

Evaluation of Technology for Multilingual Information Access: the Next Step

Carol Peters
ISTI-CNR, Pisa, Italy

There is general consensus that well-organised worldwide evaluation initiatives contribute significantly to the building of strong research communities, advancement in state-of-the-art, and industrial innovation in a given domain. In the Information Retrieval (IR) field, there are currently three major internationally recognized activities of this type: the Text Retrieval Conference¹ (TREC) series organized by the US National Institute of Standards and Technology (NIST); the NTCIR Evaluation of Information Access Technologies² organized by the National Institute of Informatics, Japan; the Cross Language Evaluation Forum³ (CLEF), recently supported by the European Commission under FP7⁴.

CLEF has been running for almost ten years now with the main goal of sustaining the growth of excellence in language processing and multilingual information access (MLIA) across language boundaries within the global context of the multilingual Web. Over the years, strongly motivated by the need to promote the study and utilisation of languages other than English on the Internet, a core network of research institutions involved with CLEF, with some support for the central coordination mainly from the DELOS Network of Excellence for Digital Libraries, has produced the following significant results:

- Creation of a very active multidisciplinary international research community, with strong interactions with both TREC and NTCIR including coordination of schedules and activities;
- Investigation of core issues in MLIA which enable effective transfer over language boundaries, including the development of multiple language processing tools (e.g. stemmers, word decomposers, part-of-speech taggers); creation of linguistic resources (e.g. multilingual dictionaries and corpora); implementation of appropriate cross-language retrieval models and algorithms for different tasks and languages;
- Creation of important reusable test collections and resources in diverse media for a large number of European languages, representative of the major European language typologies;
- Significant and quantifiable improvements in the performance of MLIA systems;

However, since CLEF began the associated technologies, services and users of multilingual IR systems have been in continual evolution, with many new factors and trends influencing the field. For example, the growth of the Internet has been exponential with respect to the number of users and languages used regularly for global information dissemination. The expectations and habits of users are constantly changing, together with the ways in which they interact with content and services, often creating new and original ways of exploiting them. Language barriers are no longer seen as inviolable and there is a growing dissatisfaction with the technologies currently available to overcome them.

This constantly evolving scenario poses challenges to the research community which must react to these new trends and emerging needs. CLEF initially assumed a user model reflecting simple information seeking behavior: the retrieval of a list of relevant items in response to a single query that could then be used for further consultation in various languages and media types. This simple scenario of user interaction has allowed researchers to focus their attention on studying core technical issues for CLIR systems and associated components.

¹ <http://trec.nist.gov/>

² <http://research.nii.ac.jp/ntcir/>

³ <http://www.clef-campaign.org/>

⁴ CLEF is currently run as an activity of the TrebleCLEF Coordination Action; TrebleCLEF is responsible for disseminating the results of CLEF to application and industrial communities, see <http://www.treblecle.eu/>

If we are to continue advancing the state-of-the-art in multilingual information access technologies, we now need to rethink and update this user model. We have to study and evaluate multilingual issues from a communicative perspective rather than a purely retrieval one. We need to examine the interactions between four main entities: users, their tasks, languages, and content in order to understand how these factors impact on the design and development of MLIA systems. It is not sufficient to successfully cross the language boundary, results must be retrieved in a form that is interpretable and reusable. Future cross-language system evaluation campaigns must activate new forms of experimental evaluation - laboratory and interactive – in order to foster the development of MLIA systems more adherent to the new user needs. We need a deeper understanding of the interaction between multicultural and information proactive users, multilingual content, language-dependent tasks, and the enabling technologies consisting of MLIA systems and their components.

At the same time, benchmarking efforts must prove their usefulness for industrial take-up; evaluation initiatives risk being seen as irrelevant for system developers if the data they investigate are not of realistic scale and if the use cases and scenarios tested do not appear valid.

Future editions of CLEF should thus introduce a new series of evaluation cycles which move beyond the current set-up, impacting on:

- Methodology definition: evolution of the current evaluation paradigm, developing new models and metrics to describe the needs and behavior of the new multicultural and multi-tasking users;
- System building: driving the development of MLIA systems and assessing their conformity with respect to the newly identified user needs, tasks, and models;
- Results assessment: measuring all aspects of system & component performance including response times, usability, and user satisfaction
- Community building: promoting the creation of a multidisciplinary community of researchers which goes beyond the existing CLEF community by building bridges to other relevant research domains such as the MT, information science and user studies sectors, and to application communities, such as the enterprise search, legal, patent, educational, cultural heritage and infotainment areas;
- Validation of technology: providing a reasonably comprehensive typology of use cases and usage scenarios for multilingual search, validated through user studies, to enable reuse of appropriate resources and to enable common evaluation schemes;
- Technology transfer: guaranteeing that the results obtained are demonstrated as useful for industrial deployment.

Achieving this goal will require further synergy between various research communities including machine translation, information retrieval, question answering, information extraction, and representatives from end user groups. Furthermore, if this programme is to be implemented, it is clear that CLEF – or any similar evaluation initiative – cannot operate only via a voluntary networking basis; a solid underlying management and coordination structure is crucial in order to ensure that the programme of activities is viable, consistent and coherent and that CLEF can successfully scale up and embrace new communities and technological paradigms.

The presentation will focus on:

- the importance of evaluation activities in the promotion of system development
- the achievements and limits of the Cross Language Evaluation Forum
- proposals for future directions that guarantee continuing progress in MLIA system research and development and effective transfer of the results to application communities.