DESIRE 2011: First International Workshop on Data infrastructurEs for Supporting Information Retrieval Evaluation

Maristella Agosti
University of Padua, Italy
agosti@dei.unipd.it

Nicola Ferro
University of Padua, Italy
ferro@dei.unipd.it

Costantino Thanos
ISTI-CNR, Pisa, Italy
costantino.thanos@isti.cnr.it

ABSTRACT

The workshop focuses on the three areas of interest to CIKM to discuss how to envisage and design evaluation infrastructures able to store, manage, and make accessible the scientific data and knowledge of interest for advancing the evaluation of information retrieval and access tools.

Main goal is to understand how to make use of the expertise of the three scientific areas in a cooperative way to avoid the duplication of efforts which may occur when addressing the problem separately in each specific area and to trigger synergies and joint actions on the issue.

Main purposes of the workshop are the identification of a roadmap and the definition of initial best practices to guide the development of the necessary evaluation infrastructures.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*Scientific databases*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search process*; H.3.4 [**Information Storage and Retrieval**]: Systems and Software—*Performance evaluation (efficiency and effectiveness)*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Benchmarking, evaluation, methodology*; D.2.8 [**Software Engineering**]: Metrics—*Complexity measures, performance measures*

General Terms

Design, Experimentation, Management, Measurement, Performance

Keywords

experimental evaluation, scientific data, evaluation infrastructure, data test collection, best practices

1. INTRODUCTION

Information Retrieval has a strong and long tradition dating back to the 1960s in producing and processing scientific data resulting from the experimental evaluation of search algorithms and search systems. This attitude towards evaluation has led to fast and continuous progress in the evolution of information retrieval systems and search engines.

DESIRE'11, October 28, 2011, Glasgow, Scotland, UK.
ACM 978-1-4503-0952-3/11/10.

However, in order to make the data test collections, that are used in the context of the evaluation activities, understandable and usable they must be endowed with some auxiliary information, i.e., provenance, quality, context. Therefore, there is a need for metadata models able to describe the main characteristics of evaluation data. In addition, in order to make distributed data collections accessible, sharable, and interoperable, there is a need for advanced data infrastructures.

In contrast, the information retrieval area has barely explored and exploited the possibilities for managing, storing, and effectively accessing the scientific data produced during the evaluation studies by making use of the methods typical of the database and knowledge management areas. Over the years, the information retrieval area has produced a vast set of large test collections which have become the main benchmark tools of the area and contribute to reproducible and comparable experiments. However, these same collections have not been organised into coherent and integrated infrastructures which make them accessible, searchable, citable, exploitable, and re-usable to all possibly interested researchers, developers, and user communities.

It is thus time for these three communities – information retrieval, databases, and knowledge management – to join efforts, meet, and cooperate to envisage and design useful infrastructures able to coherently manage pertinent data collections and sources of information, and so take concrete steps towards developing them.

To address the workshop issues, experts have been invited to give two keynote addresses and six relevant papers have been accepted, among the submitted ones, for presentation. The two following sections are reporting the addressed specific topics.

2. KEYNOTE ADDRESSES

The keynote address by Norbert Fuhr of the University of Duisburg-Essen in Germany[1], entitled *An Infrastructure for Supporting the Evaluation of Interactive Information Retrieval*, addresses the presentation of a testbed for the evaluation of interactive information access. Starting with the INEX[2] interactive track in 2004, the group lead by professor Fuhr developed the Daffodil (now ezDL) framework, providing an experimental framework for interactive retrieval, that allows for easy exchange or extension of the system components. Moreover, this framework also contains tools for

[1] `http://www.is.informatik.uni-duisburg.de/staff/fuhr.html.en`
[2] `https://inex.mmci.uni-saarland.de/`

organizing laboratory experiments. Besides extensive logging (including the possibility to exploit eye tracking data), the system allows for presenting questionnaires at all stages of a search session (pre-/post- task/session), as well as the scheduling of search tasks and monitoring task time.

The keynote address by Maurizio Lenzerini of the Sapienza University of Rome, Italy[3], entitled *Ontology-based data management*, addresses how the ontology-based data management aims at accessing and using data by means of a conceptual representation of the domain of interest in the underlying information system. The talk provides an introduction to ontology-based data management, by illustrating the main ideas and techniques for using an ontology to access the data layer of an information system. Then, it describes an architecture for ontology-based data access and discussed the issue of choosing the appropriate language for expressing the various components of the architecture, by illustrating the main advantages one gains in managing the information system through the ontology. Finally, the issue of developing methodologies and tools for the design and usage of ontology-based data management solutions have been explained.

3. POSITION AND COMMUNICATIONS PAPERS

The paper *Principles for Robust Evaluation Infrastructure* by Justin Zobel (Department of Computer Science and Software Engineering, The University of Melbourne, Australia), William Webber (Department of Computer Science and Software Engineering, The University of Melbourne, Australia), Mark Sanderson (School of Computer Science and Information Technology, RMIT University, Australia), and Alistair Moffat (Department of Computer Science and Software Engineering, The University of Melbourne, Australia) makes reference to the standard "Cranfield" approach to the evaluation of information retrieval systems that has been used and refined for nearly fifty years. Over the last few years, investigation of the strengths and limitations of this approach have led to identification of serious flaws in some experiments. Since the knowledge of these flaws can prevent their perpetuation into future work and informs the design of new experiments and infrastructures, the authors review relevant aspects of evaluation and, based on their research and observations over the last decade, outline principles on which new infrastructures should rest.

The paper *A Lightweight Framework for Reproducible Parameter Sweeping in Information Retrieval* by Richard Eckart de Castilho (Ubiquitous Knowledge Processing Lab, Technical University of Darmstadt, Germany) and Iryna Gurevych (Ubiquitous Knowledge Processing Lab, Technical University of Darmstadt, Germany) introduces a lightweight framework for parameter sweep experiments geared towards evolution, efficiency and reproducibility of experiments running on a single machine. To reduce the computational effort of running an experiment with many different parameter settings, the framework uses the tasks and the dataflow dependency information to maintain and reuse intermediate results whenever possible.

The paper *Evaluation with the VIRTUOSO platform* by Gérard Dupont (CASSIDIAN, Elancourt, France), Gaël de Chalendar (CEA, Fontenay-aux-Roses, France), Khaled Khelliff (CASSIDIAN, Val de Reuil, France), Dmitri Voitsekhovitchy (CEA, Fontenay-aux-Roses, France), Géraud Canet (CEA, Fontenay-aux-Roses, France), and Stéphan Brunessaux (CASSIDIAN, Val de Reuil, France) describes a software architecture for providing an open technical framework for the integration of tools for collection, processing, analysis and communication of open source information. The integration of heterogeneous components is implemented in a way that also permit the comparison of capabilities of multiple tools. The platform that supports the evaluation framework has been named VIRTUOSO. It supports an evaluation framework that allows to deploy and run evaluation kits for different use-cases.

The paper *Use Cases as a Component of Information Access Evaluation* by Jussi Karlgren, Anni Järvelin, Preben Hansen, and Gunnar Eriksson, all authors at the Swedish Institute of Computer Science, Sweden, argues that use cases for information access can be written to give explicit pointers towards benchmarking mechanisms and that if use cases and hypotheses about user preferences, goals, expectation and satisfaction are made explicit in the design of research systems, they can more conveniently be validated or disproved - which in turn makes the results emanating from research efforts more relevant for industrial partners, more sustainable for future research and more portable across projects and studies.

The paper *PatOlympics - An Infrastructure for Interactive Evaluation of Patent Retrieval Tools* by Mihai Lupu (Vienna University of Technology, Austria) presents the infrastructure behind the PatOlympics interactive evaluation campaign. This infrastructure, consisting of a relational database back-end, a Java processing core and a JavaScript interface, makes it possible for real users and researchers to interact in a competitive environment, while maintaining, to the extent possible, the evaluation procedures of standard information retrieval campaigns.

The paper *Infrastructure and Workflow for the Formal Evaluation of Semantic Search Technologies* by Stuart N. Wrigley (Department of Computer Science, University of Sheffield, UK), Raùl García-Castro (Facultad de Informàtica Universidad Politécnica de Madrid, Spain), and Càssia Trojahn (INRIA and LIG, Montbonnot Saint Martin, France) describes an infrastructure for the automated evaluation of semantic technologies and, in particular, semantic search technologies. For this purpose, an evaluation framework is introduced which follows a service-oriented approach for evaluating semantic technologies and uses the Business Process Execution Language (BPEL) to define evaluation workflows that can be executed by process engines.

Acknowledgments

We would like to thank those institutions and individuals who have made this workshop possible: the ACM CIKM 2011 Conference organisation, the Program Committee members, and the Department of Information Engineering of the University of Padua, Italy.

The DESIRE 2011 workshop has been supported by the PROMISE network of excellence[4] (contract n. 258191), as part of the 7th Framework Program of the European Commission.

[3] http://www.dis.uniroma1.it/~lenzerin/index.html/

[4] http://www.promise-noe.eu/