

Leveraging User Interaction and Collaboration for Improving Multilingual Information Access in Digital Libraries

Juliane Stiller

Berlin School of Library and Information Science, Humboldt University
Dorotheenstr. 26
10117 Berlin, Germany
juliane.stiller@ibi.hu-berlin.de

ABSTRACT

The goal of interactive cross-lingual information retrieval systems is to support users in formulating effective queries and selecting the documents which satisfy their information needs regardless of the language of these documents. This dissertation aims at harnessing user-system interaction, extracting the added value and integrating it back into the system to improve cross-lingual information retrieval for successive users. To achieve this, user input at different interaction points will be evaluated. This will, among others, include interaction during user-assisted query translations, implicit and explicit relevance feedback and social tags. To leverage this input, explorative studies need to be conducted to determine beneficial user input and the methods of extracting it.

Categories and Subject Descriptors:

H.3.7. [Information Storage and Retrieval]: Digital Libraries -*user issues*; H.3.7. [Information Storage and Retrieval]: Information Search and Retrieval

General Terms: Design, Human Factors, Languages

Keywords: Interactive CLIR, Digital Libraries, Social tags

1. INTRODUCTION

Digital Libraries which aggregate and provide access to collections in different languages are confronted with problems of multilingual information access (MLIA). The most challenging aspect is the search capability of the system, which is referred to as cross-lingual information retrieval (CLIR). CLIR is characterized by differences in query and document language [3]. The main difficulties for CLIR are the disambiguation of the query term in the source and target language and the identification of the query language. Tools for CLIR such as dictionaries are not universally available in every language needed or in every domain covered in digital libraries.

This dissertation aims at harnessing user input for improving and extending existing CLIR processing tools and consequently CLIR systems in their entirety. The goal is to bridge a gap between research on interactive cross-lingual information retrieval and studies focusing on establishing sustainable ways to leverage user input.

In this context, user interaction and collaboration comprises social tagging and other system interaction components like query reformulations and relevance feedback.

2. HARNESSING USER INPUT

In CLIR systems, interactive components are crucial to accomplish search tasks [2]. There are two main scenarios where the user input could be incorporated into the system to enhance multilingual information retrieval:

1. Harnessing multilingual tags for enriching metadata and disambiguating query terms and
2. Improving existing dictionaries by adding translations entered into the system by the user, e.g. directly from the user or via log files analysis.

In this project, interactive CLIR focuses on input from previous user-system interactions which can offer improvements for CLIR tools. This requires the identification of different interaction points, where users are encouraged to leave their feedback in the form of alternative translations, relevance feedback or tags. This input will be collected and integrated back into the CLIR system. To gather data, the user input at different points of user interaction will be analyzed to identify the added value. A method to harness this input and to integrate it back into the system will be determined accordingly, e.g. evaluation of query logs [1].

Europeana - a multilingual digital library - will serve as main research object for explorative studies. Unifying the different results from these experiments will provide a basis for determining the benefits of user interaction for CLIR in digital libraries. These results will serve as use cases for interaction which enhances CLIR and can easily be prototyped. The project will include a final analysis of the effectiveness of different interactive features within CLIR systems.

3. REFERENCES

- [1] A. Bosca and L. Dini. Cacao project at the logclef track. In *Working notes of the Cross Language Evaluation Forum (CLEF)*, 2009.
- [2] J. Gonzalo. Scenarios for interactive cross-language retrieval systems. In *Proceedings of the Workshop of Cross-Language Information Retrieval: A Research Roadmap Workshop held at the 25th Annual International ACM SIGIR Conference*, 2002.
- [3] D. W. Oard. Multilingual information access. In M. J. Bates and M. N. Maack, editors, *Encyclopedia of Library and Information Sciences*. Taylor & Francis, 3rd edition, 2009.