

Association for Information Systems AIS Electronic Library (AISeL)

AMCIS 2005 Proceedings

Americas Conference on Information Systems
(AMCIS)

2005

Searching for Non-English Web Content: An Empirical Study of the Spanish Business Intelligence Portal

Wingyan Chung

The University of Texas at El Paso, wchung@utep.edu

Follow this and additional works at: <http://aisel.aisnet.org/amcis2005>

Recommended Citation

Chung, Wingyan, "Searching for Non-English Web Content: An Empirical Study of the Spanish Business Intelligence Portal" (2005).
AMCIS 2005 Proceedings. 228.

<http://aisel.aisnet.org/amcis2005/228>

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2005 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Searching for Non-English Web Content: An Empirical Study of the Spanish Business Intelligence Portal

Wingyan Chung

Department of Information and Decision Sciences
The University of Texas at El Paso
wchung@utep.edu

ABSTRACT

As non-English-speaking online populations grow rapidly, there are increasing needs to support searching for non-English Web content. Prior research has assumed English to be the primary language for Web searching, but this is not the case for many non-English-speaking regions. For example, Latin America will have the fastest growing population in the coming decades but existing Spanish search engines lack search, browse, and analysis capabilities. In this paper, we have proposed a language-independent approach to supporting non-English Web searching. Based on the approach, we have developed the Spanish Business Intelligence Portal (SBizPort) to support searching, browsing, summarization, categorization, and visualization of Web information. Results from an empirical study involving Spanish subjects show that the portal achieved significantly better user ratings on information quality, cross-regional search capability, and overall satisfaction than the benchmark search portal. This study thus contributes to human-computer interaction research on non-English Web searching.

Keywords

Web searching, Web browsing, non-English Web searching, business, summarization, categorization, visualization, empirical study, human-computer interaction.

INTRODUCTION

A report published in September 2004 shows that the majority (64.8%) of the world's online population consists of non-English speakers (Global Reach, 2004b). Moreover, that population was estimated to have grown significantly in the near future to 820 million while the size of English-speaking online population was predicted to remain at 300 million (Global Reach, 2004a). For instance, the Spanish-speaking online population has exceeded 9 million and Latin America is estimated to have the fastest growing population in the world in the coming decades (Caramelli, 2003).

Despite growing needs for non-English Web searching, most existing technologies have been developed for English-speaking users and fail to address the needs of non-English Web searching. Current search engines in Spanish, for example, lack search and analysis capabilities and provide information of varying quality. Meanwhile, research about human-computer interaction (HCI) aspects of non-English Web searching is scarce. Knowledge of these aspects, if available, would provide system developers with insights to enhance non-English Web searching.

To address these problems, we have developed a language-independent approach to building intelligent portals to support non-English Web searching. We have used this approach to develop a Web search portal in the Spanish business domain. Specifically, we studied the HCI aspects of non-English Web searching in an empirical experiment involving Spanish subjects. The results are expected to benefit both theoretical understanding and practical Web applications.

LITERATURE REVIEW

English has been the dominant language for communication on the Web since its inception. However, as more non-English-speaking users have adopted Internet technologies, other languages have gained popularity. It is useful to review previous HCI research on the Web and to review developments of Web search technologies for the Spanish-speaking regions.

HCI Research on the Web

Researchers in HCI who have studied information seeking on the Web have described the process of information seeking as consisting of various stages of problem identification, problem definition, problem resolution, and solution presentation (Wilson, 1999). Variations of this process model can be found in the literature (Kuhlthau, 1998; Marchionini, 1995; Sutcliffe

and Ennis, 1998). Two major information-seeking activities are searching and browsing. In directed searching, the user first decomposes his goal into smaller problems, then expresses his needs as concepts and higher level semantics, formulates queries using such supports as Boolean query languages and syntax directed editors, and finally evaluates the results by serial search or systematic sampling. In exploratory browsing, the user first transforms his general information need into a problem. He then articulates that need as search terms or hyperlinks that appear on the system interface, searches using the terms or explores the hyperlinks using such browse supports as automatic summarization, clustering and visualization tools, and Web directories and finally evaluates the results by scanning through them.

As the Internet evolves as a major information-seeking platform, its HCI aspect has been addressed in recent research. Two approaches are found in previous research, namely, a system-centered approach and a user-centered approach.

The system-centered approach aims to use information technologies to assist human beings in their information-seeking process. The use of information retrieval systems (most notably search engines) is one major strategy. Various techniques have been incorporated into search engines, including meta-searching (Chen, Fan, Chau and Zeng, 2001), summarization (McDonald and Chen, 2002), and visualization (Lin, 1997; Marshall, McDonald, Chen and Chung, 2004). Despite the potential advantages of these techniques, they rarely have been applied to developing non-English search engines.

The user-centered approach to information seeking concerns the behavioral and cognitive aspects of the information seekers. In this approach, human information-seeking has been described as a behavior that includes questions, dialogue, and social and cognitive situations that are associated with a user's interaction with an information retrieval system (Kuhlthau, 1993; Kuhlthau, Spink and Cool, 1992). Previous research has dealt with issues relating to user cognitive structure (Ingwersen, 1992) and factors affecting user-intermediary interaction process (Saracevic, 1996). However, relatively little research was done to study the perception of information seekers in the context of non-English Web searching. In particular, the quality of information sources and regional impacts deserve more attention. Information quality, a multifaceted concept, is considered to be an important aspect of evaluating the quality of a Web site (Loiacono, 2002) and is one that has been explored by Wang and Strong (Wang and Strong, 1996), who evaluated information quality using a set of 16 dimensions that they had developed and had been tested in (Pipino, Lee and Wang, 2002). Regional impacts from different cultural, social and economic environments may arise due to diverse use of a language in different regions or countries. For example, Spink et al. (Spink, Ozmutlu, Ozmutlu and Jansen, 2002) found that FAST users (who are mostly Europeans) input queries more frequently than Excite users (who are mainly Americans), who focused more on e-commerce topics.

Spanish Search Engines

Major search engines have been developed for Spanish, the second most popular language in the United States and the primary language for Spain and some 22 Latin American countries. Terra (<http://www.terra.com/>) offers its services to more than 3.1 million Internet users in Europe and the Americas. A Gallup poll in 2002 reported Terra to be the most popular search engine in Spain, Wanadoo (<http://www.wanadoo.com/>), a subsidiary of France Telecom, was rated second (Gallup, 2002). Currently, Terra serves more than 3 million Internet users in Spain, Latin America, the United States, and many European countries. Wanadoo is currently the leading Internet service provider in France and the United Kingdom with 9.3 million customers in June 2004. Yahoo Español (Spain, <http://espanol.yahoo.com/>), the Spanish version of Yahoo, provides a human-compiled Web directory developed by about 150 editors who categorized over one million listed sites. YahooES also supplements its results with those from Inktomi and Google. Inktomi matches also appear to users after all YahooES matches have first been shown. Established in 1995, BIWE (<http://www.biwe.com/>) is one of the earliest search engines for searching Spanish information on the Web. BIWE supports searching of news, products, images, and other information and provides a variety of services including a Web directory, email, entertainment, and market information for Hispanics. Headquartered in the United States, Quepasa (<http://www.quepasa.com/>) was launched in 1997 and is a bilingual Web portal (Spanish and English) serving Hispanic populations in the United States and Latin America. It uses proprietary Web search technologies to reduce the number of irrelevant results by utilizing terms most frequently used and documents most frequently viewed (Peterson, 2002). Quepasa also offers other services such as news, email, online radio, chat, online translation, forums, and Web hosting.

Although different types of information are provided, these search engines typically present results as a long textual list and lack post-retrieval analysis capabilities. Moreover, except for some large portals such as Yahoo Español and Terra, most Spanish search engines serve a few regions rather than an entire Spanish speaking community.

Summary

Because existing search engines in Spanish typically lack analysis capabilities, they limit users' ability to understand retrieved results. The collections searched by these search engines are often region-specific, so they do not provide a

comprehensive understanding of the environment where they are operating. Research in these areas should enhance understanding of HCI issues in non-English Web searching. In this research, we address two questions: (1) How can a language-independent approach support Web searching in non-English languages that are widely used across different regions? (2) What are the user satisfaction, cross-regional searching capability, and information quality of our portals compared with existing search engines?

AN APPROACH TO NON-ENGLISH WEB SEARCHING

In this section, we describe a language-independent approach to supporting non-English Web searching. We used the approach to build the Spanish Business Intelligence Portal (SBizPort) providing domain-specific collections for Spanish Web searching and post-retrieval analysis for the Spanish business domain. Given the growing Spanish-speaking populations in the United States, Spain, and Latin America, businesses actively expand their opportunities by seeking information on the Web. Our approach consists of two major steps, as illustrated in the screen shots shown in Figure 1.

System Development

Step 1 – Searching and Collection Building

On the search page, a user can input keywords and choose whether to search, organize, or visualize the results. The user can input multiple keywords separated by line breaks and can choose among a number of carefully selected information sources by checking the boxes. The result page lists search results according to the information sources selected by the user.

To provide high-quality Spanish business information, we manually analyzed the existing information sources from key business categories such as e-commerce, international business, and competitive intelligence and obtained more than 183 seed URLs. A Web crawler then followed these URLs to collect pages automatically. The pages were then automatically indexed and stored in our database. In addition to domain spidering, we performed meta-spidering of six major search engines (Yahoo ES, Ahijuna, Conexcol, Ambdirecto, Auyantepui, and Teoma) using queries translated from English queries that previously had been used to build an English business intelligence search portal (Marshall et al., 2004). We chose these search engines because of their rich Spanish business content. The Spanish business collection obtained from this method contained more than 476,084 Web pages covering more than 22 countries. In addition, SBizPort supports meta-searching two domain-specific databases (SBizPort collection and AMBDirecto) and six Spanish general search engines (Yahoo Español, Terra, Ahijuna, Auyantepui, Bacan, and Ascinsa).

Step 2 – Post-retrieval Analysis

SBizPort provides post-retrieval analysis capabilities in the form of Web page summarization, categorization, and visualization. The SBizPort summarizer was modified from an English summarizer that uses sentence-selection heuristics to rank text segments (McDonald et al., 2002). To summarize a Web page, the summarizer automatically performs sentence evaluation, segmentation or topic identification, and segment ranking and extraction. Users can invoke it by clicking the number of sentences for summarization under each result. Then, a new window is activated, that displays the summary and the original Web page.

The SBizPort categorizer organizes Web pages (related to the input query) into twenty (or fewer) folders labeled by the key phrases appearing most frequently in the page summaries or titles. Each categorizer relies on a phrase lexicon in the relevant language to extract phrases from Web page summaries obtained from meta-searching or searching our collections. The phrase lexicon was automatically created from a large number of Spanish Web pages using the mutual information approach (used in (Chung, Zhang, Huang, Wang, Ong and Chen, 2004)). In addition, SBizPort supports visualization of Web pages retrieved using a Kohonen self-organizing map (SOM) algorithm (Kohonen, 1995) to categorize and place Web pages onto a two-dimensional jigsaw map. The larger the size of a region on the map, the more the Web pages are assigned to it. Users can click on a region to see a list of pages on the right and can open pages by clicking the link-embedded titles.

HCI Implications

Because the proposed approach has incorporated various advanced techniques for Web searching and analysis, it offers a language-independent way to building non-English Web portals. To customize the search engine development to a specific language, a system developer needs only to provide the regional information such as seed URLs, major information sources for meta-spidering and meta-searching, and interface requirements. To our knowledge, not many existing portals support searching and analysis in non-English domains (such as the Spanish business domain), where rapid future growth is expected.

From an HCI standpoint, our approach helps bridge the gap between the large amount of non-English information on the Web and the need for searching and analysis of this information.

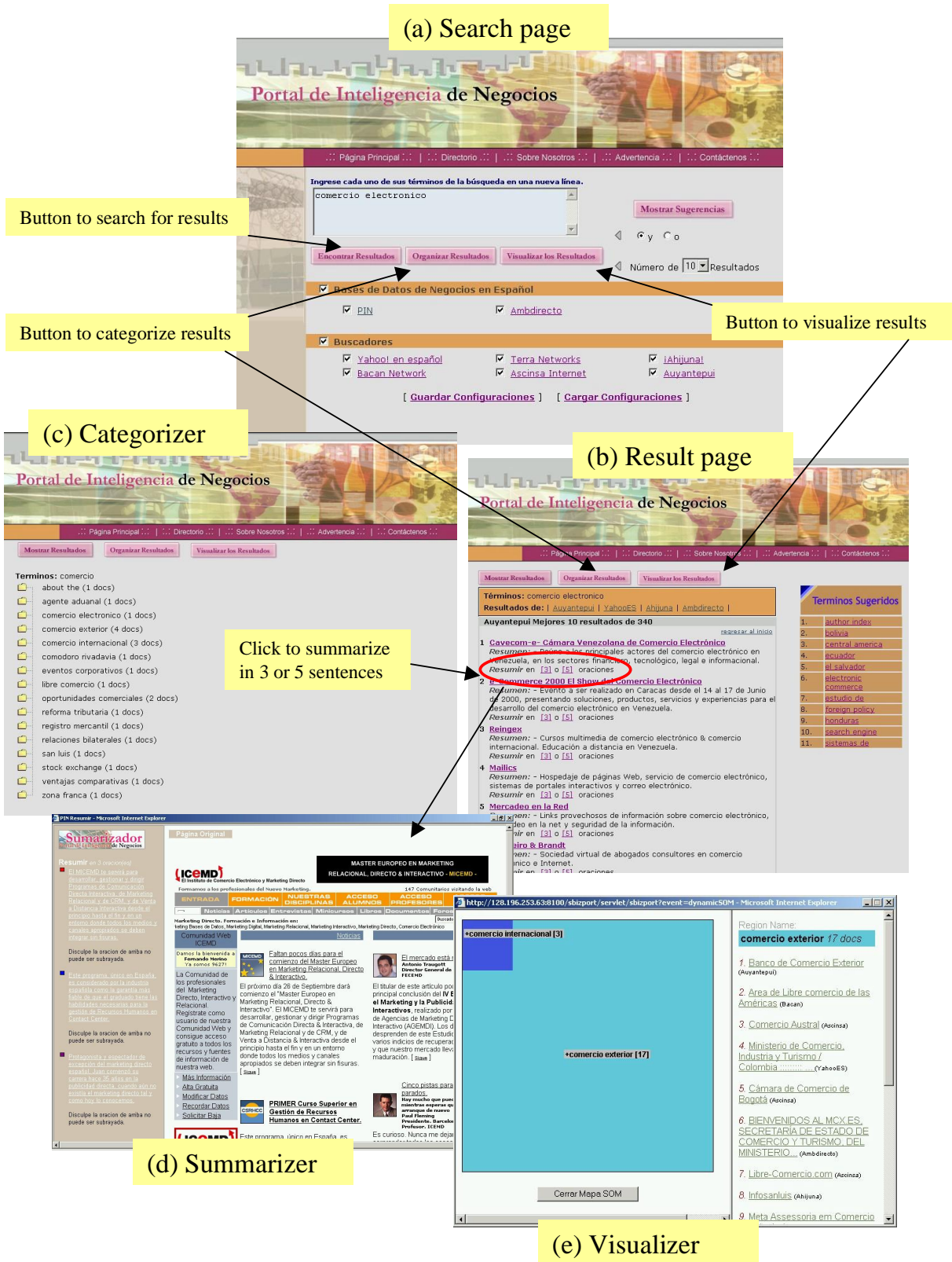


Figure 1. Screen shots of SBizPort**EXPERIMENTAL DESIGN AND FINDINGS**

In this section, we describe an experiment we conducted to evaluate SBizPort and report the findings. We invited native Spanish speakers to search and browse the Spanish business domains using SBizPort and BIWE, the benchmark search portal chosen to compare with our portal. BIWE (Buscador en Internet para la web en Español, <http://www.biwe.com/>) is a major Spanish search engine providing information for the Spanish-speaking community. It also has a detailed Web directory for users to browse topics in which they are interested. Compared with other Spanish search engines, BIWE's services are more comprehensive and target more closely to Hispanics.

Experimental Design

We designed scenario-based search and browse tasks consistent with Text Retrieval Conference standards (Voorhees and Harman, 1997) to evaluate the performance of our Web portal. For example, a scenario was "America Online (AOL) in Latin America," where a search task was "When was AOL Latin America launched in the United States?" and a browse task was "Find the URLs of financial portals where you can find stock quotes on America Online." To further validate the relevance of tasks, before conducting the actual experiment we did a pilot test with three subjects.

Nineteen Spanish students recruited from undergraduate and graduate courses at a university in the United States participated as volunteer subjects to evaluate the portal. Each subject was paid a fixed amount of money as an incentive for their participation. In a one-hour experiment, we introduced the two systems (our portal and the benchmark system) to a subject and randomly assigned different scenarios to evaluate the systems. The experiment and all the questionnaires involved were administered in Spanish. Before the subject used a system, the experimenter provided a demonstration of the system by showing how to perform a search task and a browse task. Each scenario contained two search tasks and one browse task. Although we did not impose any time limit on completing the tasks, we found that each subject spent an average three minutes to finish a search task and eight minutes to finish a browse task. The order in which the systems were used was randomly assigned to avoid bias due to sequence of use. After using a system, a subject filled in a post-session questionnaire about his ratings and comments on the system. The experimenter recorded all verbal comments or behavioral observations that were later analyzed using protocol analysis (Ericsson and Simon, 1993).

Performance Measure

Upon finishing the study, each subject also filled in a post-study questionnaire to rate each system in terms of information quality, cross-regional search capability, and overall satisfaction. A seven-point Likert scale was used in the rating. To measure information quality, we modified the 16-dimension construct developed in (Wang et al., 1996) by dropping the "security" dimension which is not relevant because the information provided by the systems is already public. To accommodate the different levels of importance in the remaining 15 dimensions, we invited a Spanish business expert to provide ratings on the relative importance of different dimensions (see Table 1). The Spanish business expert is a senior executive of a management consulting company in Mexico. Being a native Spanish speaker, he had 24 years of experience in business development, raising capital, negotiations, finance, and strategic planning. He also worked as the Vice President of Business Development for the Gallup Organization in Mexico. The subjects also provided demographic information, which was kept confidential in accordance with the Institutional Review Board Guidebook (Penslar, 2001).

Hypothesis Testing

Because SBizPort encompasses Web resources from different Spanish regions, we believed that it would provide richer content and higher usability than those of benchmark systems. Users could thus find relevant results more quickly from our portal. We tested the following hypotheses:

H1: SBizPort provides higher information quality than a benchmark search engine.

H2: SBizPort provides better cross-regional search capability than a benchmark search engine.

H3: SBizPort users achieve a higher overall satisfaction than users of a benchmark search engine.

Experimental Results

Table 2 and Table 3 respectively show the statistical results of hypothesis testing and subjects' demographic profile. Subjects rated SBizPort more favorably than BIWE in terms of information quality, cross-regional search capability, and overall

satisfaction. The mean differences between the two systems' ratings ranged from 0.6 to 1.5 and were all statistically significant at a 5% alpha-error level. Subjects were very satisfied with SBizPort. We believe that several aspects of SBizPort contributed to its superior performance: the high-quality meta-searchers and domain-specific collection used in SBizPort, the useful post-retrieval analysis tools, and the comprehensive cross-regional coverage.

Table 1. Definitions of 15 dimensions of information quality and expert ratings

Dimension	Definition	Expert Rating*
Presentation quality and clarity		
Accessibility	The extent to which information is available, or easily and quickly retrievable	3
Concise Representation	The extent to which information is compactly represented	3
Consistent Representation	The extent to which information is presented in the same format	3
Ease of Manipulation	The extent to which information is easy to manipulate and apply to different tasks	3
Interpretability	The extent to which information is in appropriate languages, symbols, and units, and the definitions are clear	2
Coverage and reliability		
Appropriate amount of information	The extent to which the volume of information is appropriate for the task at hand	2
Believability	The extent to which information is regarded as true and credible	2
Completeness	The extent to which information is not missing and is of sufficient breadth and depth for the task at hand	3
Free-of-error	The extent to which information is correct and reliable	2
Objectivity	The extent to which information is unbiased, unprejudiced, and impartial	2
Usability and analysis quality		
Relevancy	The extent to which information is applicable and helpful for the task at hand	3
Reputation	The extent to which information is highly regarded in terms of its source or content	3
Timeliness	The extent to which information is sufficiently up-to-date for the task at hand	3
Understandability	The extent to which information is easily comprehended	3
Value-Added	The extent to which information is beneficial and provides advantages from its use	3

* Expert rating: 3 = extremely important, 2 = very important, 1 = important

Table 2. Statistical Results of Hypothesis Testing

Hypothesis	Measure	SBizPort		Benchmark		p-value	Result
		Mean ¹	S.D.	Mean ¹	S.D.		
H1	Information Quality (Overall)	2.1	0.66	2.9	1.07	0.005*	Confirmed
	- Presentation quality and clarity	2.3	0.78	2.9	1.3	-	-
	- Coverage and reliability	2.2	0.63	3.0	1.1	-	-
	- Usability and analysis quality	1.98	0.76	2.9	1.1	-	-
H2	Cross-regional searching capability	1.7	0.81	3.2	1.6	0.009*	Confirmed
H3	Overall satisfaction	1.8	0.76	3.1	1.7	0.002*	Confirmed

*The range of rating is from 1 to 7, with 1 being the best.

Table 3. Subjects' demographic profile

Demographic information	Spanish subjects (total: 19)
Country of origin	Mexico (12), USA (3), Panama (1), Puerto Rico (1), Colombia (1), Peru (1)
Education	Undergraduate (13), bachelor earned (2), master earned (3), doctorate earned (1)
Age range	18-25 (14), 26-30 (2), 31-35 (2), 41-50 (1)
Gender	Female (10), Male (9)
Hours of using computer per week	< 5 (1), 5-10 (2), 10-15 (1), 15-20 (3), 20-25 (9), 30-35 (1), > 40 (2)

The subjects also provided many positive comments on SBizPort's search and browse capabilities. Twelve subjects agreed that SBizPort was very useful for searching Spanish business information. For instance, subject #s10 said that SBizPort "is very useful for searching," and "(the information) is clear." The subjects also liked the browse support tools provided by SBizPort. A majority of seventeen subjects commented positively on it. For example, subject #s6 said that SBizPort was "really nice to have different functions and have a catalog." Regarding the cross-regional search performance, fifteen subjects commented that SBizPort did a good job or has a greater variety than the benchmark search engine. For example, subject #s7 said: "(SBizPort) gives lots of pages related to what I look for from different countries." However, five subjects complained about the low speed of the system, especially when retrieving information from many meta-searchers.

On the other hand, the subjects were unhappy with BIWE's lack of relevance and clarity in searching and browsing. For example, subject #s7 said that BIWE "gives irrelevant pages (of) other countries I'm not interested in." Subject #s9 said that it was "time-consuming" to use BIWE. Moreover, most users did not like the presence of pop-up advertisements when using

BIWE. Nevertheless, six subjects said that BIWE was useful for searching Spanish business information. Three subjects commented that the system was easy to use and fast.

Implications of Results

The encouraging results from our experiment demonstrate the superior usability of the proposed approach to supporting non-English Web searching. We believe that the approach ensured high information quality, comprehensiveness in content coverage, useful functionality, and user-friendly interface. These important components help users who need to search for information from widely scattered regions in a language used by a multitude of countries and places. Given that the Internet will likely become more and more internationalized (O'Neill, Lavoie and Bennett, 2003), the proposed approach is expected to benefit a wide range of domains and users. As previous research has implicitly assumed English to be the major language on the Web, this research provides new implications to the study of HCI on the Web. Rapidly emerging issues on non-English Web searching such as search engine development, information quality, and browse support have been explored in this study.

CONCLUSIONS AND THE FUTURE

In this paper, we have proposed a language-independent approach to supporting non-English Web searching. Based on the approach, we developed SBizPort for searching information in the Spanish business domain and conducted an experiment using native Spanish speakers. Experimental results show that our portal significantly outperformed the benchmark search engine in terms of information quality, cross-regional search capability, and overall satisfaction. Subjects expressed a strong preference toward our portal. *We therefore conclude that the proposed approach has superior usability in supporting non-English Web searching.* This research thus contributes to developing a useful approach to non-English Web searching and achieving a better understanding of HCI in non-English Web searching.

As the notion of a "multilingual Web" continues to draw attentions, we are developing scalable techniques to collect and analyze information in different languages meaningfully to relate diverse content to produce intelligence. For instance, multinational corporations (MNCs) typically provide Web site information in different languages. Analyzing MNC's relationships with their multinational stakeholders could help provide a holistic picture of how they stand in the international arena. Also, the sample size and the number of Spanish-speaking regions represented by the subjects in this study are limited because we had difficulty recruiting more subjects. It may affect the generalization of our results and we will address this issue in the future. Furthermore, we will develop and validate new visualization techniques to support browsing and comprehending massive multilingual information on the Web. Relevant HCI issues also will be explored.

ACKNOWLEDGMENTS

This research was partly supported by funding from the National Science Foundation Knowledge Discovery and Dissemination (KDD) program #9983304 (June 2003 – March 2004 and October 2003 – March 2004) and funding from The University of Texas at El Paso URI Research Grant Program 2004-2005. We thank all the contributors of system development and user study, and all the human subjects in the experiment.

REFERENCES

1. Caramelli, P. (2003) "The current and future rapid growth of older people in Latin America: implications in psychogeriatrics (keynote presentation)," *Proceedings of the Eleventh International Congress*, International Psychogeriatric Association, Chicago, IL.
2. Chen, H., Fan, H., Chau, M. and Zeng, D. (2001) "MetaSpider: meta-searching and categorization on the Web," *Journal of the American Society for Information Science and Technology*, 52, 13, 1134-1147.
3. Chung, W., Zhang, Y., Huang, Z., Wang, G., Ong, T.-H. and Chen, H. (2004) "Internet searching and browsing in a multilingual world: An experiment on the Chinese Business Intelligence Portal (CBizPort)," *Journal of the American Society for Information Science and Technology*, 55, 9, 818-831.
4. Ericsson, K.A. and Simon, H.A. (1993) *Protocol analysis: verbal reports as data* MIT Press, Cambridge, MA.
5. Gallup (2002) "Encuesta sobre portales 2002," http://aui.es/estadi/gallup/gallup_portales_2002.htm.
6. Global Reach (2004a) "Evolution of non-English online populations," <http://global-reach.biz/globstats/evol.html>.
7. Global Reach (2004b) "Global Internet Statistics (by Language)," <http://www.glreach.com/globstats/>.
8. Ingwersen, P. (1992) *Information retrieval interaction* Taylor Graham, London.
9. Kohonen, T. (1995) *Self-organizing maps* Springer-Verlag, Berlin.
10. Kuhlthau, C. (1993) "A principle of uncertainty for information seeking," *Journal of Documentation*, 49, 4, 339-355.

11. Kuhlthau, C. (1998) "Longitudinal case studies of the information search process of users in libraries," *Library and Information Science Research*, 10, 3, 257-304.
12. Kuhlthau, C., Spink, A. and Cool, C. (1992) "Exploration into stages in the information search process in on-line IR: Communication between users and intermediaries," *Proceedings of the Annual Meeting of the American Society for Information Science*, 29, 67-71.
13. Lin, X. (1997) "Map displays for information retrieval," *Journal of the American Society for Information Science*, 48, 1, 40-54.
14. Loiacono, E. (2002) "WebQual™: A Web Site Quality Instrument," *Proceedings of International Conference on Information Systems (ICIS) Doctoral Consortium*, Charlotte, NC, USA.
15. Marchionini, G. (1995) *Information seeking in electronic environments* Cambridge University Press, New York.
16. Marshall, B., McDonald, D., Chen, H. and Chung, W. (2004) "EBizPort: Collecting and Analyzing Business Intelligence Information," *Journal of the American Society for Information Science and Technology*, 55, 10, 873-891.
17. McDonald, D. and Chen, H. (2002) "Using sentence selection heuristics to rank text segments in TXTRACTOR," *Proceedings of the second ACM/IEEE-CS Joint Conference on Digital Libraries*, ACM/IEEE-CS, Portland, OR, USA, 28-35.
18. O'Neill, E.T., Lavoie, B.F. and Bennett, R. (2003) "Trends in the Evolution of the Public Web 1998 - 2002," *Digital Library Magazine*, 9, 4.
19. Penslar, R.L. (2001) "Institutional Review Board Guidebook," Office for Human Research Protection, U.S. Department of Health and Human Services, http://ohrp.osophhs.dhhs.gov/irb/irb_guidebook.htm.
20. Peterson, J. (2002) "Quepasa announces agreement to acquire Vayala Corporation," in: *Hispanic PR Wire - Business Wire*, Phoenix.
21. Pipino, L.L., Lee, Y.W. and Wang, R.Y. (2002) "Data quality assessment," *Communications of the ACM*, 45, 4, 211-218.
22. Saracevic, T. (1996) "Modeling interaction in IR. Review and proposal.," *Proceedings of the Annual Meeting of the American Society for Information Science*, 3-9.
23. Spink, A., Ozmutlu, S., Ozmutlu, H.C. and Jansen, B.J. (2002) "U.S. versus European Web Searching Trends," *SIGIR Forum*, 36, 2.
24. Sutcliffe, A.G. and Ennis, M. (1998) "Towards a cognitive theory of Information Retrieval," *Interacting with Computers (Special Edition on HCI & Information Retrieval)*, 10, 321-351.
25. Voorhees, E. and Harman, D. (1997) "Overview of the Sixth Text Retrieval Conference (TREC-6)," *NIST Special Publication 500-240: The Sixth Text Retrieval Conference (TREC-6)*, National Institute of Standards and Technology, Gaithersburg, MD, USA.
26. Wang, R.Y. and Strong, D.M. (1996) "Beyond accuracy: what data quality means to data consumers," *Journal of Management Information Systems*, 12, 4, 5-34.
27. Wilson, T.D. (1999) "Models of information behavior research," *Journal of Documentation*, 55, 3, 249-270.