Adding MultiLingual Information Access to The European Library TEL

Martin Braschler
Institute of Applied Information Technology
Zurich University of Applied Sciences
Winterthur, Switzerland

Nicola Ferro
Department of Information Engineering (DEI)
University of Padova
Padova, Italy

Abstract

A feasability study was conducted within the confines of the DELOS Network of Excellence with the aim of investigating possible approaches to extend The European Library (TEL) with multilingual information access, i.e. the ability to use queries in one language to retrieve items in different languages. While the academic community has in recent years made substantial progress in the two related fields of multilingual information access and cross-language information retrieval, the TEL system poses some specific challenges which force adaptation of the mechanisms which have originally been developed for more general settings. Specifically, TEL uses a loose coupling of different search systems, and deals with very short information items. We address these two characteristics with two different approaches: the "isolated query translation" approach, and the "pseudo-translated expanded records" approach. The former approach has been studied together with its implications on the user interface, while the latter approach has been evaulated using a test collection of over 150,000 records from the TEL central index. We find that both approaches address the specific characteristics of TEL well, and that there is considerable potential for a combination of the two alternatives.

Categories and Subject Descriptors
H.3 Information Storage and Retrieval; H3.3 Information Search and Retrieval; H3.7 Digital Libraries

Free Keywords
Multilingual Information Access, Cross-Language Information Retrieval, Interactive Term Selection, Relevance Feedback

1 Introduction

This paper reports on a feasability study ([1], [8], [14]) conducted in collaboration between DELOS, the European Network of Excellence on Digital Libraries and The European Library (TEL). TEL is a service fully funded by the participant members (national libraries) of the Conference of European National Librarians (CENL). It aims at providing a co-operative framework for integrated access to the major collections of the European national libraries. The aim of the study was to provide a solid basis for the integration of multilingual information access into TEL.

2 Multilingual Information Access and Cross-Language Information Retrieval

By multilingual information access (MLIA) we denote procedures for search on collections of information items (in the context of this paper, bibliographic records) that are potentially stored in multiple languages. Usually the term is used for situations in which the user is allowed to query the collection across languages, i.e. retrieving information items formulated in a language that is different from the language used by the user to formulate his/her information need. The term "cross-language information retrieval" (CLIR) is often used to describe this definition of MLIA, distinguishing it from monolingual access to information in multiple languages (which is already implemented in TEL).

Today, mainstream research on cross-language information retrieval in Europe is carried out in the confines of the *Cross-Language Evaluation Forum* (CLEF) campaign [16]. The campaign gives researchers the possibility to compare different approaches to CLIR in a common setting and provides tools for in-depth analysis of the experimental result [2], [3]. Most of the experiments in CLEF concentrate on retrieval on lengthy,

unstructured full-text documents using a general vocabulary. An overview of the recent achievements in CLIR can be found in [7], [12], and [13]. Generally, there is a growing sense among the academic community that the CLIR problem as applied to such lengthy, unstructured full-text documents from a general domain is fairly well understood from an academic standpoint [5], [6].

Unfortunately, the situation in the TEL system is substantially different from the ideal "mainstream setting" for CLIR. When searching on some of the catalogs of the member libraries, there is only a loose coupling of systems, i.e. the search is forwarded to the system of that library. In such cases, translation and retrieval cannot be tightly integrated. We address this problem with the "isolated query translation approach" (see Section 4). Furthermore, the large majority of information items are very short. Similarly, the expressions of information needs by the users, i.e. the queries, tend to be very short (average query length in TEL is 2.2 words). These contradictions to the general CLIR setting are addressed by our "pseudo-translation on expanded records" approach (see Section 5).

3 TEL Architecture and Functioning

Figure 1 shows the architecture of the TEL system. The TEL system allows easy integration of national libraries [17] by extensively using the *Search/Retrieve via URL* (SRU)[1] protocol in order to search and retrieve documents. In this way, the user client can be a simple browser, which exploits SRU as a means for uniformly accessing national libraries.

With this objective in mind, TEL is constituted by three components:

- a Web server: which provides users with the TEL portal;
- a central index: which harvests catalogue records from national libraries which support the *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH)[2] and provides integrated access to them via SRU;
- a gateway between SRU and Z39.50[3]: which allows national libraries that support only Z39.50 to be accessible via SRU.

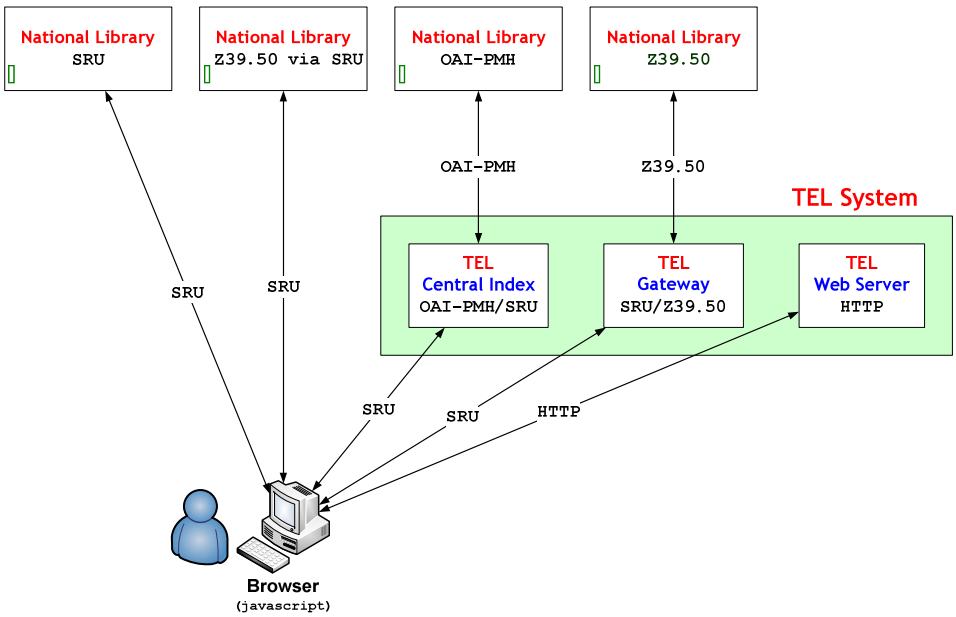

Figure 1: Present architecture of the TEL system.

This setup directly influences how MLIA/CLIR can be integrated into TEL.

3.1 Problems and Approaches

The architecture and functioning of the TEL system pose some problems when planning to introduce MLIA.

[1] http://www.loc.gov/standards/sru/
[2] http://www.openarchives.org/OAI/openarchivesprotocol.html
[3] http://www.loc.gov/z3950/agency/

The TEL system has no control on queries sent to the national libraries, since the client browser directly manages the interaction with national library systems via SRU. As a consequence, introducing MLIA functionalities into the TEL system would have no effect on the national library systems.

Thus, in order to achieve full MLIA functionalities, not only the TEL system but also all the national library systems would have to be modified. However, this is an unviable option as it would require considerable effort and disregards the "low barrier of entry" criteria adopted when designing the TEL system.

In order to avoid the problem discussed above, while still offering some MLIA functionalities, we have investigated the possibility of adding an "isolated query translation" step in the query processing.

On the other hand, the TEL central index harvests catalogue records from national libraries, which in addition to catalogue metadata may contain other information useful for applying MLIA techniques, such as an abstract. Since the central index is completely under the control of the TEL system, we have shown how to extend its functionality by adding a component that pseudo-translates the catalogue records in order to perform MLIA on them ("pseudo-translation of expanded records"). This approach takes into account that the situation in the TEL system, mainly due to the brevity of the information items involved, is substantially different from the ideal "mainstream setting" for CLIR.

4 Isolated Query Translation

4.1 Situation

As mentioned, the loose coupling of the systems of the individual national libraries in TEL has consequences for the design of MLIA functionalities. The problems are addressed by the use of an isolated query translation step in the query processing. Figure 2 shows the modifications necessary to the current TEL architecture to support this feature.

The new "isolated query translation" component can be directly accessed by the client browser using the SRU protocol. "Isolated query translation" can be considered as a sort of pre-processing step where the translation problem is treated as completely separate from retrieval.

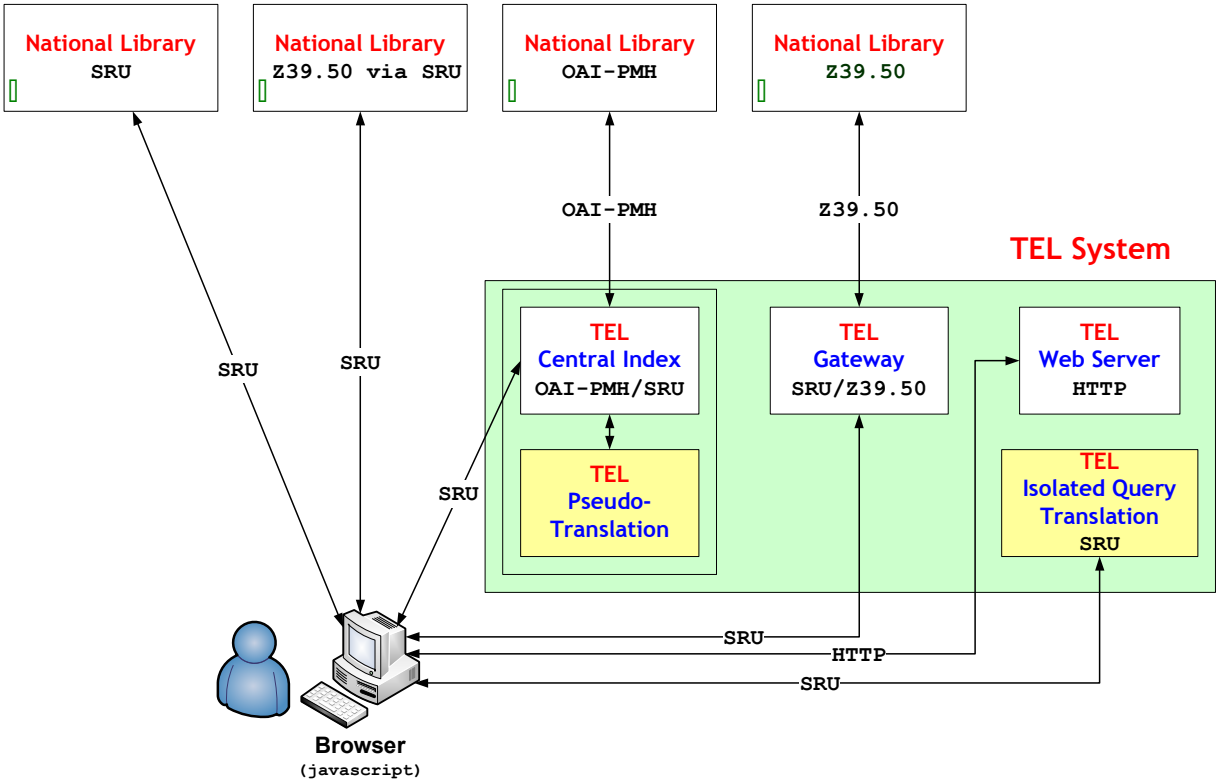

Figure 2: Architecture of the TEL system with the new "Isolated Query Translation" and "Pseudo-Translation" functionalities.

The approach works as follows:

- before actually submitting the query, the user asks the browser to translate it;
- the browser sends the query via SRU to the "isolated query translation" component, which takes care of translating it and, if necessary, applies query expansion techniques to reduce the problem of missing translations;
- at this point, the user can interactively select the translation which best matches his needs or can change some query term to refine the translation. In this latter case, the translation process may be iterated.
- Once the desired translation of the query is obtained, the retrieval process is initiated, using both the translated query and the original one.

The main advantage of this solution is its easiness of implementation and its compliance with the "low barrier of entry" approach of TEL. The national library systems do not require any modification and this new functionality can be transparently applied when querying them, even though it is actually performed within the TEL central system.

"Isolated query translation" requires some user interaction, because the user may need to choose among multiple translations of the same term in order to disambiguate them or may need to modify the original query if the translated query does not match his needs.

The main drawback of this approach is that, as the translation is separated from the retrieval process, relevant documents may be missing in the result set and thus the performance may be low. Moreover, huge linguistic resources, such as dictionaries, are needed since the vocabulary used in queries is expected to be very large; this has to be repeated for each pair of source/target language the system is going to support. Finally, the query expansion mechanism has to be generic and cannot be tailored on the collections queried, since the "isolated query translation" component does not interact with the national library systems.

4.2 Modifications to the TEL System User Interface for the "Isolated Query Translation" Feature

In this section, we discuss our proposal on how the current user interface of the TEL system can be modified to introduce the "isolated query translation" feature. In particular, we focus our attention on the simple search functionality.

First, the TEL user interface is extended with an additional link to the "Isolated Query Translation" feature.

When the user clicks on the "suggest query in other languages" link, as shown in Figure 3, a box with the supported target languages for the translation appears below the search input box. The user can now check the languages for which he wants a translation of the query. Note that the behaviour of this box is similar to that of the "virtual keyboard" link, already implemented in the TEL system.

Moreover, on the left of the search input box, a list with the possible source languages of the query is now shown, so that the user can specify the language of his original query. This list differs from the list of the languages for the user interface in the upper right corner of the user interface mainly for two reasons. First, users may opt not to select their own language for the user interface because they are able to understand English and want to avoid the "extra-clicks" needed to change the language. As a consequence, the actual selection in the list of languages for the user interface could be an erroneous indicator of the source language of the query. Second, the "isolated query translation" component could support the translation functionalities for a partially different set or a subset of languages with respect to the list of the languages available for the user interface. Thus, for both these reasons we need two separate language lists: one for the user interface and the other for the source language of a query.

As shown in Figure 4, for each target language selected by the user, a new text input box appears below the search input box containing the translation of the query in that language.

There are different possibilities for managing the user interaction when the translation of the query is shown. A first possibility would be to add a button "Suggest" so that the user presses it and the input boxes with the translation of the query appear. In this way, the user needs to explicitly request the translation. Another possibility would be a more *Asynchronous JavaScript Technology and XML* (AJAX) style of interaction where the input box with the translation appears as soon as the user selects the language in the list of the target languages. In this way, the user does not need to explicitly request the translation and the system would be more proactive. In any case, both ways of interaction comply with the current approach of the TEL system in developing the user interface, which already exploits AJAX.

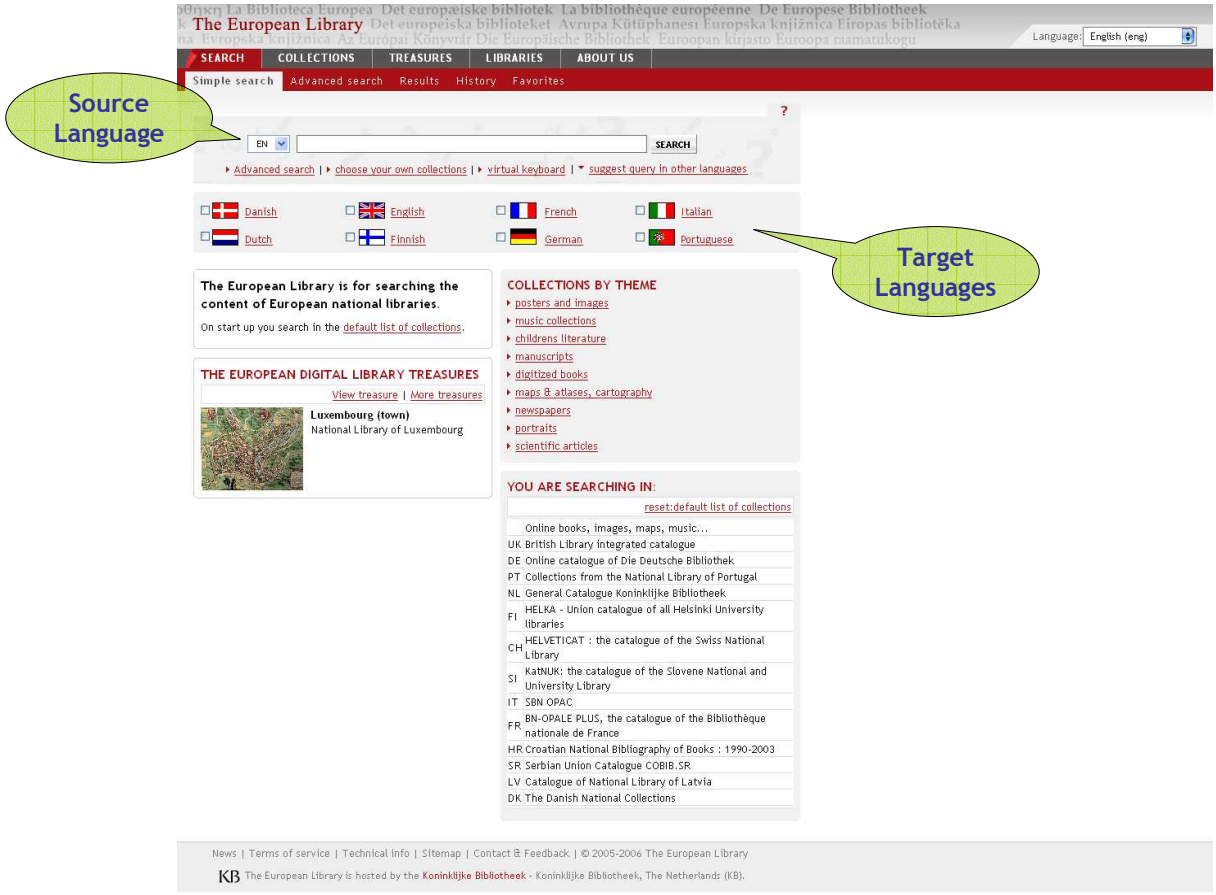

Figure 3: Selection of the source and target languages in the simple search.

Once the translations have been obtained, there is the problem of managing the user interaction if the user needs to modify or refine the translations. Since the users of the simple search probably prefer an easy and intuitive way of interacting with the system, the translation refinement step should also be as simple as possible, even if some precision or some expressive power is lost. For this purpose, two alternatives can be offered to the user:

1. the user could directly edit each text input box in order to modify the translated query until it meets his needs. This means that the user can delete or add words to the translation.
2. Note that if the user has no knowledge of a target language, he will not be able to refine the translated query and he will have to use the query suggested by the system without modifications.
3. In addition, some attention has to be paid when multiple translations are possible because they have all to be listed in the input box and thus some visual clue should be provided to help the user in distinguishing between multiple alternatives.
4. if the translation greatly differs from user's expectations, there is the possibility of modifying the source query by adding or deleting words to it thus obtaining a new translation in the target languages, which can be modified as described above.

Once the various translations of the query have been approved, the user can click the "Search Multilingual" button to perform a search in both the original and the selected target languages.

Figure 4: Query suggestions in other languages in the simple search.

5 Pseudo-Translation of Expanded Records

5.1 Situation

Today, the large majority of all records available through the search facility in TEL contain bibliographical metadata only (no abstracts or full text). Only short segments of text are thus available for free-text search by the user (such as the "title" field).

While potentially a problem in monolingual search as well, the brevity of the available text exacerbates the problems usually encountered in multilingual information access situations.

5.2 Expansion Techniques

The solution that was chosen for overcoming the lack of textual content in the information items is expansion of the content fields. The approach used is derived from techniques used in classical information retrieval for query expansion, such as relevance feedback [15]. These techniques extract new statistically related search terms from items that are ranked highly in initial searches. While often used interactively, involving a user picking relevant

items from a result set, these techniques can also be applied in an automated fashion. In such cases, the system automatically assumes that the top items returned in response to a user requests are relevant. Such a technique is called "blind feedback", and has been proven to be beneficial in many CLIR settings [4].

It is possible to use the same techniques independently of specific information needs, by expanding the information item itself instead of the query. While usually not applicable to retrieval on lengthy documents, we expected potential for such an approach in the case of the very short records present in the TEL collection.

By using expansion techniques, we intended to address both problems of vocabulary coverage and word sense ambiguity, as usually experienced during translation. Additional terms added during expansion tend to be from a more general vocabulary, as term frequency influences term selection. The new, longer representation of the record also makes it less likely that none of the terms can be translated.

5.3 Pseudo-Translation

In this proposed solution, we cross the language boundary by translating the "document", i.e. the complete record (see Figure 2 for integration of this approach into the TEL architecture). Document translation has been found to be very competitive [9] in some general cross-language retrieval settings, although query translation is more prevalent. The main reason for the scarce adoption of the document translation techniques can be attributed to problems of scalability. It is currently extremely resource-consuming to translate large sets of lengthy documents. However, this problem is much less pronounced in the case of TEL; as the brevity of the records should make document translation applicable even to large numbers of records, e.g. in the order of multiple millions of records. The approach is mainly suitable for integration with the TEL central index; since a new index needs to be built for the translated documents. However, the same approach could also be deployed in additional search systems of the national libraries that are accessed remotely via TEL.

Using the translated records for matching with queries only, and not for presentation, means that we can use "pseudo-translations", i.e. to potentially leave terms untranslated or translate them into multiple different terms in the target language. There is no need to produce a syntactically correct or semantically unambiguous document as the translation will remain hidden to the end user. This approach of using "rough" translations for retrieval is both cheaper to implement and often more effective for retrieval, as multiple translation alternatives can be retained during the process.

5.4 Outline of Approach

The outline of this approach, called in the following "pseudo-translation of expanded records", or "pseudo-translation" for short, is thus:

- the unstructured content fields of the record (usually only the title, as well as some keyword fields that can be treated as unstructured content) are expanded by additional terms that are statistically similar to the original terms in those fields
- these additional terms are derived by searching for records that are similar in content to the record that is to be expanded, and then extracting from these additional records the best terms according to well-known blind feedback techniques
- the expanded records are translated in their entirety using a simple translation resource
- retrieval takes place on the expanded, translated records

5.5 Experiment Setup

With retrieval experiments on test collections, we have aimed to demonstrate how expanded records could be represented, how they would look in their pseudo-translated state and to analyze whether they could be expected to be usable for implementing CLIR in the TEL system.

A full evaluation on a sample of 151,700 bibliographical records in English from the British library (part of the TEL central index) was carried out. We used 99 queries in English derived from three months of logfiles to represent typical information needs. Queries are a mix of one-word statements and longer formulations. The queries were manually translated into German for later cross-language retrieval experiments.

The experiments follow the so-called Cranfield paradigm [11] for retrieval tests. The retrieval system used was Terrier[4], an open-source information retrieval system developed by the University of Glasgow. Note that much of the procedures described would be implemented off-line in an operational system. Translation of the

[4] Terrier is available under the Mozillla Public License.

records was done using the PROMT[5] off-the-shelf machine translation system. Again, a variety of different translation resources could be used in support for the chosen CLIR approach.

We expanded the 151,700 records by using each record in turn as a query and running it against the whole collection to determine the set of most similar records. Based on this set of records, the statistically most discriminating terms were determined.

Different parameters for expansion were tried, with using the 10 best-ranked items to produce a maximum of 5 expansion terms leading to the most promising results.

Not all records benefit from this expansion process. For some records, no new statistically associated terms can be found, and the records remain unexpanded. In all, 44,647 out of 151,700 records (~29%) were not expanded. We expect this ratio to drop if more records were added to the "document" base.

We pseudo-translated all expanded records from English to German using PROMT. Please note that the resulting translation is not suitable for presentation: the short titles and lists of keywords contained in the records lead to non-grammatical output.

The translation suffered from aggressive compound formation by the PROMT machine translation software. Since we did not have a German compound splitter available for the Terrier system, retrieval effectiveness may have been negatively affected (For the effect of "decompounding" on retrieval effectiveness, see e.g. [10]).

The following is an example of a pseudo-translated record from our test collection:
English record, original:

```
<srw_dc:dc>
<recordPosition>103899</recordPosition>
<title>Private power : Multinational corporations for the survival
our planet.</title>
</srw_dc:dc>
```

German record, pseudo-translated, expanded.

```
<srw_dc:dc>
<recordPosition>103899</recordPosition>
<title>Private Macht : Multinationale Vereinigungen für das Überleben
unser Planet.</title>
<extendedTerms>Entwicklung, Welt, </extendedTerms>
</srw_dc:dc>
```

The resulting 151,700 pseudo-translated records were loaded into the Terrier system for retrieval.

5.6 Retrieval

The 99 queries were hand-translated into German and used to retrieve the top 10 records for each query from the pseudo-translated German records. This constitutes a cross-language retrieval experiment, as each pseudo-translated record can clearly be matched with the original English version it represents in the search index. As a baseline for comparison, we ran the same 99 queries in their original English version against the original English records.

5.7 Analysis of results

To evaluate retrieval effectiveness, usually recall and precision figures are calculated. The precondition for these calculations are manual relevance assesements, indicating which documents are relevant to each particular query. Clearly, it was not feasible to do extensive relevance assessments for all 99 queries in our study (resulting in 151,700 * 99 assessments). We used so-called "overlap analysis" as a viable alternative. The monolingual baseline of using English queries on English records represents the same information needs as the cross-language experiment of using the German queries on German pseudo-translated records. It is therefore possible to use the monolingual English baseline as a "gold standard", assuming that the results from that retrieval experiment have sufficient quality.

Using this "gold standard", any retrieval result for a query from the cross-language experiment that is sufficiently similar to the monolingual result is assumed to be acceptable. We have considered the top 10 ranked

[5] http://www.e-promt.com/

records to determine the similarity between the monolingual and the cross-language experiment. We have used this technique to exclude such queries from further analysis, therefore driving down overall workload. In all, 30 of the 99 queries had sufficiently similar results, and thus the cross-language results were considered to match the monolingual baseline. These queries were excluded from further analysis.

The remaining 69 queries have results that significantly differ from the monolingual baseline. This, however, does not necessarily indicate that these queries have poor performance. For further analysis, four cases need to be distinguished:

Case 1: good monolingual result; good, but different, cross-language result
Case 2: good monolingual result; bad cross-language result
Case 3: bad monolingual result; good cross-language result
Case 4: bad monolingual result; bad, but different, cross-language result.

The four cases can be supplemented with the previously described case:

Case 0: monolingual and cross-language result similar; assumed to be good

We have attempted to classify the remaining 70 queries (one query was accidentially duplicated) to Cases 1-4 based on relevance assessments of the top 10 records for both the monolingual and cross-language case. This would have in total meant 70 * 20 = 1400 judgments. In combination with the actual analysis of the results, it was not possible to process all 70 remaining queries. A total of 18 queries had to be excluded from further processing due to lack of resources. We thus analyzed a grand total of 52 queries, giving a categorization for 82 queries.

We argue that Case 0, 1, and 3 provide evidence for good retrieval results, whereas Case 4 at least indicates that the cross-language result is not necessarily worse than the monolingual result .

Case	# of queries
Case 0	30
Case 1	13
Case 2	14
Case 3	2
Case 4	23
Not evaluated	18

Table 1. Summary of evaluation of queries. Of the original 99 queries, 81 were categorized (one query was accidentially duplicated)

In all, using this methodology we found that 55% of queries analyzed showed evidence of good retrieval results, and 83% of queries showed evidence that they did not suffer significantly from the cross-language setup when compared to the monolingual baseline (note that for some of these queries there simply will be no relevant records in the collection!). The latter number is encouraging, being in line with what has been reported as state-of-the-art for CLIR in the CLEF campaign on lengthy documents [6]. Please note, however, that the number has to be treated with care, owing to the limitations described above. On the other hand, we believe that the approach will actually benefit in terms of effectiveness when scaling up to larger collections, such as would occur when implementing the approach in the actual TEL system.

6 Conclusions

We have described the results and the findings of a feasibility study carried out to determine how multilingual information access functionalities could be added to the TEL system. We have proposed two different approaches for introducing MLIA functionalities in the TEL system: the first one, called “isolated query translation”, performs a pre-processing step to translate the query and then routes the translated query to the national library systems. The second one, called “pseudo-translation”, involves only queries sent to the TEL central index but merges the translation process with the retrieval one in order to offer more effective MLIA functionalities. It should be noted that the two proposed approaches are neither separate nor mutually exclusive. On the contrary, they aim at addressing two aspects of the TEL system that can be considered as distinguishing features. On the one side the TEL capability of directly querying national libraries can be enhanced with MLIA

functionalities by using the "isolated query translation" approach; on the other side the search functionalities of the TEL central index, which harvests catalogue records from national libraries, can be extended with MLIA capabilities by adopting the "pseudo-translation" approach.

As a consequence, the two proposed approaches complement each other and offer a coherent multilingual extension of the TEL characteristics mentioned above. The implementation is facilitated as they share common components at the architectural level. For example, the translation engine or the translation resources, whether Machine Translation, Machine Readable Dictionaries, or a combination of methods, can be shared by both approaches in order to reduce the development effort.

A combination of the two approaches would lend itself naturally to a system using an interlingua setup. In such cases, instead of having support and multilingual resources for all the possible pairs of source and target languages, one language is selected as *pivot* and all the translations are made to and from this pivot language. For example, if we need to translate from Portuguese to Bulgarian, instead of performing a direct translation, we may choose English as the interlingua, and perform a translation from Portuguese to English and from English to Bulgarian. This solution would allow us to avoid an exponential growth of the multilingual resources, if many different partners join TEL. In a combination approach, one of the two translation steps would be implemented using "pseudo-translation", while the other translation step is carried out during retrieval using "isolated query translation". This would address issues with translation errors ("noise") that are typically amplified when multiple translation steps are used. The expansion and the user interaction, respectively, can be expected to mitigate such noise problems.

On the whole, we can envision the following evolutionary scenario for implementing of MLIA in TEL:

- short-term: the "isolated query translation" solution is a first step for adding MLIA functionalities to TEL and represents a quick way to give TEL users and partners a multilingual experience. In parallel, steps should be undertaken to take more advantage of the growing central index in TEL and prepare for additional CLIR components such as "pseudo-translation"
- mid-term: the implementation and deployment of a "pseudo-translation" solution is a second step which allows to better exploit the information directly managed by the TEL central index;
- long-term: the adoption of an inter-lingua approach allows for scaling up the system, provides the means for better integrating the "isolated query translation" and the "pseudo-translation" solution and facilitates sharing components and resources among them. The implementations can be specifically tailored to address the needs of TEL.

Acknowledgments

We thank Marco Dussin for his work on the study of the modifications to the TEL system user interface for introducing the "isolated query translation" feature. Many thanks are also due to Bill Oldroyd of the British Library for his assistance in obtaining the set of records used for the experiments on pseudo-translated, expanded records. Thanks also go to Eric van der Meulen and Julie Verleyen of the TEL office for their help with the current TEL architecture. Thomas Arni helped with running the pseudo-translation experiments, and Carol Peters provided corrections to the paper.

References

[1] Agosti, M., Braschler, M. & Ferro, N. (2006). A Study on how to Enhance TEL with Multilingual Information Access. In C. Thanos, editor, DELOS Research Activities 2006, pp. 115-116, ISTI-CNR At Gruppo ALI, Pisa, Italy.

[2] Agosti, M., Di Nunzio, G. M., and Ferro, N. (2006). A Data Curation Approach to Support In-depth Evaluation Studies. In Gey, F. C., Kando, N., Peters, C., and Lin, C.-Y., editors, Proc. International Workshop on New Directions in Multilingual Information Access (MLIA 2006), pages 65–68. http://ucdata.berkeley.edu/sigir2006-mlia.htm [last visited October 2006].

[3] Agosti, M., Di Nunzio, G. M., and Ferro, N. (2006). Scientific Data of an Evaluation Campaign: Do We Properly Deal With Them? In Nardi, A., Peters, C., and Vicedo, J. L., editors, Working Notes for the CLEF 2006 Workshop. http://www.clef-campaign.org/2006/working_notes/-workingnotes2006/agostiCLEF2006.pdf [last visited October 2006].

[4] Ballesteros, L. & Croft, W. B. (1997). Phrasal translation and query expansion techniques for cross-language information retrieval. In N. J. Belkin, D. Narasimhalu, P. Willett (Eds.), Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 84-91).

[5] Braschler, M. & Peters, C. (2004). Cross-Language Evaluation Forum: Objectives, Results, Achievements. In Information Retrieval, Volume 7, Issue 1/2, 7-31, Kluwer Academic Publishers.
[6] Braschler, M. (2004) Robust Multilingual Information Retrieval. Dissertation. Institut Interfacultaire d'Informatique, Université de Neuchâtel.
[7] Braschler, M., Di Nunzio, G. M., Ferro, N., and Peters, C. (2005). CLEF 2004: Ad Hoc Track Overview and Results Analysis. In Peters, C., Clough, P., Gonzalo, J., Jones, G. J. F., Kluck, M., and Magnini, B., editors, Multilingual Information Access for Text, Speech and Images: Fifth Workshop of the Cross–Language Evaluation Forum (CLEF 2004) Revised Selected Papers, pages 10–26. Lecture Notes in Computer Science (LNCS) 3491, Springer, Heidelberg, Germany.
[8] Braschler, M., Ferro, N. & Verleyen, J. (2006). Implementing MLIA in an existing DL system. In F. C. Gey, N. Kando, C. Peters, and C.-Y. Lin, editors, Proc. International Workshop on New Directions in Multilingual Information Access (MLIA 2006), pp- 73-76, http://ucdata.berkeley.edu/sigir2006-mlia.htm [last visited October 2006].
[9] Braschler, M. (2004). Combination Approaches for Multilingual Text Retrieval. In Information Retrieval, Volume 7, Issue 1/2, 183-204, Kluwer Academic Publishers.
[10] Braschler, M. & Ripplinger, B. (2004). How Effective is Stemming and Decompounding for German Text Retrieval?. In Information Retrieval, Volume 7, Issue 3/4, 291-306, Kluwer Academic Publishers.
[11] Cleverdon, C. W. (1967). The Cranfield tests on index language devices. Aslib Proceedings, 19, 173-192. Reprinted in (Sparck Jones and Willett, 1997)
[12] Di Nunzio, G. M., Ferro, N., Jones, G. J. F., and Peters, C. (2006). CLEF 2005: Ad Hoc Track Overview. In Peters, C., Gey, F. C., Gonzalo, J., Jones, G. J. F., Kluck, M., Magnini, B., Müller, H., and de Rijke, M., editors, Accessing Multilingual Information Repositories: Sixth Workshop of the Cross–Language Evaluation Forum (CLEF 2005). Revised Selected Papers, pages 11–36. Lecture Notes in Computer Science (LNCS) 4022, Springer, Heidelberg, Germany.
[13] Di Nunzio, G. M., Ferro, N., Mandl, T., and Peters, C. (2006). CLEF 2006: Ad Hoc Track Overview. In Nardi, A., Peters, C., and Vicedo, J. L., editors, Working Notes for the CLEF 2006 Workshop. http://www.clef-campaign.org/2006/working_notes/workingnotes2006/dinunzioOCLEF2006.pdf [last visited October 206].
[14] Ferro, N., Braschler, M., Arni, T. & Peters, C. (2006). Deliverable D8.3.1 – Feasibility study on Multilinguality. DELOS, A Network of Excellence on Digital Libraries – IST-2002-2.3.1.12, Technology-enhanced Learning and Access to Cultural Heritage
[15] Frakes, W. B. & Baeza-Yates, R. (1992). Information Retrieval. Data Structures & Algorithms. Prentice-Hall.
[16] Peters, C. & Braschler, M. (2001) European Research Letter: Cross-language system evaluation: The CLEF campaigns, Journal of the American Society for Information Science and Technology, Volume 52, Issue 12, 1067-1072, John Wiley & Sons.
[17] van Veen, T. & Oldroyd, B. (2004). Search and Retrieval in The European Library. A New Approach. D-Lib Magazine, 10(2), February 2004.