Multilingual Adaptive Search for Digital Libraries

M. Rami Ghorab[1], Johannes Leveling[2], Séamus Lawless[1], Alexander O'Connor[1],
Dong Zhou[1], Gareth J. F. Jones[2], Vincent Wade[1]

[1] CNGL, Knowledge and Data Engineering Group, School of Computer Science & Statistics,
Trinity College Dublin, Dublin 2, Ireland,
{ghorabm,seamus.lawless,alex.oconnor,dong.zhou,
vincent.wade}@scss.tcd.ie
[2] CNGL, School of Computing, Dublin City University, Dublin 9, Ireland
{jleveling, gjones}@computing.dcu.ie

Abstract. This paper describes a framework for Adaptive Multilingual Information Retrieval (AMIR) which allows multilingual resource discovery and delivery using on-the-fly machine translation of documents and queries. Result documents are presented to the user in a contextualised manner. Challenges and affordances of both Adaptive and Multilingual IR, with a particular focus on Digital Libraries, are detailed. The framework components are motivated by a series of results from experiments on query logs and documents from The European Library. We conclude that factoring adaptivity and multilinguality aspects into the search process can enhance the user's experience with online Digital Libraries.

1 Introduction

Estimates show that 60% of World Wide Web (WWW) users are non-English speakers[3] and the content available in non-English languages is growing fast. The variety in users and language means that regardless of which languages a user speaks, there is a large volume of content which users cannot easily discover, consume or comprehend. Moreover, on the specific scope of European countries, information access systems that cater for European users must take into consideration the different linguistic and cultural backgrounds of their users. Online Digital Library (DL) portals, such as The European Library (TEL)[4] support a variety of languages by allowing the user to type in queries in different languages, select different interface languages, or view result documents written in different languages. However, in this paper we argue that there is yet more potential in exploiting adaptive techniques to improve the user's experience with multilingual search in DLs, for example through query adaptation and result adaptation.

In this paper, we describe a framework for Adaptive Multilingual Information Retrieval (AMIR), which comprises techniques from Information Retrieval (IR) and Adaptive Hypermedia (AH)[1]. Traditional IR approaches focus on the retrieval and presentation of documents in a ranked list, from which the user chooses the best documents to view. AH enables the aggregation of content in dynamically tailored hypertextual presentations, ensuring that results are adapted to the user's preferences, goals, and context.

---

[3] http://www.internetworldstats.com
[4] http://www.theeuropeanlibrary.org/

The AMIR framework supports multilingual resource discovery and delivery using on-the-fly machine translation of documents and queries. This allows to bridge the gap between the user's language and the document's language. Furthermore, adaptive techniques are applied to ensure that users are presented with content that is most relevant to them. A key reason for combining AH and IR is that adaptivity can contextualise IR.

The framework will contribute to three different strands of research: i) user-centric research, determining the user requirements and consequently the required functionality of the search system, content processing, and result presentation; ii) system-centric research, targeting the study and development of system components for indexing, annotation, retrieval and presentation of multilingual documents; and iii) evaluating the system and its components in user studies and formal evaluation benchmarks and investigating the combination of approaches from AH and IR.

The rest of this paper is organised as follows: Section 2 discusses related work; Section 3 describes the AMIR framework and presents component-level evaluation for result adaptation; and Section 4 presents conclusions and future work.

2   Related Work

A digital library (DL) constitutes a network of federated information sources, interfacing to different types of electronic documents made available by these sources. Users can submit queries to the DL to search for documents which are stored at the local repositories of one or more federated sources [10].

The number of DL research projects and systems indicates that there is growing interest in DL. For example, Europeana[5] is a virtual library providing access to millions of digital items, including music, films, and text. It aims to facilitate access to Europe's cultural and scientific resources. The MILE project[6] promotes European cultural heritage and makes images accessible by improving their metadata annotation. One of the major areas of investigation in MILE was concerned with improving search based on metadata. The MultiMatch search engine[7] focuses on information from cultural heritage institutions, identifying relevant documents regardless of the language. It can organise and display search results in an integrated, user-friendly manner, allowing users to access and exploit the retrieved information regardless of language barriers. The CA-CAO Project[8] offers an integrated approach for accessing, understanding and navigating multilingual documents in library catalogues, enabling users to better exploit electronic content. Three thematic portals were created as part of the project, realising multilingual book search on History, Mathematics and Geography. ezDL[9] (formerly DAFFODIL) is a user-oriented front-end for DLs which supports proven search strategies, integrating different DLs [5]. DAFFODIL was adaptive towards different user wishes, regarding preferences concerning content and system involvement. This was achieved via a profile of user's interests, created from items in a personal library [4].

---

[5] http://www.europeana.eu/portal/
[6] http://www.mileproject.eu/
[7] http://www.multimatch.eu/aboutmultimatch.html
[8] http://www.cacaoproject.eu/
[9] http://ezdl.de/

TEL offers access to digital and bibliographical resources of major European national libraries. TEL provides a virtual collection of information resources from many domains. The LADS task (Log Analysis for Digital Societies), which is part of the LogCLEF track at CLEF[10] aims at investigating user actions in multilingual search systems. The LADS task is involved with different experimental datasets, including log files of user interactions with TEL.

While the AMIR framework presented in this paper can be customised to cater for search in DLs, it is not restricted to a particular search domain; the framework can in fact be customised to run on different corpora, and it also allows plugging-in alternative components to carry out translation, query adaptation, and result adaptation. The following sections will discuss the components of the AMIR framework. Moreover, for two of the components, a component-level evaluation is presented.

3   Framework for Adaptive Multilingual IR

The motivation for the AMIR framework originates from the observations that i) users from different linguistic or cultural backgrounds behave differently in search; ii) there are identifiable patterns in user actions; and iii) user queries and action patterns reflect the mental model or prior knowledge of a user about a search system [3]. These observations call for the need to address adaptivity and multilinguality aspects in DL search.

The framework's architecture is shown in Figure 1. The framework is fully functional and has been demonstrated at a public showcase event. It is implemented in Java and follows the Model-View-Controller (MVC) architecture. The framework has a clear separation of roles in terms of the following component: 1) controllers which govern the sequence of execution of the retrieval and adaptation processes; 2) components which hold the algorithms for query adaptation, result list merging, and result list adaptation; and 3) components which deal with external services such as Web search engines, machine translation Web services and automatic language identification. The components of the framework support a spectrum of functionality regarding Multilingual and Adpative IR. The framework components are illustrated below. The *User Modelling* component gathers user and usage information about the users and organises this information in user models. This component can make use of both, implicit and explicit information gathering approaches in order to populate the models. The implicit approach involves processing the user's search history . The explicit approach involves Web forms that ask the users to engage in two activities: 1) supply personal/demographic information; and 2) scrutinise the inferred search interests in the user model. The information from the user model is used to adapt the queries, the results, the interface language, and the appearance of the search application.

In the *Query Adaptation & Translation* component, the user's query is adapted along two stages: pre-translation query expansion and post-translation expansion. Query expansion can be based on terms obtained from blind relevance feedback, the user model, or both [2]. Blind relevance feedback involves extracting terms from top-ranked documents retrieved with the original query or with the translated query. These terms are then added to the query so that more relevant results can be retrieved.

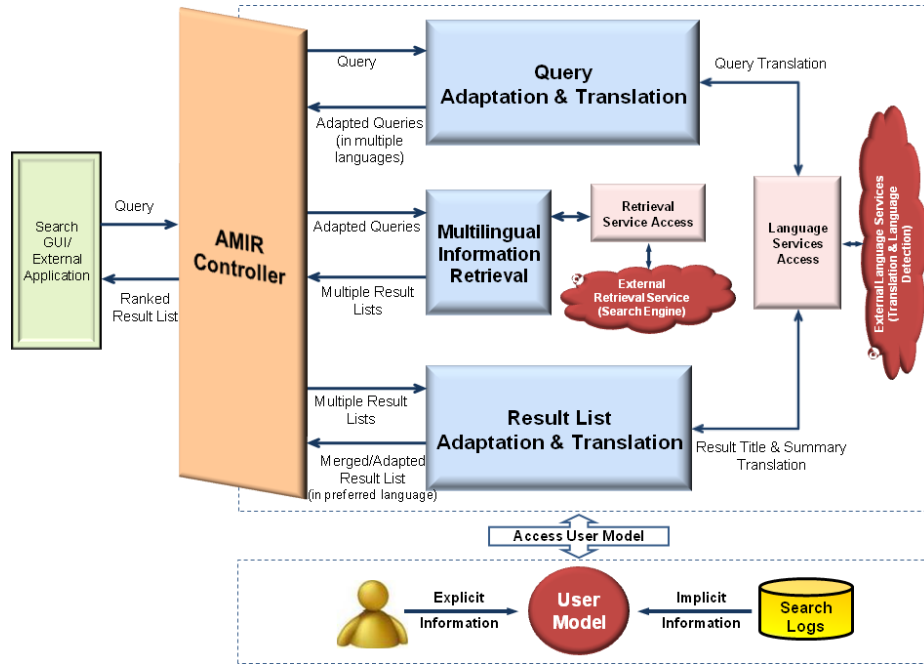[10] Cross-Language Evaluation Forum: `http://www.clef-campaign.org/`

Fig. 1. Framework Architecture.

The *Multilingual Retrieval* component performs multiple monolingual IR processes, one for each target language, and then passes multiple result lists to the Result List Adaptation & Translation component. Currently, the framework supports automatic translation of queries and documents using Web services such as Google Translate[11] or MicrosoftTranslator[12]. In addition, translation by OpenMaTrEx[13], an open-source machine translation toolkit, is supported for domain-specific translations.

The *Result List Adaptation & Translation* component merges results from multiple languages into a single list that is then translated to the user's preferred/native language. Furthermore, this component performs result list re-ranking, which can be based on latent semantic information induced from the documents or based on the user's search interests. In addition to this, in environments where the user is presented with a list of candidate collections that group the results together (such as in TEL, where users can select library collections in the results page), the component can re-rank the list of collections based on the user's language or country. Given the importance of this component with respect to DLs, result adaptation approaches are partially discussed in the next subsections.

---

[11] http://translate.google.com/

[12] http://www.microsofttranslator.com/

[13] http://openmatrex.org/

3.1 Collection Re-ranking based on User Models

In several DL websites, the user is presented with search results that are grouped under different library collections. This experiment aimed at improving retrieval effectiveness by re-ranking the list of collections based on user information, so that collections that match the user's language or country are ranked higher in the list. The experiment was conducted as part of our participation in the LogCLEF track of CLEF 2010[7]. It was based on search logs of TEL where all user interactions with the portal (including user queries and clicks on collections) were recorded[8].

The user interface of TEL shows a list of collections on the left side and a list of results from the selected collection on the right side. A collection is either a library catalog or an online resource that is associated with a certain country. The list of collections is presented to the user in alphabetical order of country acronyms. For this experiment we associated collections with languages, based on the official languages that are spoken in the collection's country. A subset of the TEL interaction logs was extracted, which contains submitted queries and clicked collections during the month of February 2007. This included approximately 566 queries from different languages.

In order to evaluate the retrieval precision over the list of collections presented to the user, we used the collections that the user clicked on as implicit relevance judgements (i.e. binary relevance assessments where the clicked collections are assumed to be the relevant ones). However, the notion of relevance in our experiments is in terms of matching language and country. Mean Average Precision (MAP) was used for evaluation where the MAP score was calculated across the queries in the selected subset of the data. The original ranked list of collections (i.e. the one presented to the user by TEL) was used as the baseline for evaluation of retrieval precision (0.580 MAP). Several alternative re-ranked lists were investigated and compared to the baseline.

The collection re-ranking mainly builds on the following three attributes: country (location from which the query was submitted), query language, and interface language. Collection scores are computed as a weighted linear combination of attributes with the collection's country and language as follows ($M_X = 1$ if matching two items succeeds, $M_X = 0$, otherwise): i) M_c: matching *user's country* and collection's country. ii) M_q: matching *query's language* and collection's language (i.e. matching with any of the official languages spoken in the corresponding country). iii) M_i: matching *interface language* and collection's language.

Each of the above attributes is multiplied by a scalar weight (W_c, W_q, W_i respectively) to control the degree of contribution of each attribute in the function. The collection list is re-ranked based on descending order of the new collection score. Table 1 shows the MAP values for some selected re-ranking runs with different weights that ranged from 0.0 to 1.0. The results showed a significant improvement of 27.4% in retrieval precision for the re-ranked collection lists (with weights: $W_c = 0.1$, $W_q = 0.3$, $W_i = 0.6$) over the baseline ranking. This improvement is statistically significant as per the T-test (with p=0.01). The results of this experiment suggest that there is opportunity for improving the user's experience with multilingual search in DLs if the user's language and country are taken into consideration when adapting the results.

Table 1. MAP for re-ranking with different weight combinations. MAP for the baseline is 0.58.

	Weights (W_c, W_q, W_i)						
W_c	0.1	0.1	0.6	0.6	0.0	0.0	1.0
W_q	0.3	0.6	0.1	0.3	0.0	1.0	0.0
W_i	0.6	0.3	0.3	0.1	1.0	0.0	0.0
MAP	**0.74**	0.73	0.72	0.71	0.71	0.68	0.60
Improvement [%]	**+27.4**	+25.3	+23.9	+22.2	+22.2	+18.1	+3.4

3.2 Multilingual Result Re-ranking based on Query-Document Features

Multilingual result adaptation can be achieved by re-ranking an initial result set. The simplest approach is to directly apply the monolingual methods on the results obtained using a translated query. The drawback of this approach is that translation errors will be propagated to the adaptation process, which may result in unsatisfactory performance. We proposed a multilingual re-ranker component which incorporates scores generated using external knowledge to enhance the semantic space produced by the latent concept method [11] through a linear combination model. This method is proposed to solve the multilingual document re-ranking problem by automatically inducing a semantic correspondence between two languages (query language and document language) using parallel corpora as training data. For the experiments described in this paper, a parallel corpus extracted from Wikipedia data has been employed. The correspondence between two languages is then used to project the query into another language in the semantic space to accomplish the re-ranking task. Formally, the latent semantic space is produced by a Latent Dirichlet Allocation (LDA) model.

The TEL corpora in English, French and German from CLEF 2009 were employed for evaluation of this component. These collections were chosen to test the scalability of the proposed method in different settings and over different languages. The corpora were provided together with queries a nd relevance judgments by the organisers of the CLEF ad-hoc retrieval task. Wikipedia documents in English, French and German were used as an explicit concept space. Only those articles that are connected via cross-language links between all three Wikipedia databases were selected. The results suggest that the method proposed in this section outperforms a baseline ranking system (standard BM25) and a previous proposed re-ranking method that only uses latent features by a statistically significant margin in many test runs. The experiments also confirm that directly applying monolingual methods into the cross-lingual applications may not always produce the most beneficial results. The results for two language pairs are shown in Table 2($*$ indicates significant improvement over the initial retrieval) The following IR evaluation metrics were used to evaluate retrieval precision: precision at N documents (P@N), Normalised Discounted Cumulative Gain (NDCG), and Mean Average Precision (MAP).

Table 2. Multilingual Re-ranker

metric	French-English			German-English		
	init.	fusion-lda	fusion-lda-cross-mix	init.	fusion-lda	fusion-lda-cross-mix
P@5	0.480	0.532*	0.540*	0.480	0.488	0.492*
P@10	0.444	0.470*	0.474*	0.416	0.466*	0.470*
P@20	0.389	0.389	0.391	0.364	0.389*	0.394*
NDCG	0.368	0.370	0.371	0.362	0.372*	0.372*
MAP	0.212	0.216	0.218*	0.210	0.223*	0.224*

3.3 The Evaluation Gap: User Studies vs. IR Evaluation

Contextual information about the user, content and environment is increasingly being used to support the tailored delivery of information in IR. The personalised discovery, retrieval and presentation of content can provide an enhanced information seeking experience to the user. While such tailored experiences can produce a more informative response than a traditional ranked list approach, there are many challenges associated with evaluating these approaches. As AMIR and response composition become more widely used, traditional approaches to IR evaluation may become less effective or applicable in isolation. The complex functionality offered by these systems and the variety of users who interact with them, mean both component-level evaluation and user-based evaluation are required to comprehensively assess the system. An approach has been proposed which combines and enhances evaluation methodologies from both the AH and IR communities [6]. In order to sufficiently evaluate the adaptive functionality and the retrieval effectiveness of an AMIR system, a hybrid approach is necessary. This involves user-centric assessment, layered evaluation of the adaptivity which has been applied, and quantitative performance metrics relating to the content delivered.

IR has traditionally been evaluated using the metrics of precision and recall and derivatives of these metrics such as MAP and the F-measure. These metrics measure the accuracy and scope of the retrieved documents. While these metrics are valuable in measuring the effectiveness of real world tasks, they are more typically used to evaluate retrieval effectiveness with test collections in laboratory IR experimental settings.

Numerous measures for the evaluation of adaptivity in adaptive systems have been proposed[9] which aim to measure the scientific performance of components and perform user-based evaluation of the adaptivity offered by the system. The approaches can be broadly divided into three categories: (i) Adaptivity Metrics; (ii) User-Interaction Metrics; and (iii) Performance Metrics. There are a set of necessary elements of a hybrid AMIR evaluation model which can be defined irrespective of the system being evaluated. The key challenge is to be able to adequately combine the data-driven approach to IR evaluation with the more user-focused approach to evaluation from AH.

4 Conclusions and Future Work

In this paper we proposed a framework that caters for adaptivity and multilinguality aspects in DL search. We have presented component-level evaluation, and highlighted

evaluation aspects with respect to the fields of IR and AH. We showed that there is scope for improving the the effectiveness of search in DLs if user and usage information is exploited with respect to the multilingual dimension. While integrating adaptivity and personalisation capabilities into search systems is desirable, the evaluation approach will have to be revised by combining evaluation approaches from AH and IR, where evaluation focusses on user and usability aspects in conjunction with retrieval precision.

A viable extension to this study can be to investigate how common query adaptation techniques from the IR field can be used within the specific domain of DLs, where the search terms used are often limited to categories like authors, book titles, etc.

Acknowledgments

References

1. Brusilovsky, P.: Adaptive hypermedia. User modeling and user-adapted interaction 11(1), 87–110 (2001)
2. Chirita, P.A., Firan, C.S., Nejdl, W.: Personalized query expansion for the web. In: SIGIR '07. pp. 7–14. ACM, New York, USA (2007)
3. Ghorab, M.R., Leveling, J., Zhou, D., Jones, G.J.F., Wade, V.: Identifying common user behaviour in multilingual search logs. In: et al., C.P. (ed.) Multilingual Information Access Evaluation I. Text Retrieval Experiments. LNCS, vol. 6241, pp. 518–525. Springer (2010)
4. Gövert, N., Fuhr, N., Klas, C.P.: Daffodil: Distributed agents for user-friendly access of digital libraries. In: Borbinha, J.L., Baker, T. (eds.) ECDL 2000. LNCS, vol. 1923, pp. 352–355. Springer (2000)
5. Kriewel, S., Fuhr, N.: Adaptive search suggestions for digital libraries. In: Goh, D.H.L., Cao, T.H., Sølvberg, I., Rasmussen, E.M. (eds.) ICADL 2007. LNCS, vol. 4822, pp. 220–229. Springer (2007)
6. Lawless, S., O'Conor, A., Mulwa, C.: A proposal for the evaluation of adaptive personalised information retrieval. In: CIRSE 2010. Milton Keynes, UK (2010)
7. Leveling, J., Ghorab, R., Magdy, W., Jones, G.J.F., Wade, V.: DCU-TCD@LogCLEF 2010: Re-ranking document collections and query performance estimation. In: CLEF 2010 LABs and Workshops, Notebook Papers, 22-23 September 2010, Padua, Italy (2010)
8. Mandl, T., Agosti, M., Di Nunzio, G., Yeh, A., Mani, I., Doran, C., Schulz, J.M.: LogCLEF 2009: the CLEF 2009 Cross-Language Logfile Analysis Track Overview. In: et al., C. (ed.) Multilingual Information Access Evaluation Vol. I. Text Retrieval Experiments. pp. 508–517. LNCS, Springer (2010)
9. Raibulet, C., Masciadri, L.: Evaluation of dynamic adaptivity through metrics: An achievable target? In: SEA 2009 (2009)
10. Rigaux, P., Spyratos, N.: Metadata inference for document retrieval in a distributed repository. In: ASIAN04. LNCS, vol. 3321. Chiang-Mai, Thailand (2004)
11. Zhou, D., Lawless, S., Min, J., Wade, V.: A late fusion approach to cross-lingual document re-ranking. In: Proceedings of the 19th ACM international conference on Information and knowledge management. pp. 1433–1436. ACM, New York, NY, USA (2010)