PROMISE – Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation

Emanuele Di Buccio, Marco Dussin, Nicola Ferro,
Ivano Masiero, and Gianmaria Silvello

Department of Information Engineering, University of Padua
{dibuccio,dussinma,ferro,masieroi,silvello}@dei.unipd.it

Abstract. Measuring is a key to scientific progress. This is particularly true for research concerning complex systems, whether natural or human-built. PROMISE will provide a virtual laboratory for conducting participative research and experimentation to carry out, advance and bring automation into the evaluation and benchmarking of complex multilingual and multimedia information systems.

1 Introduction and Motivation

With a population of over 500 million in its 27 states in which more than 80 indigenous and many more immigrant languages are found, the citizens and companies of the EU demand information systems that allow them to interact with the culturally and politically diverse content that surrounds them in multiple media. Moreover, with the advance of broadband access and the evolution of both wired and wireless connection modes, users are now not only information consumers, but also information producers: language and media barriers are no longer seen as inviolable and they are constantly crossed and mixed to provide content that can be accessed on a global scale within a multicultural and multilingual setting.

The technology and research behind multilingual and multimedia information systems are, today, in the position of intercepting these emerging trends but their design and development is becoming increasingly complex and needs proper means for ensuring that they meet the expected user requirements and provide the desired effectiveness.

2 Scientific Challenges

We consider experimental evaluation both laboratory and interactive a key means for supporting and fostering the development of multilingual and multimedia information systems which are more adherent to the new user needs in order to ensure that they meet the expected user requirements, provide the

M. Agosti et al. (Eds.): IRCDL 2011, CCIS 249, pp. 140–143, 2011.

desired effectiveness and efficiency, guarantee the required robustness and reliability, and operate with the necessary scalability. PROMISE[1] (Participative Research labOratory for Multimedia and Multilingual Information Systems Evaluation) aims at advancing the experimental evaluation of complex multimedia and multilingual information systems in order to support individuals, commercial entities, and communities who design, develop, employ, and improve such complex systems. The overall goal of PROMISE is to deliver a unified environment collecting data, knowledge, tools, methodologies, and the user community which are involved in the experimental evaluation.

PROMISE is a network of excellence (contract n. 258191), funded in the 7th Framework Programme of the European Commission, with 10 academic and industrial partners, who are leaders in the field: University of Padua, Italy (coordinator); Swedish Institute of Computer Science (SICS), Sweden; University of Amsterdam, The Netherlands; Sapienza University of Rome, Italy; University of Applied Sciences Western Switzerland, Switzerland; Vienna University of Technology, Austria; Zurich University of Applied Sciences, Switzerland; Humboldt-Universitt zu Berlin, Germany; Evaluations and Language resources Distribution Agency (ELDA), France; Centre for the Evaluation of Language Communication Technologies (CELCT), Italy.

The PROMISE project lasts three years, starting from September 2010 and ending in August 2013.

To achieve its goals, PROMISE will pursue the following activities:

- **Foster the adoption of regular and thorough experimental evaluation activities:** it will carry on the successful and renowned CLEF[2] (Cross-Language Evaluation Forum) evaluation campaigns further pushing the evaluation exercises to tackle realistic tasks, use cases, and data sets;
- **Bring automation into the experimental evaluation process:** it will propose methods and provide software infrastructure to create larger experimental collections; increase the number and size of the experiments conducted; and develop distributed, asynchronous, and loosely-coupled evaluation protocols, moving experimental evaluation from handicraft process to a mostly automatic one.
- **Promote collaboration and re-use over the acquired knowledge-base:** it will curate, preserve, and enrich the collected experimental data; provide the means for an easy comparison with and a meaningful interpretation and visualisation of the experimental results; and facilitate the discussion and collaboration among all the interested stakeholders.
- **Stimulate knowledge transfer and uptake:** PROMISE will disseminate know-how, tools, and best practices about multilingual and multimedia information systems; facilitate uptake and participation by commercial entities and industries; and give rise to multidisciplinary competencies and expertises.

[1] `http://www.promise-noe.eu/`
[2] `http://www.clef-campaign.org/`

3 Key Technologies and Contribution by Italian Research Community

Large-scale evaluation initiatives, such as Text REtrieval Conference (TREC) in the United States, the CLEF in Europe, and the NII-NACSIS Test Collection for IR Systems (NTCIR) in Asia, contribute significantly to advancements in research and industrial innovation in the information retrieval sector, and to the building of strong research communities. A study conducted by NIST reports that "for every $1 that NIST and its partners invested in TREC, at least $3.35 to $5.07 in benefits accrued to IR researchers. The internal rate of return (IRR) was estimated to be over 250% for extrapolated benefits and over 130% for unextrapolated benefits".

Large-scale evaluation campaigns produce a huge amount of extremely valuable scientific data which provides the foundations for subsequent scientific production and system development and constitutes an essential reference for literature in the field. This data is also economically valuable, due the considerable effort devoted to its production: the NIST study estimates in about 30 million dollars the overall investment in TREC.

Nevertheless, little attention has been paid over the years to modelling, managing, curating and accessing the scientific data produced by evaluation initiatives, despite the fact that the importance of scientific data in general has been highlighted by many institutional organizations, such the European Commission, the US National Scientific Board, and the Australian Working Group on Data for Science.

Our goal is to deliver a unified infrastructure and environment for data, knowledge, tools, methodologies, and the user community in order to advance the experimental evaluation of complex multimedia and multilingual information systems[1]. The evaluation infrastructure will:

- manage and provide access to the scientific data produced during evaluation activities;
- support the organization of evaluation campaigns;
- increase the automation of the evaluation process;
- provide component-based evaluation;
- foster the usage and understanding of the scientific data;

A user-centered design approach will be adopted involving the different stakeholders, e.g. scientists, evaluation campaign organizers, system developers, students, in the development of the infrastructure.

The evaluation infrastructure is not limited only at managing the experimental results but aims also at exploring how we can improve the comprehension of and the interaction with the experimental results by researchers and system developers. We imagine the following scenarios: (i) a researcher or a developer is attending the workshop of one of the large-scale evaluation campaigns and s/he wants to explore and understand the experimental results as s/he is listening at the presentation discussing them; (ii) a team of researchers or developers is working on tuning and improving a system and they need tools and applications

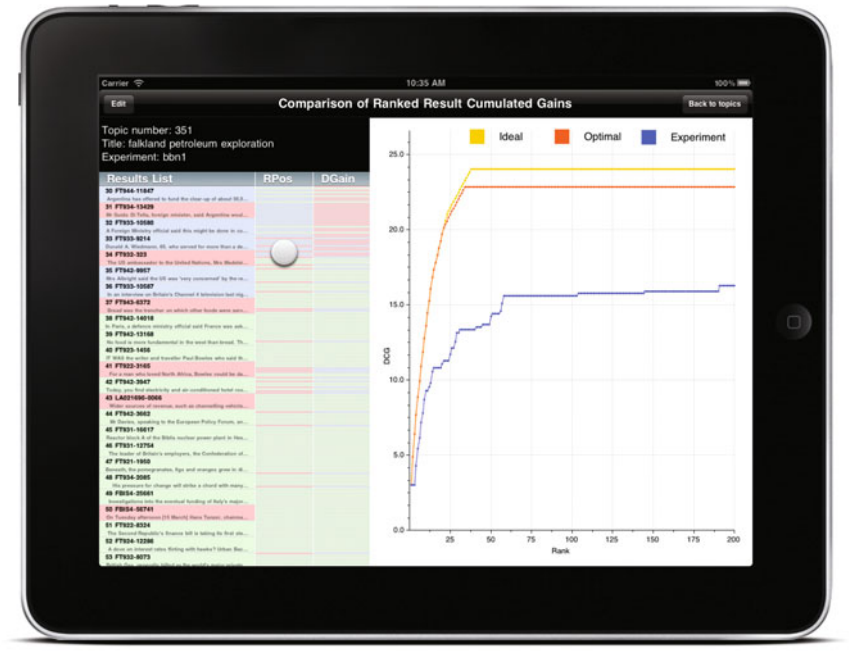

Fig. 1. Prototye iPad application for interactively exploring the experimental results of an information access system

that allow them to investigate and discuss the performances of the system under examination in a handy and effective way. To this end, we are investigating the adoption of innovative devices, such as the iPad, which can allow for a natural and easy interaction with the experimental results and the scientific data [2,3], as shown in Figure 1.

References

1. Ferro, N., Hanbury, A., Müller, H., Santucci, G.: Harnessing the Scientific Data Produced by the Experimental Evaluation of Search Engines and Information Access Systems. Procedia Computer Science 4, 740–749 (2011)
2. Di Buccio, E., Dussin, M., Ferro, N., Masiero, I., Santucci, G., Tino, G.: Interactive Analysis and Exploration of Experimental Evaluation Results. In: Wilson, M.L., Russell-Rose, T., Larsen, B., Kalbach, J. (eds.) Proc. 1st European Workshop on Human-Computer Interaction and Information Retrieval (EuroHCIR 2011), pp. 11–14 (2011), `http://ceur-ws.org/Vol-763/`
3. Ferro, N., Sabetta, A., Santucci, G., Tino, G.: Visual Comparison of Ranked Result Cumulated Gains. In: Miksch, S., Santucci, G. (eds.) Proc. 2nd International Workshop on Visual Analytics (EuroVA 2011). Eurographics Association, Geneve (2011)