PROMISE Winter School 2012
Information Retrieval
meets
Information Visualization

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

Tiziana Catarci
Sapienza University of Rome, Italy
catarci@dis.uniroma1.it

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

Henning Müller
University of Applied Sciences Western Switzerland,
Sierre, Switzerland
henning.mueller@hevs.ch

Giuseppe Santucci
Sapienza University of Rome, Italy
santucci@dis.uniroma1.it

1   Introduction

In the context of the European Union (EU) funded research project PROMISE, a winter school was organized in the small ski resort of Zinal, Valais, Switzerland from January 23-27, 2012[1].

The title of the winter school was *From Information Retrieval to Information Visualization* and the goal was to bring together these two research domains that are currently quite separated but have an important potential to help each other in advancing the fields.

Indeed, the school has been attended by participants who came from one domain or the other and offered them the possibility of starting to acquire cross-disciplinary competencies. Interestingly enough, the school turned out to be a brainstorming and discussion opportunity also for the lecturers, since they had the occasion of meeting colleagues from a quite different field with

---

[1]http://www.promise-noe.eu/events/winter-school-2012/

1

their own perspectives on a ground of shared topics and issues, such as how to envision models and design systems around user needs, how to consider the user interaction and context, how to conduct evaluation, an so on.

17 high quality lecturers from academia and industry were invited to speak on a large variety of topics from introduction talks to hot topics such as crowd sourcing and social media. 62 participants from 25 countries and all continents but Australia and four organizers followed the courses and helped to create many lively discussions and an open atmosphere with many questions. Also most of the speakers stayed for the entire week and enriched the discussions as well.

Being in a remote place has the inconvenience to complicate organization but has the advantage that people stay together beyond the courses, so in the free time during the day and the evenings. This included many lively discussions accompanied sometimes by a beer. On the other hand, over 20 participants had to be rejected due to a lack of hotel room and several of the participants had to share rooms to optimize the number of participants.

2 Lectures

A total of 17 lecturers presented each time 90 minutes on a specific topic. The goal was to have every day aspects of information retrieval (IR) and information visualization (IV), so as to mix the topics and interests of the participants as much as possible.

Introduction to Information Retrieval – Keith van Rijsbergen, University of Glasgow, Scotland, UK

The introduction to information retrieval explained the main concepts and history of more than 50 years of the research domain of information retrieval, taking into consideration both modeling and experimentation. After detailing the main concepts and providing reference to many interesting papers, prof. van Rijsbergen moved to more advanced topics, such as the impact of the notion of relevance on its probabilistic modeling and the application of geometry and quantum probability notions to develop an IR model.

Introduction to Information Visualization – Alan Dix, Lancaster University, UK

Similar to the introduction to information retrieval, a basic lecture in information visualization was given by prof. Dix in a very lively fashion. He discussed the notion of visualization and its relation with direct sensory experience, the benefits of exploiting visualization to represent data and information, and provided several example of both static and interactive visualizations that have been developed over the time.

HCI Principles – Maria Francesca Costabile, University of Bari, Italy

In this lecture, prof. Costabile covered the definition of Human-Computer interaction (HCI) and its multidisciplinary nature. She also talked about the notion of usability, by also providing several examples of it. Then, she illustrated how to exploit HCI principles to design and develop usable systems with particular reference to the notion of user experience. Finally, she introduced how to apply HCI principles to the evaluation of systems.

HCI View of Information Retrieval Evaluation – Tiziana Catarci, Sapienza, University of Rome, Italy

Prof. Catarci tried to combine the domains of information retrieval, human-computer interaction, and usability to come up with a HCI viewpoint on IR evaluation where user interface techniques, interaction measures, usability models and usability evaluation can actually help in carry out evaluation of IR systems. To this end, she presented several IR evaluation metrics and measures of information needs with the goal of sketching a framework for usability evaluation in IR.

User-oriented Information Retrieval – Elaine Toms, University of Sheffield, UK

This lecture went into the details of the different models of information which describe the information use environment and position information retrieval within the real word context. In particular, prof. Toms covered the issue of describing the problem space entailed by information retrieval in context and discussed the notion of "search task" as one step in the processing of performing a larger "work task" that is part of "work system" that may be impacted by situation.

Evaluating User-oriented IR, Kalervo Järvelin, University of Tampere, Finland

Prof. Järvelin explained the basic concepts of evaluation, ranging from batch laboratory evaluation to user-centered evaluation, linking this concepts to the previous lectures on user-oriented IR and HCI view on IR evaluation. He introduced the notion of simulation, as a kind of bridge between laboratory evaluation and user-centered evaluation. Finally, he presented the evaluation of operational systems and discussed how evaluation both needs a theory of the system being evaluation and helps to construct a theory such system.

Multilingual Information Retrieval – Jacques Savoy, University of Neuchâtel, Switzerland

Prof. Savoy presented the many aspects and challenges of multilingual information retrieval and also the differences of meaning in the same languages. He discussed the problems related to indexing multilingual documents – e.g. encoding, word decompounding, stemming – the translation problem and different alternatives for it, and matching queries and documents in a multilingual scenario. Many practical examples made it clear how difficult the multilingual and multicultural aspects of information retrieval are.

Multimedia Information Retrieval – Stéphane Marchand-Maillet, University of Geneva, Switzerland

The presentation on multimedia retrieval showed many of the difficulties to extract useful data from multimedia data such as images and videos and use this alone or in combination with text for retrieval. The presentation also explained many of the currently used techniques including machine learning applications that have helped to leverage multimedia retrieval and make it a useful tool in many specialized application domains. Multimedia data analysis is very different from text analysis and the semantic gap, the mismatch between user interpretation and machine comprehension of multimedia raw data still persists.

Bibliometrics/Scientometrics and IR – Peter Ingwersen, Royal School of Library and Information Science, Denmark

The lecture introduced the concepts and definitions of bibliometrics, scientometrics, and Webometrics and how IR is necessary to support the data collection and analyses of these disciplines. Then, prof. Ingwersen discussed how to conduct reference and citation analysis and provided several examples of visualizations which allow us to study the evolution of a research fields, e.g. in terms of its authors and their relationships.

TREC Style Evaluation– Donna Harman, National Institute of Standards and Technology, USA

The presentation provided many insights on the motivation, history, and benefits of laboratory evaluation. Starting from the early work of Cleverdon for the Cranfield experiments, dr. Harman discussed the whole methodology concerning test collection creation and the related issues, such as completeness and robustness of the relevance judgments, and provided many examples concerning the TREC evaluation campaign. Moreover, she provided students with a series of how-tos about participation in evaluation campaigns, re-use of existing collections in your own experimentation, and design of your own test collection.

Metrics, Statistics, Tests – Stephen Robertson, Microsoft Research, Cambridge, UK

Prof. Robertson provided many insights about metrics and the type of variables you can observe and went into the details of the IR evaluation tradition. He covered metrics based on relevance and assumptions about them, he presented work on commercial Web search engines, and discussed on questions of statistical significance. Finally, he gave an outlook of possible future directions beyond Cranfield, such as considering session rather than single queries or focusing on information rather than documents.

Number Visualization – Giuseppe Santucci, Sapienza, University of Rome, Italy

Prof. Santucci discussed on how to present and visualize quantitative information via tables and graphs. In particular, he provided several examples of good and bad graphs in order to distill a set of guidelines and rules to be followed. Then, he presented the basic rules of quantitative perception and the role of interaction in information visualization and provided two examples of application of such concepts to the visualization of experimental results coming from the IR evaluation.

Visual Analytics – Silvia Miksch, Vienna University of Technology, Vienna, Austria

The lecture presented the motivations for visual analytics, meant as the science of analytical reasoning facilitated by interactive visual interfaces. Prof. Miksch discussed how visual analytics intersects three areas, namely visualization and interaction, computation and mining, and human perception and cognition. She introduced several examples of visual and interactive prototypes with a particular focus on text and document visualizations and visual analytics of time-oriented data.

Web Personalisation – Owen Conlan and Vincent P. Wade, Trinity College Dublin, Ireland

The lecture concerned the motivations behind personalization in the Web and its potential impact. Dr. Conlan defined personalisation, presented several of its dimensions, the techniques for adapting content, and the approaches for user modelling. The whole problem of personalization has been then framed in the wider context of the history and evolution of adaptive hypermedia and case studies from the CULTURA and AMAS projects have been discussed.

Log File Analysis – Maarten de Rijke, University of Amsterdam, The Netherlands

Prof. de Rijke discussed log analysis from the standpoint of the behaviour – observable activities of a person, a team, a system – we can detect and record in log files. He highlighted how logs allow us to trace data and actions and not just to have access to isolated snapshots. He presented the impact of log analysis, in terms system design and optimization and improvement of models of user interaction and behaviour. Several examples of logs have been introduced and several techniques for analysing them have been explained. Finally, a number of uses of log files has been discussed such as, for example, query expansion and suggestion, simulation, learning to rank, and interleaved comparison.

Crowdsourcing – Gareth Jones, Dublin City University, Ireland

Prof. Jones talked about crowdsourcing as a form of human computation and how the availability of crowdsourcing services is now making human computation easily available to the research community. He provided several examples of implicit or explicit crowdsourcing systems and alternative platforms which are available to develop and implement them. Then, he discussed the issues involved in crowdsourcing, such as recruitment, reputation, payment and incentives, and spam detection. Finally, he applied crowdsourcing to IR evaluation and presented MediaEval as a case study.

Retrieval from Social Media – Alejandro Jaimes, Yahoo! Research Barcelona, Spain

Dt. Jaimes presented an intriguing view of what social media are, how they are shaping our lives, and why it is important to leverage on them both from a research and a commercial perspective. He discussed several examples of social media and application built over them, as well as how to exploit them for retrieval purposes. Finally, he discussed some future directions and possible interesting research issues on this topic.

3   Poster session

All participants had the possibility to present their own work during the first day of the winter school at the evening welcome reception that took two hours and started many discussions among the participants.

An evaluation and selection of the posters has been performed and a symbolic best poster award – a bottle of excellent Swiss wine – has been attributed to the three best posters.

4   Conclusions

The fact that participants were remaining close together during all five days of the winter school and had many possibilities to meet with the other participants and the lecturers gave place to many discussions and to a stimulating environment for both the participants and the lecturers.

Altogether the PROMISE winter school can be seen as a great success in connecting two research domains and allowing a large number of participants to get in contact with high quality lecturers and give them hopefully a better view of the research domains and also on the ways that they can evaluated their own research and profit from tools of visualization that are available. Most participants gave a very positive feedback and hopefully the proceedings of the winter school will also help to keep the main outcomes of the winter school available for the future and persons who could unfortunately not participate.

An analysis of the evaluation forms after the winter school highlighted that most students very much enjoyed (more than 90% of the participants) the winter school and the atmosphere among participants and with the lecturers. Most presentations were liked (about 90% of the participants) but sometimes the introductory presentations were regarded as too simply for a majority of PhD students. The students were generally interested in the different topics offered by the school (about 90% of the participants) The place in a remote alpine valley was very much appreciated but the lack of a professional conference room and of limited infrastructures as lunch time were also regarded as problematic. Lectures of 90 minutes were regarded as too long and maybe short breaks after 45 minutes would have been a better option.

The proceedings of the lectures of the winter school are currently under preparation and will be published in the Springer Tutorials series.

Acknowledgments

---

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

[3]http://www.khresmoi.eu/

[4]http://www.cultura-strep.eu/

[5]http://www.elias-network.eu/