Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2005 Proceedings

Americas Conference on Information Systems (AMCIS)

2005

Building Web Directories in Different Languages for Decision Support: A Semi-Automatic Approach

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

Guanpi Lai
The University of Arizona, guanpi@email.arizona.edu

Alfonso Bonillas
The University of Arizona, aabonill@email.arizona.edu

Theodore Elhourani
The University of Arizona, telhoura@email.arizona.edu

Tzu-Liang (Bill) Tseng
The University of Texas at El Paso, btseng@utep.edu

See next page for additional authors

Authors

Wingyan Chung, Guanpi Lai, Alfonso Bonillas, Theodore Elhourani, Tzu-Liang (Bill) Tseng, and Hsinchun Chen

Building Web Directories in Different Languages for Decision Support: A Semi-Automatic Approach

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

Guanpi Lai
Artificial Intelligence Lab
The University of Arizona
guanpi@email.arizona.edu

Alfonso Bonillas
Artificial Intelligence Lab
The University of Arizona
aabonill@email.arizona.edu

Theodore Elhourani
Artificial Intelligence Lab
The University of Arizona
telhoura@email.arizona.edu

Tzu-Liang (Bill) Tseng
Department of Mechanical and Industrial
Engineering, The University of Texas at El Paso
btseng@utep.edu

Hsinchun Chen
Artificial Intelligence Lab
The University of Arizona
hchen@eller.arizona.edu

ABSTRACT

Web directories organize voluminous information into hierarchical structures, helping users to quickly locate relevant information and to support decision-making. The development of existing Web directories either relies on expert participation that may not be available or uses automatic approaches that lack precision. As more users access the Web in their native languages, better approaches to organizing and developing non-English Web directories are needed. In this paper, we have proposed a semi-automatic approach to building domain-specific Web directories in different languages by combining human precision and machine efficiency. Using the approach, we have built Web directories in the Spanish business (SBiz) and Arabic medical (AMed) domains. Experimental results show that the SBiz and AMed directories achieved significantly better recall, F value, and satisfaction rating than benchmark directories. These encouraging results show that the approach can be used to build high-quality Web directories to support decision-making.

Keywords

Web directories, decision support, non-English languages, Spanish business, Arabic medicine, semi-automatic approach.

INTRODUCTION

Although the Internet has greatly facilitated searching for information, users are often overwhelmed with a large amount of semi-structured and unstructured information. Aside from searching, browsing well-structured Web directories may help users to explore interesting yet unfamiliar domains. Examples of general Web directories include the Yahoo (http://dir.yahoo.com/) and DMOZ (http://dmoz.org/) directories. Relying on the directory's labels and categorization, users can spend less time in finding information and experience a better Web navigation. Nevertheless, building a high-quality Web directory without much expert knowledge and extensive human efforts has challenged developers of Web portals.

On the other hand, as the Internet grows in popularity worldwide, more users access Web content in their native languages. A report published in September 2004 shows that the majority of the total global online population (64.8%) lives in non-English-speaking areas (Global Reach, 2004). Better approaches to building non-English Web directories will improve the browsing experience of a large number of people in the world.

In this paper, we have proposed a semi-automatic approach to building high-quality Web directories in different languages by combining human preciseness and machine efficiency. Human knowledge of domains and languages was used to guide the establishment of the directory frameworks. Meta-searching of high quality information sources allowed us to efficiently fill the frameworks with meaningful and relevant items. To demonstrate the usability of the approach, we have built two domain-specific Web directories in different languages: Spanish business and Arabic medical directories. Results of our user evaluation studies involving native speakers show that our directories provide more relevant information than existing Web directories in the corresponding domains.

LITERATURE REVIEW

Theoretical Models for Information Seeking and Web Browsing

Researchers 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 literature (Kuhlthau, 1998; Marchionini, 1995; Sutcliffe and Ennis, 1998). Apart from searching the Web, Internet users frequently engage in browsing. Marchionini and Shneiderman defined browsing as "an exploratory, information seeking strategy that depends upon serendipity" (Marchionini and Shneiderman, 1988). Chang and Rice stated that browsing is a direct application of human perception in information seeking (Chang and Rice, 1993). Spence defined "browse" as the registration of content into a human mental model (Spence, 1999). Having compared various definitions, Chung defined "browsing" as an exploratory information seeking process characterized by the absence of planning, with a view to forming a mental model of the content being browsed (Chung, Chen and Nunamaker, 2005). These theoretical models and definitions are useful for understanding decision support using information seeking on the Web. Nevertheless, how decision making can be supported by browsing Web directories has not been widely studied.

Building Web Directories

Previous work in building Web directories falls into two categories: (1) extensive manual identification and categorization of Web resources; and (2) automatic construction of directories using machine learning or Web mining techniques. We review related work in each category below.

Manual Categorization

Manual identification and categorization have been used in general search engines and domain-specific Web portals. The Open Directory Project, also known as Directory Mozilla (DMOZ, http://dmoz.org/), is developed and maintained by a large, global community of volunteer editors. The rationale of DMOZ is to use extensive human work to combat growth of human-created Web resources, the amount of which often increases with the size of online population. However, the quality of the directory constructed by this method depends highly on the volunteer editors' domain knowledge, which usually varies from person to person. Moreover, the approach is not scalable because many Web resources are generated automatically, making its growth more rapid than the growth of online population.

Other examples of manually-created Web directories include the Yahoo! Directory (http://dir.yahoo.com/), the Librarian's Index to the Internet (LII, http://lii.org/) and the UMLS Semantic Network (http://semanticnetwork.nlm.nih.gov/). The Yahoo! Directory is built and maintained by a team of paid editors who organize Web sites into 14 categories and subcategories, each having around 20 to 45 subcategories. Despite Yahoo's popularity, the small team of editors, with limited knowledge and time, may be ineffective and inefficient in identifying and evaluating Web resources. LII provides a searchable, annotated subject directory of more than 12,000 Internet resources selected and evaluated by librarians for their usefulness to users of public libraries. Over 100 contributors participate in the index building and updating process, which is facilitated by a Web-based system (Leita, 2004). Developed by the National Library of Medicine, the UMLS Semantic Network provides a categorization of all concepts represented in the UMLS Metathesaurus. Although both LII and UMLS Semantic Network are highly regarded in their respective domains, their construction relies heavily on expert participation and their domain knowledge that may be difficult to obtain.

Automatic Construction of Web Directories

Beside manual methods, automatic approaches to constructing taxonomy and ontology have been proposed in previous research. Sato and Sato developed an automated editing system that generates a Web directory from a given category word without human intervention (Sato and Sato, 1999). Although the system was efficient for generating a directory from a category label, the generated directory only had one level that restricted its use in more complicated browse tasks. In another research, Chuang and Chien propose a query-categorization approach to facilitate the construction of Web taxonomies (Chuang and Chien, 2003). The approach requires search engine log data that are typically not accessible by outsiders. The predefined taxonomy structure also does not suit domains having relatively smaller coverage on the Web. On the other hand, Kumar's approach of using both ontologists' knowledge and a search engine's efficiency yielded interesting results, but the quality of these results depends highly on the ontologist's limited knowledge (Kumar, Raghavan, Rajagopalan and Tomkins, 2001). Their use of only one system to search for related links also limits the coverage of results severely.

A SEMI-AUTOMATIC APPROACH

Considering the limitations found in previous research, we herein propose a semi-automatic approach to Web directory construction. Our approach combines human knowledge and machine efficiency, while incorporating various information sources to ensure a high quality of content. We used meta-searching to provide a high quality of Web directory content as it has been shown to reduce the bias from using only a small number of search engines (Mowshowitz and Kawaguchi, 2002). Manual filtering also helped to remove irrelevant content. In the following, we explain our approach in the context of building the Spanish Business Intelligence (SBiz) and Arabic Medical Intelligence (AMed) Web directories. We chose these two domains as our testbeds because of the popularity of Spanish (Caramelli, 2003) and Arabic (Norton, 2001) in their geographic regions and their growing Web contents. These domains also have a relatively smaller coverage on the Web compared with their English counterparts. The approach consists of three steps described below.

Identification and Modification of Category Labels

In this step, our domain experts identified an existing Web directory as the base directory and modified its category labels as queries for meta-searching (in the next step). For building the SBiz Web directory, we used DMOZ directory as the base directory because it provides a representative business directory used by many other search engines. Since DMOZ directory is mainly used in English-speaking countries, we removed 546 nodes from the original 779 nodes of its business sub-directory, leaving 233 nodes in the directory. To ensure that our resulting directory contains items specific to the Spanish business domain (rather than English business domain), we reviewed a Spanish business directory provided in BIWE (Buscador en Internet para la Web en Español, `http://www.biwe.es/`) to add 81 nodes to the 233 nodes, resulting in a 314-node Spanish business directory framework. Based in Spain, BIWE is a major general Spanish search engine serving more than four million Web pages per month for the Spanish-speaking community. Its Web directory, with 16 main categories and 633 sub-categories, was chosen as it is more comprehensive, and contains more Spanish resources than existing Spanish Web directories such as Degerencia (`http://www.degerencia.com/`) and Gestiopolis (`http://www.gestiopolis.com/`).

A similar procedure was taken to create an Arabic medical directory framework. We also used DMOZ directory as the base directory because of its comprehensiveness in the English medical domain. We removed 46 nodes from the original 356 nodes of its medical sub-directory, leaving 310 nodes in the directory. Then, we manually added 11 nodes by including cultural specific items such as Islamic medicine, resulting in a 321-node Arabic medical directory framework.

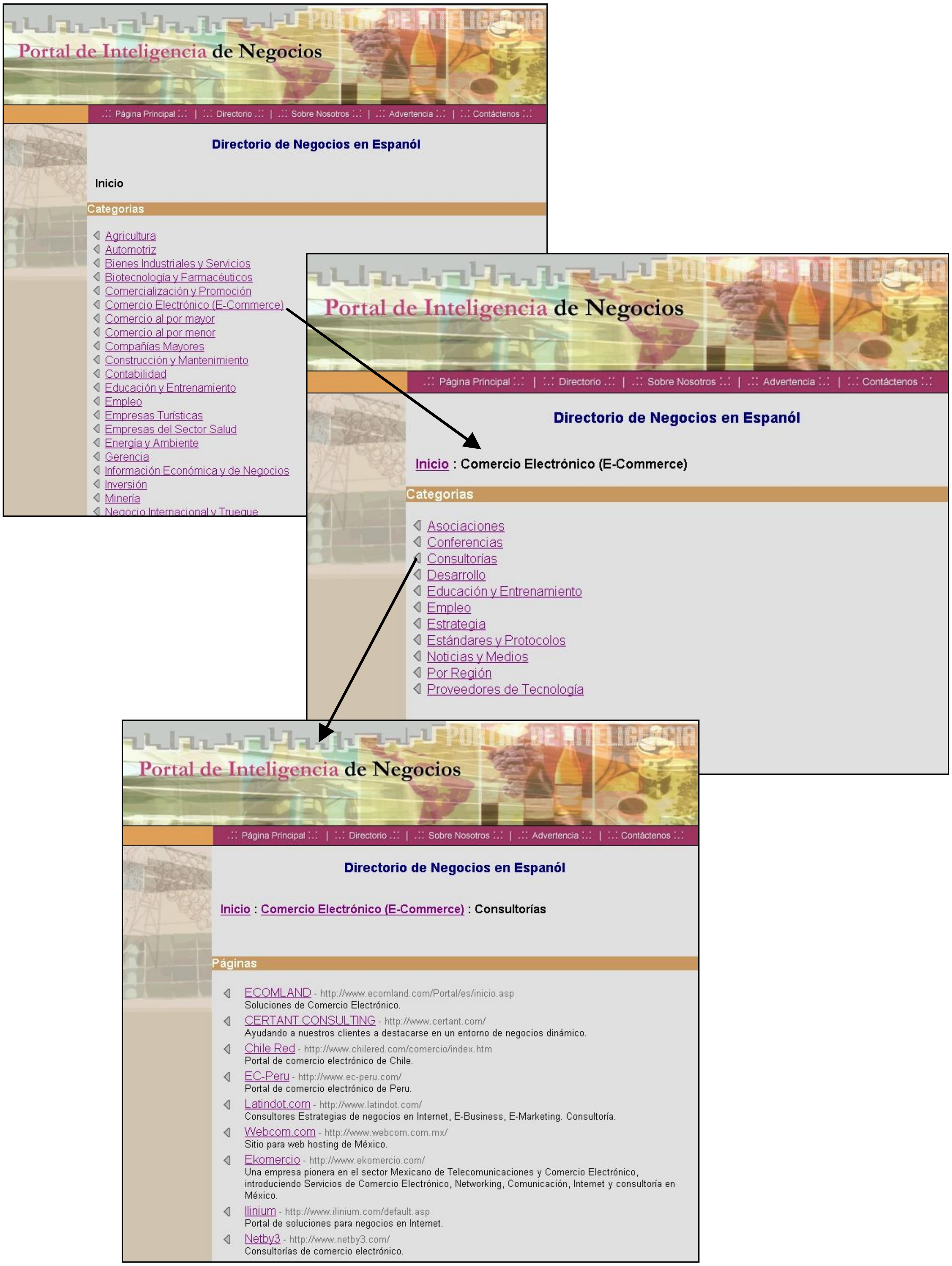
Automatic Generation of Taxonomy Items

The purpose of this step is to fill in each directory framework (from the previous step) with items obtained by meta-searching. To fill in the Spanish business directory, we used seven major search engines (Yahoo Español, Auyantepui, Teoma, Conexcol, Ambdirecto, Terra, and Ahijuna) as meta-searchers and the 314 category labels of the framework (from the previous step) as input queries. From our survey, these meta-searchers provide the richest Spanish resources on the Web. The top ten results obtained from each meta-searcher were collated, with duplicate results removed. Using this meta-search program, we obtained 12,234 unique Web URLs related to 296 category labels (non-empty nodes) out of the 314 nodes. The maximum depth of the resulting taxonomy is 5.

To fill in the Arabic medical directory, we used six major search engines (Ba7th, Arabmedmag, Google, Ayna, Sehha, and ArabVista) as meta-searchers and the 321 category labels of the framework (from the previous step) as input queries. From our survey, these meta-searchers provide the richest Arabic resources on the Web. Running the similar program as described above, we obtained 8,040 unique Web URLs related to 292 category labels (non-empty nodes) out of the 321 nodes. The maximum depth of the resulting taxonomy is 5.

Manual Filtering and Enhancement

This step aims to enhance the quality of the automatically generated Web directory (from the previous step) by filtering out irrelevant items and by adding in relevant items that were missed. URLs were removed if (a) they were not relevant to the topic; (b) they are not related to the domain (i.e., Spanish business or Arabic medical domains) being considered. Empty nodes were also removed. Table 1 shows the statistics of the two resulting Web directories and their screen shots are shown in Figure 1 and Figure 2.


PORTAL DE INTELIGENCIA
Portal de Inteligencia de Negocios
.:: Página Principal .:: | .:: Directorio .:: | .:: Sobre Nosotros .:: | .:: Advertencia .:: | .:: Contáctenos .::

Directorio de Negocios en Espanól

Inicio

Categorias

- Agricultura
- Automotriz
- Bienes Industriales y Servicios
- Biotecnología y Farmacéuticos
- Comercialización y Promoción
- Comercio Electrónico (E-Commerce)
- Comercio al por mayor
- Comercio al por menor
- Compañías Mayores
- Construcción y Mantenimiento
- Contabilidad
- Educación y Entrenamiento
- Empleo
- Empresas Turísticas
- Empresas del Sector Salud
- Energía y Ambiente
- Gerencia
- Información Económica y de Negocios
- Inversión
- Minería
- Negocio Internacional y Trueque

PORTAL DE INTELIGENCIA
Portal de Inteligencia de Negocios
.:: Página Principal .:: | .:: Directorio .:: | .:: Sobre Nosotros .:: | .:: Advertencia .:: | .:: Contáctenos .::

Directorio de Negocios en Espanól

Inicio : Comercio Electrónico (E-Commerce)

Categorias

- Asociaciones
- Conferencias
- Consultorías
- Desarrollo
- Educación y Entrenamiento
- Empleo
- Estrategia
- Estándares y Protocolos
- Noticias y Medios
- Por Región
- Proveedores de Tecnología

PORTAL DE INTELIGENCIA
Portal de Inteligencia de Negocios
.:: Página Principal .:: | .:: Directorio .:: | .:: Sobre Nosotros .:: | .:: Advertencia .:: | .:: Contáctenos .::

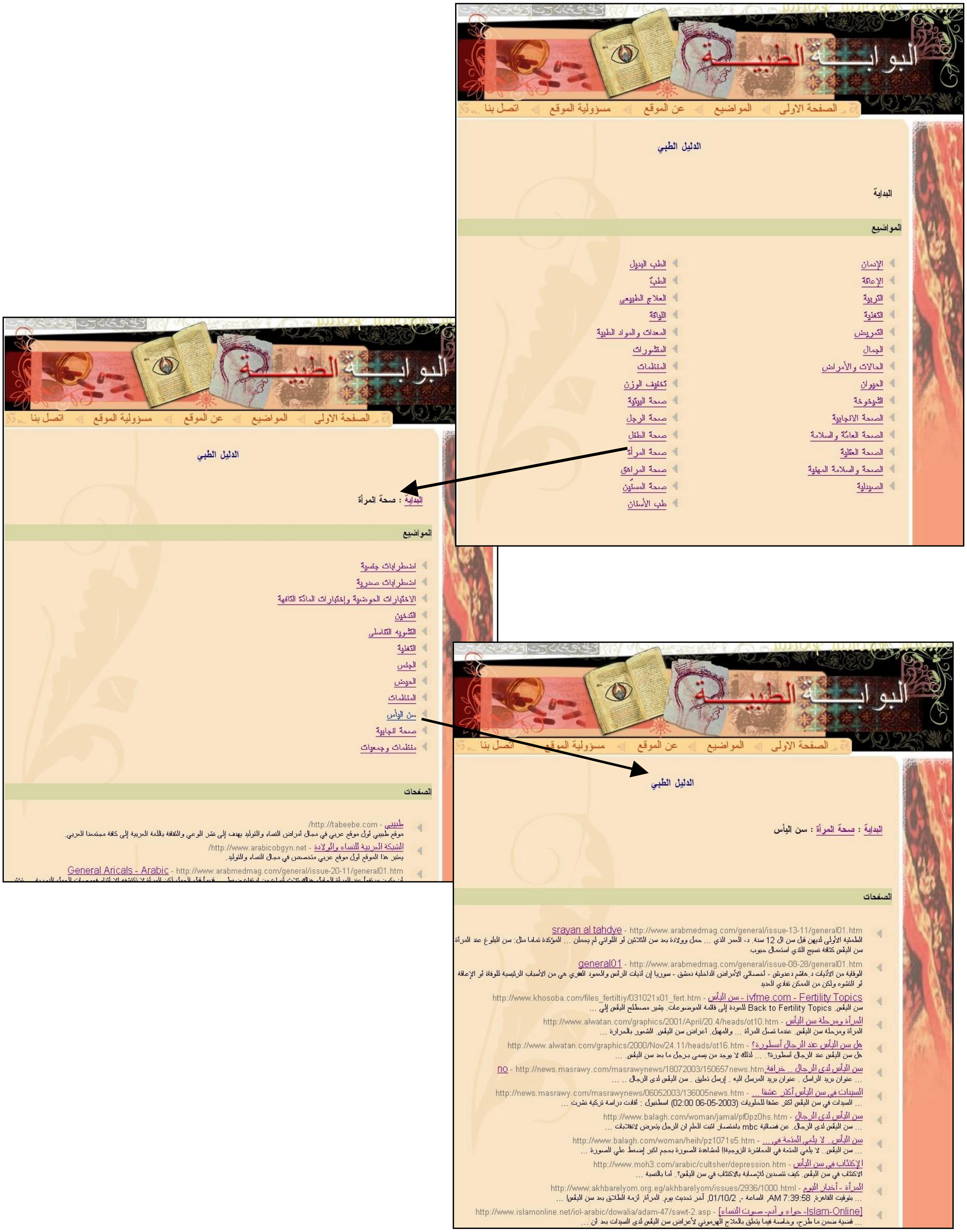
Directorio de Negocios en Espanól

Inicio : Comercio Electrónico (E-Commerce) : Consultorías

Páginas

- ECOMLAND - http://www.ecomland.com/Portal/es/inicio.asp
  Soluciones de Comercio Electrónico.
- CERTANT CONSULTING - http://www.certant.com/
  Ayudando a nuestros clientes a destacarse en un entorno de negocios dinámico.
- Chile Red - http://www.chilered.com/comercio/index.htm
  Portal de comercio electrónico de Chile.
- EC-Peru - http://www.ec-peru.com/
  Portal de comercio electrónico de Peru.
- Latindot.com - http://www.latindot.com/
  Consultores Estrategias de negocios en Internet, E-Business, E-Marketing. Consultoría.
- Webcom.com - http://www.webcom.com.mx/
  Sitio para web hosting de México.
- Ekomercio - http://www.ekomercio.com/
  Una empresa pionera en el sector Mexicano de Telecomunicaciones y Comercio Electrónico, introduciendo Servicios de Comercio Electrónico, Networking, Comunicación, Internet y consultoría en México.
- Ilinium - http://www.ilinium.com/default.asp
  Portal de soluciones para negocios en Internet.
- Netby3 - http://www.netby3.com/
  Consultorías de comercio electrónico.

Figure 1. Screen shots of the SBiz Web directory



الدليل الطبي

البداية

المواضيع

الإدمان
الإعاقة
التربية
التغذية
التمريض
الجمال
الحالات والأمراض
الحيوان
الشيخوخة
الصحة الانجابية
الصحة العامّة والسلامة
الصحة العقلية
الصحة والسلامة المهنية
الصيدلية

الطب البديل
الطبّ
العلاج الطبيعى
اللياقة
المعدات والمواد الطبية
المنشورات
المنظمات
تخفيف الوزن
صحة البيئية
صحة الرجل
صحة الطفل
صحة المرأة
صحة المراهق
صحة المسنّين
طب الأسنان

الدليل الطبي

البداية : صحة المرأة

المواضيع

اضطرابات جنسية
اضطرابات صدرية
الاختبارات الحوضية وإختبارات المادّة الدافهة
التدخين
التشويه التناسلى
التغذية
الجنس
الحيض
المنظمات
سن اليأس
صحة انجابية
منظمات وجمعيات

الصفحات

طبيبي - /http://tabeebe.com
موقع طبيبي أول موقع عربي في مجال أمراض النساء والتوليد يهدف إلى نشر الوعي والثقافة باللغة العربية إلى كافة مجتمعنا العربي.

الشبكة العربية للنساء والولادة - /http://www.arabicobgyn.net
يعتبر هذا الموقع أول موقع عربي متخصص في مجال النساء والتوليد.

General Aricals - Arabic - http://www.arabmedmag.com/general/issue-20-11/general01.htm

الدليل الطبي

البداية : صحة المرأة : سن اليأس

الصفحات

srayan al tahdye - http://www.arabmedmag.com/general/issue-13-11/general01.htm
الطمثية الأولى لديهن قبل سن ال 12 سنة. د- العمر الذي ... حمل وولادة بعد سن الثلاثين أو اللوائي لم يحملن ... المؤكدة تماما مثل: سن البلوغ عند المرأة سن اليئس كثافة نسيج الثدي استعمال حبوب

general01 - http://www.arabmedmag.com/general/issue-08-28/general01.htm
الوقاية من الأذيات د.هاشم دعدوش - أخصائي الأمراض الداخلية دمشق - سوريا إن أذيات الرأس والعمود الفقري هي من الأسباب الرئيسية للوفاة أو الإعاقة أو التشوه ولكن من الممكن تفادي العديد

سن اليأس - ivfme.com - Fertility Topics - http://www.khosoba.com/files_fertiltiy/031021x01_fert.htm
سن اليئس. Back to Fertility Topics للعودة إلى قائمة الموضوعات. يشير مصطلح اليئس إلى ...

المرأة ومرحلة سن اليأس - http://www.alwatan.com/graphics/2001/April/20.4/heads/ot10.htm
المرأة ومرحلة سن اليئس. عندما تصل المرأة ... والمهبل. أعراض سن اليئس. الشعور بالحرارة ...

هل سن اليأس عند الرجال أسطورة؟ - http://www.alwatan.com/graphics/2000/Nov/24.11/heads/ot16.htm
هل سن اليئس عند الرجال أسطورة؟. ... لذلك لا يوجد من يسمى برجل ما بعد سن اليئس. ...

no - http://news.masrawy.com/masrawynews/18072003/150657news.htmسن اليأس لدى الرجال .. خرافة
... عنوان بريد الراسل . عنوان بريد المرسل اليه . إرسل تعليق . سن اليئس لدى الرجال .. ...

السيدات فى سن اليأس أكثر عشقا ... - http://news.masrawy.com/masrawynews/06052003/136005news.htm
... السيدات في سن اليئس أكثر عشقا للحلويات (2003-05-06 02:00) اسطنبول : أفادت دراسة تركية نشرت ...

سن اليأس لدى الرجال - http://www.balagh.com/woman/jamal/pf0pz0hs.htm
... سن اليئس لدى الرجال. عن فضائية mbc باختصار اثبت العلم ان الرجل يتعرض لانقلابات ...

سن اليأس.. لا يلغى المتعة فى ... - http://www.balagh.com/woman/heih/pz1071s5.htm
... سن اليئس.. لا يلغي المتعة في المعاشرة الزوجية!! لمشاهدة الصورة بحجم اكبر إضغط علي الصورة ...

الإكتئاب فى سن اليأس - http://www.moh3.com/arabic/cultsher/depression.htm
الاكتئاب في سن اليئس. كيف تتصدين للإصابة بالاكتئاب في سن اليئس؟. أما بالنسبة ...

المرأة - أخبار اليوم - http://www.akhbarelyom.org.eg/akhbarelyom/issues/2936/1000.html
... بتوقيت القاهرة, 7:39:58 AM, الساعة -, 01/10/2, آخر تحديث يوم. المرأة, أزمة الطلاق بعد سن اليئس! ...

[Islam-Online- حواء و آدم- صوت النساء] - http://www.islamonline.net/iol-arabic/dowalia/adam-47/sawt-2.asp
... قضية ضمن ما طرح، وخاصة فيما يتعلق بالعلاج الهرموني لأعراض سن اليئس لدى السيدات بعد أن ...

Figure 2. Screen shots of the AMed Web directory

Table 1. Summary statistics of the two taxonomies

Statistics	SBiz taxonomy	AMed taxonomy
Total number of categories	295	232
Total number of Web pages	4,753	5,107
Average number of pages per categories	16.1	22.1
Maximum depth	5	5

EXPERIMENTAL DESIGN AND FINDINGS

In this section, we describe an experiment we conducted to evaluate the SBiz and the AMed Web directories and report the findings. The objective was to study how the proposed approach could assist human browsing and decision-making on specialized domains on the Web. Native Spanish and Arabic speakers participated in the experiment to evaluate each of the two Web directories by comparing it with a benchmark Web directory. We selected BIWE and Ajeeb as benchmarks to compare against our two directories respectively. BIWE has a more comprehensive Web directory than many existing Spanish Web directories (see description in the preceding section). Ajeeb (http://www.ajeeb.com/) has a general Arabic Web directory that has in its health related part 113 topics, 589 Web sites, and a depth of three levels. To our knowledge, these directories cover the most comprehensive resources in their respective domains on the Web.

Experimental Design

We recruited 19 Spanish students and 11 Arab students as volunteer subjects to evaluate the browse performance of the SBiz and AMed Web directories respectively. In the half-hour experiment, each subject performed a browse task using our directory (SBiz or AMed Web directories, depending on the subject's native language) and another browse task using the benchmark directory. We designed scenario-based browse tasks consistent with Text Retrieval Conference standards (Voorhees and Harman, 1997) to evaluate the performance of the Web directories. For example, a scenario for testing SBiz Web directory was "e-commerce in Latin America" and a browse task was "Find e-commerce sites from different Spanish speaking regions or countries." A scenario for testing AMed Web directory was "Prevention and treatment of cancer" and a browse task was "Find articles about healthy diet and cancer prevention." In each task, the subject used the Web directory to find addresses (represented by URL links) of relevant Web sites or pages. To further validate the relevance of tasks, we did a pilot test with three subjects for each Web directory before conducting the actual experiment.

Performance Measure

Although we did not impose any time limit on completing the tasks, we found that each subject spent an average 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 satisfaction ratings and comments on the system. A seven-point Likert scale was used in the rating.

We recruited a Spanish business consultant and an Arab microbiology researcher to act as experts to provide answers to the tasks. The Spanish business expert is a senior executive of a management consulting company in Mexico. Being a native Spanish speaker, he had had 24 years of experience in business development, raising capital, negotiations, finance, and strategic planning. Born and raised in Lebanon, the Arab medical expert was pursuing a Ph.D. degree in a tier-one research university in the United States. We measured the effectiveness by calculating precision, recall, and F value using the following formulae.

$$\text{Precision} = \frac{\text{Number of relevant URLs identified by the subject}}{\text{Number of all URLs identified by the subject}}$$

$$\text{Recall} = \frac{\text{Number of relevant URLs identified by the subject}}{\text{Number of relevant URLs identified by the expert}}$$

$$\text{F value} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$$

Precision reflects how well a Web directory helps a user to find relevant Web pages (expressed as URLs) to answer the questions in the tasks. Recall measures how well a Web directory helps a user to find the Web pages identified by the expert. A Web page can be an article or can represent a Web site. F value provides a balanced view between precision and recall. These are standard performance measures used in the information retrieval field. The subjects also provided demographic information, which was kept confidential in accordance with the Institutional Review Board Guidebook (Penslar, 2001).

Hypothesis Testing

We tested four hypotheses about the effectiveness and usability of using the Web directories. Because the Web directories developed by the proposed approach encompassed Web resources from different Spanish or Arabic regions, we believed that they should provide richer content than that of benchmark directories, which often have pop-up advertisements hindering browsing. Users could thus find relevant results more quickly from our directories. The hypotheses are listed below.

H1: The SBiz Web directory enables users to achieve higher effectiveness than the BIWE Web directory in performing browse tasks.

H2: The AMed Web directory enables users to achieve higher effectiveness than the Ajeeb Web directory in performing browse tasks.

H3: The SBiz Web directory achieves a higher user satisfaction rating than the BIWE Web directory.

H4: The AMed Web directory achieves a higher user satisfaction rating than the Ajeeb Web directory.

Experimental Results

Table 2 and Table 3 respectively show the statistical results of hypothesis testing and subjects' demographic profile. We found that both SBiz and AMed directories achieved better precision, recall, F value, and satisfaction rating than the benchmarks directories. Pairwise t-tests show that SBiz was significantly more effective than BIWE in terms of mean recall, F value, and satisfaction rating. AMed Web directory was significantly more effective than Ajeeb in terms of mean recall, F value, and satisfaction rating. H3 and H4 were thus confirmed.

Table 2. Statistical Results of Hypothesis Testing

Hypothesis	**Measure**	**SBizPort**		**BIWE**		***p*-value**	**Result**
		Mean	**S.D.**	**Mean**	**S.D.**		
H1	Precision	**0.79**	0.33	0.75	0.40	0.633	Partially confirmed
	Recall	**0.29**	0.14	0.17	0.13	0.002*	
	F value	**0.41**	0.19	0.16	0.12	0.000*	
H3	Satisfaction Rating[1]	**1.7**	0.73	3.0	1.67	0.009*	Confirmed
Hypothesis	**Measure**	**AMedPort**		**Ajeeb**		***p*-value**	**Result**
		Mean	**S.D.**	**Mean**	**S.D.**		
H2	Precision	**0.82**	0.13	0.45	0.52	0.059	Partially confirmed
	Recall	**0.39**	0.13	0.03	0.04	0.000*	
	F value	**0.51**	0.12	0.06	0.08	0.000*	
H4	Satisfaction Rating[1]	**2.2**	1.17	4.5	1.51	0.003*	Confirmed

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

* Alpha error = 0.05

Table 3. Subjects' demographic profile

Demographic information	**Spanish subjects (total: 19)**	**Arab subjects (total: 11)**
Country of origin	Mexico (12), USA (3), Panama (1), Puerto Rico (1), Colombia (1), Peru (1)	Lebanon (7), Morocco (1), Iraq (1), Mauritania (1), Jordan (1)
Education	Undergraduate (13), bachelor earned (2), master earned (3), doctorate earned (1)	Undergraduate (3), associate degree (1), bachelor earned (2), master earned (5)
Age range	18-25 (14), 26-30 (2), 31-35 (2), 41-50 (1)	18-25 (6), 26-30 (3), 36-40 (1), 41-50 (1)
Gender	Female (10), Male (9)	Female (3), Male (8)
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)	5-10 (1), 10-15 (3), 15-20 (1), 20-25 (2), 25-30 (1), 30-35 (1), > 40 (2)

However, we found no significant difference in the precisions between the two Spanish or the two Arabic Web directories. We believe that the filtering process of SBiz Web directory needs to be improved to remove irrelevant items and to make relevant items more apparent. Although the mean precision of AMed Web directory is much higher than that of Ajeeb directory, both directories had high variations of performance, making the difference insignificant. Yet the precisions are significantly different at a 6% alpha-error level. Therefore, H1 and H2 were partially confirmed.

Subjects also rated our Web directories to be significantly better than the benchmark Web directories, showing the superior usability of our approach in building Web directories. When asked about which Web directory they preferred to use to browse Spanish business-related Web resources, 13 out of 19 Spanish subjects said they would choose the SBiz Web directory. When asked about which Web directory they preferred to use to browse Arabic medical Web resources, all the 11 Arab subjects said they would choose the AMed Web directory.

These encouraging results show that our Web directories could help users locate relevant and high-quality information more effectively than existing Web directories, thus leading to more effective Web browsing and decision support. We believe that several aspects of our approach contributed to its superior performance: the high quality and comprehensive coverage of information sources, the cross-regional coverage of meta-searching, and the precise categorization.

CONCLUSIONS AND THE FUTURE

Building a high-quality Web directory without much expert knowledge and extensive human efforts has challenged developers of Web portals. As more users browse the Web in their native languages, better approaches to building Web directories in non-English languages are needed. In this paper, we have proposed a semi-automatic approach to building Web directory in different languages. The approach combines machine efficiency and human domain knowledge to assist in Web browsing and decision-making. Based on the approach, we developed the SBiz and AMed Web directories for the Spanish business and Arabic medical domains respectively. Experimental results show that the Web directories developed by the approach significantly outperformed existing Web directories in effectiveness and usability. We therefore conclude that *our approach is highly usable and can support construction of high-quality domain-specific Web directories in different languages*. These directories should bring value to end-users, system developers, and researchers of Web information seeking. Our future directions include refining the SBiz and AMed Web directories and testing our approach in other domains and languages. Also, the numbers of Spanish and Arab subjects in the experiment are limited because we had difficulty recruiting more subjects. We will expand the sample sizes to address this issue in the future.

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. Chang, S.J. and Rice, R.E. (1993) "Browsing: a multidimensional framework," in: *Annual Review of Information Science and Technology*, M.E. Williams (ed.), Information Today, Inc., Medford, NJ, 231-276.
3. Chuang, S.-L. and Chien, L.-F. (2003) "Enriching Web taxonomies through subject categorization of query terms from search engine logs," *Decision Support Systems*, 35, 1, 113-127.
4. Chung, W., Chen, H. and Nunamaker, J.F. (2005) "A visual framework for knowledge discovery on the Web: An empirical study on business intelligence exploration," *Journal of Management Information Systems*, 21, 4, 57-84.
5. Global Reach (2004) "Global Internet Statistics (by Language)," http://www.glreach.com/globstats/.
6. Kuhlthau, C. (1998) "Longditudinal case studies of the information search process of users in libraries," *Library and Information Science Research*, 10, 3, 257-304.
7. Kumar, R., Raghavan, P., Rajagopalan, S. and Tomkins, A. (2001) "On semi-automated Web taxonomy construction," *Proceedings of the Fourth International Workshop on the Web and Databases*, ACM Press, Santa Barbara, California, USA, 91-96.
8. Leita, C. (2004) "LII - Behind the Scenes," Librarian's Index to the Internet, http://lii.org/search/file/behind.
9. Marchionini, G. (1995) *Information seeking in electronic environments* Cambridge University Press, New York.
10. Marchionini, G. and Shneiderman, B. (1988) "Finding facts vs. browsing knowledge in hypertext systems," *IEEE Computer*, 21, 1, 70-80.
11. Mowshowitz, A. and Kawaguchi, A. (2002) "Bias on the Web," *Communications of the ACM*, 45, 9, 56-60.
12. Norton, L. "The Expanding Universe: Internet Adoption in the Arab Region," World Markets Research Centre, 3.
13. Penslar, R.L. (2001) "Institutional Review Board Guidebook," Office for Human Research Protection, U.S. Department of Health and Human Services, http://ohrp.osophs.dhhs.gov/irb/irb_guidebook.htm.
14. Sato, S. and Sato, M. (1999) "Automatic generation of Web directories for specific categories," *Proceedings of the AAAI Workshop on Intelligent Information Systems*, AAAI Press, Orlando, FL.
15. Spence, R. (1999) "A framework for navigation," *International Journal of Human-Computer Studies*, 51, 5, 919-945.

16. 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.
17. 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.
18. Wilson, T.D. (1999) "Models of information behavior research," *Journal of Documentation*, 55, 3, 249-270.