Literature Review of Interactive Cross Language Information Retrieval Tools

Farag Ahmed and Andreas Nurnberger
Faculty of Computer Science, Otto-von-Guericke-University, Germany

Abstract: *The unprecedented rise of multilingual resources afforded by the exponential web growth demands the development of communication technologies in order to eliminate the barriers between languages. More comprehensive tools to overcome such barriers, such as machine translation and cross-lingual information retrieval applications, are nowadays in strong demands. In this paper, we present an overview of the literature of interactive Cross Language Information Retrieval (CLIR) tools and discuss their limitations. In addition, possible future research directions of interactive CLIR tools are discussed.*

Keywords: *CLIR, CLIR interaction tools, query disambiguation.*

Received July 6, 2010; accepted October 24, 2010

1. Introduction

In a time of wide availability of communication technologies, language barriers are a serious issue to world communication and to economic and cultural exchange. More comprehensive tools to overcome such barriers, such as machine translation and cross-lingual information retrieval applications, are currently, in strong demand. The increasing diversity of the Internet web sites has created millions of multilingual resources in the World Wide Web. At a first glance, it therefore seems that more information can be retrieved by non-English speaking people. However, even users that are fluent in different languages face difficulties when they try to retrieve documents that are not written in their mother tongue or if they would like to search documents in all languages they can speak in order to cover more resources with a single query. The issues of Cross Language Information Retrieval (CLIR) have been discussed for several decades. In the early seventies experiments for retrieving information across languages were first initiated by [11]. Currently, CLIR issues are addressed in several evaluation forums such as TREC, CLEF and NTCIR, while each of them covers different languages: TREC includes Spanish, Chinese, German, French, Italian, and Arabic; CLEF includes French, German, Italian, Swedish, Spanish, Dutch, Finnish, and Russian; and NTCIR includes Japanese, Chinese and Korean. Finding the most effective way to bridge the language barrier between queries and documents is the central challenge in CLIR [12]. CLIR systems allow the user to submit the query in one language and retrieve the results in different languages and thus provide an important capability that can help users to meet this challenge. In addition to the classical IR tasks, CLIR also requires that the query (or the documents) be translated from one language into another. Query translation is the most widely used technique for CLIR due to its low computationally cost for translation compared to the effort of translating a large set of documents. However, the query translation approach suffers from translation ambiguity as queries are often short and do not provide rich context for disambiguation. In the following, we give detailed overview of the CLIR interaction tools.

2. CLIR Interaction Tools

The design of a multilingual system faces specific challenges regarding the best way of handling multiple languages, best query translations and presentational requirements for the user. In the past, most research has been focused on the retrieval effectiveness of cross language systems through IR test collection approaches [4], whereas few researchers focused on the user interface requirements with respect to the multilingual retrieval task [9]. Despite the clear effort which has been directed toward retrieval functionality and effectiveness, only little attention was paid to developing multilingual interaction tools, where users are really considered as an integral part of the retrieval process. One potential interpretation of this problem is that users of CLIR might not have sufficient knowledge of the target languages and therefore they are usually not involved in multilingual processes [10]. However, the involvement of the user in CLIR systems, by reviewing and amending the query, have been studied, e.g., the New Mexico State University Keizai system [9], the German Research Center for Artificial Intelligence's (DFKI) MULINEX system [5], a Multilingual Information Retrieval Tool UCLIR [1], the Maryland Interactive Retrieval Advanced Cross-Language Engine (MIRACLE) system [8],

MultiLexExplorer [6] and recently our CLIR Interactive Tool "multi Searcher" [2]. In the following we describe these CLIR tools in detail.

2.1. Mulinex

Mulinex supports cross-lingual search by giving the users possibilities to formulate, expand and disambiguate queries. Furthermore, the users are able to filter the search results and read the retrieved documents by using only their native language. Mulinex performs the multilingual functionality based on a dictionary-based query translation. Besides the cross lingual functionality, where the query is submitted in one language and the retrieved documents are presented in another language, Mulinex provides the automatic translation of documents and their summaries. In Mulinex, three languages are supported, French, German, and English. In Mulinex, the CLIR process is fully supported by the translation of the queries, documents and their summaries. Hereby, users do not need to have any knowledge about the target language. Mulinex provides a lot of functionality to support the retrieving of the documents in multilingual collections. Examples of these functionalities are translation of the user's query, interactive disambiguation of the query translation, interactive query expansion, on-demand translation of summaries and search results, etc., The Mulinex interface is available in three languages English, German, and French. Since the search engine queries are usually between 2.4 and 2.7 in length [6] which typically does not provide enough context for automatic disambiguation, Mulinex using "query assistant" provides an opportunity for interactive query translation disambiguation. This task is performed by the "query assistant" by performing the backtranslation. The translated query terms are translated back into the original query language. However, this approach has some clear limitations. When no synonyms can be found in the dictionary, the technique is not helpful; and significant homonymy in the target language can result in confusing back translations [8]. In Mulinex, the backtranslation concept is used for expanding the original query with potentially relevant terms. The query term translation is translated back to the original query language; the result of this step is having a list of possible translation in the query's original language. The user, in this case, can select some of these translation alternatives, in order to expand the user query. For example, the user submits the query, "fair", in English. The system provides the user with alternative translations in French and German. For French, the system provides the following translations: ("blond", "moral", "marché", "kermesse", "juste", "foire" and "équitable"). For German the system provides the following eight translations: ("Jahrmarket", "Messe", "blond", "gerecht", "hübsch", "mittelmäßig", "ordentlich" and "schön") shown in Figure 1. In order to expand the query, the system translates back the translated user query terms. The result of this step is having translation alternatives in the user's original query language. For example, the back translation alternatives for the French translation "marché" are ("bazaar", "walked", "sales activities", "marketplace", "market" and "fair") and the back translation alternative for the French translation "foire" are ("bazaar", "trade fair", "market" and "fair"). Based on the translation alternatives provided by the system, the translation "sales activities" and "trade fair" can be selected by the user as relevant expanded terms to the original query "fair".

Figure 1. MULINEX query assistance.

2.2. Keizai

The goal of the Keizai project is to provide a Web-based cross-language text retrieval system that accepts the query in English and searches Japanese and Korean web data. Furthermore, the system displays English summaries of the top ranking retrieved documents. In Keizai the query terms are translated into Japanese or Korean languages along with their English definitions and thus this feature allows the user to disambiguate the translations shown in Figure 2. Based on the English definitions of the translated query terms, the user who does not understand the Japanese or Korean language can select the appropriate translation, out of several possible translations. Once the user selects those translations whose definitions are consistent with the information needed, the search can be performed. Only documents that are relevant to the selected translations will be retrieved. For each retrieved document in Japanese or Korean, an English summary along with a target document language summary will be displayed in the Keizai interface. Keizai investigates the effectiveness of representing the retrieved documents together with small images, which they call "Document Thumbnail Visualizations". Using this document representation, the retrieved documents are

retained with a familiar shape and format and thus the user can see how the query terms are distributed in the retrieved documents. Using this technique the authors investigated the potential advantage of the representation of the documents as one image within the context of different interactive text retrieval tasks. In Keizai, the authors could show that the visualization improved recall and efficiency.

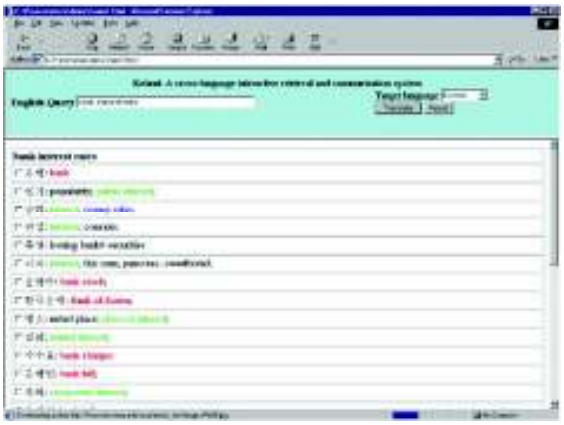

Figure 2. Keizai query term selection.

2.3. UCLIR

In UCLIR, the Arabic language was included. The system performs its task in any of the following three different modes: the first mode, using a multilingual query (query can consist of terms of different languages), the second mode using an English query without user involvement in the multilingual query formulation, the third mode using an English query with user involvement in the formulation of the multilingual queries. The first system mode: Multilingual query, in this mode the system accepts a query which consists of terms of various languages. The system will retrieve the relevant documents regardless of the query term language. The documents in the entire multilingual collection, those relevant to one of the query terms, will be retrieved. The second system mode: English query: noninteractive approach, this mode is based on the use of a set of bilingual dictionaries for translating an English query into the different target languages. First, for the English query term a set of possible translations will be obtained from the bilingual dictionaries. Second, the set of possible translations will be compared with an index word list (obtained from the system's entire multilingual resource); the translations which are not in the index word list will be eliminated from the query. The filtered query then can be used to retrieve the relevant documents from the system's entire resource and these retrieved documents are then displayed to the user in the system interface. The third system mode: English query: interactive approach, in this system mode, the user is involved in the selection of appropriate translations. The same as in the second mode, a set of possible translations will be obtained from the bilingual dictionaries and compared with the index word list; the translations which are not in the index word list will be eliminated from the query. The rest will be kept and presented to the user in the system interface along with their English translation beside other information e.g., part of speech shown in Figure 3. At the end, the user selects the appropriate translation out of the filtered translation list. The selected multilingual terms then can be used to form the multilingual query which is then submitted to retrieve the relevant documents from the system's entire multilingual resource. After the retrieval process is performed, the relevant retrieved documents can be then translated into English. To perform the document translation, two approaches are used. The first approach is word-level translation, where the user can click on the selected word and this word will be translated using the dictionary and displayed as a pop-up view to the user with its lexical information. The second approach is a document-level translation, where the whole retrieved document, using a translation system, is translated into English. Similar to Keizai, UCLIR uses "Document Thumbnail Visualizations". The retrieved documents are retained with familiar shape and format which make it possible for the user to see how the query terms are distributed in the retrieved documents. Although the system in the second mode automates the process of the appropriate translation selection by comparing a set of possible translations with an index word list (the translations which are not in the index word list will be eliminated from the query). However, this can include an irrelevant translation to the user query since it is possible that not all translations can be relevant to the original query term.

Figure 3. UCLIR document thumbnail visualizations.

2.4. MIRACLE

In order to support the interactive CLIR, the system uses the user-assisted query translation. The user assisted-query translation feature supports the user to select the correct translation. However, there might be a case when the user might delete a correct translation. The system reacts, in that the searcher can see the

effect of the choice and have possibilities to learn better control of the system. This is done by providing the following features, the meaning of the translation (loan word or proper name), using back translation, a list of possible synonyms are provided. Translation examples of usage are obtained from translated or topically-related text.

In MIRACLE, there are two types of query translations, fully automatic query translation (using machine translation) and user-assisted query translation. In fully automatic translation the user can be involved only once. After the system translates the query and retrieves the search results, the user can refine the query if he/she isn't satisfied after examining the search results. In the user-assisted query translation, four possible refinement steps give the user an opportunity to be involved in the translation process. First, based on evidence about the meanings of the proposed translations by the system, the user has an opportunity to deselect some of the proposed translations before the search can be performed. Second the user can reform the query based on evidence about the meanings of the proposed translations. Third, the user can reform the query based on examining the search results. Fourth, in case the search result doesn't satisfy the user's needs, the user has a possibility to deselect/reselect the translations. In other words, the user submits his query; the system provides him/her with translation alternatives. Before the search can be performed, the user has an opportunity to deselect some of the proposed translations. The user has an opportunity to refine his/her query based on evidence about the meanings of the proposed translations by the system. After the search is performed, the system provides the user with the search results shown in Figure 4. If the user is satisfied with the search result then there will be no further actions by the system. In contrast, based on examining the search result, the user has two opportunities: refine his/her query and perform a new search or deselect/reselect a translation out of the translation alternatives proposed by the system.

Figure 4. MIRACLE query assistance.

The interaction between the system and the user, gives the user possibilities to see the effect of his/her decision (selection, deselection of the translation or query refinement) in that the user can cycle the search till it satisfies his/her needs. A very important aspect in MIRACLE, is that the system provides the user with immediate feedback in response to any action, which gives the user an important opportunity to refine his/her search. The rapid integration of new languages was taken into account in the design of the MIRACLE system. The query language is always English, in MIRACLE. However, language resources that are available for English can be leveraged, regardless of the document language. Currently, MIRACLE works with a simple bilingual term list. However, it is designed to readily leverage additional resources when they are available.

2.5. MultiLexExplorer

The goal of the MultiLexExplorer tool is to support multilingual users in performing their web search. Furthermore, the MultiLexExplorer supports the user in disambiguating word meanings by providing the user with information about the distribution of words in the web. The tool allows users to explore combinations of query term translations by visualizing EuroWordNet[1] relations together with search results and search statistics obtained from web search engines. Based on the EuroWordNet, the tool supports the user with the following functionality:

- Exploring the context of a given word in the general hierarchy.
- Searching in different languages, e.g., by translating word senses using the interlingual index of EuroWordNet.
- Disambiguating word sense for combinations of words.
- Provide the user with the possibility to interact with the system i.e., changing the search word and the number of retrieved documents.
- Expanding the original query with extra relevant terms.
- Automatically categorizing the retrieved web documents.

As Figure 5 shows, the different parts of the user interface are labelled. In (label a1) the user expresses his/her needs. In addition, in (label a1) the user can select the source language he/she would like to use with the help of the language resource to explore the context of the query. In (label e), the user has the possibility to interact with the tool in modifying the query context by selecting different linguistic relations i.e., Hypernym or Hyponym. In order to conduct a cross lingual search, the user can select the target

[1]http://www.illc.uva.nl/EuroWordNet/

language in (label d). The tool then automatically provides translations of all possible source language senses in the target language.

This translation is performed, based on the interlingual entries of EuroWordNet. After the translation is performed, the tool retrieves the number of relevant documents. The number of documents is then presented to the user in a visualization manner (circle visualization, which shows the distribution of document hits of the translations). The larger the number of retrieved documents is, the bigger the circle is (see label c). The tool automatically searches for all combinations between all senses including synonyms. With a mouse click, the user can display the relevant documents to the selected translation on the tool interface (label f), based on the displayed "circles". In (label c1) the user has the possibility to change the search context. For example, with a right mouse click, the user can select a new word (given by the linguistic relation) and replace it with the originally searched word. For example, the original query was (haus tür), with a right mouse click the user can select a new term (gebäude). In this case, the tool reacts by automatically repeating the same process which was done for the original query. This will involve translation, disambiguation and the visualization of the searched terms. Furthermore, in (label c) another important aspect in the MultiLexExplorer, is that the user is given the possibility of removing any term/terms that are not of interest. In addition, the user can select any desired term/terms as expansion term/terms to the original query. These expanded term/terms will be presented along with the original query terms in (label a2). As shown in (label g), the tool provides the user with different categorization techniques to categorize the huge search results for better navigation. In MultiLexExplorer, very useful aspect was taken into account. The information is expressed in a visually attractive manner, which makes the user's task easier. For example, in the retrieved document hits, the user doesn't need to check numbers, instead he just checks the "circle" (the bigger the circle, the greater the retrieved document hits are) that expresses the retrieved document hits.

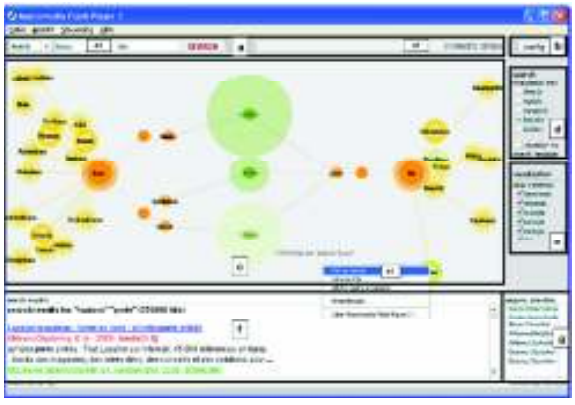

Figure 5. MultiLexExplorer interface.

2.6. Multi Searcher

The "multi Searcher" tool provides users with interactive contextual Information that describes the translation in the user's own language, in order for him/her to have a certain degree of confidence about the translation. In order to consider users as an integral part of the retrieval process, the tool provides the users with possibilities to interact with it, where they can select relevant terms from the contextual information in order to improve the translation and thus improve the CLIR process. "multi Searcher" deals with two issues concerning the CLIR. Firstly, there is translation ambiguity, where one word in one language can have several meanings in another language. "multi Searcher" make the use of automatic translation with disambiguation (in the previously mentioned tools, the disambiguation done by the user). Secondly, the user's lack of knowledge in the target language, where the tool supports the user by providing him/her with interactive contextual information that describes the translation in his/her own language. Based on the interactive scenario, the user has possibilities of improving the translation and thus improving the CLIR process. Due to the availability of the language resources needed (dictionary[2] and parallel corpora[3]) English was selected as test languages. In "multi Searcher", the query is automatically translated. The automatic translation method consists of two main steps: First, using an Arabic analyzer, the query terms are analyzed and the senses (possible translations) of the ambiguous query terms are identified. Second, the most likely correct senses of the ambiguous query terms are selected based on co-occurrence statistics. In multi searcher, we used cohesion scores for possible translation-candidate pairs (translation combinations) to resolve the translation ambiguity of the user query terms. The translation process starts by translating the query terms; a set of possible translations of each of the query terms are obtained from the dictionary. Based on the translation sets of each term, sets of all possible combinations between terms in the translation sets are generated. Using co-occurrence data extracted from monolingual corpora[3], the translations are then ranked based on a cohesion score computed using Mutual Information [3]. After Translation, the translated user query is looked up in the target language documents index (one translation after the other), in order to obtain the relevant documents (contextual information), for the translation. In order to get the equivalent documents in the source language, the parallel corpora is queried. Since it is possible that some retrieved documents will be very similar-which would result in duplicate contextual information-the

[2]http://www.nongnu.org/aramorph/
[3]www.ldc.upenn.edu/

documents retrieved from the source language are automatically grouped and contextual information is selected only once from each cluster. Using the Contextual Information Analyzer, the final selected contextual information is not provided to the user as raw text, but instead, it will be presented as a classified representation of each contextual information term will: each term of the contextual information is color-coded according to its related type and can be selected as a disambiguating term (the user's query terms are in green, suggest terms, by the tool based on highly frequent co-occurrences in the context of the query are in bold blue and underlined, all remaining terms are blue except stop words that are not selectable and are black). Let's consider the Arabic query "دين الحكومة" "dyen alh.Kwmh". The query term "الحكومة" "alh.Kwmh" has two translations "the government" or "the administration", while the other term "دين" "dyen" has several possible translations e.g., "Religion" or "Debt". Based on the MI score translation alternatives are displayed in ranked order together with their contextual information. Thus the user has the possibility to select the suitable translation. Here, the translations provided by the system "the government religion" and "the government debt" are correct even though they are used in a different context. This is due to the fact that "government" appears frequently in the context of "religion" or "debt". As shown in Figure 6, the user is interested in the second ranked translation "debt government". Using the contextual information, the user can select one or more terms to improve the translation. To simplify the user's task, the tool automatically proposed relevant terms highlighted in bold blue and underlined, e.g., "payment", "financial", "lending", "loan". Once the user selects, for example, the interactive term "إقراض, ¯aqr¯ad" "loan or lending", the tool re-translates the modified query and displays the new translations "debt government loan", "debt government lending" and "debt administration loan", to the user. Finally using search engine integrated web services, the user can, with a simple mouse click, confirm the translation which will then be sent to his favorite search engine, retrieving the results and displaying them to the user on the tool interface.

Figure 6. "multi Searcher": The translation alternatives with their contextual information.

3. Discussion

We studied in detail the state-of-the art CLIR interaction tools that can be used to support the user to perform his/her cross lingual search. All of the previously mentioned tools consider the user as an integral part of the retrieval process, in that the user can plays an essential role in improving the search. One notices that there are some insufficiencies in supporting the user when he/she wants to retrieve documents written in a language which differs from the language he/she speaks. A possible reason for this deficit is that the user is requested to perform the translation disambiguation process. For example, using Keizai, MULINEX or UCLIR, the user is requested to check all translation alternatives for each query term with the dictionary definition, in order to select the correct translation. However, the disambiguation process needs full concentration from the user, in that the user has to scroll up all translation alternatives in order for him/her to select relevant expanded terms. This can be very laborious especially for query terms that have abundant possible translations e.g., based on the given example in MULINEX, the user has to check 82 translation alternatives in order to select the appropriate translations out of them. In addition, the previously mentioned tools rely on the use of a bilinligual dictionary or WordNet for translation as well as for disambiguation. However, bilingual dictionaries or WordNet in which the definitions of source language are available for each translation for the target languages are very rare and very laborious. Despite the good visual and functional design of MultiLexExplorer, it relies on the use of WordNet, which only employs a limited number of languages. Furthermore, no automatic translation is integrated into the tool to assist the user in checking many word sense combinations. In "multi Searcher" we minimize the task of the user by checking all translation alternative for each query term by the automatically translation of the user query. The translation as well as the disambiguation process is done automatically through the use and the enhancement of statistical methods e.g., mutual information where the tool provides the user with ranked translation alternative with their contextual information. Using this contextual information, which is displayed in the user's own language, the user can have a certain degree of confidence about the translation. Since bilingual dictionaries or WordNet, in which the definitions of the source language are available for each translation for the target languages are very rare since the creation of it is very laborious, "multi Searcher" used a parallel corpora which can be obtained easily and for free. The ideal candidates mainly consist of texts available on the Internet which are growing consistently. For example, two sources that meet these requirements is the united nation text collection that can be downloaded from the united

nation web site[4]. These text collections are documents written in six languages (Arabic, Chinese, English, French, Russian and Spanish) and belong to different institutions in the United Nation, i.e., Security Council, Economic and Social Council. The other source is the European Parliament Proceedings Parallel Corpus[5] which is available for 11 European languages and is consistently growing.

4. Conclusion and Future Work Directions

In this paper, we presented an overview of the literature of cross language interaction tools. Based on this review, one notices that there is a clear lack of support when the user wants to retrieve documents written in a language which differs from the language he/she speaks. A possible reason for this deficit is that the user is requested to perform the translation disambiguation process. For example, using Keizai, MULINEX or UCLIR, the user is requested to check all translation alternatives for each query term with the dictionary definition, in order to select the correct translation. In multi Searcher, the disambiguation process is done automatically through the use of statistical methods e.g., mutual information. Using contextual information, which is displayed in the user's own language, the user can have a certain degree of confidence about the translation. Since bilingual dictionaries usually lack in detecting named entities, numbers, technical terms and acronyms, using Named Entity Recognition approaches e.g., [14] can enhance the performance of multi searcher.

Future research, for improving the CLIR interaction tools, can focus on improving the automatic translation so that the user can rely on the translation alternatives displayed by the system. Since the user is considered an integral part of the CLIR tools, there is an urgent need to involve deeper visualization techniques to display the translation as well as the contextual information in smooth way. This will help the user to perform the task efficiently, in that this can be an encouraging factor for the user, to use these tools. All of the CLIR tools rely totally on the use of language resources, which are not available for all languages. The absence of these translation resources between the source and the target language is a real obstacle in doing more research in this direction. Recent research on machine translation focuses on using transitive translation [7], which is said to be useful when direct translation from the source to the target language is unavailable. However, transitive translation can produce some fatal errors introduced by each translation phase, e.g., high ambiguity due to several phases of translation or due to the selection of the pivot language. Different important aspects related to transitive translation need to be researched e.g., which criteria need to be studied in order to select the suitable pivot language. These criteria can be grammatical relatedness and word level ambiguity for the pivot language/languages. Detailed studies need to be conducted in order to deduce which features in the pivot language leads to good translation quality. Using more than one pivot language can be beneficial and can lead of reducing the ambiguity. For example, translating from German to Catalan, using Spanish and Portuguese as pivot languages, the German phrase is first translated into Spanish and Portuguese, ranked translations will be obtained then the Spanish and Portuguese translation is translated individually, to Catalan. Only the intersection found in both ranked translations will be selected and the unique translations will be omitted.

[4]http://www.un.org
[5]http://www.statmt.org/europarl/

References

[1] Abdelali A., Cowie J., Farwell D., and Ogden W., "Uclir: A Multilingual Information Retrieval Tool," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Arti ficial*, vol. 8, no. 22, pp. 103-110, 2003.

[2] Ahmed F. and Nürnberger A., "Can we Support People to Get Information From Text They Can't Read or Understand?," *in Proceedings of the 33rd Annual ACM SIGIR conference in Research and Development in Information Retrieval*, Switzerland, pp. 837-838, 2010.

[3] Ahmed F., Nürnberger A., and Luca E., "A Corpus-Based Approach to Improve Arabic/English Cross-Language Information Retrieval," *in Proceedings of the Corpus Linguistics Conference*, Uk, pp. 195-213, 2009.

[4] Braschler M., Peters C., and Schäuble P., "Cross-Language Information Retrieval Clir Track Overview," *in Proceedings of the 8th Text Retrieval Conference*, pp. 25-33, 2000.

[5] Capstick J., Diagne A., Erbach G., Uszkoreit H., Leisenberg A., and Leisenberg M., "A System for Supporting Cross-Lingual Information Retrieval," *International Journal of Information Processing and Management,* vol. 36, no. 2, pp. 275-289, 2000.

[6] Luca E., Hauke S., Nürnberger A., and Schlechtweg S., "MultiLexExplorer-Combining Multilingual Web Search with Multilingual Lexical Resources," *in Proceedings of the Combined Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken Dialogue Systems*, Germany, pp. 17-21, 2006.

[7] Gabrilovich E., Broder A., Fontoura M., Joshi A., Josifovski V., Riedel L., and Zhang T., "Classifying Search Queries Using the Web As a

Source of Knowledge," *Journal of ACM Transactions on the Web.* vol. 3, no. 2, pp. 1-28, 2009.

[8] Mayfield J. and McNamee P., "Triangulation Without Translation," *in Proceedings of The 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, USA, pp. 490-491, 2004.

[9] Oard D., He D., and Wang J., "User-Assisted Query Translation for Interactive Cross-Language Information Retrieval," *International Journal of Information Processing and Management,* vol. 44, no. 1, pp. 181-211, 2008.

[10] Ogden C. and Davis W., "Improving Cross-Language Text Retrieval with Human Interactions," *in Proceedings of the 33rd Hawaii International Conference on System Sciences, IEEE Computer Society*, USA, vol. 3, pp. 3044, 2000.

[11] Petrelli D., Beaulieu M., Sanderson M., Demetriou G., Herring P., and Hansen P., "Observing Users, Designing Clarity: A Case Study on The User-Centered Design of A Cross-Language Information Retrieval System," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 10, pp. 923-34, 2004.

[12] Salton G., "Experiments in Multi-Lingual Information Retrieval," *Technical Report*, Computer Science, Cornell University, 1972.

[13] Yang Y. and Ma N., "Cmu in Cross-Language Information Retrieval At Ntcir-3," *in Proceedings of The 3rd NTCIR Workshop on Research in Information Retrieval-Automatic Text Summarization and Question Answering*, 2002.

[14] Benajiba Y., Diab M., and Rosso P., "Using Language Independent and Language Specific Features to Enhance Arabic Named Entity Recognition," *The International Arab Journal of Information Technology*, vol. 6, no. 5, pp. 463-471, 2009.

Farag Ahmed is working as a scientific researcher in the group of Data and Knowledge Engineering in the Faculty of Computer Science, Otto-von-Guericke-University in Magdeburg, Germany since February 2006. He holds two Masters degrees, a Masters in computer science from Heidelberg University of Applied Science, Heidelberg, Germany and a Masters of Science Data and Knowledge Engineering from Otto-von-Guericke-University in Magdeburg, Germany. Besides his research work, he is teaching assistant on Master's level, gives seminar courses and is a PhD student in his last semester.

Andreas Nürnberger studied computer science and economics at the Technical University of Braunschweig, Germany, received his PhD in computer science from the University of Magdeburg, Germany in 2001 and worked after this as postdoctoral researcher at UC Berkeley on adaptive soft computing and visualisation techniques for information retrieval systems. Since 2003 he was assistant professor for information retrieval at the University of Magdeburg and in 2007 he was called to a chair of a tenured professor for 'Data and Knowledge Engineering' at the same university. He was visiting researcher at the University of Melbourne, Australia, and visiting professor at Université Pierre et Marie Curie, Paris. His current research interest focuses on smart user support in interactive information systems.