Literature Review

Compiled by Magda El-Sherbini

The following is a list of scholarly works that were written on cross-lingual information retrieval, multi-lingual name and subject access, information seeking behavior in multi-lingual environment, and social tagging (folksonomies).

On cross-lingual information retrieval

A Low Cost Machine Translation Method for Cross-Lingual Information Retrieval. By: Bracewell, David B.; Ren, Fuji; Kuroiwa, Shingo. Engineering Letters, 2008, Vol. 16 Issue 1, p160-165, 6p, 4 charts, 2 diagrams;

> Abstract: In one form or another language translation is a necessary part of cross-lingual information retrieval systems. Often times this is accomplished using machine translation systems. However, machine translation systems offer low quality for their high costs. This paper proposes a machine translation method that is low cost while improving translation quality. This is done by utilizing multiple web based translation services to negate the high cost of translation. A best translation is chosen from the candidates using either consensus translation selection or statistical analysis. Which to use is determined by a heuristic rule that takes into account that most web based translation services are of similar quality and that machine translation still produces relatively poor results. By choosing the best translation the method is able to increase translation quality over the base systems, which is verified by the experimentation.

An effective and efficient results merging strategy for multilingual information retrieval in federated search environments. By: Luo Si; Callan, Jamie; Cetintas, Suleyman; Hao Yuan. Information Retrieval, Feb2008, Vol. 11 Issue 1, p1-24, 24p, 10 charts, 1 diagram;

> Abstract: Multilingual information retrieval is generally understood to mean the retrieval of relevant information in multiple target languages in response to a user query in a single source language. In a multilingual federated search environment, different information sources contain documents in different languages. A general search strategy in multilingual federated search environments is to translate the user query to each language of the information sources and run a monolingual search in each information source. It is then necessary to obtain a single ranked document list by merging the individual ranked lists from the information sources that are in different languages. This is known as the results merging problem for multilingual information retrieval. Previous research has shown that the simple approach of normalizing source-specific document scores is not effective. On the other side, a more effective merging method was proposed to download and translate all retrieved documents into the source language and generate the final ranked list by running a monolingual search in the search client. The latter method is more effective but is associated with a large amount of online communication and computation costs. This paper proposes an effective and efficient approach for the results merging task of multilingual ranked lists. Particularly, it downloads only a small number of documents from the individual ranked lists of each user query to calculate comparable

document scores by utilizing both the query-based translation method and the document-based translation method. Then, query-specific and source-specific transformation models can be trained for individual ranked lists by using the information of these downloaded documents. These transformation models are used to estimate comparable document scores for all retrieved documents and thus the documents can be sorted into a final ranked list. This merging approach is efficient as only a subset of the retrieved...

Creating and Exploiting a Comparable Corpus in Cross-Language Information Retrieval. By: Talvensaari, Tuomas; Laurikkala, Jorma; Järvelin, Kalervo; Juhola, Martti; Keskustalo, Heikki. ACM Transactions on Information Systems, Feb2007, Vol. 25 Issue 1, p1-21, 21p, 8 charts, 1 diagram, 3 graphs;

Abstract: We present a method for creating a comparable text corpus from two document collections in different languages. The collections can be very different in origin. In this study, we build a comparable corpus from articles by a Swedish news agency and a U.S. newspaper. The keys with best resolution power were extracted from the documents of one collection, the source collection, by using the relative average term frequency (RATF) value. The keys were translated into the language of the other collection, the target collection, with a dictionary-based query translation program. The translated queries were run against the target collection and an alignment pair was made if the retrieved documents matched given date and similarity score criteria. The resulting comparable collection was used as a similarity thesaurus to translate queries along with a dictionary-based translator. The combined approaches outperformed translation schemes where dictionary-based translation or corpus translation was used alone.

Crossing language barriers in Europe: linking LCSH to other subject heading languages By: MacEwan, Andrew. Cataloging & Classification Quarterly, 2000, Vol. 29 Issue 1/2, p199-207, 9p;

Abstract: Describes research conducted by a study group representing four European national libraries (Swiss, German, French, and British) to examine the possibility of establishing multilingual thesaural links between the headings in the Library of Congress Subject Headings (LCSH) authority file and the authority files of the German indexing system SWD/RSWK and the French indexing system RAMEAU. Results demonstrate a high level of correspondence in main headings, as well as a number of issues requiring further investigation. The study group's findings led to recommendations on the scope for the development of a prototype system for linking the three Subject Heading Languages in the databases of the four institutions

Cross-Language Evaluation Forum - CLEF 2008. By: Peters, Carol. D-Lib Magazine, Nov/Dec2008, Vol. 14 Issue 11, p13-13, 1p;

Abstract: The article offers information on the 2008 Cross Language Evaluation Forum (CLEF) that was held on September 17-19, 2008 in Aarhus, Denmark. It states that the forum's objective is to promote research in multilingual system development. Topics include retrieval of multilingual textual document, multiple language question answering,

and retrieval of cross-language geographical information. Participants in the event include groups from Europe, South America, and Africa.

Multi-lingual name and subject access

Cross-lingual Name and Subject Access: Mechanisms and Challenges. By: Park, Jung-Ran. Library Resources & Technical Services, Jul2007, Vol. 51 Issue 3, p180-189, 10p, 2 charts, 1 diagram;

> Abstract: This paper considers issues surrounding name and subject access across languages and cultures, particularly mechanisms and knowledge organization tools (e.g., cataloging, metadata) for cross-lingual information access. The author examines current mechanisms for cross-lingual name and subject access and identifies major factors that hinder cross-lingual information access. The author provides examples from the Korean language that demonstrate the problems with crosslanguage name and subject access.

Images: indexing for accessibility in a multi-lingual environment -- challenges and perspectives. By: Ménard, Elaine. Indexer, Jun2009, Vol. 27 Issue 2, p70-76, 7p, 1 chart;

> Abstract: Elaine Ménard presents the results of a research project that sought to identify the differences, in a multilingual environment, between indexing a set of images representing common objects using either a controlled vocabulary approach, or a free or uncontrolled vocabulary approach. The differences are explored from the terminological, perceptual and structural points of view, Consideration is given to the factors that may have influenced the choice of indexing terms, such as language, indexing experience and indexing guidelines, and suggestions are made for further research.

Improving access to online multilingual resources by adopting the My Language Portal in the City of Greater Dandenong Libraries. By: Bogdanovic, Marijana; Johanson, Graeme. Australian Library Journal, May2007, Vol. 56 Issue 2, p135-151, 14p;

> Abstract: This paper reports on the implementation of ‘My Language Portal’ in the City of Greater Dandenong Libraries (CGDL), Melbourne, Victoria, through the development of a ‘My Language Portal Project Plan’ in 2006. It discusses how the aims of the designers of My Language Portal (MLP) are fulfilled in the exceptional, changing demographics of Dandenong. It provides a rationale for the adoption of MLP, by evaluating census and library statistics, and through local assessment of usability features. Wide consultation led to the creation of a user guide in the form of a fact sheet for users of CGD Libraries, and collaborations with IT and marketing staff are strongly recommended to facilitate smooth implementation. MLP offers a powerful online multilingual resource that bridges the inevitable gap in collections caused by an inability to immediately provide in-house appropriate resources for recently-arrived and diverse migrant communities. Analysis of service provision in Dandenong highlights the need for extra resources, for improved information literacy training, and marketing, and for increasing the number of public access terminals.

Library of Congress subject heading period subdivisions for West Asia and the Near East in general: some proposed additions By: Studwell, W E; Aggarwal, N K. Cataloging & Classification Quarterly, Fal 1983, Vol. 4 Issue 1, p35-52, 18p;

> Abstract: This essay is the third in a trilogy of studies on subject heading period subdivisions for Asia. Like the two previous essays, this one presents the concept that LC period subdivisions need considerable amplification to meet the needs of medium-sized, large, and specialized libraries. The set of three essays covers all areas of Asia except Asia in the Soviet Union. The same techniques used in the first two studies are continued here. Given in the left column are the period subdivisions, by multi-country region or country, that are printed in LC's subject heading publications through the June 1981 period. If there are no period subdivisions, the form subdivisions implied by LC policy are given. Given in the right column are suggested additions. The reasons and/or documentation for all additions follow each area. Excluded are the headings for all specifically named events which are not given under place names except by cross reference, e.g., Iran Hostage Crisis, 1979-1981. If an area, e.g., Turkey, already has many period subdivisions established by LC and few additions are proposed, only the subdivisions directly affected are given. If an area does not need subdivision for any reason, it is excluded with a brief explanation.

Multilingual Subject Access: The Linking Approach of MACS. By: Landry, Patrice. Cataloging & Classification Quarterly, 2004, Vol. 37 Issue 3/4, p177-191, 15p;

> Abstract: The MACS (Multilingual access to subjects) project is one of the many projects that are currently exploring solutions to multilingual subject access to online catalogs. Its strategy is to develop a Web-based link and search interface through which equivalents between three Subject Heading Languages-SWD/RSWK (Schlagwortnormdatei/Regeln für den Schlagwortkatalog) for German, RAMEAU (Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié) for French, and LCSH (Library of Congress Subject Headings) for English-can be created and maintained, and by which users can access online databases in the language of their choice. Factors that have led to this approach will be examined and the MACS linking strategy will be explained. The trend to using mapping or linking strategies between different controlled vocabularies to create multilingual access challenges the traditional view of the multilingual thesaurus.

Information seeking behavior in multi-lingual environment

Information seeking and use of high school students with diverse linguistic and cultural backgrounds in learning contexts. By: Kim, Sung Un. Information Research, Dec2008, Vol. 13 Issue 4, p24-24, 1p;

> Abstract: The article presents a study on the information seeking behavior of high school students with diverse linguistic and cultural backgrounds in the learning environment. The study was done at a public high school in New Jersey. Students' demographic information, such as origin, length of time residing in the U.S. or other countries, and languages spoken, was included in a questionnaire for the students. The study observes the high school students' information search habits as well as their researching ability

Learning the Languages. By: West, Jessamyn. Computers in Libraries, Nov/Dec2008, Vol. 28 Issue 10, p40-41, 2p, 3 diagrams;

Abstract: The article presents technical help for librarians on how to make libraries more welcoming to patrons who speak limited or no English. Software support for multiple languages is described for Microsoft, Apple, and Linux operating systems. The author looks at two aspects of languages and Internet use: the capabilities of Internet browsers and how websites display non-Western alphabets. Suggestions are offered on using online translation tools to make library signs and PC browser homepages for different languages. The author notes the value of knowing the library staff's language competencies and keeping up to date with contact information for libraries around the world. The article also includes a list of links to useful resources.

Speaking in tongues, Part 2: Foreign language KM technologies. (cover story) By: PEPUS, GREG. KM World, Jan2009, Vol. 18 Issue 1, p1-9, 3p, 3 color;

> Abstract: The article presents the second of a two-part series on foreign language knowledge management technologies that can help meet foreign language challenges. It focuses on BBN Technologies' Broadcast Monitoring System and some of the underlying components that make it work. It is a suite of technologies that have been integrated to provide a comprehensive capability to monitor, search and supply alerts based on the specific content in streaming audio and video.

Subject Access for Readers' Advisory Services: Their Impact on Contemporary Spanish Fiction in Selected Public Library Collections. By: Hall-Ellis, Sylvia D.. Public Library Quarterly, 2008, Vol. 27 Issue 1, p1-18, 27p, 4 charts;

> Abstract: Study findings suggest that access to Spanish language adult fiction through bilingual records in the OPAC is mutually beneficial for RA librarians and patrons. Subject access depends on local cataloging policies regarding enhancements for bibliographic records and catalogers' Spanish language proficiencies. Without incentives to enhance bibliographic records, local bilingual cataloging will continue but may not be shared. Reader advisory can be improved with the multicultural RA tools, multilingual RA websites, incentives to libraries for enhancements to non-English records, and linking individual bibliographic records in OPACS to reviews and comments for titles in languages other than English.

Subject heading syntax and 'natural language' nominal compound syntax By: Eichman, Thomas Lee. Subject heading syntax and 'natural language' nominal compound syntax, 1978;

> Abstract: Subject headings consisting of words or strings of words are interpreted by the user as a form of natural language. This pragmatic approach must be taken into account by producers of indexes using such subject headings

Survey on subject heading languages used in national libraries and bibliographies By: Heiner-Freiling, Magda. Cataloging & Classification Quarterly, 2000, Vol. 29 Issue 1/2, p189-198, 10p;

> Abstract: Surveys conducted during the last four years under the auspices of the International Federation of Library Associations and Organizations (IFLA) reveal that the Library of Congress Subject Headings (LCSH) is heavily used in national libraries outside the US, particularly in English-speaking countries. Many other countries report

using a translation or adaptation of LCSH as their principal subject heading language. Presents an analysis of the IFLA data, which also includes information on the classification schemes used by the libraries and whether or not the libraries have produced a manual on the creation and application of subject headings. Concludes with an Appendix showing the complete data from the 88 national libraries that responded to the surveys.

The RAMEAU/KABA Network: An Example of Multi-Lingual Cooperation. By: Kotalska, Barbara. Slavic & East European Information Resources, 2002, Vol. 3 Issue 2/3, p149, 8p;

Abstract: In 1991 work began at Warsaw University Library to establish an online catalog. One of the first issues faced was choosing an indexing language and subject heading system that would meet the requirements of, and use the capabilities of, an automated library system with authority control. From a variety of types of indexing languages reviewed, a subject heading system was chosen, and it was decided to make it compatible with the two most important and widely-used foreign systems, the American Library of Congress Subject Headings (LCSH) and the French Répertoire d'autorité-matière encyclopédique et alphabétique unifié (RAMEAU). The new Polish subject heading system was called KABA, "Katalogi automatyczne bibliotek akademickich" (Academic libraries' automated catalogs), and is now used by many major academic libraries in Poland. KABA headings, written in USMARC format, can be automatically translated into RAMEAU or LCSH headings. The author describes the procedure for creating and verifying new headings by member libraries.

Translating the Libraries: A Multilingual Information Page for International Students. By: McClure, Jennifer; Krishnamurthy, Mangala. Southeastern Librarian, Spring2007, Vol. 55 Issue 1, p26-31, 6p;

Abstract: The article discusses the difficulties of international students in American campuses and the effectiveness of multilingual translations of an information page to enhance communication to and the success of international students. The University of Alabama created an International Students' Web page offering library tours/orientation in Chinese and Spanish, 2 of the most prominent foreign languages by hiring international student translators while editing was done by teaching faculty.

Translation disambiguation for cross-language information retrieval using context-based translation probability. By: Kishida, Kazuaki; Ishita, Emi. Journal of Information Science, 2009, Vol. 35 Issue 4, p481-495, 15p;

Abstract: Disambiguation between multiple translation choices is very important in dictionary-based cross-language information retrieval. In prior work, disambiguation techniques have used term co-occurrence statistics from the collection being searched. Experimentally these techniques have worked well but are based upon heuristic assumptions. In this paper, a theoretically grounded alternative is proposed, one which uses sense disambiguation based upon context terms within the source text. Specifically this paper introduces the concept of translation probabilities incorporating a context term and extends the IBM Model 1 for estimating context-based translation probabilities from

a sentence-aligned bilingual corpus. Experimental results in English to Italian bilingual searches show significant performance improvement of the context-based translation probabilities over the case without any disambiguation

Unearthing Archaeology: A Study of the Recent Coverage of Selected English-Language Archaeology Journals by Multi-Subject Indexes and by Anthropological Literature. By: Tyler, David C.; Yang Xu; Nimsakont, Emily Dust. Behavioral & Social Sciences Librarian, 2009, Vol. 28 Issue 3, p100-144, 45p, 2 charts, 3 graphs;

Abstract: Librarians, faculty, and professional researchers, and students already encounter difficulties in locating journal articles for the field of archaeology, yet, in the current budgetary climate, librarians needing to reduce subscription costs may be tempted to cancel smaller, discipline-specific indexes in favor of large multi-subject indexes with broad coverage. This study examines and compares the coverage provided to 208 archaeology and archaeology-related journals and magazines by six multi-subject indexes and by anthropology's primary index, Anthropological Literature, over a twenty year period

Social tagging or folksonomies

Arch, Xan, "Creating the Academic Library Folksonomy: Put Social Tagging to Work at Your Institution," C&RL News, 68, no. 2 (Fall 2007): 80-1.

Beard, Kurt A., comment on "When Tags Work and When They Don't: Amazon and LibraryThing," Thingology Blog, comment posted February 21, 2007, http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php (accessed July 8, 2007).

Blood, Rebecca Blood, "I've Noticed a Slight Problem with the Technorati Tagging System," Rebecca's Pocket, http://www.rebeccablood.net/archive/2005/01.html#11technorati (accessed July 13, 2007).

Boyd, Danah, "Issues of Culture in Ethnoclassification/Folksonomy," Many 2 Many Blog, posted January 28, 2005, http://many.corante.com/archives/2005/01/28/issues_of_culture_in_ethnoclassificationfolksonomy.php (accessed July 9, 2007).

Dye, Jessica, "Folksonomy: A Game of High-tech (and High-stakes) Tag", EContent, 29 no. 3 (April 2006): 38-43.

Golder, Scott A. and Bernardo A. Huberman, "The Structure of Collaborative Tagging Systems," http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf (accessed July 9, 2007).

Guy, Marieke and Emma Tonkin, "Folksonomies: Tidying up Tags?" D-Lib Magazine, 12, no.1 (January 2006), http://www.dlib.org/dlib/january06/guy/01guy.html (accessed July 8, 2007).

Hammond, Tony, Timo Hannay, Ben Lund and Joanna Scott, "Social Bookmarking Tools (I)," D-Lib Magazine 11 no. 4 (April 2005): 4.

Kome, Sam H., "Hierarchical Subject Relationships in Folksonomies," University of North Carolina at Chapel Hill School of Information and Library Science, (MS Thesis, University of North Carolina at Chapel Hill, 2005) http://hdl.handle.net/1901/238 (Accessed March 2008).

Lambe, Patrick, "Folksonomies and Rich Serendipity," Green Chameleon Blog,"comment posted on October 20, 2006, http://www.greenchameleon.com/gc/blog_detail/folksonomies_and_rich_serendipity/ (accessed July 13, 2007).

Lawson, Steve, comment on "When Tags Work and When They Don't: Amazon and LibraryThing," Thingology Blog, comment posted February 21, 2007, http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php (accessed July 8, 2007).

Mathes, Adam, "Folksonomies – Cooperative Classification and Communication Through Shared Metadata," Graduate School of Library and Information Science, http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html (accessed July 9, 2007).

Sierra, Tito, comment on "When Tags Work and When They Don't: Amazon and LibraryThing,"

Thingology Blog, comment posted February 21, 2007, http://www.librarything.com/thingology/2007/02/when-tags-works-and-when-they-dont.php (accessed July 8, 2007).

Spiteri, Louise F., "The Use of Folksonomies in Public Library Catalogues," The Serials Librarian, 51, no. 2 (2006): 75-89.

University of Pennsylvania Libraries, "What is PennTags?" University of Pennsylvania Libraries, http://tags.library.upenn.edu/help/ (accessed on July 13, 2007).

West, Jessamyn, "Subject Headings 2.0: Folksonomies and Tags," Library Media Connexion, 25, no. 7 (April / May 2007): 58-9.