

Exploring patterns in website content structure

Svetlana Symonenko

School of Information Studies

Syracuse University

ssymonen@syr.edu

Abstract

A pilot study was conducted for the dissertation research on the indications of conventionalization in the observable structure of website content, i.e. in the way information is displayed to and perceived by users. The pilot applied qualitative content analysis methods, guided by the genre theory, to the sample of the top three structural levels of fifteen websites of three types (universities, governmental, and business). Because of the small sample, the pilot results should be treated as preliminary, but they do point to certain type-dependent patterns in organization of information on websites. In addition, analysis of page titles and link labels identified some naming conventions for particular content categories. The analysis of content structure also appears potentially informative about major “lines of business” of an actual entity, or entity type, behind the website.

Introduction

Since its advent, the Internet has evolved from government-funded research of technologies for interlinking networks to pervasive, though intangible, part of our life. In 2005, over 15% of the global population were online (Internet World Stats); according to MorganStanley estimates, this number exceeded 1 billion (Internet World Stats, ; Meeker, 2005). The volume and range of Web-based activities is overwhelming. Yet, this enormous diversity has a common ground in pursuit and exchange of information, which can be quite a frustrating process: it is estimated that only 21% of users find the desired information more than 80% of the time (Feldman, 2004). Low productivity of *browsing* the Web and websites, in particular, contributed to making *search* the most common method of finding information on the Internet. Keyword search works fairly well when the target information is clearly defined, but what if it is not? What if someone is interested in employment opportunities in a particular sector and would like to review and compare information from different companies all at once? Specialized portals offer seemingly similar services, but there the user is not an active searcher, but rather a passive consumer of what the information provider chooses to offer. So, has the Web reached the state when it can empower a more proactive information seeking mode?

In the physical environment, ease of use is often related to predictability and intuition: it is when, finding ourselves in a new area, we can figure out our way around by utilizing our knowledge and past experiences, with little or no extra help. That implies the existence of certain patterns in the real world, which we, as users, have taken notice of and are taking advantage of. How well do ideas of predictable and intuitive design apply to organization of information on the Web? This is what the dissertation research, the pilot for which is presented here, is set to determine.

Study background and related work

This paper describes a pilot study for ongoing dissertation research on the indications of conventionalization in the observable structure of website content, i.e. in the way information is displayed to and perceived by users as they navigate and explore the site. For the purposes of this study, a website is defined as an integrated information package created by the site sponsor with a particular audience in mind (for example, a university website is geared toward prospective and current students, faculty, alumni, potential sponsors with interest in research, etc.). Collectively, the sponsor and the audience represent the site’s discourse community, which, according to genre theory (Orlikowski and Yates, 1994; Swales, 1990), shapes the site as a communication artifact, in terms of the kind of content presented there, as well as the way it is presented.

The agenda for this research evolved from an initial informal review of a large set of websites, which suggested developing patterns in organizing information on a site for user access, depending on the type of site sponsor (e.g. university vs. governmental agency). This trend is also apparent in emergent conventions for presenting to the user the structure of information on a website: by means of site maps or directories (Gillenson et al., 2000); with the help of topological or semantic clues (Dill et al., 2002; Nielsen, 2004). In addition, there is empirical evidence indicative of the existence of certain expectations of websites’ content structure, which users develop, based on their experience in online searching and browsing (Nielsen, 2000, 2005; Vaughan and Dillon, 2005).

Until now, research has mainly focused on interface aspects of Web-based information (*Research-Based Web Design & Usability Guidelines*, 2002; Ivory and Medgraw, 2005; Morville and Rosenfeld, 2002) or on its genre-related aspects at the page level (Bates and Lu, 1997; Dillon and Gushrowski, 2000; Haas and Grams, 2000; Ihlström and Ekesson, 2004; Rehm, 2002), with insufficient attention to patterns in visible structure of information within the scope of an entire website. To fill this gap, the dissertation will explore regularities in websites' content structure and in users' expectations of such structure.

The study presented here piloted one dissertation phase, which will assess the regularities in the structure of information on websites of two types by means of qualitative content analysis of elements of website content structure, such as page title and hyperlinks. A website's type is derived from the type of its sponsor (Shneiderman, 1997). To control for cultural differences, the scope of the study is limited to sites sponsored by U.S.-based entities.

Data Collection

A sample for the pilot comprised fifteen websites of three types: *universities* (EDU), drawn from a list of U.S. colleges (U.S.News); *business* (COM), sampled from a list of telecommunications companies (*OneSource*); and *government* (GOV), selected from a list of state portals (*State and Territorial Governments*).

Elements of website content structure (*Page Title, Link URL, Link Label, and Level*) were extracted using a Perl script. The script traversed the site in a breadth-first manner; so if a page was linked to from multiple points within the site, each link was collected and assigned the respective Level number, but the page itself was assigned the earliest possible level and stripped only once. For example, on EDU sites, the link to *Maps* was encountered on a page with resources about the university (*About, Level_2*) and on a page with information for visitors (*Visit, Level_3*). The script collected two *Maps* links (one with Level_2 and the other one with Level_3 attributes), but stripped the *Maps* page only once at Level_2.

Data collection restrictions were applied to keep the dataset within the limits manageable for a pilot. First, only the first three levels of a website were stripped. This limit was derived from the literature (Brinck et al., 2001; Larson and Czerwinski, 1998; Morville and Rosenfeld, 2002) that recommends a 3-5 level hierarchy as the maximum depth, to which a user is willing to "trace" a link. Analysis of the pilot data showed some compliance of sites' architectures with these guidelines: three levels usually comprised the major content of COM sites; the 3rd level also seemed to be a threshold, at which EDU and GOV sites began to lose their institutional uniformity and break out into sub-sites of individual departments.

A second restriction limited pages visited to only those within the 2nd-level domain. Links on a webpage, by destination, fell under 3 major categories: staying within the original 2nd-level domain of a homepage (e.g.: <http://www.harvard.edu/academics/catalogs.html>); going to the 3rd-level of the homepage domain (e.g.: <http://atwork.harvard.edu/>); and pointing to an *external domain*, a website of an unrelated entity (e.g.: <http://www.flymanchester.com/>). Analysis of links in the pilot sample showed that COM sites were more centralized, with the majority of links staying within the 2nd-level domain, as opposed to EDU and GOV sites, which resembled portals to organizations within the main entity and, thus, had a significant portion of links to 3rd-level domains (Table 1). Therefore, this restriction did not greatly affect the content extracted from COM sites, but caused some content "gaps" for EDU and GOV sites. For the full study, the script will be modified to collect data from 3rd-level domain pages.

Level	Level 1 (Homepage)			Level 2			Level 3		
	2 nd -l. Domain	3 rd -l. Domain	External Domain	2 nd -l. Domain	3 rd -l. Domain	External Domain	2 nd -l. Domain	3 rd -l. Domain	External Domain
COM	111	1	3	1506	28	254	2098	45	268
EDU	72	59	4	1358	824	111	6857	2865	597
GOV	116	130	32	2149	1508	539	9194	3457	1344

Table 1. Distribution of link types in the pilot data set, by type of site.

Finally, links to pages with dynamically generated content were excluded. Review of the pilot data showed that, on some websites, such links pointed to important content, e.g. entire site sections, not just news items or event calendar. For the full study, decisions on whether to follow such links, will be made on a case-by-case basis. Table 2 shows the resulting dataset for the pilot.

	Sites	Lev_1 links	Lev_2 pages	Lev_2 links	Lev_3 pages	Lev_3 links	Total
COM	5	122	58	1926	79	2489	4552
EDU	5	165	65	2591	297	11864	14817
GOV	5	288	90	5106	416	14777	20389

Table 2. Pilot study collection.

Data Analysis

Page titles and link labels from the sample were coded with the help of the Atlas.ti content analysis software. The coding followed the analytic induction approach, wherein content categories (codes) emerge from the data, i.e. are inferred by the coder (Miles and Huberman, 1994). Data analysis focused on type-dependent distribution of content categories at different levels within the site, for particular kinds of links and pages, as well as on naming conventions for particular categories. To assess category distribution, conventionalization degrees (Nielsen, 2004) were extended in such a way that a content category was considered *standard*, if it was present on over 80% of websites; *conventional*, if it appeared on 50-79% of websites; and *unconventional*, if it was employed by under 49% of websites.

Given the small sample size and data collection restrictions, the pilot results should be treated as preliminary. Nevertheless, they do point to certain patterns in the way information is organized for the user on websites of different types.

Global navigational links

For example, global navigational links, commonly displayed as a bar on the top or on the side of a webpage, not only support user's orientation within the website, but also represent the top levels of a site's information organization scheme and signify the most important content in the "mind" of the site sponsor. In the pilot data, navigational links constituted, on average, 1/5 of all links, with almost no difference between EDU (16%) and COM (15%) sites and a slightly higher share for GOV (23%) sites. As Table 3 shows, a link to a *homepage* is, unsurprisingly, the most common navigational element, regardless of the site type. But some categories do appear to have type-specific importance. In the GOV context, standard navigational links included resources for those visiting the state (*visit*) or conducting business (*business_resources*) there; whereas on EDU sites, navigational links typically pointed to information about the university (*about_entity*), resources of interest to perspective students (*admissions*) or past ones (*alumni*), as well as to potential employees (*employment_info*); and on COM sites, information *about_entity* or ways to *contact* unanimously received navigational importance. Less common links also seemed to possess distinct site type "flavor": a password-protected section with resources for *customer_only* was likely to be indicated by a global navigational link for a COM site, but not for the other two types. In turn, information about *schools* or *community* programs tended to be universally accessible on EDU sites; and the same was true for *e-government services* and *resources for state residents* on GOV sites.

	EDU	GOV	COM
standard	homepage	homepage	homepage
	about_entity	business_resources	about_entity
	admissions	directory	contact
	alumni	visit	
	employment_info		
	news		
conventional	community	about_gov	customer_only
	contact	e_gov_services	privacy
	copyright	employment_info	vacancy_any
	schools	policy	
	search	resident_resources	
		site_map	

Table 3. Navigational Links: content categories by site type and level of conventionalization.

Content structure at a particular level

Review of content categories accessible from Level_2 (one click from a homepage) also suggested the presence of type-specific patterns. One striking difference is in the variety of available information: compared to COM sites, EDU and GOV sites are more loaded with diverse content (COM sites featured

36 categories, EDU — 136, and GOV — 156). For space reasons, Table 4 shows only selected standard content categories. The table includes both navigational and ad hoc (i.e. non-navigational hyperlinks found on a page) links.

Even with such variety, some categories of information were typically found at Level_2 *only* on EDU sites (e.g.: *alumni*, *admissions*, *research*), *only* on GOV sites (e.g.: *about_state*, *business_resources*, *driver_services*, *e_gov_services*), or *only* on COM sites (e.g.: *partners*, *products_services*). On the other hand, some categories (*home*, *about*, *contact*) appeared to be “universal”, being present on every site regardless of its type. It is hardly coincidental that the latter categories are among the ones, which, for a number of years, have been strongly recommended by information architecture guidelines (*Research-Based Web Design & Usability Guidelines*, 2002; Morville and Rosenfeld, 2002; Nielsen, 2000).

GOV	EDU	COM
homepage	homepage	homepage
about_gov	about_org	about_org
contact	contact	contact
about_state	academics	business_partners
accessibility	admissions	products_services
area	alumni	
arts_culture	area	
attractions	arts_culture	
business_resources	community	
calendar	continued_education	
career_services	copyright	
community	course_info	
directory	department_academic	
driver_services	directory	
e_gov_services	employment_info	

Table 4. Selected standard content categories for Level_2, by site type.

Content structure of pages of selected categories

Content analysis of links on pages which are topically similar, but located on sites of different types (that is, produced within different discourse communities), revealed some audience-specific patterns as well. For example, *About_entity* pages at EDU sites tended to link to general *academic* information, resources for *alumni*, information about *community*-related programs, as well as resources for visitors (*visit*). In COM context, *about_entity* pages were more likely to feature information about *products_services* and about *partners* (Table 5).

EDU	COM
about_entity	about_entity
academics	contact
alumni	homepage
community	business_partners
employment_info	products_services
history	
homepage	
news	
visit	

Table 5 Standard content categories for *About_entity* page: EDU vs. COM sites.

Comparison of *Visit* pages in EDU vs. GOV context (Table 6) suggested that typical content of such a page on EDU site would include information about the university (*about_entity*), *news*, *sports*, *maps*; *employment*-related information; as well as content of interest to *alumni* and area residents (*community*). A *Visit* page of a GOV site would, more likely, feature information about the state: cultural life (*arts_culture*) and *recreation*; resources and services for businesses and individuals residing (*business_* and *resident_resources*) in the state, as well as for visitors (*visit*); and information to encourage and help people or businesses to *relocate* to the state.

EDU	GOV
about_entity	about_state
alumni	arts_culture

EDU	GOV
community	attractions
employment_info	business_resources
events	directory
homepage	e_services
maps	gov_state
news	homepage
sports	recreation
visit	resident_resources
	visit

Table 6. Standard content categories for Visit page: EDU vs. GOV sites

Creating a typical profile of a website information structure

The analysis of distribution content categories can be summarized and translated into a profile (observable structure) of information for each site type. Figure 1 shows an excerpt from a sample profile for an EDU site. Due to space constraints, only selected sections are presented: global navigational links (*Homepage* shows their detailed breakdown; at Level 2 and Level 3 they are indicated by the *Global_Navigation* boxes) and pages *About* and *Visit*. As this profile shows, the *About* page was typically observed to be just one click from the *Homepage*, accessed via a global navigational link; it also typically contained an ad hoc link to the *Visit* page. From this profile, one could also infer that, for example, information *about* university or about *employment* was commonly accessible from any point within the top three levels of the sites. Or that the link to resources for those *visiting* the university was typically found on *About* page, one click from the *Homepage*; and the *Maps* page was likely to be linked to from the *About* or from the *Visit* page. A few points seem worth emphasizing. First, this is a structure of website content as it reveals itself to the user browsing the site from the homepage downwards. Second, this is an abstraction of a few (five, in this case) concrete structures, that captures only the most typical components of website content at their most typical level.

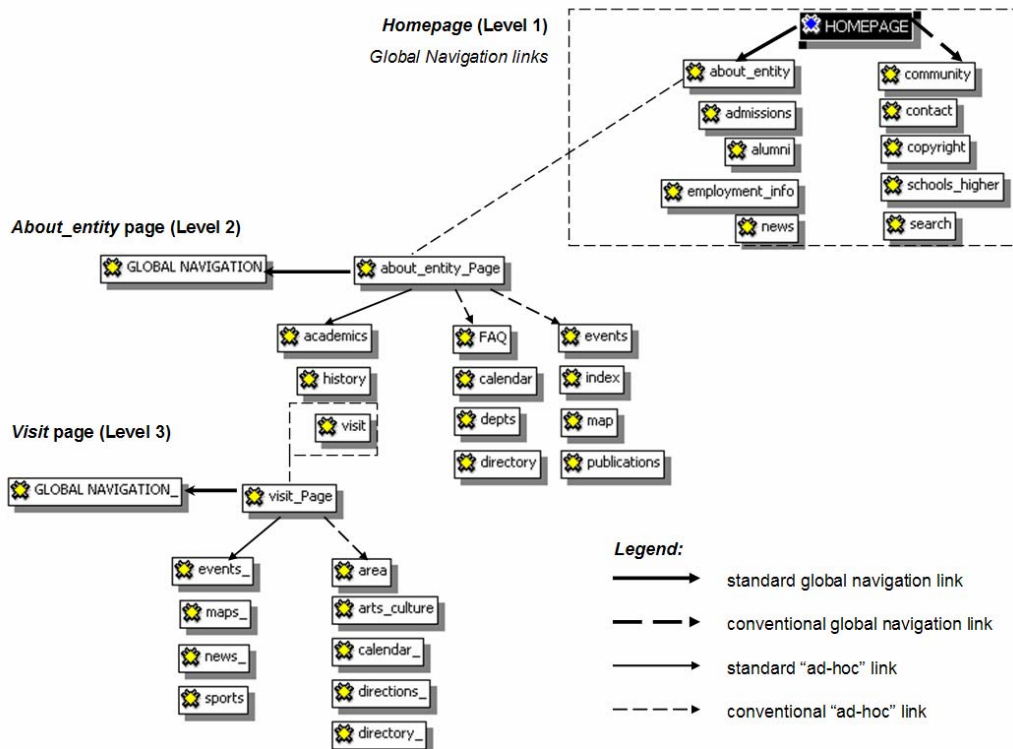


Figure 1. A sample profile of EDU site.

This profile reflects the actual, observed structure of information on a website. But do people possess certain shared perceptions or expectations of such structure? And if so, how do these expectations match the actual state of things? A good match would make the structures more predictable and, thus, easier for the user to navigate. But what happens when they do not match; how can this affect users' information seeking experience? These are the questions that the second phase of the dissertation research will seek to answer by means of the qualitative user study piloted in the summer of 2004 (Symonenko, 2005).

Naming conventions

Analysis of page titles and link labels identified certain naming conventions for some content categories. For example, Table 7 shows all labeling variants for information *about* organizations encountered at EDU and COM sites, with their respective frequencies. Here, a prevalent pattern is quite obvious: ‘About’ or ‘About (Us, [Entity_Name])’.

EDU									
<i>Caltech</i>		<i>Harvard</i>		<i>Stanford</i>		<i>WUSTL</i>		<i>Yale</i>	
About Caltech	50	About Harvard	36	About Stanford	55	About the University	6	About Yale	73
At a Glance	3	Harvard Guide	5	Stanford Facts 2005	5	About WUSTL	3	Facts About Yale	8
Caltech 101 (.pdf format)	2	About Harvard University	1	Who We Are	5	Facts Brochure	2	About	5
Annual Report (.pdf format)	1	General information	1	Back to Stanford Facts 2005	4	Facts Guide	1	Facts & Stats	4
Honors & Awards	1	Guide to Harvard University acronyms	1	Stanford Facts	4			Images of Yale	1
Nobel & Crafoord Laureates	1	Harvard Fact Book	1	Introduction to Stanford	3			Reaccreditation at Yale	1
		Quick Facts and Figures	1	An Introduction To Stanford	2			Yale and the World	1
				Landmarks	2				
				Approaching Stanford	1				
				Campus Sights	1				
				lingo, catch-phrases and acronyms	1				
								
.COM									
<i>Connectus</i>		<i>Arvig</i>		<i>Connectme</i>		<i>Aptela</i>		<i>Datapath</i>	
ABOUT US	6	About Us	63	About	26	Company	30	About DataPath	3
				More Info	2			About Us	1
								CEOs Letter	1

Table 7. About_entity lexicon by site type

Similar review of labels indicating *employment opportunities* (Table 8) identified such word patterns as ‘Employment (Career, Job) Opportunities’ or simply ‘Jobs’, ‘Careers’.

EDU									
<i>caltech</i>		<i>stanford</i>		<i>wustl</i>		<i>yale</i>			
	13		19		21		23		
Employment Opportunities, academic	7	Employment Opportunities	2	Job Opportunities	3	Employment Opportunities	9		
Employment Opportunities, staff	6	Employment	1	Employment	1	Internal Applicant	1		
		Faculty Openings	1	Job Opportunities - Washington	1	Job Opportunities	1		
		Faculty Recruitment Office	1			Job Posting	1		
		Open Faculty Positions	1						
		faculty positions	3						
COM									
<i>connectus</i>		<i>arvig</i>		<i>datapath</i>					
	1		2		7				
CAREER OPPORTUNITIES	6	Jobs@ACS	63	Careers	2				
		Current job listings	1	Satellite Communications Specialists	2				
				DataPath Careers Asia Openings	1				
				DataPath Careers Europe Openings	1				
				DataPath Careers US Openings	1				
				DataPath Careers	1				
				Field Engineers	1				
				Satellite Communications Specialists	1				
				Sales Engineer	1				
				Technical Trainer	1				
GOV									
<i>hawaii</i>		<i>delaware</i>		<i>florida</i>		<i>alabama</i>		<i>alaska</i>	
	1		2		3		4		5
State of Hawai i Job Search	2	State Job Postings	3	Find a State Job	7	Alabama JobLink	7	Jobs	21
America s Career Infonet	1	Job Hotline	2	Jobs	1	Alabama s Job Bank	1		
America s Job Bank	1	Postings	2	Joining the Bush Team	1	America s Job Bank	1		
Federal Job Listings	1	- Apply for a State Job	1	Working in Florida	1	Jobs with the State of Alabama	1		
Hawai i s Job Bank	1	24-hour State Job Line	1			Virtual One Stop Job Search	1		

Table 8. Job_Opportunities lexicon by site type

Again, comparing these labeling patterns to what people tend to call these content categories will show how well the actual and the expected labels match, which, at least in part, indicates how intuitive (predictable) the structure of the site content is.

Type-dependent features in overall website architecture

In addition to type-related patterns in organization of information site-wide, site types also showed differences in their overall architecture, consistent with what was observed by Amitay et al. (2003). EDU and GOV strongly resembled portals to sites of individual organizations (schools, state or county agencies, etc.) within the main entity. For example, the textual component of many Level_2 pages on *delaware.gov* and *florida.gov* consisted almost exclusively of links to various state and county agencies or state resources (Figure 2).

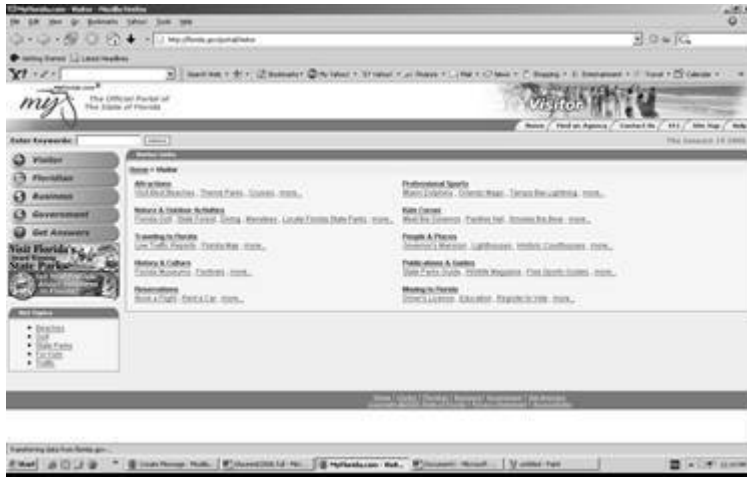


Figure 2. A sample page from Florida.gov

Some sites or sub-sites of individual departments, linked to from a state portal, exhibited not only distinctive visual appearance, but also differences in elements of the content structure: they would use name variants for links to global content or use local navigational links. On such sites, institutional identity would be typically maintained on the top 2-3 levels, but further down, would break into a kaleidoscope of divisions. COM sites demonstrated a much more uniform look and product-focused architecture, with prominent product- or customer-oriented sections and very few links leading off the site. They also seemed to prefer shallow site structure of at most 3-4 levels deep.

Also, compared to COM sites, EDU and GOV sites appeared much more verbose, in terms of the number of pages and links comprising an average site. This observation, made on the pilot data, was confirmed on a larger data set, stripped from 123 EDU and 73 COM sites, including 3rd-level domain pages: 89% of COM sites did not reach 0.5 MB in size (Figure 3), compared to only 7% of EDU sites; on the other hand, most (73%) of EDU sites were of sizes between 0.5 and 3 MB (Figure 4).

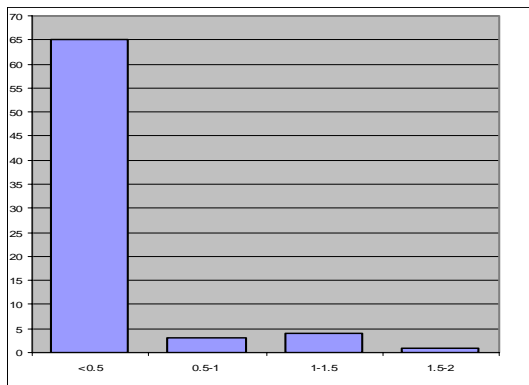


Figure 3. 73 COM sites, by size, MB

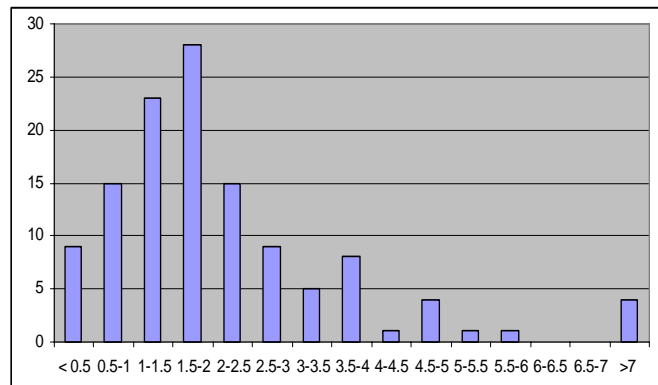


Figure 4. 124 EDU sites, by size, MB

Although the analysis focused on the visible structure of website content, the structure appeared to be quite revealing about the actual entity, or entity type, behind the site. Analysis of standard content categories for different site types could indicate their major “lines of business”, such as: education and research for EDU; particular products or services for COM; representation of three branches of government, services for businesses and individuals, promotion of the state image for GOV, and so on.

Conclusions and future research directions

In addition to testing the methodology and tools for data collection and analysis for the larger study, the pilot study yielded results suggestive of the presence of patterns in the organization of information on a website depending on the site type. Building on the pilot results, the full-fledged dissertation study will focus on the content structure of a larger sample of EDU and COM sites. The outcome of the study will be type-specific profiles of website content structure and lexicons of naming conventions for the most common content categories.

Given the exploratory nature of the proposed research, the most immediate contribution of the study will be in enhancing the current state of knowledge about organization of information on the Web, as well as in setting a research agenda and testing the methodology for further investigation of websites’ content structure, for example, across country or language barriers. The study will also provide empirical

evidence of consistent patterns in the content organization of sites; as well as of users' expectations of such organization, both conditioned on site type. The pragmatic contribution of the study will be in facilitating the online search and browsing processes. The acquired knowledge of persistent similarities in organization of website content and, in particular, of factors contributing to such similarities, as well as of structural features, through which such similarities reveal themselves, can be applied to organize Web-based content in a more intuitive (and, thus, *findable*) way to the user; or to design a tool to assist the user's exploration of the Web.

References

- Amitay, E., Carmel, D., Darlow, A., Lempel, R., and Soffer, A. 2003. The Connectivity Sonar: Detecting Site Functionality by Structural Patterns. *Journal of Digital Information*, 4(3).
- ATLAS.ti. *ATLAS.ti* <http://www.atlasti.com>.
- Bates, M. and Lu, S. 1997. An Exploratory Profile of Personal Home Pages: Content, Design, Metaphors. *ONLINE & CDROM Review*, 21(6): 331-340.
- Brinck, T., Gergle, D., and Wood, S. D. *Usability for the Web: Designing Web Sites that Work* (2001). Morgan Kaufmann.
- Dill, S., Kumar, R., McCurley, K., Rajagopalan, S., Sivakumar, D., and Tomkins, A. 2002. Self-Similarity in the Web. *ACM Transactions on Internet Technology*, 2(3): 205-223.
- Dillon, A. and Gushrowski, B. 2000. Genres and the Web: Is the personal home page the first uniquely digital genre? *Journal of the American Society for Information Science*, 51(12): 202-205.
- Feldman, S. 2004. The high cost of not finding information. *KMWorld*, 13(3).
- Gillenson, M., Sherrell, D., and Chen, L. 2000. A Taxonomy of Web Site Traversal Patterns and Structures. *Communications of AIS*, 3(17): 2-27.
- Haas, S. W. and Grams, E. S. 2000. Readers, authors, and page structure: A discussion of four questions arising from a content analysis of Web pages. *Journal of the American Society for Information Science*, 51(2): 181-192.
- Ihlström, C. and Ekesson, M. 2004. *Genre Characteristics — A Front Page Analysis of 85 Swedish Online Newspapers*. 37th Annual Hawaii International Conference on System Sciences (HICSS'04).
- Internet World Stats. <http://www.internetworldstats.com/stats.htm>.
- Ivory, M. Y. and Medgaw, R. 2005. Evolution of web site design patterns. *ACM Transactions on Information Systems (TOIS)*, 23(4): 463-497.
- Larson, K. and Czerwinski, M. 1998. *Web Page Design: Implications of Memory, Structure and Scent for Information Retrieval*. Proceedings of the SIGCHI conference on Human factors in computing systems, Los Angeles, CA.
- Meeker, M. 2005. *Global Technology / Internet Trends*: MorganStanley <http://www.morganstanley.com/institutional/techresearch/gsb112005.html>.
- Miles, M. B. and Huberman, A. M. *Qualitative Data Analysis* (1994).(2nd ed.): Sage Publications.
- Morville, P. and Rosenfeld, L. *Information Architecture for the World Wide Web: Designing Large-scale Web Sites* (2002).(2nd ed.): O'Reilly.
- Nielsen, J. *Designing Web Usability: The Practice of Simplicity* (2000).(1st ed.): New Riders Press.
- Nielsen, J. 2004. The Need for Web Design Standards. *Jakob Nielsen's Alertbox*.
- Nielsen, J. 2005. Mental Models For Search Are Getting Firmer. *Jakob Nielsen's Alertbox*. *OneSource*. <http://www.onesource.com/>
- Orlikowski, W. and Yates, J. 1994. Genre repertoire: The structuring of Communicative Practices in Organizations. *Administrative Science Quarterly*, 33: 541-574.
- Rehm, G. 2002. *Towards automatic web genre identification -- a corpus-based approach in the domain of academia by example of the academic's personal homepage*. HICSS'02, Hawaii.
- Research-Based Web Design & Usability Guidelines*. (2002) <http://www.usability.gov/pdfs/>
- Shneiderman, B. 1997. Designing information-abundant web sites: issues and recommendations. *International Journal of Human-Computer Studies*, 47: 5-29.
- State and Territorial Governments*. http://first.gov/Agencies/State_and_Territories.shtml
- Swales, J. M. *Genre Analysis : English in Academic and Research Settings* (1990). Cambridge University Press.
- Symonenko, S. 2005. *No Stranger in the Strange Land: Users' Expectations of Website Content Organization*. IA Summit 2005, Montreal, Canada.
- U.S.News America's Best Colleges 2005. http://www.usnews.com/usnews/edu/college/rankings/ranknatudoc_brief.php
- Vaughan, M. and Dillon, A. 2005. Why structure and genre matter for users of digital information: A longitudinal experiment with readers of a web-based newspaper. *International Journal of Human-Computer Studies*, In Press.