# 7th Russian Summer School in Information Retrieval (RuSSIR 2013)

Pavel Braslavski
Kontur Labs &
Ural Federal University, Russia
*pbras@yandex.ru*

Nikita Zhiltsov
Kazan Federal University, Russia
*nikita.zhiltsov@gmail.com*

Stefan Rüger
The Open University, UK
*stefan.rueger@open.ac.uk*

Yana Volkovich
Barcelona Media, Spain
*yana.volkovich@barcelonamedia.org*

## 1 Introduction

The 7th Russian Summer School in Information Retrieval (RuSSIR 2013) was held on September 16-20, 2013 in Kazan, Russia.[1] The school was co-organized by the Kazan Federal University[2] and the Russian Information Retrieval Evaluation Seminar (ROMIP)[3].

The RuSSIR school series started in 2007 and has developed into a renowned academic event with solid international participation. Previously, RuSSIR took place in Yekaterinburg, Taganrog, Petrozavodsk, Voronezh, Saint Petersburg, and Yaroslavl. RuSSIR courses were taught by many prominent international researchers in IR and cognate areas.

Kazan, about 800 km east of Moscow, has the population of about 1.1 million people and is one of the oldest Russian cities. Geographical and cultural aspects of both Europe and Asia come together here. Founded in 1804, Kazan University is the third oldest university in Russia. It has over 40,000 full-time students in 180 major degree programmes. Kazan University is the alma mater and life-long affiliation of the Russian mathematician Nikolai Lobachevsky (1792–1856), the developer of non-Euclidean geometry, a fact that was symbolised by a hyperbolic triangle on RuSSIR t-shirts.

In 2013, the RuSSIR programme featured a track on audio and music IR alongside core information retrieval topics. This led to fruitful discussions among participants coming from different domains and allowed students to learn cross-disciplinary competencies. The school programme consisted of a plenary invited course, six courses running in two parallel sessions, two sponsor presentations, as well as the RuSSIR Young Scientist Conference. Music Hackathon, a co-located event with a focus on hands-on development, was an innovation of this year.

[1] http://romip.ru/russir2013/
[2] http://kpfu.ru/
[3] http://romip.ru/

The school welcomed 93 participants that were selected based on their applications. The majority of students came from Russia, but there were also 7 students from Europe, and 2 from the rest of the world. The RuSSIR audience comprised of undergraduate, graduate, and doctoral students, as well as industrial developers. The total number of participants including students, sponsor representatives, lecturers and organisers was 134.

Thanks to support from the sponsors, participation in the school was free of charge. In addition, 19 accommodation grants were awarded to participants and 6 European students were fully funded through an European Science Foundation grant.

# 2 Courses

The RuSSIR programme was compiled based on reviewing of submitted course proposals by the programme committee. Each course proposal was reviewed by at least six PC members. In total, eleven course prospects were submitted, six of which were selected for the school programme. Additionally, there was an invited course. Each of the seven courses consisted of five 90-minute lectures taught in five subsequent days. The invited course ran as a plenary session, the other six courses ran in two parallel sessions.

**Spoken Content Retrieval: Challenges, Techniques and Applications – Gareth Jones, Dublin City University, Ireland**

In his invited course, Prof. Jones introduced main challenges and technologies of spoken content retrieval. He reviewed the history of the discipline to date, its core technologies, its relationship to text information retrieval, critical system design issues, and application domains. The course also had an overview of initiatives for evaluation of speech retrieval, available research and development resources in the area of spoken content retrieval, and open challenges going forward.

**Introduction to Information Retrieval Models – Massimo Melucci, University of Padua, Italy**

This course surveyed the core information retrieval models – vector-space model, the classical probabilistic model, and the language models, as well as latent semantic analysis.

**Large Scale Information Retrieval – Paolo Boldi, University of Milan, Italy**

The course covered topics crucial for modern IR. Special attention was paid to the impact of the size of data on design of algorithms and data structures for IR. The course focused on link analysis for web search, graph compression techniques, perfect hashing, web crawling, clustering computation in large graphs, as well as analysis of social graphs.

**Novel Representations and Methods in Text Classification – Manuel Montes-y-Gomez and Hugo Jair Escalante, National Institute of Astrophysics, Optics and Electronics (INAOE), Mexico**

The course explained main traditional text document representation and classification techniques. Moreover, the lectures covered recent developments in text classification – for example, classification techniques accounting for document context and automated construction of classification models.

**Content- and Context-based Music Similarity and Retrieval – Markus Schedl and Peter Knees, University of Linz, Austria**

The course surveyed methods that extract features both from the music content and music context data via web and social media. Based on these features, students could learn methods to compute similarities between songs and music artists, a key ingredient of music retrieval systems.

**Adaptivity in Audio and Music Retrieval – Andreas Nürnberger and Sebastian Stober, Otto von Guericke University Magdeburg, Germany**

This course provided an overview and in-depth discussion of selected ideas and concepts for the adaptation of systems to search and explore audio collections. I particular, the course dealt with approaches that allow to structure and visualise sets of audio objects under specific user interests in different scenarios.

**Query by Singing/Humming and Audio Fingerprinting as Two Successful Paradigms of Music Information Retrieval – Jyh-Shing Roger Jang, Taiwan University, China**

The course introduced two of the most successful paradigms of content-based music information retrieval: QBSH (Query by Singing/Humming) and AFP (Audio Fingerprinting). The components of the approach – feature extraction, comparison methods, speedup techniques, optimisation strategies, etc. – were considered in detail.

Additionally, school sponsors gave two presentations. Natalia Ostapuk, Tatiana Lando, and Sergey Zubkov presented a tutorial on Tomita, a free fact-extracting tool from Yandex. Alexey Voropaev gave a presentation on the crawler architecture implemented by Mail.Ru and related issues.

# 3 Young Scientist Conference

For the seventh time the RuSSIR Young Scientist Conference (RuSSIR YSC 2013) was held within the school. The goal of the conference was to provide a platform for ideas and knowledge exchange between all RuSSIR participants. The conference ran over two consecutive evenings and consisted of two parts, oral presentations and poster sessions. There were two types of submissions: full papers that underwent a thorough reviewing process and short poster notes. Out of 18 submitted full papers seven were accepted for oral presentation at the conference and subsequent published in Kazan University Proceedings:

- Artem Churkin and George Mazurkevich "Substitutions Acquisition Method for an Informational Retrieval System"
- Sergey Ermakov and Liana Ermakova "Linguistic approach to suicide detection"
- Nikolay Glazyrin "Mid-level Features for Audio Chord Estimation using Stacked Denoising Autoencoders"
- Dmitry Kostyrev, Sergey Anishchenko, and Mikhail Petrushan "Time Invariant Hand Gesture Recognition for Human-Computer Interaction"

- Svetlana Popova and Ivan Khodyrev "Ranking in keyphrase extraction problem: is it useful to use statistics of words occurrences?"
- Alexey Raskin "Comparison of Chains Clustering Techniques"
- Dmitry Ulyanov "NMF-based method for drum separation"

During the poster session all participants of the school were provided with an opportunity to discuss their current research results and ideas. In total about 70 posters were displayed. Similar to the previous years, the YSC poster session was among main RuSSIR highlights.

# 4 Social Programme

The Welcome Reception took place in the evening of the first day of the school, on September 16. Wine and hors d'oeuvre were served in the restaurant of the Hayal hotel located near the main school site. After a short opening ceremony participants had a perfect chance to get to know each other in a relaxed atmosphere.

A sport event at university sport center on Wednesday, September 18 offered participants a welcome change from lectures and conference sessions. Participants had to choose between futsal, volleyball, or just cheering on the players.

The School party took place on Thursday night, September 19, in Zheltaya Kofta (Yellow Shirt), a club whose name and atmosphere refer to the poet Vladimir Mayakovsky and Russian avant-garde at the beginning of the $20^{\text{th}}$ century in general. School participants were offered snacks, wine and beer. The party finished with a disco. As a precaution, the lectures on Friday morning started two hours later.

# 5 Hackathon

This year we decided to complement school courses with a co-located event of a different type – a Music Hackathon[4]. The Hackathon started on Friday evening and ran for 24 hours. The event was open for RuSSIR students as well as for participants registered with Hackathon only. The task of participating teams was to develop a music mock-up application of any kind.

Seven projects took part in Hackathon (including a project by coaches that was out of competition), while 34 people were involved in the activity in total.

The TwiMoMusic project won this contest. The application analyses geotagged tweets for two facets: sentiment and mentions of music artists, albums, and songs, thus creating a geographical map of both mood and music listened by the people locally. The Pintrack project received a special mention of the jury. Pintrack is a browser extension that allows to assign a music track to any web page, so the page visitor will listen to it while surfing.

[4] `https://www.hackerleague.org/hackathons/russir-music-hackathon-2013/`

# 6 Conclusions

The 7th Russian Summer School in Information Retrieval was a very successful event: It brought together participants with diverse backgrounds from Russia and abroad and facilitated cross-disciplinary exchange of experience and ideas. Students had an unique opportunity to learn new material that is not normally otherwise presented in university curricula and got feedback from peers and teachers during the poster sessions and informal communications. The event contributed to supporting a lively IR community in Russia and establishing ties with international colleagues. We received very positive feedback from attendees on all the different aspects of the school.

The 8th RuSSIR will be held in Nizhny Novgorod in August 2014, organised by the Nizhny Novgorod chapter of the Higher School of Economics.

# 7 Acknowledgments

We thank all Local Organizing Committee members (namely, Airat Khasianov, Valery Solovyev, Vladimir Ivanov, Rinat Gareev, Danila Zaikin, Alexander Kirillovich, Marsel Sidikov, Fedor Nikolaev) for their committment, which made the school possible, all Programme Committee members for their time and efforts ensuring a high level of quality for the RuSSIR 2013 programme and, in particular, all the lecturers and students who came to Kazan and made the school such a success. We also thank student volunteers who contributed to school organisation on-site. Our special gratitude goes to Maxim Gubin who was responsible for legal and financial matters.

We appreciate generous financial support from our sponsors – Yandex[5], Mail.Ru[6], and Dream Industries[7] (gold level); Russian Foundation for Basic Research[8] (silver level); Google[9], ABBYY[10], and STEL CS[11] (bronze level). We also thank the ELIAS network[12] of the European Science Foundation and The Open University[13] for their support.

[5] http://yandex.com
[6] http://go.mail.ru/
[7] http://dreamindustries.co/
[8] http://www.rfbr.ru/rffi/eng
[9] http://google.com/
[10] http://abbyy.com/
[11] http://stel.ru/en/
[12] http://www.elias-network.eu/
[13] http://www.open.ac.uk/