

TrebleCLEF: Evaluation, Best Practices and Collaboration for Multilingual Information Access

Carol Peters

Istituto di Scienza e Tecnologie dell'Informazione (ISTI-CNR), Pisa, Italy
carol.peters@isti.cnr.it

Abstract. The objectives of the TrebleCLEF Coordination Action for multilingual information access are outlined, the results obtained in the first twelve months are presented, and the activities planned for 2009 are described.

1 Introduction

The popularity of Internet and the consequent global availability of networked information sources and digital libraries has led to a strong demand for multilingual access and communication technologies. Such technologies should support the timely and cost-effective provision of knowledge-intensive services for all members of linguistically and culturally diverse communities. This is particularly true in the multilingual setting of Europe. However, despite recent research advances, there are still very few operational systems in existence, and these are limited to the most widely used languages. The challenge to be faced is how to best transfer the research results to the wider market place. The objective of the two-year TrebleCLEF project is to build on and extend the results already achieved by the existing Cross Language Evaluation Forum. The aim is not only to support the development and consolidation of expertise in the multidisciplinary research area of multilingual information access (MLIA) but also to disseminate this knowhow to the application communities.

Treble-CLEF thus has three main goals:

1. **Organisation of Annual Evaluation Campaigns and Analysis of Results:** Promoting the academic excellence represented by the Cross Language Evaluation Forum (CLEF) by coordinating research aimed at finding solutions for key problems in the multilingual information access area and providing an infrastructure to evaluate the results and to conduct in-depth analyses.
2. **Technology Transfer and Best Practices:** Launching a concerted action of technology transfer and dissemination of knowhow, tools, resources and best practice guidelines aimed at system developers and application communities.
3. **Dissemination, Community-Building and Training:** Encouraging community building and collaboration through the provision of a discussion forum, by making the scientific data and results produced during the evaluation campaigns publicly available, and by organizing workshops, and training activities.

TrebleCLEF began in January 2008. In the following sections we report on the results obtained during the first twelve months and the activities planned for the second year.

2 Evaluation Campaign

The CLEF 2008 evaluation campaign was a great success with a large number of participants. There were seven main evaluation tracks plus two pilot tasks. The aim, as usual, was to test the performance of a wide range of MLIA systems or system components. Bearing in mind the objectives of TrebleCLEF, in addition to research-oriented tasks, such as the new VideoCLEF track, tasks were organised (i) to meet the needs of a specific application community, e.g. digital libraries with the TEL task using the data of The European Library, the medical image processing sector with ImageCLEFmed activities, and (ii) to examine user behaviour, e.g. the interactive iCLEF track which used a multilingual search game, based on the Flickr database of images, to study user interaction in this context via questionnaires and log analyses. 100 groups, mainly but not only from academia, participated in the campaign. Most of the groups were from Europe but there was also a good contingent from North America and Asia plus a few participants from South America and Africa. In 2008, the evaluation infrastructure was considerably strengthened with the further design and development of the DIRECT system, providing participants with new means of access to the test collections and the campaign results [1]. Full details regarding the design of the tracks, the methodologies used for evaluation, and the results obtained can be found in the on-line Working Notes [2] or the post-campaign Proceedings [3].

The results of the campaign were presented at a two-and-a-half day workshop held in Aarhus, Denmark, 17-19 September 2008 and attended by 150 researchers and system developers. The annual workshop, held in conjunction with the European Conference on Digital Libraries, plays an important role by providing the opportunity for all the groups that have participated in the evaluation campaign to get together comparing approaches and exchanging ideas.

The focus on the user, on log analysis, and on the needs of application communities is being increased in 2009. In line with the TrebleCLEF philosophy, CLEF2009 includes three new tracks focused on analysing user behaviour in a multilingual context (LogCLEF), on studying the requirements of multilingual patent search (CLEF-IP), and on improving our understanding of MLIA systems and their behaviour with respect to languages (GridCLEF) [4].

2 Best Practices and Guidelines

Three best practices white papers will be produced by the project. The first focuses on recommendations for the language resources needed in the development of MLIA systems. A preliminary version of this report was released at the end of 2008 and a final version in Spring 2009 [5]. It provides a survey of the state-of-the-art and presents an assessment of the priority requirements in this sector resulting from consultations with system developers, language industry and communication players. The first part of the report focuses on the description of a generic MLIA system, describing the required modules (morphological analyzers, part-of-speech taggers, syntactic parsers, cross-language components, etc.) An important part of the deliverable is dedicated to a survey of existing language resources that are used and/or are useable for MLIA. On

the basis of a series of inventories and community feedbacks, the report proposes an action plan for developing a set of Language Resources for all technologies related to MLIA. This minimal set of resources and basic tools/modules constitutes our contribution to the setting up of a Basic LAngeuage Resource Kit (BLARK) for MLIA.

The survey also focused on new trends to improve the existing resources for MLIA, both in terms of quality and language coverage. A section is devoted to the description of needs as expressed by key developers. Many developers are working on exploiting the combination of various modalities (e.g. text and image, audio-visual documents, etc.) as well as deploying MLIA for specific domains (e.g. patents). The languages of interest to most respondents were identified: the major European languages are the most cited but less-used European languages, Chinese, Japanese, Arabic, Russian and several others are also named.

The other two white papers will regard "Best Practices in System and User-oriented MLIA" and "Best Practices for Test Collection Creation". In order to obtain input for the first one, which will be released in June 2009, two workshops were organised on the needs of MLIA system developers and MLIA users in 2008 [6,7]. Issues raised at the system developers workshop included the realism of the evaluation resources with respect to use cases and scenarios, performance measures, verifiability and robustness of results, presentation of multilingual results and user-oriented evaluation, and the transferability of results to operational settings. At the User Communities workshop consensus emerged around (i) features that any MLIA system should have from a user's perspective, and (ii) possible strategies to bring MLIA research closer to real world needs and to effectively transfer MLIA technologies to society. The white paper will aim at providing recommendations for successful MLIA implementations, generalizing as much as possible from the mostly academic papers and experiences as there are few "packaged" cross-language offerings in the marketplace. It is hoped that the recommendations will be found to be useful for commercial developers/implementers of complex search applications, who intend to build their own components tailored to their specific applications.

A TrebleCLEF Workshop focused on novel methodologies for evaluation in information retrieval held at ECIR (European Conference on Information Retrieval) on March 30, 2008 in Glasgow, United Kingdom [8] and a first draft survey of test collection evaluation practises in information retrieval have produced input for the Best Practices for Test Collection Creation, Evaluation Methodologies and Language Processing Technologies deliverable due for release in October 2009.

In addition to completing the best practice guidelines, TrebleCLEF will also build a portal with pointers to relevant resources and tools to aid implementers and developers in adopting the recommendations. We hope to influence future academic research, addressing the gaps which have been identified during these studies.

3 Dissemination and Training

Dissemination activities in this first year have mainly concentrated on the scientific community with the organisation of the workshops mentioned above, a good number of presentations at conferences and invited talks, and a large number of publications. In

2009 the dissemination training activities will be directed more consistently towards the application communities with the presentation of the Best Practice Guidelines and at the MLIA Technology Transfer Day, planned for December 2009. A major training activity will be the TrebleCLEF Summer School to be held in June 2009. The aim of the Summer School is to give participants a grounding in the core topics that constitute the multidisciplinary area of Multilingual Information Access. Both theoretical and practical issues are addressed. The focus of the school is on "How to build effective multilingual information retrieval systems and how to evaluate them" [9].

4 Conclusions

It is still too early to assess to what extent TrebleCLEF will be able to realise all its objectives. While the continually increasing impact of the evaluation campaign activity is demonstrated by the rise in numbers of participants and the increase in complexity of the tasks, this is still mainly concentrated in the academic circles. It is more difficult to breakthrough and influence industry. However, the response to our workshops and surveys has been encouraging and from the input received it is clear that the need for guidelines on how to put MLIA theory into commercial practice is becoming a more widely acknowledged reality.

References

1. Ferro, N. Operational Scientific Digital Library. TrebleCLEF D2.2 <http://www.trebleclef.eu/publications.php>
2. http://clef-campaign.org/2008/working_notes/CLEF2008WN-Contents.html
3. Peters, C., Deselaers, T., Ferro, N., Gonzalo, J., Jones, G.J.F., Kurimo, M., Mandle, T., Peñas, A., Petras, V. Evaluating Systems for Multilingual and Multimodal Information Access, 9th Workshop of the Cross-Language Evaluation Forum, Aarhus, Denmark, September 17-19, 2008, Springer LNCS 5706, in print.
4. <http://www.clef-campaign.org/2009.html>
5. Moreau, N. Best Practices in language Resources for MLIA. TrebleCLEF D5.2 <http://www.clef-campaign.org/>
6. Braschler, M., Clough, P.D. Bringing Multilingual Information Access to Operational Systems. TrebleCLEF System Developers Workshop. D3.3 <http://www.trebleclef.eu/publications.php>
7. Gonzalo, J., Peñas, A., Verdejo, F., Peters, C. Workshop on Best Practices for the Development of Multilingual Information Access Systems: the User Perspective TrebleCLEF User Communities Workshop D3.2 <http://www.trebleclef.eu/publications.php>
8. Sanderson, M. Workshop on Novel Methodologies for Evaluation in Information Retrieval. In SIGIR Forum, June 2009, 43(1), pp59-62. <http://www.sigir.org/forum/2009J/2009j-sigirforum-sanderson.pdf>
9. <http://www.trebleclef.eu/summerschool.php>