A Study on how to Enhance TEL with Multilingual Information Access

Maristella Agosti, University of Padua, agosti@dei.unipd.it
Martin Braschler, Zurich University of Applied Sciences Winterthur, martin.braschler@zhwin.ch
Nicola Ferro, University of Padua, ferro@dei.unipd.it

This work reports on the study conducted in collaboration between DELOS and *The European Library* (TEL)[8], a service fully funded by the participant national libraries members of the *Conference of European National Librarians* (CENL)[9], which aims at providing a co-operative framework for integrated access to the major collections of the European national libraries.

The aim of the collaboration, started in April 2006, is to conduct a feasibility study which identifies the main issues involved in implementing full *MultiLingual Information Access* (MLIA) functionalities in TEL. By full MLIA we mean the possibility for users of TEL to access and search the federated libraries in their own (or preferred) language, retrieve documents in other languages, have the results presented in an interpretable fashion (e.g. possibly with a summary of the contents in their chosen language).

The TEL project aims at providing a "low barrier of entry" in the TEL system to the national libraries which want to join it. This easiness of integration is achieved by extensively using the *Search/Retrieve via URL* (SRU)[10] protocol in order to search and retrieve documents from national libraries. 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 *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) and provides integrated access to them via SRU;
- a gateway between SRU and Z39.50: it allows national libraries which support only Z39.50[11] to be accessible via SRU.

The architecture and functioning of the TEL system pose some problems when planning to introduce MLIA. 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 should be modified and this is an unviable option that would require a very big effort and disregards the "low barrier of entry" guideline adopted in designing the TEL system.

In order to avoid the problem discussed above, still offering some MLIA functionalities, we are considering to introduce an *isolated query translation* step in the query processing. In addition, the TEL central index harvests catalogue records from national libraries, which beside 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 are considering to extend its functionalities by adding a component able to translate the catalogue records in order to perform MLIA on them. We call this approach *pseudo-translation*.

[8] http://www.theeuropeanlibrary.org/

[9] http://www.cenl.org/

[10] http://www.loc.gov/standards/sru/

[11] http://www.loc.gov/z3950/agency/

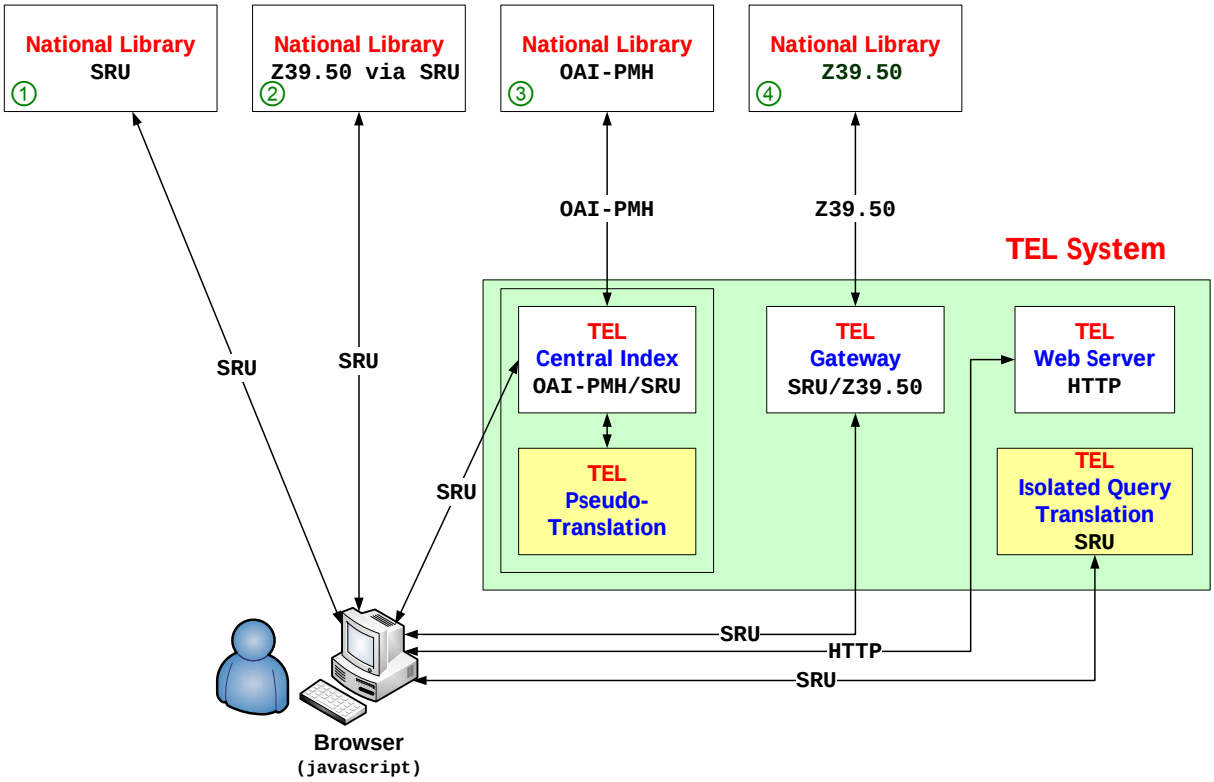

Figure 1: Architecture of the TEL system with the new MLIA functionalities.

Figure 1 shows the architecture of the TEL system with two new components: the first one is going to perform the "isolated query translation", while the second one is going to be responsible for the "pseudo-translation". Note that the "isolated query translation" component can be directly accessed by the client browser by using the SRU protocol and thus the interaction with this new component is actively used by the final user. On the other hand, the "pseudo-translation" component is not directly accessed by the client browser but it represents an extension of the TEL central index, which would be enhanced with MLIA functionalities.

References

[1] Agosti M., Di Nunzio G.M. and Ferro N. 2006. A data curation approach to support in-depth evaluation studies. In F.C. Gey, N. Kando, C. Peters and C.-Y. Lin (eds.), *Proceedings of the International Workshop on New Directions in Multilingual Information Access (MLIA 2006).*

http://ucdata.berkeley.edu:7101/sigir2006-mlia.htm

[2] Braschler M., Di Nunzio G.M., Ferro N. and Peters C. 2004. CLEF 2004: ad hoc track overview and results analysis. In C. Peters, P. Clough, J. Gonzalo, G. J. F. Jones, M. Kluck, and B. Magnini (eds.), *Multilingual Information Access for Text, Speech and Images: Revised Selected Papers from the 5th Workshop of the Cross-Language Evaluation Forum (CLEF 2004) Papers. Bath (UK), 15-17 September 2004*. Lecture Notes in Computer Science 3491. Berlin-Heidelberg: Springer. 10-26.

[3] Braschler M., Ferro N. and Verleyen J.. Implementing MLIA in an existing DL system. In F.C. Gey, N. Kando, C. Peters, and C.-Y. Lin (eds.), *Proceedings of the International Workshop on New Directions in Multilingual Information Access (MLIA 2006).*

http://ucdata.berkeley.edu:7101/sigir2006-mlia.htm

[4] Di Nunzio G.M., Ferro N., Jones G.J.F. and Peters C. 2005. CLEF 2005: ad hoc track overview. In C. Peters, F.C. Gey, J. Gonzalo, G.J.F. Jones, M. Kluck, B. Magnini, H. Müller and M. de Rijke (eds.), *Accessing Multilingual Information Repositories. Revised Selected Papers from the 6th Workshop of the Cross-Language Evaluation Forum (CLEF 2005)*. Lecture Notes in Computer Science 4022. Berlin-Heidelberg: Springer.