# Report on ESSIR 2019: the 12th European Summer School in Information Retrieval

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

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

Gabriella Pasi
University of Milano-Bicocca, Italy
*gabriella.pasi@unimib.it*

Marco Viviani
University of Milano-Bicocca, Italy
*marco.viviani@unimib.it*

**Abstract**

This paper reports on the 12th edition of the *European Summer School in Information Retrieval* (ESSIR 2019), held in Milan, Italy, from 15 to 19 July 2019.

## 1 Overview

The European Summer School in Information Retrieval (ESSIR)[1] is a scientific event founded in 1990, which has given rise to a series of Summer Schools held on a regular basis to provide high quality teaching of Information Retrieval (IR) and advanced IR topics to an audience of researchers and research students. ESSIR is typically a week-long event consisting of guest lectures and seminars from invited lecturers who are recognised experts in the field.

The 12th European Summer School in Information Retrieval (ESSIR 2019)[2] marks the 30th anniversary of ESSIR. It was held at the University of Milan Bicocca, Milan, Italy, jointly organized by the Information Retrieval Laboratory (IR Lab) at the University of Milano-Bicocca and the Information Management Systems (IMS) research group at the University of Padua.

ESSIR 2019 has been a 5-day event (15-19 July, 2019) that offered a high quality teaching on IR and related research topics, in a friendly atmosphere. The 9th edition of the PhD Symposium on Future Directions in Information Access (FDIA)[3] was also held in conjunction with ESSIR 2019, providing a forum for early researchers to present their research in a friendly environment, whilst among senior researchers.

The mission of the school is to enable students to learn about modern research challenges and methods in information retrieval and related disciplines, to stimulate scientific research and

[1] `https://en.wikipedia.org/wiki/European_Summer_School_in_Information_Retrieval`
[2] `http://www.ir.disco.unimib.it/essir2019/`
[3] `https://easychair.org/cfp/FDIA2019`

collaboration in these fields, and to grow a community of researchers, students, and industry professionals working on information retrieval with collaborations all around the world.

38 participants from 9 countries attended the courses (2.5% BCs students, 8% MSc students, 79% PhD students, 2.5% academic, 8% industrial). 7 ESSIR2019 scholarships have been granted to students to attend the school; of the scholarships, 5 were supported by the SIGIR Friends program,[4] 1 was supported by AI*IA: Associazione Italiana per l'Intelligenza Artificiale,[5] and 1 was supported by AIUDC: Associazione per l'Informatica Umanistica e la Cultura Digitale.[6] Many of the other participants were able to participate thanks to a scholarship from their own institution or national institutions.

The multidisciplinarity of the participants and lectures helped to create many lively discussions and a friendly atmosphere with a large number of questions. Also most of the speakers stayed for the entire week and enriched the discussions as well. 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 different field with their own perspectives on a ground of shared topics and issues.

At the end of the school, we conducted a survey among students to gather their feedback and suggestions. An analysis of the evaluation forms compiled after the school highlighted that most students very much enjoyed it (97% of the participants) and the atmosphere among participants and lecturers.

In the following sections we briefly summarize the lectures and the panel delivered during the school. All teaching material that was used and made available during the week can be downloaded from the ESSIR 2019 Web site.[7]

The last day of the school there was a "student challenge" where attendees were asked to work in groups and pitch a research idea, related to their own interests and the topics covered during the school. This assignment served also as small exam to award 1.5 ECTS (European Credit Transfer and Accumulation System) to the attendees.

# 2 Lectures

## Information Retrieval as Interaction – Maarten de Rijke, University of Amsterdam, The Netherlands

Modern Information Retrieval (IR) systems, such as search engines, recommender systems, and conversational agents, are best thought of as interactive systems. And their development is best thought of as a two-stage development process: offline development followed by continued online adaptation and development based on interactions with users. The lecture sketched a rich landscape of offline and online topics that any student interested in IR system development should be familiar with. The lecture discussed IR scenarios, such as search, recommender systems, conversational interaction, and topics, like query and interaction mining and understanding, offline and online evaluation, and offline and online learning to rank.

[4] https://sigir.org/general-information/funding-for-sigir-related-events/
[5] https://aixia.it/en/
[6] http://www.aiucd.it/
[7] http://www.ir.disco.unimib.it/essir2019/program/

## Approaches to Research in IR – W. Bruce Croft, University of Massachusetts, Amherst, USA

The lecture conveyed lessons learned from 40 years of working with graduate students doing research in information retrieval. It started by giving a brief overview of IR research, focusing on the historical context, the characteristics that make IR unique, and what distinguishes good IR research. Then it covered a variety of areas that are important for IR research, including how to choose a research topic, what to publish, where to publish, and how to present the research. Finally, some of current projects at the Center for Intelligent Information Retrieval (CIIR) have been used as examples for this discussion. Although the lecture was aimed at graduate students, the material should be useful for researchers both in academia and industry.

## Foundations of Evaluation 1 – Nicola Ferro, University of Padua, Italy

The lecture introduced the basic notions on how to evaluate information retrieval systems. After discussing why we need evaluation and presenting the whole evaluation spectrum, the lecture focused on system-oriented evaluation and the Cranfield paradigm. In particular, the lecture covered experimental collections, namely corpora of documents, topics, and relevance judgments with some insights about the pooling process. The notions behind evaluation measures have been introduced, dealing with both set-based and rank-based measures. Finally, some ideas about statistical significance testing have been presented in order to allow for a sound comparison of system performance.

## Foundations of Evaluation 2 – Julio Gonzalo, UNED, Spain

The lecture aimed at underlining two things. First, that evaluation might be the most crucial aspect of high-quality research in Information Access and pervades all aspects of the research cycle, rather than being a final step to assess how well we solved our target problem. Second, that proper evaluation – the one that leads to valid, relevant, unbiased and generalizable results – can be very challenging. In fact, proper evaluation methodologies are still an open issue for many Information Access topics. While the first ESSIR talk on evaluation had an affirmative nature, covering the essential evaluation tools and good practices, this second talk had a more interrogative edge, and focused on all the things that can go wrong, how they may lead to biased or flawed results, and how they can, occasionally, even drive an entire research community into unproductive or misleading roads.

## Foundations of User-oriented IR 1 – Nick Belkin, Rutgers University, USA

The lecture provided an overview of the development of research in user-oriented information retrieval, with special focus on the relationships between research in information science on models of information seeking and search, and research in computational information retrieval on system design. The lecture investigated theories about why people engage in information seeking; user-centered models of information retrieval; and the implications of these for the nature of research in

information retrieval, and the design and evaluation of information retrieval systems. The lecture concluded with some thoughts about the implications of the ubiquitous information environment for information retrieval research and practice.

## Foundations of User-oriented IR 2 – Diane Kelly, University of Tennessee, USA

This lecture continued the previous lecture on user-oriented information retrieval by examining the history and evolution of user interfaces that support people's access to, and interactions with, information. Special care was taken to connect the design of some of these interfaces to the models and theories presented in the previous lecture. The lecture also reviewed some of the more common user-centered methods and measures researchers use to evaluate and study interfaces and information interactions, and highlights some of the challenges of conducting user-centered studies.

## Foundations of Machine Learning for IR – Claudia Hauff, TU Delft, The Netherlands

Passage retrieval, question answering, and language generation – those are some of the tasks developed for MS MARCO, one of the very few leaderboard-based benchmarks that have a heavy focus on IR tasks. The large-scale nature of the data naturally leads to machine learning (ML) based solutions. The lecture took a closer look at some of the best performing ML approaches which are based on learning to rank and neural IR, the latter often heavily inspired by the most recent advances in NLP. The lecture concluded with a critical discussion of the recent neural net wave and its applicability to a wide(r) range of IR tasks.

## Foundations of Contextual IR – Gabriella Pasi, University of Milano-Bicocca, Italy

To overcome the "one size fits all" behavior of most search engines, in recent years a great deal of research has addressed the problem of defining techniques aimed at tailoring the search outcome to the user context in order to improve the quality of search. The main idea is to produce context-dependent and user-tailored search results. Search tasks are subjective and often complex. The user-system interaction, based on keyword-based querying and on the presentation of search results as a list of Web pages ordered according to their estimated relevance, is often unsatisfactory. The lecture presented an overview of the main issues related to contextual search.

## Efficiency and scalability in IR – Nicola Tonellotto, ISTI, CNR, Italy

Typically, techniques that benefit effectiveness of information retrieval (IR) systems have a negative impact on efficiency. Yet, with the large scale of Web search engines, there is a need to deploy efficient query processing techniques to reduce the cost of the infrastructure required. The lecture provided a detailed overview of the infrastructure of an IR system devoted to the efficient

yet effective processing of user queries. The lecture guided the attendees through the main ideas, approaches, and algorithms developed in the last 30 years in IR query processing.

## Mining Social Media – Carlos Castillo, Universitat Pompeu Fabra, Spain

Social media holds the promise of a vast dataset that represents "what people think" about various issues. The lecture covered several social media applications in various domains, including economics, politics, public health, and emergency management. A particular focus were sudden-onset events such as natural disasters or mass convergence events. The lecture also discussed various biases that affect this research and ethical boundaries that delimit it.

## Medical (text and image) IR – Henning Müller, HES-SO Valais, Switzerland

Medical information is produced in massive quantities both targeting health professionals and also for patients, for example on the Internet. Information quality can vary widely and finding trustable information of high quality is not easy. Health topics are some of the most frequently searched topics on the web. Clinical texts are often semi-structured but contain much variability in terms of wording, spelling mistakes and also unusual abbreviations. In addition to free text, medical data include images, videos, signals, and structure data, and often several languages. This means that almost all challenges in information retrieval also occur for medical text.

The lecture covered medical data production focusing on images and text. It explained information needs and how these needs are situation dependent and subjective. Based on the medical ImageCLEF challenges several scenarios and techniques were explained to fulfill medical information needs and analyze medical data to create knowledge and help clinical decision making.

## Design and Evaluation of Recommender Systems – Paolo Cremonesi, Politecnico di Milano, Italy

Recommender Systems (RSs) help users search large amounts of contents by allowing them to identify the items that are likely to be more attractive or useful. RSs play an important role in many domains (e.g., e-commerce, e-tourism, entertainment), as they can potentially augment the users' trust towards an application and orient their decisions or actions towards specific directions. The goal of the lecture was to give participants a solid background on how to design and evaluate RSs, with a focus on user experience aspects, and to provide pragmatic guidelines to perform these activities more effectively. The lecture was structured into two parts. The first part provided a general overview on recommender systems and their design issues. The second part analyzed "off-line" (system-centric) evaluation techniques.

### Understanding & Inferring User Tasks and Needs – Emine Yilmaz, University College London, UK

Search behavior, and information behavior more generally, is often motivated by tasks that prompt search processes that are often lengthy, iterative, and intermittent, and are characterized by distinct stages, shifting goals and multitasking. Current search systems do not provide adequate support for users tackling complex tasks due to which the cognitive burden of keeping track of such tasks is placed on the searcher. Developing a comprehensive understanding of user's tasks would help in providing better support and recommendations to users based on their contextual information and as a result, help users accomplish the task. The lecture discussed the traditional methods used for inferring user "intent" and then focused on the recent advancements towards building task-based IR systems and present analytical results which highlight the importance of considering tasks as the focal unit of modeling search behavior. Additionally, the lecture considered the challenge of extracting tasks from a given collection of search log data and present some recently proposed task extraction techniques which rely on recent advancements in Bayesian non-parametrics, word embeddings, and deep learning. The lecture went beyond traditional web search scenarios, and characterized user tasks with conversational agents and digital assistants, including the recently introduced voice only assistants. The lecture additionally presented a detailed overview of task-based evaluation techniques. Finally, it presented applications of task inference techniques.

### Biases on Web Search and Recommender Systems – Ricardo Baeza-Yates, NTENT / Northeastern University, USA

The lecture covered all biases that affect search and recommender systems. They include biases on the data, the algorithms as well as the user interaction, in particular, the ones related to relevance feedback loops (e.g., ranking). In each case, the lecture covered the known techniques to ameliorate them as well as biases that might be the product of the evaluation methods used.

## 3 Panel on Teaching and Learning IR in Academia and Industry

The European Summer School in Information Retrieval (ESSIR) is a scientific event established in 1990 that promotes a series of Summer Schools which provide high-quality training on Information Retrieval, covering advanced topics. ESSIR was started at a time that proved especially favorable for Information Retrieval. In fact, the Schools of Cornell University in the USA (headed by Gerard Salton) and Cambridge University (with Karen Spark Jones, Steve Robertson, and Keith van Rijsbergen) in Europe paved the way for a new discipline which started flourishing in the 1980s to emerge globally as an independent area of computer science. While the discipline was rapidly growing from a methodology perspective – with search engines being designed and implemented once the Web was invented – its teaching within university curricula did not move on as swiftly. For years ESSIR has helped fill this gap, preparing students in researching Information Retrieval. As the situation rapidly evolved over the years, the discipline has grown significantly and has had

to face diversified issues. The panel of the Summer School, therefore, was centered on what, those who are starting scientific research, need in terms of methodological and cultural background.

Panel Moderator:

- Maristella Agosti, University of Padua, Italy.

Panelists:

- Ricardo Baeza-Yates, NTENT / Northeastern University, USA,
- W. Bruce Croft, University of Massachusetts, Amherst, USA,
- Diane Kelly, University of Tennessee, USA,
- Henning Müller, HES-SO Valais, Switzerland.

# 4 Sponsorship

ESSIR 2019 Summer School has been sponsored by:

- ACM Special Interest Group on Information Retrieval (SIGIR),[8]
- Associazione Italiana per l'Informatica e il Calcolo Automatico (AICA),[9]
- Associazione Italiana per l'Intelligenza Artificiale (AI*IA),[10]
- Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD),[11]
- Bloomberg,[12]
- British Computer Society Information Retrieval Specialist Group (BCS IRSG),[13]
- Google,[14]

and by the following institutional sponsors:

- Bicocca Sostenibile,[15]
- Dipartimento di Informatica, Sistemistica e Comunicazione (DISCo) / UNIMIB,[16]
- Dipartimento di Ingegneria dell'Informazione (DEI) / UNIPD,[17]

[8] https://sigir.org/
[9] https://www.aicanet.it/
[10] https://aixia.it
[11] http://www.aiucd.it
[12] https://www.bloomberg.net/
[13] https://irsg.bcs.org/
[14] https://www.google.com
[15] https://www.instagram.com/bicoccasostenibile/
[16] https://www.disco.unimib.it/en
[17] https://www.dei.unipd.it/en/

- Università degli Studi di Milano-Bicocca (UNIMIB),[18]
- Università degli Studi di Padova (UNIPD).[19]

# 5 Acknowledgements

We express our gratitude to the supporting organisations and institutions, because, without their support, ESSIR 2019 could not have taken place.

We also like to thank all the lecturers and the students who participated in the school, for useful discussions and opportunities to exchange ideas and to imagine the future of IR.

[18] https://www.unimib.it/unimib-international
[19] https://www.unipd.it/en/