

10th European Summer School in Information Retrieval (ESSIR 2015)

Theodora Tsikrika, Symeon Papadopoulos, Stefanos Vrochidis, Yiannis Kompatsiaris

Information Technologies Institute, Center for Research and Technology Hellas (ITI-CERTH)

{theodora.tsikrika papadop, stefanos, ikom}@iti.gr

http://mklab.iti.gr

Abstract

The 10th European Summer School in Information Retrieval (ESSIR 2015) was held in Thessaloniki, Greece between August 31 and September 4, 2015. The summer school offered high quality lectures on 13 topics in Information Retrieval and related areas, a new edition of the Symposium on Future Directions in Information Access (FDIA), group activities, and a rich social programme. This report provides an overview of a successful international summer school that attracted a total of 51 participants from a wide range of countries in Europe and North America.

1 Introduction

The European Summer School in Information Retrieval (ESSIR) is a scientific event founded in 1990, which has given rise to a series of summer schools held on a regular (currently biennial) basis to provide high quality teaching of Information Retrieval (IR) and advanced IR topics to an audience of research students and early stage researchers. ESSIR is typically a week-long event consisting of guest lectures and seminars from invited lecturers who are recognised experts in the field.

The 10th European Summer School in Information Retrieval¹ (ESSIR 2015) was held in Thessaloniki, Greece between August 31 and September 4, 2015, hosted by the Multimedia Knowledge and Social Media Analytics Laboratory² (MKLab) of the Information Technologies Institute³ (ITI) at the Centre for Research and Technology Hellas⁴ (CERTH).

Thessaloniki (also known as Thessalonica or Salonica) is the second-largest city in Greece and the capital of the Central Macedonia region. At about a million inhabitants, including a large and vibrant student community, it is considered Greece's cultural capital, renowned for its festivals, events and diverse cultural life. Importantly, it is also a city with a continuous 3,000 year old history, including relics of its Roman, Byzantine, Ottoman past, and of its formerly dominant Jewish population. In

¹ <http://mklab.iti.gr/essir2015/>

² <http://mklab.iti.gr/>

³ <http://www.iti.gr/iti/index.html>

⁴ <http://www.certh.gr/>

fact, many of its Byzantine churches and also a whole district of the city are included in UNESCO's World Heritage list. A few of the key sights that attract many of the city's visitors include the White Tower, the Arch of Triumph of Galerius, the 10km long seafront promenade (Figure 1), the Byzantine citadel, the Roman Forum excavations, the Byzantine churches of Agios Demetrios and Agia Sophia, the Turkish public baths of Bey Hamam and Bezesteni, and the Jewish museum. Thessaloniki is also considered to be the culinary capital of Greece, featuring hundreds of restaurants and food spots, covering a wide range of styles, cuisines, and size.



Figure 1: The White Tower (left), the Arch of Triumph of Galerius (center), and the Promenade (right).

The ESSIR 2015 programme consisted of a single-track lecture programme that offered high quality teaching on 13 topics in IR and related areas, a new edition of the Symposium on Future Directions in Information Access (FDIA), group activities, and a rich programme of social events and activities.

The school welcomed 51 participants, mainly post-graduate and doctoral students, with the majority coming from Europe, but with three students coming from the USA and Canada (Figure 2). Thanks to our generous sponsors, 25 scholarships covering the registration fees were offered to ESSIR 2015 participants: 22 scholarships were offered by the ELIAS⁵ (Evaluating Information Access Systems) ESF Research Networking Programme and three scholarships by the ACM Special Interest Group on Information Retrieval⁶ (ACM SIGIR). The total number of attendees including participants, lecturers, organizers, and volunteers was 81.

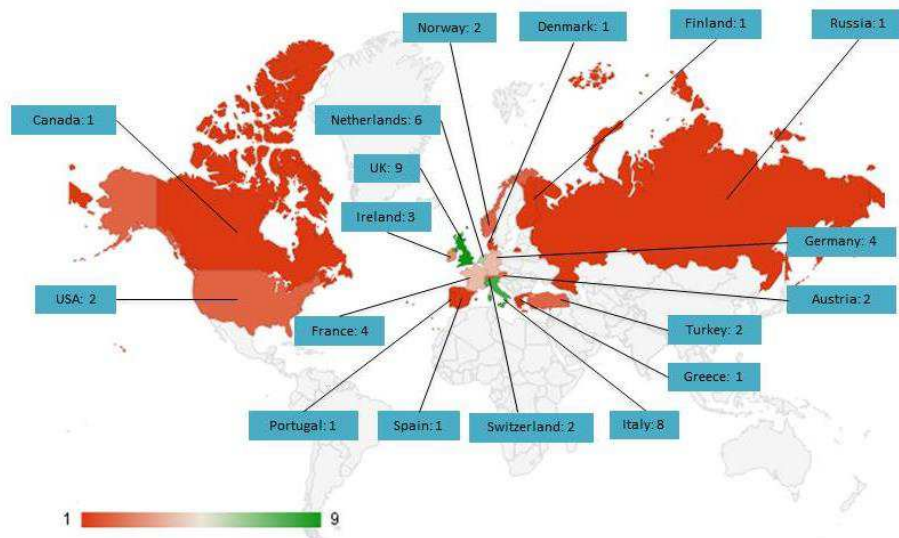


Figure 2: Countries of affiliations of ESSIR 2015 participants.

⁵ <http://elias-network.eu/>

⁶ <http://sigir.org/>

2 Lectures

The ESSIR 2015 programme comprised a wide range of lectures: (i) *foundational topics* that provided a broad overview of key concepts upon which the IR field is built, (ii) *recent and emerging topics* on the current and future developments in IR, and (iii) *special topics* on the inherently interdisciplinary nature of IR. A short description of the lectures is provided below, while slides and videos of all lectures are available online⁷.

Foundational topics

IR Foundational Concepts & Models by Professor Maarten de Rijke (University of Amsterdam)

In this lecture, the canonical architecture of a modern search engine was presented and the important concepts, results, and key research issues around this architecture were introduced. The first part focused on the offline stages of the canonical information retrieval architecture including crawling, document enrichment, and aggregation of external sources. The second part focused on the online stages of the canonical information retrieval architecture including query auto-completion, query understanding, ranking and retrieval, and result page generation. In addition, the main components of a solid experimental evaluation framework were outlined.

IR Evaluation & Experimental Design consisting of four parts as follows:

Axiometrics – An Axiomatic Approach to Evaluation Metrics by Professor Stefano Mizzaro (University of Udine): In this talk, some of the most common IR metrics were introduced and an axiomatic approach to metrics was proposed. A general framework based on Measurement Theory that can be used to express axioms was then presented, thus making explicit the desirable metric properties, and that will hopefully allow to better understand them.

Evaluation Metrics for Clustering and Filtering Tasks by Dr. Julio Gonzalo (Universidad Nacional de Educacion a Distancia): This talk reviewed the most popular evaluation metrics for clustering and filtering tasks and specified a few intuitive formal constraints for each task that every suitable metric should satisfy. These were analyzed in the context of the framework presented by Stefano Mizzaro and the results of the analysis provide useful guidance to select the most adequate evaluation metric for each application scenario.

Evaluation Metrics for Document Retrieval by Dr. Enrique Amigo (Universidad Nacional de Educacion a Distancia): During the first part of this talk, a set of basic formal constraints that a desirable metric should satisfy were defined and analyzed in the context of the framework presented by Stefano Mizzaro. This analysis indicated that only the most recent evaluation metrics are able to satisfy every formal constraint. In the second part of this talk, two complementary evaluation measures, reliability and sensitivity, were described; these satisfy more formal constraints than previously existing evaluation metrics for the ranking, clustering and filtering tasks.

Experimental Design for Collection-based Comparative Evaluation of Search Engines by Dr. Evangelos Kanoulas (University of Amsterdam): Given the great interest in creating test collections that better model the variability encountered in real-life search scenarios, this talk discussed different ways of incorporating user behavior in batch experimentation focusing on how to model the variance introduced to measurements of effectiveness and on how to extend the statistical significance test arsenal to allow for comparing search algorithms.

⁷ http://mklab.itl.gr/essir2015/?page_id=22, http://mklab.itl.gr/essir2015/?page_id=734

Effectiveness and Efficiency in Web Retrieval Systems by Dr. Barla Cambazoglu (Yahoo! Labs)

This lecture provided an overview of the techniques employed by the state-of-the-art web retrieval systems. First, a system-centric view was provided by discussing document indexing and query processing in web retrieval systems. Then, a user-centric view was presented by discussing various evaluation metrics employed for estimating user satisfaction with search results. Finally, the talk focused on efficiency and scalability issues in web retrieval and concluded by briefly discussing the implications of efficiency improvements on users' engagement with the retrieval system.

Human Information Interaction & Retrieval by Professor Diane Kelly (University of North Carolina)

This lecture provided a historical overview of human information interaction and retrieval (HIIR) research. Key ideas and concepts, the (d-)evolution of search interfaces, and methods of evaluation were presented with the goal to provide students with a greater understanding of, and appreciation for, HIIR studies, the kinds of research questions with which they are concerned, and the types of methods that are useful for addressing these questions.

Information Retrieval Infrastructures by Dr. Craig Macdonald (University of Glasgow)

Starting from the traditional information retrieval system architecture, this lecture provided an overview of important IR advances that enhance a search system in four aspects: (i) increasing system efficiency (e.g., through caching strategies, dynamic pruning techniques, and index compression), (ii) ensuring that it copes with big data (e.g., through distributed architectures of retrieval), (iii) processing modern real-time data streams (e.g., Twitter through using frameworks such as Storm or Spark), and (iv) facilitating more effective ranking using state-of-the-art learning to rank techniques that require new search infrastructures to both train and deploy effective supervised search systems.

Recent topics and emerging trends**Social Media Mining & Retrieval** by Dr. Carlos Castillo (Sapienza University of Rome)

During this lecture, methods for processing social media data were introduced with particular focus on two of the conditions that should be met in the context of such applications. First, the data should be processed in real time or with low latency; this creates important algorithmic challenges in areas such as event detection, online classification/clustering, and credibility/veracity assessment, among others. Second, the application designer must be well acquainted with the application domain; this is usually achieved by working in interdisciplinary teams. The introduced methods were exemplified through concrete applications, particularly in the areas of emergency response and disaster relief.

Models for Information Retrieval & Recommendation by Professor Arjen de Vries (CWI & Delft University)

Based on the observation that methods used in online recommendation systems are closely related to the models developed in the information retrieval area, a unified approach to information retrieval and collaborative filtering was presented in this lecture. It was then demonstrated how this will allow for the conversion of a standard information retrieval system into a state-of-the-art recommendation system.

Text Classification, Sentiment Analysis & Opinion Mining by Dr. Fabrizio Sebastiani (Qatar Computing Research Institute)

First, this lecture discussed the main steps towards the construction of a text classifier, from the

generation of vector representations of the texts, to training a classifier from examples, to evaluating its accuracy on benchmark datasets. Then, there was a discussion of how the text representation techniques are adopted in state-of-the-art sentiment classification systems, with particular emphasis on systems that tackle text arising within social media.

Graph-of-words: Boosting Text Mining with Graphs by Professor Michalis Vazirgiannis (LIX, Ecole Polytechnique)

This lecture presented the graph-of-words model, an alternative approach to the bag-of-words model that capitalizes on a graph representation of documents and challenges the word independence assumption by taking into account words' order and distance. In this context, the application of the graph-of-words model in various tasks, such as ad-hoc information retrieval, single document keyword extraction, text categorization, and sub-event detection in textual streams, was discussed.

Special topics

Machine Learning for Information Retrieval by Dr. Katja Hofmann (Microsoft Research Cambridge)

This lecture discussed and exemplified the dual role of IR as both a consumer of Machine Learning (ML) technology, and as a driver towards new challenging ML problems. First, an overview of typical ML applications to IR was presented, including an overview of learning to rank approaches. Next, the recent trends towards online learning approaches that allow continuous learning from user interactions were presented, including existing solutions, open questions, and directions for future research.

Multimedia Analysis & Retrieval by Dr. Vasileios Mezaris (Centre for Research and Technology Hellas – Information Technologies Institute)

This lecture focused on the analysis and retrieval of multimedia content, most notably images and videos. First, the data granularities that are useful for retrieval were introduced and then methods for the temporal segmentation of video to shots and scenes were presented. Then, the problems of image and video indexing with low-level features, high-level concepts, and complex event labels were discussed, along with the machine learning methods that are in the core of multimedia indexing.

Integrating Language, Semantics, & Multimedia for IR by Professor Marie-Francine Moens (Katholieke Universiteit Leuven)

This lecture focused on recent advances in semantic processing of language and more specifically on entity and event recognition. In this context, the joint processing of language and visual data, where language data forms weak annotations to recognize information in images and video, was discussed in depth. A substantial focus was also given on representation learning and the use of topic models and neural embeddings.

Multilingual Summarization by Professor Leo Wanner (Universitat Pompeu Fabra)

This lecture provided an overview of the state of the art in summarization, with a special focus on multilingual techniques. In the first part of the lecture, the traditional distinction between extractive and abstractive summarization was introduced. In the second part, multilingual summarization was described. The third part was dedicated to the presentation of the evaluation measures used to assess the quality of summarization techniques. The lecture concluded with a discussion on how summaries can be taken advantage of in IR.

3 Future Directions in Information Access (FDIA) Symposium

The Symposium on Future Directions in Information Access (FDIA) aims to provide a forum for participants of the summer school to share their research and interact with senior researchers in an informal and relaxed atmosphere. Aimed towards PhD students, researchers new to the field, and post-doctoral researchers, its objectives are: (i) to provide an accessible forum for new researchers to discuss their research and projects, (ii) to help foster formative and tentative research ideas, (iii) to encourage discussion and successful progression, and (iv) to share outcomes of doctoral work.

Now in its sixth edition, FDIA was held as part of the 10th European Summer School in Information Retrieval (ESSIR 2015). Its programme committee was chaired by Dr. Leif Azzopardi (University of Glasgow) and Dr. Max L. Wilson (University of Nottingham). Following an open call for papers, FDIA received 23 submissions, of which 19 (90%) were accepted.

Leif Azzopardi opened this year's FDIA programme with an engaging talk that urged summer school participants to question, read, do, argue, and think more! To this end, he encouraged students to search more for information and strive more to acquire knowledge. Based on the Foraging Theory that aims to understand the rules that shape the foraging behavior of animals, he discussed the ways in which people modify their strategies to maximize their rate of gaining valuable information, and the Information Patch Model that describes how foragers move between and within information patches. He then presented an online fishing game⁸, based on the principles underlying these theories and models, that aims to measure how good people are at searching. His analysis of the results for the ESSIR 2015 participants that played the game indicated that there is definitely room for improvement in their searching strategies and that they should try to dig more for information, ultimately aiming to be able to theorize and model more.

The rest of the programme comprised two parts: a series of 19 fast-paced presentations, followed by a poster presentation session. During the presentation phase, students gave a five-minute talk explaining their research in succinct and engaging manner, while during the poster phase presenters and participants could discuss the research in detail, form acquaintances, and receive advice and mentorship from senior IR attendees. The programme featured a variety of novel and emerging topics including: temporal and location-based information retrieval, visualizing user models, topic centric classification of tweets, opinionated learners, gamification of searching and learning, along with work on social media analysis, emotion aware recommender systems, energy efficiency systems, and reputation management systems. The 6th FDIA papers are listed below:

- Energy Efficiency in Web Search Engines – Matteo Catena (National Research Council of Italy)
- Query Recommendation as Query Generation – Matthew Mitsui (Rutgers University)
- Heterogeneous Information Access through Result Composition – Horatiu Bota (University of Glasgow)
- Temporal Information Retrieval Revisited: A Focused Study on the Web – Yue Zhao & Claudia Hauff (Delft University of Technology)
- Two-dimensional Point Set Pattern Matching with Horizontal Scaling – Antti Laaksonen (University of Helsinki)
- Syntactic and Semantic Structures for Relation Extraction – Duc-Thuan Vo & Ebrahim Bagheri (Ryerson University)

⁸ <http://maxxsol.pythonanywhere.com/gofish/#/>

-
- Word-Context Matrix based Query Expansion in Information Retrieval for Turkish Text – Emre Şatır, Adil Alpkoçak & Deniz Kılınç (Dokuz Eylul University)
 - A Framework for Enhanced Text Classification in Sensitivity and Reputation Management – Graham McDonald (University of Glasgow)
 - Whether a CQA User is a Medical Professional? Work in Progress – Alexander Beloborodov (Ural Federal University)
 - Different Tools for Handling Geographic Information Retrieval Problems – Yisleidy Linares Zaila (University of Bologna)
 - Improving Information Retrieval Evaluation via Markovian User Models and Visual Analytics – Maria Maistro (University of Padua)
 - Sentiment Analysis via Fractal Dimension – Symeon Symeonidis (Democritus University of Thrace)
 - Topic-centric Classification of Twitter User’s Political Orientation – Anjie Fang, Iadh Ounis, Philip Habel, Craig Macdonald & Nut Limsopatham (University of Glasgow)
 - Explanatory Opinions: To Whom or What is All the Fuzz About? – Filipa Peleja, Ioannis Arapakis & Joao Magalhaes (Yahoo! Labs & NovaLincs)
 - The Opinionated Learner – Stephen Bradshaw, National University of Ireland Galway
 - A Framework for Emotion-aware Recommender Systems supporting Decision Making – Marco Polignano (University of Bari Aldo Moro)
 - User’s Location Prediction in Location-based Social Networks – Jarana Manotumruksa, M-Dyaa Albakour & Craig Macdonald (University of Glasgow)
 - Investigating Search Behavior and Performance using Personal and Social Context Signals – Dongho Choi (Rutgers University)
 - A Gamification Framework for Enhancing Search Literacy – Ioannis Karatassis (University of Duisburg-Essen)

The proceedings of the 6th FDIA Symposium have been published by the BCS Electronic Workshops in Computing⁹ (eWiC) series.

4 Group activities

The school group activities focused on two key emerging areas of interest for IR research: a) veracity of online content, and b) online disclosure of personal information. Students were split into four work groups and were given assignments. The group activity descriptions, presentations and resulting datasets are available on a dedicated Web page on the summer school site¹⁰.

The first two groups worked on the issue of fake content distributed through social media. They were given a set of 20 tweets and they were asked to assess individually and as a group whether each of the tweets was truthful or misleading. Later, they were provided with the “ground truth” for these tweets and they discussed what kind of approach they followed during their verification and the types of information they used for making their decision.

The other two groups worked on the issue of online disclosure of personal information. They first filled in an online questionnaire, where they indicated the type of information (out of eight categories, such as demographics, political views, health condition, etc.) they consider sensitive, and whether

⁹ <http://ewic.bcs.org/category/18533>

¹⁰ http://mklab.itl.gr/essir2015/?page_id=885

they believe that such information could be inferred from their online behavior and shared content. Then, they continued with the analysis of the questionnaire results and exploration of questions related to information disclosure and privacy in an “always connected” world.

5 Social programme

The summer school included a rich programme of carefully selected social activities:

- welcome reception at the Aldebaran restaurant¹¹, located at the top floor of the new Music Hall building, with a view to the city seafront (Monday evening);
- interactive city tour by Dot2Dot and Gamecraft¹² that provided the opportunity to explore through a highly engaging role playing game the city of Thessaloniki and get a good glimpse into its history and culture (Tuesday afternoon);
- social dinner at Aigli Geni Hamam restaurant¹³ located at a well-preserved Ottoman Hamam (bath) constructed during the 16th century (Wednesday evening);
- social dinner at Rio¹⁴, a Greek tavern located by the beach (Thursday evening).

6 Feedback and recommendations

Most participants and lecturers gave a very positive feedback both through personal communication with the organizers and also in the evaluation questionnaire provided after the end of the summer school. The questionnaire was filled in by 43 ESSIR 2015 attendees (35 participants and 8 lecturers-FDIA chairs) and some of the results of this analysis are depicted in Figure 3.

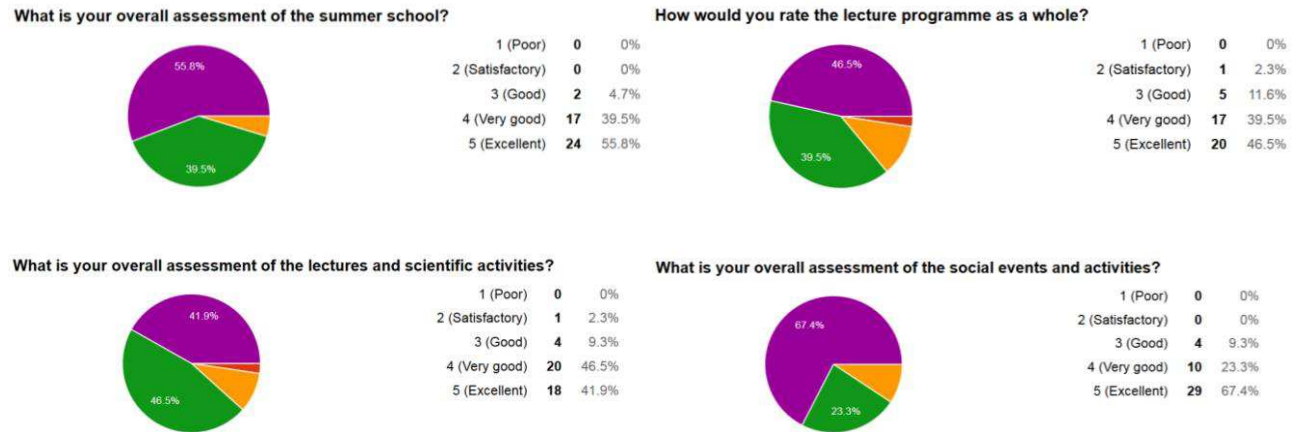


Figure 3: Results of the ESSIR 2015 evaluation questionnaire.

In their overall assessment, the overwhelming majority of the attendees (95%) rated the summer school as excellent or very good. The lecture programme as a whole was rated highly by the majority (86%) with the quality and depth of the lectures considered excellent or very good by 93% and 63% of the attendees, respectively, indicating that while they all appreciated the high quality of the lecture programme, some would find beneficial some more depth in the lecture material. Regarding the

¹¹ <http://www.aldebaranm2.gr/>

¹² <http://www.dot2dot.gr/>, <https://www.facebook.com/gamecraft.gr>

¹³ <https://www.facebook.com/aigligenihamam>

¹⁴ <https://foursquare.com/v/rio/4e429a12b61ca5ba3b27c739>

overall scientific programme (i.e., lecturers, FDIA sessions, and group activities), 88% of attendees rated it as excellent or very good. Finally, 91% highly enjoyed the social events and activities.

The questionnaire also allowed ESSIR 2015 attendees to provide their recommendations. One point that came up a couple of times was that students would also appreciate some hands-on sessions where they would be made familiar and gain practical experience on the material taught and IR tools introduced during the lecturers. They consider that this would be particularly useful given that most of the participants were first-year PhD students and this would allow them to benefit from the guidance and experience of senior IR researchers. This desire for more practical activities was also reflected in other recommendations that indicated that they would be interested in spending more time in group activities.

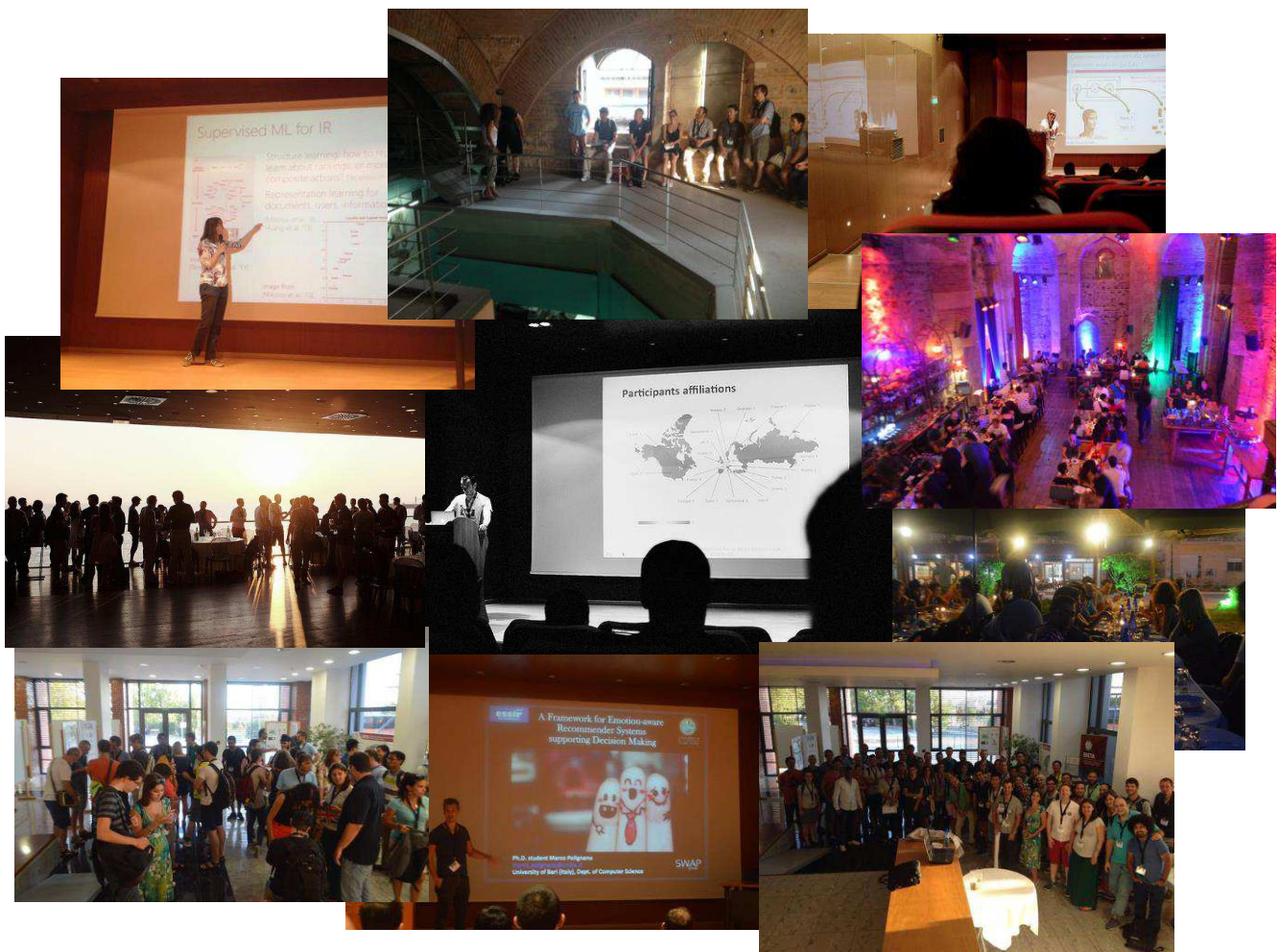


Figure 4: ESSIR 2015 moments.

7 Conclusions

The 10th European Summer School in Information Retrieval (ESSIR 2015) was a highly successful event that brought together participants with a broad range of research interests and offered them the opportunity to attend high-quality lectures covering a wide spectrum of foundational and emerging topics in IR and related areas, to interact both with their peers and senior IR researchers, and also get acquainted with the life and rich culture of the city of Thessaloniki. The online material (slides and videos from the lectures) can be found on the summer school website, while photos (Figure 4) and impressions of the event are available on its social streams¹⁵.

Overall, we received very positive feedback from attendees on all the different aspects of the school and we hope that the sentiments of the participants are echoed by the words provided by one of the attending students: *“Superb course. Got a real understanding of where I stood in relation to my peers. Great opportunity to network, quiz the lecturers. The friendliness and helpfulness of everyone staff/organisers/students was second to none. Came home feeling refreshed and ready to work for my PhD. Thanks a million to all involved and would go again next year in a heartbeat.”*

8 Acknowledgements

We would like to thank all lecturers, FDIA chairs, and participants to ESSIR 2015 who made the school a success through their extremely valuable contributions. Special thanks also go to the summer school volunteers who contributed to the smooth running of the school activities during the whole week.

We are also very grateful to our sponsors: the ELIAS (Evaluating Information Access Systems) ESF Research Networking Programme, the ACM Special Interest Group on Information Retrieval (ACM SIGIR), Yahoo! Labs¹⁶, Alpha Bank¹⁷, and the Information Retrieval Specialist Group of the British Computer Society¹⁸ (BCS-IRSG) that also supported the publishing of the FDIA proceedings through its BCS eWIC series.

¹⁵ <https://www.facebook.com/essir2015>, <https://twitter.com/essir2015>

¹⁶ <https://labs.yahoo.com/>

¹⁷ <http://www.alpha.gr>

¹⁸ <http://irsg.bcs.org/>
