Comparing Communication Technology on Chinese, English, and Spanish Diabetes Web Sites

Yanko Michea MD, Karen Pancheri BSN, RN, Yang Gong MD, MS, Elmer Bernstam MD, MSE
University of Texas Health Science Center at Houston, School of Health Information Sciences

ABSTRACT

Technological and cultural factors influence access to health information on the web in multifarious ways. We evaluated structural differences and availability of communication services on the web in three diverse language and cultural groups: Chinese, English, and Spanish. A total of 382 web sites were analyzed: 144 were English language sites (38%), 129 were Chinese language sites (34%), and 108 were Spanish language sites (28%). We did not find technical differences in the number of outgoing links per domain or the total availability of communication services between the three groups. There were differences in the distribution of available services between Chinese and English sites. In the Chinese sites, there were more communication services between consumers and health experts. Our results suggest that the health-related web presence of these three cultural groups is technologically comparable, but reflects differences that may be attributable to cultural factors.

INTRODUCTION

The web is an important source of health information and social support for the general public[1-3]. Moreover, people use the web not only as a health resource, but also as a source of expert opinion in the setting of virtual communities[4]. Web users seek advice on therapy, diagnosis, prognosis, and education[3]. Recent research raises questions about the magnitude of the differences between English and Spanish sites. The Berland study[5] compared English and Spanish web sites, finding a decrease in search efficiency and an increase in reading skills required to access information on Spanish web sites.

Differences in the level of web presence of diverse languages result from a combination of multiple factors: technological, historical, and cultural. While English represents 43% of the existing web sites, Chinese and Spanish together constitute no more that 17%[6]. Several factors can play a role in explaining the differences between English and Spanish sites. Industrialized countries enjoy a more extensive telecommunication infrastructure[7]. Additionally, a lower level of interconnectivity of peripheral nodes on the web can affect a search engine's ability to index sites in non-English domains, which can contribute to decreased visibility of web sites on search engines [8].

However, the differences on the web cannot be fully explained by technological disparity. Cultural idiosyncrasies or preferences can influence what is seen on the web. For instance, cultural identity is explicitly represented in virtual communities and software design[9-11]. Cultural preferences and beliefs can also influence health behavior. These can affect attitudes towards pain[12], willingness to undergo preventive exams[13, 14], and the use of social resources and health information[15]. Use of language can be an important determinant of the outcomes of inter-cultural health interventions[16, 17].

Interaction between individuals using the internet depends on the availability of communication services. A description and analysis of the characteristics of the services offered by web sites can help policy makers, health educators, and web designers to tailor their interventions to a specific culture.

Building web pages involves the design of information spaces and technological architectures. Technologies like chatroom, email, and forum are digital building blocks that support the development of virtual communities. Changes in communication technologies have influenced our perception of time (synchronous/asynchronous) and space. Social interaction is now supported and, to some degree, controlled by technical capabilities and limitations[18, 19].

We evaluated structural and communicational differences in a sample of web sites. The web sites represent three languages: Chinese, English, and Spanish. These languages are identified with different cultures and important ethnic groups. In addition, these groups represent different levels of technological development and presence on the web.

We hypothesized that the technological and cultural differences in these groups would be reflected in a measurable difference in web traffic, site structure, and distribution of communication services. An understanding of these differences will influence the design and implementation of web-based health interventions. Policy planners and public health officials can use this information to choose an adequate

media, tailored to the target population needs and idiosyncrasies. This knowledge will help web developers not only to direct technology to a definite group, but also to implement special support and training in specific technological areas.

METHODS

In order to select web sites for analysis, we used the search term "diabetes" on two search engines (Google and AltaVista). "Diabetes" was selected because of the chronic nature of the disease and its prevalence in the three cultures. Approximately five million people per language group suffer from this disease [20]. Another advantage of "diabetes" is that it exists as a well-defined, unambiguous synonym in the three languages.

We chose Google and AltaVista because they are popular web search engines that allow search and retrieval in the three languages. We evaluated the first 300 URLs in each language group as returned by the two search engines. Therefore, 600 URLs were evaluated per language, with a total of 1800 (600 x 3) URLs evaluated.

Once the URLs were established, we evaluated the language of the web site. Even though the search was restricted to a specific language, we found that primary sites and internal pages were sometimes in other languages. For instance, NIH is primarily an English site that has internal pages in Spanish. Since our primary focus is on evaluating the web site, we counted NIH only as an English site.

After we verified the language, we excluded non health-related web sites, incidental newspaper articles, and animal-related sites. News sites with dedicated health sections were included. These selection criteria constrained our evaluation to sites that were associated with both diabetes and health care. Also, we noted the number of domain extensions in each language.

In order to measure the structural differences between our language groups, we counted the number of internal links (number of pages) as well as the number of outgoing links (connectivity). We used Astra SiteManager 2.0 (available at http://www-heva.mercuryinteractive.com) to analyze the web sites. SiteManager creates a map of a site starting from the selected URL down through the tree-like structure of a site. SiteManager ignores levels of navigation above the selected URL. In order to standardize our measures, we decided to use the site's root (home) page to conduct the tests. The analysis of web site structure was limited to two levels of depth because of software and memory restrictions.

We used Alexa (available at http://www.alexa.com) to compare the traffic ranking of the web sites and the domain registration dates. The web page ranking reflects the traffic that each site receives. This ranking is based on the total number of Alexa visits over a rolling six-month period. Sites with the highest number of visits receive the higher ranking while low traffic pages receive the lowest ranking. For example, the popular search engine Google ranks fourth while a personal bulletin board ranks in the millions. We calculated the domain age by determining the difference between the domain registration date and January 1, 2002

We then evaluated the types of communication tools available on the site and the intended agents of communication. The types of communication tools were email, chatroom, forum, and newsletter. The communication agents were the web manager, health consumer, and health expert. We used the term 'consumer' to refer to any user of the web site that was not specifically identified as an expert or web site manager. 'Experts' were defined as health experts (usually doctors, nurses or psychologists), who were identified as such on the web site. 'Web managers' corresponded to any non health-expert web site personnel. Thus, each identified communication tool was tagged as 'consumer-web manager', 'consumer-expert', or 'consumer-consumer'. Communications on chatrooms and forums were evaluated for an expert's participation. The difference in the technological support of communication services was quantified as the total count of services (mail, chat, forum, and newsletter).

Additionally, we calculated the number of available services per communication recipient (web manager, expert, or another consumer). We compared these variables between Chinese, English, and Spanish web sites, and within the individual language groups.

Thus, we compiled the statistics of age, ranking, internal links, outgoing links, quotient of internal/outgoing links, as well as the total of individual communication tools and communication agents for each language. We then applied the test of normality and nonparametric ANOVA (Kruskal-Wallis Test) to compare the means of each group.

We used Spearman's correlation statistical test to measure the correlation between the internal links (number of pages per site) and the total number of services available in each language group. In addition, we tested the correlation between the outgoing links and total communication services in each language group. If the number of links relate to the availability

of communications services, a correlation between these variables should be present.

Finally, Spearman's correlation test was performed to compare the internal links with the traffic ranking as established by Alexa. Likewise, the number of outgoing links was compared with the Alexa traffic ranking.

RESULTS

Our search resulted in the evaluation of 382 web sites garnered from the initial 1800 URLs. The excluded URLs (1418) correspond to sites in other languages (148), duplicated URLs (800), or sites that did not fulfill our selection criteria (470).

Table 1 reflects that from the 382 web sites, there are 144 English language sites (38%), 129 Chinese language sites (34%) and 108 Spanish language sites (28%). Most of these sites domain extensions are ".com", ".edu", and ".gov".

Table 1. Number of web sites per domain extension.

Domain	Chinese	English	Spanish
".com"	89	73	69
".gov"	2	8	0
".edu"	3	3	0
".org"	9	46	10
".net"	22	9	10
others	4	5	20
Total	**129**	**144**	**109**

Table 2 summarizes the results of our analysis of the websites. Since all of the variables failed to pass the test of normality, we used the nonparametric ANOVA (Kruskal-Wallis Test) to compare the means between the three populations. When the level of significance is noteworthy, it is stated in Table 2.

After calculating the domain age of each site, we found that the English web sites are older than Chinese and Spanish web sites. As shown in Table 2, there is no statistical difference between the ages of Chinese and Spanish web sites.

The web site rankings in our sample are highly variable, ranging from 1,691 to 9,453,482. Table 2 details the values for each language group. There is no statistical significance in the difference of means of this variable. Alexa consistently provided information on 95% of the sites in all three languages.

In our sample, the mean count of pages per website in the Chinese web sites (684) is significantly greater than English and Spanish sites. The Spanish (232) and English (236) sites have similar mean values. On the other hand, Chinese (122), English (199) and Spanish (143) sites do not present significant differences in the number of outgoing links.

In Table 2, the comparison of the differences between the connectivity designs shows that the Internal/Outgoing quotient decreases as the number of outgoing links per web page increases. A lower quotient relates to a higher connectivity. In our sample, English sites have the highest level of connectivity. There is no statistically significant difference between Spanish and Chinese sites.

The data collected about the availability of email, chatrooms, and forums is also shown in Table 2. There is no significant difference in the means of total email services available in each of our groups. In general, there are fewer chatrooms than email links. The differences between language groups are not statistically significant. In contrast, Chinese sites present significantly more web forums than English or Spanish sites. The availability of newsletters is homogeneously distributed between the three language groups.

Next, we quantified the differences in the total communication services available. The difference in the total availability of communication tools between our three language groups is not significant, suggesting similar level of support for communication.

Our comparison of the number of available services per communication recipient (web-manager, expert, or consumer) shows that there are significantly more communication services between the consumer and web manager available on the English sites. On the other hand, Chinese web sites have more communication with an expert than the English sites. The communication services available to connect consumers with other consumers are not different between the three language groups.

Table 3 reflects the relationship between the complexity of the web site expressed as the number of internal links (pages per site) and the total number of communication services available. A significant correlation does exist in the Chinese web sites. Our evaluation of the external links versus total communication services demonstrates this same relationship in the Chinese sites. The Spanish and English sites have neither of these correlation levels. A consistent correlation does not exist between a major structural feature, the number of web pages, and the availability of web communication services.

The Spearman test shows a correlation between the internal links (number of pages) and traffic ranking, also, in the outgoing links (connectivity) and traffic ranking. As shown in Table 4, higher complexity and a lower ranking correspond to higher traffic flow. We found similar results for each of the three groups.

Table 2: Summary of Results

	Mean			Significance		
	English	Spanish	Chinese	En-Ch*	En-Sp*	Ch-Sp*
Age (days)	1750	985	878	P<0.001	P<0.001	NS
Ranking	1560875	1076611	1347964	NS	NS	NS
Internal Links	235.96	232.17	684.24	P<0.001	NS	P<0.05
Outgoing Links	198.69	143.45	122.85	NS	NS	NS
Internal/Outgoing	9.00	15.15	18.62	P<0.001	P<0.01	NS
Mail Total	1.03	1.11	1.10	NS	NS	NS
Chat Total	0.26	0.19	0.17	NS	NS	NS
Forum Total	0.40	0.35	0.70	P<0.05	NS	P<0.05
Communication Total	2.77	2.29	2.55	NS	NS	NS
Consumer-Consumer	0.40	0.41	0.51	NS	NS	NS
Consumer-Web Manager	1.92	1.42	1.52	P<0.01	NS	P<0.001
Consumer-Expert	0.23	0.40	0.49	P<0.01	NS	NS

*English-Chinese (EN-Ch), English-Spanish (En-Sp), and Chinese-Spanish (Ch-Sp)

Table 3. Spearman's correlation between number of links and total communication services (TCS)

	Internal Links vs. TCS		Outgoing Links vs. TCS	
	r	significance	r	significance
English	0.10	NS	0.13	NS
Spanish	0.15	NS	0.09	NS
Chinese	0.38	<0.0001	0.36	<0.0001

Table 4. Spearman's correlation of Internal Links and Outgoing Links with traffic ranking

	Internal Links vs. Ranking		Outgoing Links vs. Ranking	
	r	significance	r	significance
English	-0.54	<0.0001	-0.34	<0.0001
Spanish	-0.41	<0.0001	-0.35	<0.0001
Chinese	-0.37	<0.0001	-0.37	<0.0001
Total	-0.43	<0.0001	-0.35	<0.0001

DISCUSSION

Use of the web for health-related, inter-personal contact and information gathering reflects cultural and technological factors. Our results show that Chinese, English, and Spanish "diabetes" sites support similar levels of inter-personal communication, but emphasize different agents of communication.

Our findings support the existence of differences in the distribution of communication services offered in these three diverse language groups of "diabetes" web sites. These differences exist independently of the structural characteristics of the evaluated web sites. We expected that the known differences in the distribution of technology and web connectivity in these cultural groups would be represented by the characteristics of their respective web sites. The general features captured by our study, including web site structure, traffic, and communication features, do not show the expected differences.

Since our assumption about the effect of technological differences on the web is not evident in our sample, we might conclude that these groups have found a way to overcome the technological gap or that our methodology did not detect the differences. On the other hand, the cultural differences between these different groups might be represented in the characteristics of the distribution of the available services. For example, the greater availability of expert contact and web forums in Chinese sites may relate to problems with access to health services or cultural differences in the status of health experts, and therefore, their availability.

The detailed evaluation of web site characteristics shows higher levels of connectivity in the English web sites. This may reflect a later stage of web evolution compared to Chinese and Spanish sites. In addition, English sites reflect higher levels of contact with web site personnel. This feature might be associated with

cultural tendencies with an emphasis on service and support of commercial sites.

We found not only a similar level of representation in search engines, but also a similar traffic level in the three groups. These observations support current research about the importance of the internet as a source of health information for diverse cultural groups. These findings may serve as a baseline for further research regarding health-related use of the web in diverse cultures.

Future research must address how these technological idiosyncrasies represent information preferences; also, how diverse virtual communities provide information and support for health consumers. We believe that efforts must be made to test these observations in other scenarios, such as different diseases and languages. Moreover, qualitative as well as quantitative methodologies should be used to evaluate the contents of communication facilitated by internet technology.

REFERENCES

1. Wellman B. An electronic group is virtually a social network. In: Kiesler S, ed. *Culture of the Internet.* Hillsdale, NJ: Lawrence Erlbaum; 1997:179-205.
2. Wellman B, Gulia M. Virtual Communities as Communities: Net Surfers Don't Ride Alone. In: Kollock P, ed. *Communities and Cyberspace*; 1998:167-194.
3. Widman LE, Tong DA. Requests for medical advice from patients and families to health care providers who publish on the World Wide Web. *Arch Intern Med.* Jan 27 1997;157(2):209-212.
4. Pennbridge J, Moya R, Rodrigues L. Questionnaire survey of California consumers' use and rating of sources of health care information including the Internet. *West J Med.* Nov-Dec 1999;171(5-6):302-305.
5. Berland GK, Elliott MN, Morales LS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. *Jama.* May 23-30 2001;285(20):2612-2621.
6. Reach G. Global Internet Statistics by Language [Web Page]. 12/01/01. Available at: http://www.glreach.com/globstats/. Accessed 03/02/02.
7. Dodge M, Kitchin R. An Atlas of the Cyberspace. 02/11/02. Available at: http://www.cybergeography.org/atlas/cables.html. Accessed 03/02/02, 2002.
8. Kleinberg J, Lawrence S. The Structure of the Web. *Science.* 2001;294:1849-1850.
9. Correll S. The Ethnography of an Electronic Bar. *Journal of Contemporary Ethnography.* 1995;24(3):270-298.
10. Burkhalter B. Reading Race Online: Discovering Racial Identity in Usenet Discussions. In: Kollock P, ed. *Communities in Cyberspace.* London: Routledge; 1998:60-75.
11. Forsythe D. New Bottles, Old Wine: Hidden Cultural Assumptions in a Computerized Explanation System for Migraine Sufferers. *Medical Anthropology Quarterly.* 1996 1996;10(4):551-574.
12. Honeyman PT, Jacobs EA. Effects of culture on back pain in Australian aboriginals. *Spine.* Apr 1 1996;21(7):841-843.
13. Barroso J, McMillan S, Casey L, Gibson W, Kaminski G, Meyer J. Comparison between African-American and white women in their beliefs about breast cancer and their health locus of control. *Cancer Nurs.* Aug 2000;23(4):268-276.
14. Lawrence VA, Streiner D, Hazuda HP, Naylor R, Levine M, Gafni A. A cross-cultural consumer-based decision aid for screening mammography. *Prev Med.* Mar 2000;30(3):200-208.
15. Sennott-Miller L, May KM, Miller JL. Demographic and health status indicators to guide health promotion for Hispanic and Anglo rural elderly. *Patient Educ Couns.* Jan 1998;33(1):13-23.
16. Carrese JA, Rhodes LA. Bridging cultural differences in medical practice. The case of discussing negative information with Navajo patients. *J Gen Intern Med.* Feb 2000;15(2):92-96.
17. Warnecke RB, Johnson TP, Chavez N, et al. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol.* Jul 1997;7(5):334-342.
18. Pankoke U, Petersen U. Dealing with phenomena of time in the age of the internet. Paper presented at: IFIP WWC 2000 - Charting a Course to the Future 7th International Federation of Information Processing (IFIP) - Working Group 9.1 (Computers and Work) Women, Work and Computerization Conference; June 8-11, 2000, 2000; Vancouver, British Columbia, Canada.
19. Mantovani G, Riva G. "Real" presence: how different ontologies generate different criteria for presence, telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments.* 1999 1999;8(5):538-548.
20. King H, Aubert R, Herman W. Global Burden of Diabetes, 1995-2025. Prevalence, numerical estimates, and projections. *Diabetes Care.* 1998;21(9):1414-1431.