Report on the 3rd MUMIA Training School on Information Retrieval and Interactive Information Access, July 21-25, 2014, FORTH, Heraklion, Crete, Greece

Michail Salampasis
Alexander Technology Educational Institute
(ATEI) of Thessaloniki, Greece
msa@it.teithe.gr

Yannis Tzitzikas
Computer Science Department
University of Crete,
and
FORTH-ICS, Greece
tzitzik@ics.forth.gr

1 Introduction

MUMIA[1] is a research network (funded by the European Cooperation in Science and Technology - COST) on Multilingual and Multifaceted Interactive Information Access. MUMIA started in 2010 and ends on December 2014. About 300 senior scientists, experienced researchers as well as early stage researchers and PhDs mostly from the IR/NLP/MT communities participate in the Action[2]. Overall, the objective of the Action is to implement and efficiently run a series of networking activities that will bring together key players from academia and industry and to enable cross fertilization of disciplines related to information access. Another objective is the exploration of the state of the art of the main disciplines involved in the model of MUMIA and to advance their mutual understanding, a key factor for the development of next generation search systems.

The education and training of new researchers is another key objective. In that context, from 21 to 25 of July 2014, the MUMIA Action in collaboration with the Information Systems Laboratory of the Institute of Computer Science of the Foundation for Research and Technology - Hellas (FORTH), organized in Heraklion (Crete) the 3rd MUMIA Training School on Information Retrieval and Interactive Information Access. This 3rd Training School was typically a week-long event consisting of a series of lectures and seminars delivered from invited lecturers recognized as experts in the field.

The school was intended for PhD/MSc students and post-doctoral researchers interested in the fields of information management, information retrieval and related fields. The aim of the school was to give grounding in the core topics that constitute the interdisciplinary area of Multilingual and Multifaceted Interactive Information Access.

The scientific coordinator of the MUMIA Cost Action for this training school (Yannis Tzitzikas), and the training school program committee collaborated designing the most appropriate set of lecturers for the summer school, contacting the potential lecturers, inviting them to participate and explaining the details concerning MUMIA rules about training schools. Also an open call was sent to more senior MUMIA members inviting them to submit proposals for lectures. The summer school was

[1] http://www.mumia-network.eu/
[2] research networks are called Actions in COST terminology

publicized by means of MUMIA website and in other IR/NLP events, reaching a broad amount of potential attendants. MUMIA provided a good number of scholarships for Master or PhD students for attending the summer school. This was very important for the success of the training school in terms of the high number of attendants, as well as for supporting students who could not have attended the training school for financial reasons. The grants covered some of the students' expenses.

Based on a seminar proposal evaluation procedure the school contained lecturers from the MUMIA network and invited speakers who are recognized experts in the field. In numbers, the school had 13 Professors and Researchers from 9 European countries. In total 54 students from 17 different countries attended the school and 11 of them also presented and discussed their ongoing research in the scheduled poster session.

2 Lectures

The program started with lectures focusing on the main topics of text-based information retrieval, i.e. retrieval models, evaluation and scalability. Then it moved to cross-language and multimedia retrieval. The next session contained lectures focusing on collaborative retrieval, recommender systems and opinion retrieval in social media. Finally, the school was completed with lectures on contextual factors, faceted search over structured data, and methods for bridging the world of documents with the web of data. Each of the thirteen lecturers had two to three hours for presenting their topics. A short description follows.

Introduction to IR: Foundation and Models - 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, providing a thorough discussion over the crucial parts of IR (relevance, evaluation, users and information needs) and taking into consideration both modeling and experimentation. It described in detail the indexing and query process, text statistics, a number of popular retrieval models (Boolean, VSM and probabilistic), as well as language model approaches to IR. During the lecture, students were provided with a number of references to works of the 50's or 60's that a number of core research questions, that are still open today, were first specified.

IR Evaluation ++, Mihai Lupu, *TU-WIEN - Austria*
This lecture covered the fundamental issues of IR evaluation, as observed in evaluation campaigns, considering the perspectives of both their supporters and critics. It gave an overview of the available collections, tools and best practices. It was entitled IR Evaluation++ because, while focusing on the large amount of know-how present in the IR community, it also touched upon other aspects of Information Access Evaluation and put the work in the context of professional environments.

Integrating IR Technologies for Professional Search - Michail Salampasis, *Alexander Technology Educational Institute (ATEI) of Thessaloniki, Greece*
This lecture briefly presented the objectives of MUMIA, and then focused on the key challenges for professional search. It presented a general framework serving as a useful topology for better understanding the design space of professional search systems and how different IR/NLP technologies can be integrated to enable rich information seeking environments where different tools can support specific objectives within a typically lengthy search process.

Task-based Information Interaction Evaluation - Kalervo Järvelin, *University of Tampere, Finland*
This lecture focused on understanding and evaluation of information systems and information behaviors for their contribution to task performance. It discussed extensions to the traditional Cranfield evaluation paradigm and user-centered evaluation, including evaluation of multiple query sessions though simulations and experiments with human participants. Then it proposed an evaluation framework for task-based information interaction based on generic activities in task-processes that can be supported by information systems.

Cross-language and Multilingual IR - Martin Braschler, *Zurich University of Applied Sciences, Switzerland*
The lecture described how to extend basic monolingual indexing and matching, and adapt them for working with other languages. It discussed language identification, tokenization/segmentation (including Asian languages), word normalization, stemming/decompounding, and how these effect the retrieval effectiveness. It discussed NTCIR (for eastern Asian languages) and CLEF (for European languages) evaluation campaigns, and then focused on the problem of Cross-Language Information retrieval (CLIR) covering the various translation strategies.

Scalability and Efficiency Challenges in Large-Scale Web Search Engines - Berkant Barla Cambazoglu, *Yahoo Research, Spain*
This lecture provided an overview of the fundamental scalability and efficiency challenges in commercial web search engines. It stressed that achieving scalability and efficiency requires making careful architectural design choices while devising algorithmic performance optimizations. Since most details about the internal functioning of commercial web search engines remain undisclosed due to their financial value and the high level of competition in the search market, the lecture aimed to bridge the existing gap between the industry and academia.

Collaborative Information Seeking and Search Interfaces - Preben Hansen, *Stockholm University, SICS, Sweden*
This lecture introduced the topic of Collaborative Information Seeking (CIS) and its background and frameworks within the research areas of IR, IS and CSCW. CIS aims at giving teams (in companies or organizations) the capability to search more deeply and more broadly. The lecture highlighted concrete examples of current CIS systems and discussed CIS in the context of work-tasks. Finally it stressed that it is still an open area of research.

Recommender Systems - Kostas Stefanidis, *FORTH-ICS, Greece*
This lecture gave an overview of the existing approaches for computing recommendations. Special emphasis was given on recent approaches that are based on building user models, specifically on approaches that organize users into clusters and then use these clusters, instead of linear scanning the database, for computing predictions more efficiently. Finally, it discussed approaches that handle different temporal aspects of recommendations.

Opinion Retrieval in Social Media - George Paltoglou, *University of Wolverhampton, UK*
This lecture introduced the basic principles of opinion search in social media, and stressed that opinion search combines methodologies from two distinct areas of research: information retrieval and sentiment analysis, where the former deals with the issues of representing, storing and providing access to information, while the latter focuses on the detection, extraction and analysis of affective content. The lecturer provided various examples, and identified the current challenges in the Opinion

retrieval field and described recent research that attempts to address them.

Interactive Content-based Multimedia Retrieval - Stefanos Vrochidis, *Informatics and Telematics Institute, Greece*
The lecture provided an overview of video and image indexing and retrieval techniques and provided an introduction to techniques for the exploitation of user interaction during multimedia retrieval tasks. For the former, it presented techniques for video segmentation and low-level feature generation from audiovisual content, as well as methods for semantic indexing of video and images using concepts and events based on supervised machine learning. For the latter, it focused on the exploitation of aggregated implicit user feedback (e.g. log files, social media interactions) in the search process. The lecturer demonstrated various image and video search systems.

Contextual Search and Contextual Factors Aggregation - Gabriella Pasi, *Milano Bicocca, Italy*
This lecture stressed that relevance is time, situation, and user specific, essentially it is a multi-dimensional notion that depends on many complex factors related to the search task, to the user, to her/his context, and to the document and search context as well. Then it described the problem of contextualizing search, by trying to outline both the research challenges and the user role and involvement in the various identified tasks.

Faceted Search - Sébastien Ferré, *University of Rennes 1, IRISA - France*
This lecture focused on Faceted Search (FS) an interactive information access paradigm offering more flexibility, and covering more search tasks, compared to search engines, database queries, and static navigational structures. The lecture stressed that FS retains the usability, guidance and feedback of navigational structures while offering more expressivity. Subsequently, it detailed how the principles of faceted search can be applied to semantic relational data and the participants had the opportunity to experiment the presented interaction model on public semantic datasets.

Bridging the Web of Documents with the Web of Data - Yannis Tzitzikas, *University of Crete and FORTH-ICS, Greece*
This lecture started by stressing the importance of serving effectively and efficiently recall-oriented information needs (especially in the context of professional search), and to this end it discussed the merits of exploratory search for such information needs. Then it described a process for exploratory search that can bridge the web of documents with the web of data at search time. The related challenges and the current state-of-the art were presented as a series of milestones each associated with one or more research questions and a prototype/ operational search system.

3 Conclusions

As the summer school reached its end, after five days of intense attendance and tutoring, the general feeling was very positive and pleasant, with students and tutors exchanging personal information to keep in touch and gratefully thanking the organizers, marking the 3rd MUMIA Training School a very successful school.

The lectures took place in the Amphitheater of FORTH while the lunches were served outside in a peristyle with a magnificent view of the Psiloritis mountain and of the central/northern part of Crete. In general, there were interesting discussions during and after each lecture. Also there were individual contacts between trainees and lecturers, which have originated research collaborations and research internships. The trainees had the opportunity to know other people working on the same,

similar or complementary research areas, which have also originated research collaborations.

Participants and lecturers made very keen comments regarding the location and organization of the training school, as well as FORTH's facilities. The organizers had arranged the transportation of the participants to the FORTH's buildings by bus, which was picking both the students and the lecturers from the city center all days. This not only made their transportation easier, but also gave the possibility to the participants to meet each other and discuss before and after each day of the school. Since all participants arrived at the scheduled time by bus, there were no starting delays in any of the lectures. The attendance percentage was almost 100% for all days (the participants signed twice, once in the morning and once in the afternoon). During the lectures the students participated actively by asking questions, that led to off-line discussions with the tutors during the coffee break or lunch time. The poster session turned out a very good idea. All lecturers had the chance to look at the posters and discuss with the students, making it a highly interactive session, where presenters were given valuable feedback from experts of the domain. The organized social events allowed them to know each other and establish friendships as evidenced by the social media (google groups and facebook) and the big number of messages and photos that were shared in those mediums.

As regards the program of the school, through some discussions with individual students we identified that there was a group of students that liked most the lectures of the first days (on core IR topics), and another one that liked most the lectures of the last days (on topics related to metadata, semantics, context). This can be explained since the former group of participants included students doing research on “pure IR” topics, while the latter students were doing research on digital libraries and semantic data management. However all students liked all lectures, and they admitted that the attendance of a school with a wide spectrum of topics aided them to understand the different perspectives of the problem of providing advanced search and access services.

At the end of the school the students filled in a questionnaire, in which they could evaluate using a scale with values from 1 to 5 each of the lectures, the organization of the school, the social events, as well as the local facilities including the hotels. After the analysis of the results provided by them, we conclude that the summer school has been overall a success in terms of the high rates given to the scientific program and organization. Another interesting finding is that most of students were either satisfied or very satisfied by the lectures (values 4 and 5 resp.), commenting that they found the school an enjoyable and great experience. They marked that the school was very informative and covered a range of themes in Information Retrieval, with an emphasis on trends and challenges that could be very helpful for their research work. Some of them commented negatively on the intensity of the school program, proposing an extra day. Most of the students commented on the very good organization of the school and the nice social events that let them make new friends.

Acknowledgements and Links

This event was supported by the COST Action MUMIA. Many thanks to the Information Systems Laboratory of FORTH-ICS for organizing and hosting the meeting, especially to Maria Moutsaki and to Panagiotis Papadakos. Many thanks also to the student volunteers from the Computer Science Department of the University of Crete.

The Web page of the school that contains the slides of all lectures is available at: http://www.mumia-network.eu/index.php/training-school-2014