

Report on the 12th Russian Summer School in Information Retrieval (RuSSIR 2018)

Yelena Mejova
ISI Foundation, Italy
yelena.mejova@gmail.com

Ekaterina Chernyak
Higher School of Economics, Russia
echernyak@hse.ru

Elena Tutubalina
Kazan Federal University, Russia
elvtutubalina@kpfu.ru

Pavel Braslavski
Ural Federal University, Russia
pbras@yandex.ru

1 Introduction

The 12th Russian Summer School in Information Retrieval (RuSSIR 2018) was held on August 27–31 in Kazan, Russia.¹ The school was co-organized by Kazan Federal University² and the Russian Information Retrieval Evaluation Seminar (ROMIP).

Starting from 2007, the RuSSIR school series has become an internationally recognized venue [1]–[5]. Previously, RuSSIRs have taken place in Yekaterinburg, Taganrog, Petrozavodsk, Voronezh, Saint Petersburg, Yaroslavl, Kazan, Nizhny Novgorod, and Saratov. Over these years, RuSSIR courses have been taught by world-renowned researchers and industry leaders in Information Retrieval and related areas.

This year, we had a record number of around 120 participants in total (counting in students, teachers, organizers, and sponsors' representatives); there were 87 students attending RuSSIR 2018. While the majority of participants were from various Russian universities, there were also representatives from the USA, Spain, Ireland, Switzerland, the United Kingdom, Australia, the Netherlands, Estonia, China, Germany, Italy, Denmark, Austria, and Finland. School participation was free of charge due to support from the school's sponsors. In addition, 10 international students and two teachers received travel support from the BCS IRSG and ACM SIGIR.

2 Courses

The special topic for RuSSIR 2018 was Information Retrieval for Good, with a special focus on applications in humanitarian, medical, and health domains. The following lectures and courses were given at the school.

¹<http://romip.ru/russir2018/>

²<https://kpfu.ru/eng/>

Crisis Informatics – Carlos Castillo (Universitat Pompeu Fabra)

Social media is an invaluable source of time-critical information during an emergency or a sudden-onset disaster. However, emergency response and humanitarian relief organizations that would like to use this information struggle with an avalanche of social media messages that exceeds human capacity to process. This talk was on computational methods to process these data and infer general parameters of a crisis, as well as determining priorities for intervention, draw from many disciplines, including natural language processing, semantic technologies, data mining, machine learning, network analysis, human-computer interaction, and information visualization.

The Biases of Social Data – Carlos Castillo (Universitat Pompeu Fabra)

Online social data such as user-generated content, expressed or implicit relationships between people, and behavioral traces are at the core of many popular web applications and platforms. Social data has been used to study a variety of domains including public policy, healthcare, economics and many social good applications. However, many academics and practitioners are also increasingly warning against the naive usage of social data. They highlight that there are biases and inaccuracies occurring at the source of the data, but also introduced during data processing pipeline; there are methodological limitations and pitfalls, as well as ethical boundaries and unexpected consequences that are often overlooked.

The Information Retrieval Challenge of Lifelogs and Personal Life Archives – Cathal Gurrin (School of Computing, Dublin City University)

This course covered the new forms of personal data being created today and explored how the current generation of IR tools can be deployed and enhanced to realize the potential of the new personal datasets. The course considered the opportunities and challenges of data access, storage, indexing, retrieval and presentation. The state of the art approaches developed for collaborative benchmarking competitions and emerging research challenges and opportunities in the domain were discussed.

Evaluation of IR systems and multi-modal retrieval in the medical domain – Henning Müller (University of Geneva)

This course covered two parts: (1) medical image retrieval and (2) benchmarking of image analysis and retrieval applications and corresponding infrastructures. In terms of medical image retrieval, application domains were explained and areas of text and visual retrieval with a focus on combining visual cues and medical text or structured data (multimodal retrieval). In terms of benchmarking, medical data have several challenges, such as data confidentiality, quickly changing nature of the data and increasingly large data sets. With the Evaluation-as-a-Service paradigm several of these challenges have been addressed, and examples of this approach were presented and explained

Conversational AI through Deep Learning – Valentin Malykh, Mikhail Burtsev (Moscow Institute of Physics and Technology)

The course was dedicated to the deep natural language processing which is a hot topic of the past few years. It consisted of the overview of current state in the field of artificial intelligence, accompanied with the real usage cases, followed by introduction to major neural network architectures – convolutional and recurrent neural networks in application to the problems of natural language processing. Deep reinforcement learning which is currently on the rise in the field of conversational AI was also covered as short introduction to the field. Course listeners received hands-on experience with intelligent chatbot creation.

Learning from User Interactions – Rishabh Mehrotra (Spotify Research)

The ability to learn from user interactions with online services promises pathways for solving a number of problems and improving user engagement and satisfaction. It involves a number of aspects from understanding user intent and tasks, to developing user models and personalization services. A user's understanding of their need and the overall task develop as they interact with the system. Supporting the various stages of the task involves many aspects of the system, e.g. interface features, presentation of information, retrieving and ranking. Beyond understanding user needs, learning from user interactions involves developing the right metrics and experimentation systems, understanding user interaction processes, their usage context and designing interfaces capable of helping users. This course covers (1) Understanding & Extracting User Tasks, (2) Learning User Representations, and (3) Behavioural Metrics & Experimentation.

Health Search – Guido Zuccon (Queensland University of Technology)

This course introduced researchers to the challenges and opportunities in health search, providing insights into current techniques and their results. It also offered a hands-on overview of tools specific to the health domain made available by the clinical informatics and natural language processing communities. In particular, it covered the different end user requirements, provided a hands-on introduction to domain-specific tools and presented resources and campaigns for evaluation in health search.

Learning to Rank and Evaluation in the Online Setting – Harrie Oosterhuis (University of Amsterdam)

This course presented the fields of Online Evaluation and Online Learning to Rank. Methods from these fields are based around user interactions, and have been proven to be reliable and efficient even when very few interactions are available. However, user interactions bring their own difficulties, as user behavior is very dependent on the systems actions, thus online methods must deal with interaction noise, and biases w.r.t. display positions, document selection, etc. Yet methods in the online setting overcome these issues and provide results that are more in line with the true user preferences than traditional methods. This course detailed the particularities of the online setting and showed how online evaluation and online learning to rank methods still provide reliable results in this setting.

Retrieving Information Interactively Using Natural Language – Prasenjit Mitra (Pennsylvania State University)

The module discussed the developments in natural language processing and how that is being leveraged to (a) pose queries in natural language, (b) send responses generated using natural language generation, and/or (c) engage in dialog with the end-user to refine questions or answers. Specifically, the course introduced the student to the models of conversation and discourse in natural language, identified desirable properties of such conversations, and introduced issues involved with extracting the semantics of queries and the information needs of the end-users who pose the queries.

Sponsor’s Lecture

Ivan Fursov from Tinkoff Bank, one of the RuSSIR 2018 sponsoring organizations, gave a talk on *Dialog Systems at Tinkoff.ru*.

3 Young Scientist Conference

Two poster sessions ran over two consecutive evenings (Monday & Tuesday), during which the participants had an opportunity to discuss and exchange their research results and ideas. In total 76 posters were displayed. As in previous years, the Young Scientist Conference was one of the main highlights of the school.

4 Social Program

On the first evening of the school, the RuSSIR Welcome Reception was held in the Main Dining Hall of the Institute of Physics, KFU. This reception helped school participants to get to know each other. Social program on the third day included a guided sightseeing tour of Kazan. Most of all attractions, the participants liked Syuyumbike Tower, Kul Sharif Mosque, and Cathedral of the Annunciation in Kazan Kremlin. Social program on the fourth day included a tour to the Nikolai Lobachevsky Museum that was opened in 2017. The museum is situated in the Rector’s House, the one where the great mathematician and Rector of Kazan University and his family resided during his tenure from 1827 to 1845. The closing party was held on Thursday in the Tugan Avylym Restaurant Village that is a small oasis of original Tatar culture in the business central part of Kazan. National rural style of the village is emphasized by buildings that are made of natural log. After the closing party, many school attendees used a chance to watch colorful fireworks near the Kazan River that were timed with the celebration of the official day of the city of Kazan and the Republic of Tatarstan.

5 Conclusions

The 12th Russian Summer School in Information Retrieval was a very successful event in many aspects. The school brought together participants with diverse backgrounds from Russia and abroad

and facilitated cross-disciplinary exchange of experience and ideas. The RuSSIR students had a unique opportunity to learn new material that is not usually present in university curricula and receive feedback from their peers and lecturers during the poster sessions and informal communications. The event contributed to supporting a lively IR community in Russia and establishing ties with international colleagues. The organizers received very positive evaluation from attendees on various aspects of the school.

6 Acknowledgments

We thank Kazan Federal University officials Airat Khasianov and Valery Solovyev for their cooperation and support. We also thank Local Organizing Committee members and school volunteers, especially Iurii Dedenev, Anita Kurova, Ilseyar Alimova, Zulfat Miftahutdinov, Gulnaz Amirhanova, Alteya Bady, Kseniya Edkova, Daniyal Salahov, Alexander Kirillovich, Timur Hamidullin, Polina Alikina, Anastasia Zhuravleva, Dinir Imameev, and Pavel Penkov.

We appreciate generous financial support from our sponsors: Tinkoff Bank³, JetBrains⁴, as well as Russian Foundation for Basic Research⁵. We also thank ACM SIGIR⁶ and BCS IRSG⁷ for providing travel grants for RuSSIR participants. Special thanks go to Jaap Kamps (University of Amsterdam) for his valuable help.

References

- [1] Pavel Braslavski, Nikita Zhiltsov, Stefan M. Rüger, Yana Volkovich: 7th Russian Summer School in Information Retrieval (RuSSIR 2013). *SIGIR Forum* 47(2): 96-100 (2013)
- [2] Pavel Braslavski, Nikolay Karpov, Marcel Worring, Yana Volkovich, Dmitry I. Ignatov: 8th Russian Summer School in Information Retrieval (RuSSIR 2014). *SIGIR Forum* 48(2): 105-110 (2014)
- [3] Pavel Braslavski, Ilya Markov, Panos M. Pardalos, Yana Volkovich, Sergei Koltsov, Olessia Koltsova, Dmitry I. Ignatov: 9th Russian Summer School in Information Retrieval (RuSSIR 2015). *SIGIR Forum* 49(2): 72-79 (2015)
- [4] Alexander Kotov, Elena Treshcheva, Leonid Bessonov, Dmitry I. Ignatov, Yana Volkovich, Maria Eskevich, Pavel Braslavski: 10th Russian Summer School in Information Retrieval (RuSSIR 2016). *SIGIR Forum* 50(2): 28–35 (2016)
- [5] Pavel Braslavski, Jaap Kamps, Julia Kiseleva, Alexander Halperin: Report on the 11th Russian Summer School in Information Retrieval (RuSSIR 2017). *SIGIR Forum* 51(3): 94–98 (2017)

³<https://www.tinkoff.ru/eng/>

⁴<https://www.jetbrains.com/>

⁵<http://www.rfbr.ru/rffi/eng>

⁶<http://sigir.org/>

⁷<https://irsg.bcs.org/>