https://doi.org/10.48009/2_iis_2007_4-19

DIVERSITY AND HOMOGENEITY IN THE CYBERSPACE: A STUDY OF WEBSITE CONTENT PREFERENCES OF SELECTED ETHNO-LINGUISTIC GROUPS

Bruce W.N. Lo, University of Wisconsin-Eau Claire, lobw@uwec.edu
Adriana Martinez Santa Cruz, University of Wisconsin-Eau Claire, martinea@uwec.edu

ABSTRACT

The Internet has been hailed as the ideal medium to facilitate expressions of cultural diversity and individualism in cyberspace. But the dominance of a few Web giants also exerts a homogenizing influence over users and audience in this space. To assess the relative influence of these two opposing forces, we examines the top ranking websites of several ethno-linguistic groups to determine whether there are any discernable differences in content preferences.

Keywords: Diversity in cyberspace, website content, top ranking websites, ethno-linguistic groups, culture.

INTRODUCTION

The Internet has revolutionized the ways we communicate and interact with each other, not only just to those who are in our geographic neighborhood but also to those who are far away from us across the globe. Thus as a communication medium, its sphere of influence far exceeds any we knew before. An intriguing question that has baffled both technologists and social scientists is, "Does the Internet empowers different ethno-linguistic groups or subgroups, to assert their individual cultural variance more effectively or does it, as a result of the dominance of selected transnational Web giants, exert an homogenizing influence that diminishes ethno-cultural uniqueness and individualism?".

The importance of this question is of concern not only to social scientists engaging in the study of the impact of the Web on cultural diversity or homogeneity in cyberspace, but also to many practitioners - for example, Web professionals who must decide to localize or standardize in the design of their global E-business websites [6,7]. This question is also important to writers and authors from different parts of the world, engaging in the collaborative development of scholarly works [8]; and to multinational business executives who used the Web to mediate their corporate decisions in a GDSS environment [14].

During the past few years the authors have been tracking the rise and falls of the top Internet sites in several ranking lists [1]. In this study, we examine the variation (or the lack thereof) in site contents among the most frequently visited websites (called the Top 100 Sites) for different language groups. If we are able to observe discernable differences in content preference among different language-based top website lists, it may lend credence to the diversity influence of the Internet on cultural heterogeneity. Otherwise, one may not be able to ignore the effect of the homogenizing influence of the Internet.

Previous Related Studies

Previous studies on the cultural impacts on website design tended to focus on format, appearance, or presentation, and many of these studies did find evidence of significant design differences [2, 3, 4, & 5]. The relatively few that claimed to investigate the content element of the website were concerned with indirect aspects of site content and not the central *content* elements. We propose to examine the phenomenon differently and focus specifically on *site content* per se.

According to Huizingh [10], *content* refers to the information, features, or services that are offered by the website. But in the past, researchers on cultural impacts on content characteristics had conducted their investigations from many different view points. Zhao et al. [5] measured contents in terms of degrees of personalization, organizational achievements, and organizational history; Cyr et al. [7], Wurtz [4], and Singh et al. [11] focused on site content features rather than site services; while Fletcher [12] examined content from a qualitative point of view.

In this study, we want to return to the more basic definition of website *content* – consider "content" in terms of the *services* that the site offers, that is, what are the main *functions* of the website? The question we now face is how do we measure the cultural impact in terms of this definition of site content?

Time Variation of Website Ranks

To search for an answer to this question, we began tracking the list of top ranking websites over a 12-month period starting from April 2006 to April 2007.

The top 100 website list was based on Alexa.com, which is probably one of the more reliable ranking providers [1]. Figure 1 shows the rise and fall of the top 100 websites world wide. Only the highest 30 are shown to avoid cluttering up the graph. Please note that the lower the rank number the closer to the top is the associated websites. Although the graph still looks rather cluttered (with apology), two trends are clear. Firstly, the positions of the top 3 websites (Yahoo, Google, & MSN) remain rather constant over the 12-month period (There was a change over for Google and MSN between June and July of 2006). Secondly, there were some rather spectacular rises among the rank. In particular, Myspace and Youtube moved into the top 5 positions (After the Google took over of Youtube). This becomes clearer if we look at Figure 2 where the time variation of the top 5 sites was traced. In this figure the rise of the two social networking sites becomes rather obvious.

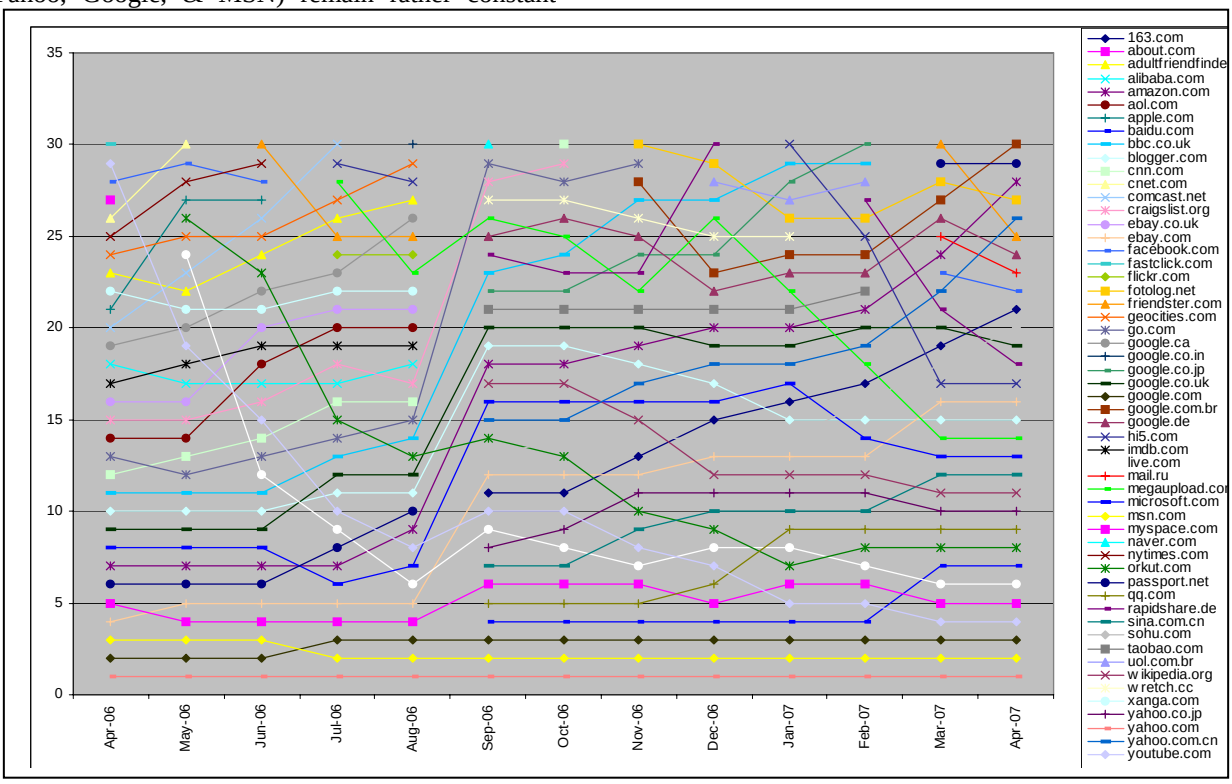


Figure 1. Rise and Fall of the Top 30 Websites World Wide

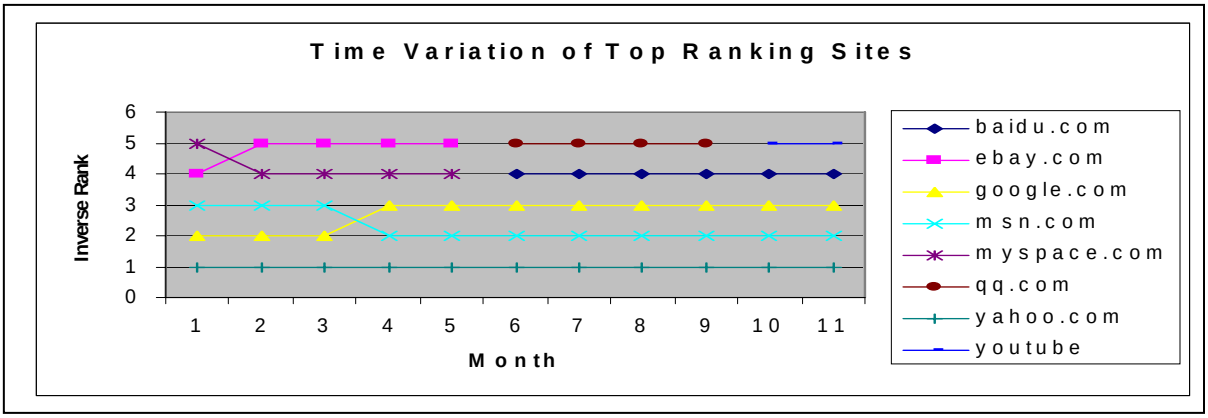


Figure 2: Time Variation of the Top 5 Ranking Sites

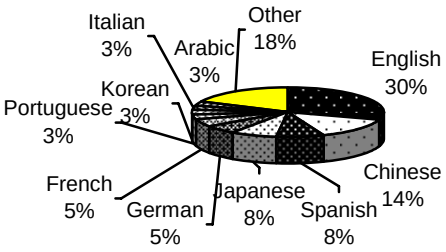
This observation lead us to the idea that rather than comparing the content or services offered by a single website from a given language group with one from another language group, we examine the variation among the top ranking websites from different language groups. In another word, we compare

ranking lists from different language groups, instead of comparing individual sites. The observed rank differences in the different site services will be regarded as an indicator of the different content preferences of the different ethno-linguistic groups. Therefore content preferences are measured in terms of differences in the distribution of the ranks associated with each category of the website.

RESEARCH METHODOLOGY

What language groups should we consider? According to the most recent World Internet Statistics [9] the top 10 languages in the Internet are shown in Figure 3 below.

Figure 3: Top 10 Internet Languages



As expected the most popular language on the Net is English at a share of 30%, followed by Chinese at 14%. Then there are two groups: at 8%, Spanish and Japanese; at 5%, German and French; and at 3% Portuguese, Korean, Italian, and Arabic. Compared with 2005 [1], we found that the Spanish language group has moved up in rank surpassing Japanese, and the Arabic group has moved into the top 10, surpassing Russian.

In view of this trend and the language capability available to the researchers, we decided to collect data for the top six language groups: At the time of writing, 4 groups: English, Chinese, Spanish, and French were available for analysis.

The main research hypothesis to be tested is:

H: The website contents of the top ranking sites for the different language groups do not exhibit any significant difference.

Once the issues of the targeted population and research hypothesis have been settled, the next methodological question relates to what does one mean by *variation* in "website contents". As mentioned earlier, an examination of the literature reveals that this is not a simple question. In this study we use "website services" as the key variable to measure website contents. In order to generate a measurable metrics, we examine the top ranking website lists from different ethno-linguistic groups to determine the ranks that are associated with the different website services.

Our next step is to classify each site on the top ranking list into different content categories. The categories used in this analysis are show in Table 1.

Table 1. Content Categories for Top Websites

Category	Explanation	Examples
Buy & Sell / Shopping	Facilitate buy and sell of goods and services over the Net, including any b2b, b2c or c2c business transactions.	eBay, amazon.com, Staples
Knowledge	Provide knowledge content, scholarly articles, reference materials, education services. Also include dictionaries and encyclopedia	ProQuest, Wikipedia, university websites
Entertainment	Focus on the provision of activities which give pleasure, recreation, or relaxation to the audience, who may actively or passively participate	Game sites for adults and kids, movie, music, video
News/media	News and media sites	CNN, Fox
Online service	Delivers specialized applications online to the user's desktop.	Online storage, Mapquest.
Portal	Primarily intended as a gateway to other website. Although this does not exclude the site to provide some service by itself.	MSN, Yahoo, Baidu
Search engine	Facilitate users to search for information on the Web.	Google, Dogpile
Social networking	Facilitate individuals or groups to communicate and interact with each other.	Myspace, youtube, eHarmony,
Corporation	Focus of the promotion of the image of a business company or non-profit organization.	Toyota, Coca Cola

It should be noted that these 9 content categories are specifically developed to classify the websites in the top ranking lists. They are not intended to be all inclusive of all websites out there. We are aware that many ranking providers also give site categories. But their classifications were found to be not appropriate in the current study, as we require the classification categories to be disjoint.

The rank associated with each site will be taken as an indication of the site visitors' preference of the site

content. A higher rank means that the particular group of site visitors attribute a higher preference to the types of services (or site content) provided by that site. Any difference in average rank and rank distribution indicates a difference in site content preferences.

Following our earlier approach [1], we investigate variations among the top ranking lists by firstly comparing the list membership composition, that is, commonality of the membership, and secondly, by comparing the differences in rank distribution. With this in mind the main hypothesis may be replaced by two sub-hypotheses:

H_1: The membership compositions of the different content categories for the different language groups show no significant difference.

H_2: The mean ranks in the different content categories for the different language groups show no significant difference.

RESULTS

At the time of the writing of this draft not all data have been analyzed. Therefore we shall provide a preliminary report of the results thus far to illustrate what may be expected from this research. We anticipate that the full report will be available in about 4 week's time.

The results below represent the comparison between the top 100 English language websites with the top 100 Spanish language websites. It is important to note that website languages do not necessarily correspond to specific geographical regions. For example, while many of top English language websites may come from the US, some of them are from Europe (e.g. UK), or even Asia (e.g. India). In the same manner, not all Spanish websites are based in Spain. In fact, many of them are from Latin America.

To address hypothesis H_1, we compare the four language groups by comparing the number of websites in each of the content categories (Thus the comparison is on membership composition). This is shown in Figure 4, where we use a polar plot to illustrate this comparison. The difference in content preferences of the different language groups is clearly evidenced. The polar plot shows three distinct spikes. The English language group shows a clear preference for "social networking" sites, the Spanish language group shows a clear preference for "search" sites, while the Chinese language group shows a the most pronounce preference for "portal" sites. Although the French language group shows a small spike for the "portal" sites, the distribution of content preference is more even.

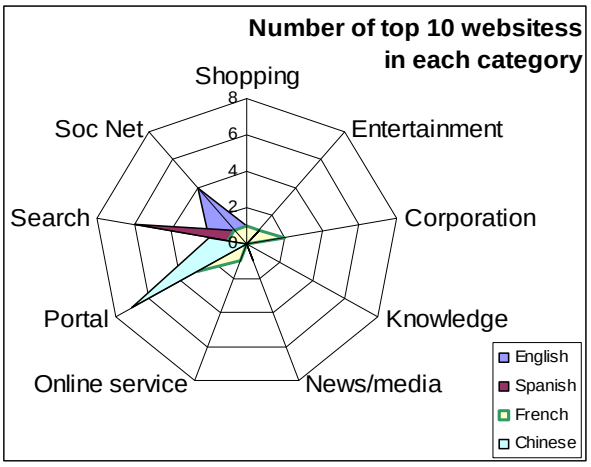


Figure 4. Polar plot showing number of top 10 sites in each category (Compare membership composition)

To ensure the observed differences in content preferences are not accidental, we also did similar polar plots for the top 20, 30, 50, 70, and 100 websites comparing the four language groups. Due to space limitation, only the case of the top 100 sites is presented here in Figure 5.

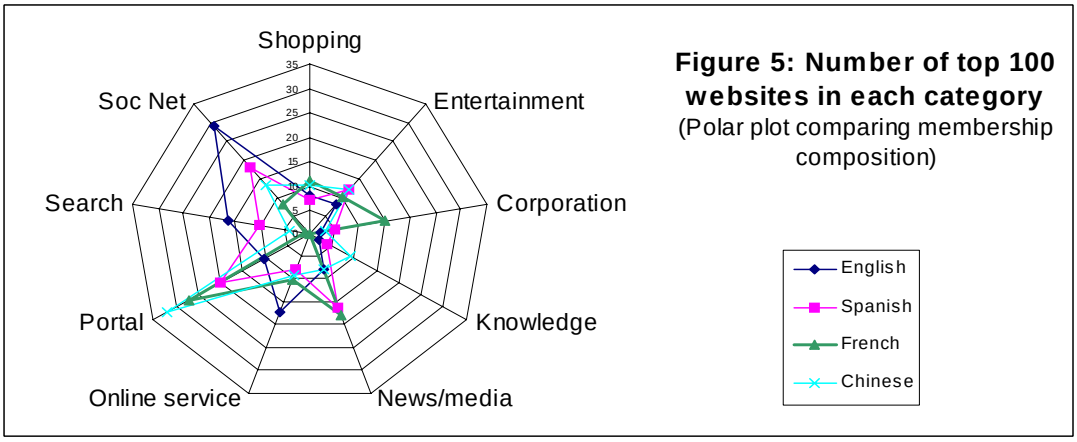


Figure 5: Number of top 100 websites in each category
(Polar plot comparing membership composition)

To show the different "spikes" for the different language groups, we use the "line" polar plot rather than "solid colors". With the large number of websites (100 as compared to 10) in the list, it is expected that the preferences will be more spread out. Again, the English language group shows a clear preference for "Social networking" websites, while the Chinese language group shows a clear preference for the "portal" sites. However, the Spanish language group's preferences are a lot more spread out, with smaller spikes in "portal", "social networking", and "new/media" sites. And the French language group also shows a strong preference for the "portal" sites. Chi square computation indicated that the difference is significant. This confirms that hypothesis H1 may be rejected/

To address the hypothesis H2, we calculated the mean and maximum ranks of each content category of the Top 100 sites for the 4 language groups. Due to space limitation, only the figures for the Spanish and English groups are reported in Table 2.

Table 2. Content Category Ranks for Language Groups

Category	Spanish		English	
	Mean	Max	Mean	Max
Shopping	32	9	50	8
Entertainment	54	8	78	20
Corporation	73	38	52	35
Knowledge	83	51	89	88
News/media	47	10	45	17
Online service	56	12	51	9
Portal	58	17	36	1
Search	18	1	50	3
Social network	52	7	47	4

It is interesting to observe that, the distributions of the mean ranks and the maximum ranks of the two language groups exhibit some degree of similarity. This is in contrast to what we observed in the previous membership comparison we made earlier.

Figure 6 shows the mean ranks for all four language groups. While many of the content categories show some degree of similarity, there appears to be a clear difference in the "Social Networking" categories. The English language group is clearly different from the other three languages.

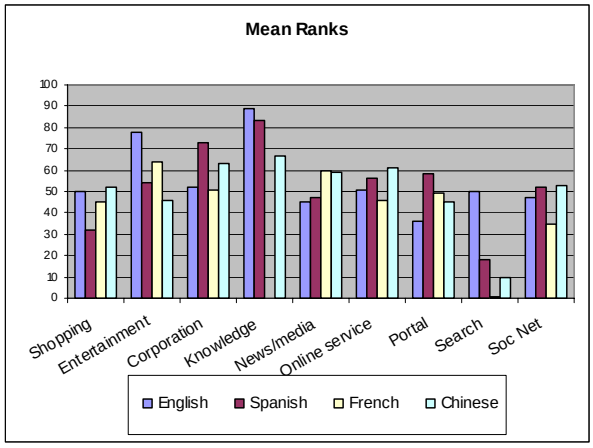


Figure 6: Mean ranks in each content category for the four language groups

The phenomenon of "Social Networking" websites is relatively new trend. This may account for the fact that these sites, e.g. MySpace, YouTube, Facebook etc, have already an establish followings in the English language, but have not attracted a great deal

of attention in the other language group yet. However, given time, it is expected that Internet users in the other language groups may become more familiar with this types of "Social Networking" sites. To test whether this is the correct interpretation, we are tracking these ranking lists in the different language group over time. Presumably, over time, the other language groups may also follow the same trend as the English language group, in terms of their content preferences.

PERLIMINARY CONCLUSIONS

It is important to note that the results here are very preliminary. There any interpretation should be considered with caution. Further investigations investigation is need to confirm some of these preliminary conclusions.

It appears that the content distribution of top website lists for different language groups does show some differences. In another word, the Internet has a diverging influence. Yet, at the same time there is also some similarity between their content preferences. It would appears therefore the Internet as a communication medium also have a homogenizing influence.

(As this is a work in progress study, we expect that as the latest results become available, further discussions will be presented.)

REFERENCES

1. Lo, B.W.N., & Sedhain, R.S. (2006). How reliable are Website rankings? Implications for E-business advertising and Internet search. *Issues in Information Systems*, 7(2), 233-238.
2. Lo, B.W.N., & Gong, P. (2005). Cultural impact on the design of E-commerce Websites: Part 1 – Site format and layout. *Issues in Information Systems*, 6(2), 182-189.
3. Callahan, E. (2005) Cultural similarities and differences in the design of university websites. *Journal of Computer-Mediated Communication*, 11(1), Article 12.
4. Wurtz, E. (2005) A cross-cultural analysis of websites from high-context cultures and low-context cultures. *Journal of Computer-Mediated Communication*, 11(1), Article 13.
5. Zhao, W., Massey, B.L., Murphy, J. & Fang. L. (2003) Cultural dimensions of website design and content. *Prometheus*, 21(1), 75-84.
6. Singh, N, Furrer, O, & Ostinelli, M. (2004) To localize or standardize on the web: empirical evidence from Italy, India, Netherlands, Spain, and Switzerland, *Multinational Business Review*, 12 (1), 69-87.
7. Cyr, D. & Trevor-Smith, H. (2004) Localization of web design: an empirical comparison of German, Japanese, and US website characteristics. *Journal of the American Society of Information Science and Technology*, 55(1), 1199-1208.
8. Pfell, U., Zaphiris, P. Ang, C.S. (2006). Cultural differences in collaborative authoring of Wikipedia. *Journal of Computer-Mediated Communication*. 12(1), Article 5.
9. WWW.Internetworldstats.com (2007). This site contains statistical data on World Internet usage. Retrieved March 12, 2007.
10. Huizingh, E. (2000). The content and design of web sites: an empirical study. *Information & Management*. 37(3), 123-134.
11. Singh, N. & Zhao, H. & Hu, X. (2003) Cultural adaptation on the web: A study of American companies' domestic and Chinese websites. *Journal of Global Information Management*: 11(3), 63-80.
12. Fletcher, R. (2006) The impact of culture on web site content, design, and structure: An international and multicultural perspective. *Journal of Communication Management*. 10(3), 259-273.
13. WWW.Alexa.com, (2006-2007), Data on the list of top ranking websites were collected over a period of 12 months from April 2006 to March 2007.
14. Hossain, L. & Wigand, R.T. (2004), ICT enables collaboration through trust. *Journal of Computer-Mediated Communication.* 10(1) Article 8.