 1994.

[7] L. F. Bitzer, "The rhetorical situation," *Philosophy and Rhetoric*, vol. 1, pp. 1–14, 1968.

[8] K. Crowston and M. Williams, "Reproduced and emergent genres of communication on the World-Wide Web," in Thirtieth Hawaii International Conference on Systems Science (HICSS–30), Maui, HI, 1997.

Appendix

Title, URLs and counts of links from sample FAQs.

Title	URL	Links	Same file	Same host	Different host
Java FAQ Archives	http://www-net.com/java/faq/	35		5	30
The hypertextualized Tolkien FAQ	http://www.daimi.aau.dk/~bouvin/tolkienfaq.html	19		19	
Trumpet FAQ	http://www.roth-music.com/trumpet/trumpet-faq.html	6		2	4
Gnus 5.x Frequently Asked Questions	http://www.miranova.com/~steve/gnus-faq.html	81	59	2	20
GIS Frequently Asked Questions and General Info List Index	http://www.census.gov/geo/www/faq-index.html	61		61	
An Anarchist FAQ Webpage	http://www.geocities.com/CapitolHill/1931	29		18	11
"The Proto Jai Maharaj Periodic Informational Posting"	http://panacea.phys.utk.edu/~dbd/minifaqs/jai.maharaj.miniFAQ				
The Caption FAQ	http://www.caption.com/capfaq/	41	4	20	17
Scientology: Just the FAQs	http://www.ncf.carleton.ca/~av282/	37		20	17
Mike Legeros Home Page - Batman Forever Goofs List/FAQ	http://www.n-vision.com/hell/per/bat.html	1			1
"[rec.games.frp.*] Frequently asked questions Part 2"	http://blues.helsinki.fi/~vviitane/rec.games.frp.Faq_Part_2.txt				
Terse Map of the *.Aquaria FAQs	http://www.cco.caltech.edu/~aquaria/Faq/	17		17	
Kempo/Kenpo Frequently Asked Questions	http://www.ultranet.com/~eclipse/kempofaq.html	119	30	43	46
Skate FAQs	http://www.skatefaq.com	43		37	6
Hootie & the Blowfish Suck FAQ	http://home.tampabay.rr.com/badwiring/hootie.htm	5		5	
Virginity FAQ	http://osf1.gmu.edu/~dwojciec/vfaq.html	3		3	
alt.polyamory Frequently Asked Questions (FAQ)	http://www.cis.ohio-state.edu/text/faq/usenet/polyamory/faq/faq.html	17		17	
alt.supermodels FAQ	http://www2.gol.com/users/graham/faq.html	1			1
Necro Fairy Faq	http://www.pitzer.edu/~zkazi/necrofaq.html	4		4	
G-Spot FAq, part 2	http://cctr.umkc.edu/user/cgladish/faq/gspot2-faq.html				
MPEG Moving Picture Expert Group FAQ	http://www.crs4.it/HTML/LUIGI/MPEG/mpegfaq.html	15		13	2
The soc.subculture.bondage-bdsm FAQ list	http://www.unrealities.com/adult/ssbb/faq.htm	62	4	42	16
Bullwhip FAQ	http://www.juggling.org/help/circus-arts/bullwhip.html	68	24	6	38
Roshambo FAQ	http://www.emf.net/~estephen/roshambo/faq.html	17	10	1	6
"Halo's Depeche Mode Page"	http://www.depeche-mode.com/devotedFAQ.txt				
Zoroastrianism and Avesta: Overview and FAQ	http://www.avesta.org/zfaq.html	28		17	11
Erasure FAQ	http://www.taynet.co.uk/users/bennett/music/erasure-faq.html	21	4	17	

Title	URL	Links	Same file	Same host	Different host
APL language FAQ	http://www.lib.ox.ac.uk/internet/news/faq/archive/apl-faq.html	54	1	5	48
Arrow's QBasic Page	http://www.geocities.com/SiliconValley/Heights/8967/faq.htm	43	12	24	7
"TEMPLE OF SET FREQUENTLY-ASKED QUESTIONS (FAQ)"	http://www.dnaco.net/~raensept/t osfaq.htm				
Mercedes Lackey Mailing List FAQ	http://www.herald.co.uk/local_info/lackey_faq.html	7		7	
FAQ: BETA Programming Language	http://www.daimi.aau.dk/~beta/FAQ/	194	163	4	27
"The Nirvana FAQ"	http://www.oswego.edu/~jmcrae/nirvana/faq				
Pearl Jam FAQ	http://www-personal.engin.umich.edu/~galvin/faq.html	72	63	4	5
Answers to Frequently Asked Questions about Poi Dog Pondering	http://reality.sgi.com/employees/relph/poi-pounders/FAQ.html	20	10	6	4
"Frequently Asked Questions about Neopaganism"	http://earthspace.net/esr/faqs/paganism.txt				
Frequently Asked Questions About Religious Science	http://www.mnsinc.com/nsomc/FAQs.html	15		9	6
Kabbalah FAQ	http://www.digital-brilliance.com/kab/faq.htm	27	26	1	
Common Questions	http://www.amfi.org/faqs.htm	24	20	4	
Bad Religion FAQ	http://thebrpage.tierranet.com/front/faq.html	3		2	1
The Jello Biafra FAQ	http://www.paranoia.com/~gallium/dk/jello.faq.html	1		1	
4AD-L FAQ	http://www.wolfson.ox.ac.uk/4ad-l/	39	1	31	7
Sleater-Kinney FAQ	http://www.napanet.net/~lina/sk/faq.htm	12	8	4	
Icon Programming Language FAQ	http://www.cs.arizona.edu/icon/www/faq.html	50	22	28	
Soc.Religion.Christian	http://geneva.rutgers.edu/src/faq.html	135	1	134	
Frequently Asked Questions About Shonen Knife	http://walden.mo.net/~mckenzie/faq.html	32	30	2	
soc.religion.quaker Answers to Frequently Asked Questions	http://www.cis.ohio-state.edu/hypertext/faq/usenet/Quaker-faq/faq.html	41		41	
Smashing Pumpkins FAQ v3.00	http://www.starla.org/faq.html	59			59
Plymouth Brethren FAQ	http://www.storm.ca/~sabigail/faqs/brethren.htm	60	45	1	14
FAQ: THE AMERICAN WAR LIBRARY	http://www.geocities.com/Tokyo/Towers/1955/faqawld.htm	19		11	8
Latter-day Saints: FAQ	http://www.athenet.net/~jlindsay/LDSFAQ/FQ_index.shtml	55		40	15
"Peter Straub - Frequently Asked Questions"	http://www.geocities.com/Athens/1916/index.html	19	14	3	2
CD Club FAQ: Breaking News	http://www.blooberry.com/cdfaq/	11	1	7	3
DC Heroes Role Playing Game and Mailing List Home Page	http://w3.one.net/~joshdm/dch.html	52	8	22	22
KAFFE - Frequently asked questions	http://www.kaffe.org/faq.html	7			7
The Horn Players' FAQ, V 0.54	http://www.io.com/~rboerger/Hornfaq.html	176	69	18	87
Branagh FAQ Menu	http://www.ultranet.com/~luvvy/kcb/kb_menu.htm	19		19	

Title	URL	Links	Same file	Same host	Different host
FAQ (Frequently Asked Questions) about the band The Police	http://www.geocities.com/SunsetStrip/2106/policex.html	84	37	23	24
General Church : Frequently Asked Questions	http://www.newchurch.org/FAQ/faq-aboutnc.html	13		13	
WWW.PHISH.NET - Phish MiniFAQ	http://www.phish.net/faq/minifaq.html	42		39	3
Welcome to the FAQ	http://www.phish.net/PhishFAQ/	20	2	13	5
comp.lang.eiffel Frequently Asked Questions (FAQ)	http://www.cis.ohio-state.edu/hypertext/faq/usenet/eiffel-faq/faq.html	44			44
comp.lang.forth FAQ: General Information (1 of 6)	http://www.cis.ohio-state.edu/hypertext/faq/usenet/computer-lang/forth-faq/part1/faq.html	24			24
Icon Programming Language FAQ	http://www.cis.ohio-state.edu/hypertext/faq/usenet/comp-lang-icon-faq/faq.html	5			5
Rolling Stones FAQ [1/4]	http://www.cis.ohio-state.edu/hypertext/faq/usenet/music/rollingstones-faq/part1/faq.html	20		1	19
Core War Frequently Asked Questions	http://www.koth.org/corewar-faq-.html	72	63		9
alt.support.asthma FAQ: Asthma -- General information	http://www.radix.net/~mwg/asthma-gen.html	57	3	52	2
Frequently Asked Questions about Bill Nelson	http://billnelson.com/other/faq.htm	4		3	1
FAQ for comp.lang.functional	http://www.cs.nott.ac.uk/Department/Staff/gmh/faq.html	130	63	10	57
Rage against the machine unofficial FAQ	http://www.accidentprone.com/ragefaq/ramfaq29.html	315	293		22