



Science Meeting – Scientific Report

Scientific report (one single document in WORD or PDF file) should be submitted online within two months of the event. It should not exceed seven A4 pages.

Proposal Title: Russian Summer School in Information Retrieval

Application Reference N°: 5041

1) **Summary (up to one page)**

The 7th Russian Summer School in Information Retrieval (RuSSIR 2013) was held on September 16-20, 2013 in Kazan, Russia (<http://romip.ru/russir2013>). The school was co-organized by the Kazan Federal University and the Russian Information Retrieval Evaluation Seminar (ROMIP). 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 so far were taught by many prominent international researchers in IR and cognate areas.

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.

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 attendees including participants, 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.

For more details, please refer to the following a SIGIR Forum report: *P Braslavski, N Zhiltsov, S Rüger and Y Volkovich: 7th Russian Summer School in Information Retrieval (RuSSIR 2013), SIGIR Forum, Dec 2013 (in press).*

2) **Description of the scientific content of and discussions at the event**

The RuSSIR programme was compiled based on reviews 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. Below is a short description of each course.

Spoken Content Retrieval: Challenges, Techniques and Applications (Gareth Jones, Dublin City University, Ireland)

This course introduced the challenges and technologies of spoken content retrieval. Prof Jones reviewed the history of spoken data search to date, its component technologies (including an introduction to speech recognition), its relationship to text information retrieval, critical system design issues, domains of application and issues of interaction with spoken content to support efficient access to relevant material. Special attention was paid to initiatives for evaluation of speech retrieval including: TREC SDR, CLEF CL-SR and MediaEval Search and Hyperlinking.

Introduction to Information Retrieval Models (Massimo Melucci, University of Padua, Italy)

This course introduced the mathematical basis of the main IR models, ie, the theory of vector spaces and probability theory. Probability space and random variable were explained in a way to make the connection to the respective notions of the abstract vector spaces explicit. Within this framework, the main retrieval models and methods of latent semantic analysis were illustrated.

Large Scale Information Retrieval (Paolo Boldi, University of Milan, Italy)

The course was a collection of topics crucial for the analysis of big data and building a modern IR search engine. In particular, Prof Boldi surveyed the following topics: design and engineering of a modern web crawler; link analysis with emphasis on centrality measures; large graph representation and compression techniques; the realization of fast and efficient dictionaries; approximating the distance distribution and the clustering coefficient in large graphs.

Novel Representations and Methods in Text Classification (Manuel Montesy-Gomez and Hugo Jair Escalante, National Institute of Astrophysics, Optics and Electronics, Mexico)

The course surveyed different approaches to two core components of any text classification system: the adopted representation for documents and the classification model. The lecturers examined traditional bag-of-words (BOW) document representation and its extensions locally weighted BOW, distributional term representations, concise and graph-based representations. In the part devoted to classification models, classification techniques accounting for document context and automated construction of classification models were covered. Evaluation approaches and measures, existing test collections and evaluation campaigns played a vital contribution of this course.

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

The course gave participants a solid understanding of basic methodology used in and typical applications of music information retrieval (MIR). The focus of the course was on similarity computation and retrieval. In particular, methods that extract features (i) from the music content by employing audio analysis techniques and (ii) from music context data via web and social media mining techniques were presented. Based on these features, participants learned methods to compute similarities between songs and between music artists, a key ingredient of music retrieval systems. Another aspect of the class was on evaluating MIR systems beyond the traditional IR-related measures and the difficulties entailed by the need for objective quantification. The lectures were accompanied by a practical exercise, in which participants learned how to build a music retrieval system that combines music content (audio) and music context (web) information.

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. In particular, selected approaches that allow to structure and visualize sets of audio objects using user specific interests were discussed in more detail. The lecturers focused on approaches that adapt the underlying similarity measures, which are a central issue in Music Information Retrieval (MIR). The course demonstrated how personalization in MIR tasks is important. Throughout the course, prototypical implementations of systems for personalized access to audio and music collections were presented.

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). These two paradigms have been used in various companies (Shazam, Soundhound, Intonow, to name a few) to create innovation and fun applications over mobile devices. The course's goal was to cover each single

step of the retrieval process: feature extraction, comparison methods, speedup techniques, optimization strategies, etc. The issue of implementation over high-computing GPU was also addressed.

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.

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 Lyanov: 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 the main RuSSIR highlights.

Hackathon

RuSSIR 2013 featured a co-located event after the Summer School: a Music Hackathon. It 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, while 34 people were involved in the activity in total. The winning application, TwiMoMusic, 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.

3) Assessment of the results and impact of the event on the future directions of the field

The School has attempted to span a bridge between a rich tradition of excellence in science, including information processing in Russia and information retrieval research in the Western World. This was achieved by international speakers and, owing to ELIAS funding, participation of students from Europe. RuSSIR 2013 was co-organised by ROMIP, the Russian Information Retrieval Evaluation Seminar (<http://romip.ru/>), which has its own tradition of evaluation events. ECIR 2013 held in Moscow and the RuSSIR 2013 meeting in Kazan have managed to blur the boundaries of research traditions in information retrieval between these two geographic areas and in this sense benefitted the exchange between these two communities.

The audience of the school consisted of advanced graduate and PhD students, post-doctoral researchers, academic and industrial researchers. The mission of the RuSSIR school series was to enable students to learn about modern problems and methods in information retrieval and related disciplines, to stimulate scientific research and collaboration in the field; and to create an environment for informal contacts between scientists, students and industry professionals. The school has hosted 134 attendees.

It has become evident that RuSSIR has started to, and continues, to attract international, particularly European, attendance benefitting both Russian and European researchers. The next and 8th RuSSIR summer school in Information Retrieval will be held in Nizhny Novgorod (previously named Gorky after the writer Maxim Gorky) in Summer 2014.

4) Annexes 4a) and 4b): Programme of the meeting and full list of speakers and participants

Annex 4a: Programme of the meeting

<http://romip.ru/russir2013/> details Russir 2013's Scientific Programme:

	16.09.2013, Mon	17.09.2013, Tue	18.09.2013, Wed	19.09.2013, Thu	20.09.2013, Fri	
9:30-11:00	<u>SCR</u> Small concert hall (UNICS)		<u>SCR</u> Small concert hall (UNICS)		<u>SCR</u> Small concert hall (UNICS)	
11:30-13:00	<u>IIRM</u> №104 (Lib)	<u>CCMSR</u> №218 (Lib)	<u>IIRM</u> №104 (Lib)	<u>CCMSR</u> №218 (Lib)	<u>IIRM</u> №104 (Lib)	<u>CCMSR</u> №101 (Lib)
14:00-15:30	<u>LSIR</u> №104 (Lib)	<u>QBSH&AFP</u> №218 (Lib)	<u>LSIR</u> №104 (Lib)	<u>QBSH&AFP</u> №218 (Lib)	<u>LSIR</u> №104 (Lib)	<u>QBSH&AFP</u> №101 (Lib)
16:00-17:30	<u>NRMTC</u> №104 (Lib)	<u>AAMR</u> №218 (Lib)	<u>NRMTC</u> №104 (Lib)	<u>AAMR</u> №218 (Lib)	<u>NRMTC</u> №104 (Lib)	<u>AAMR</u> №101 (Lib)
18:00-19:30	<u>Yandex</u> №109 (Building II)	<u>YSC</u> №109 (Building II)	<u>YSC</u> №109 (Building II)	<u>Mail.ru</u> №109 (Building II)	<u>SCR</u> №109 (Building II)	
	welcome party 20:00		sport evening 20:00	RuSSIR party 20:30		

SCR: Spoken Content Retrieval: Challenges, Techniques and Applications

IRM: Introduction to Information Retrieval Models

LSIR: Large Scale Information Retrieval

NRMTC: Novel Representations and Methods in Text Classification

CCMSR: Content- and Context-based Music Similarity and Retrieval

AAMR: Adaptivity in Audio and Music Retrieval

QBSH&AF: Query by Singing/Humming and Audio Fingerprinting

YSC: Young Scientist Conference

Hackathon: After closure of the Summer School for a duration of 24 hours

Annex 4b: Full list of speakers and participants

Speakers

Mr Paolo Boldi, Milano, IT
Mr Hugo Jair Escalante, Puebla, MX
Mr Jyh-Shing Roger Jang, Taipei, TW
Mr Gareth Jones, Dublin, IE
Mr Peter Knees, Linz, AT
Mr Massimo Melucci, Padova, IT
Mr Manuel Montes-y-Gómez, Puebla, MX
Mr Andreas Nürnberger, Magdeburg, DE
Mr Markus Schedl, Linz, AT
Mr Sebastian Stober, Magdeburg, DE

Participants

Mr Anton Bakhtin, Moscow, RU
Mr Igor Baltiyskiy, St Petersburg, RU
Ms. Yulia Baranova, Moscow, RU
Mr Alexander Batalschikov, Moscow, RU
Mr Alessio Bazzica, Delft, NL
Mr Alexander Beloborodov, Yekaterinburg, RU
Mr Yerlan Bissenov, Moscow, RU
Ms Anastasia Bodrova, St.Petersburg, RU
Mr Alexander Bragin, Moscow, RU
Mr Pavel Braslavski, Yekaterinburg, RU
Mr Sergiu Chelaru, Hannover, DE
Mr Ilia Chetviorkin, Moscow, RU
Mr Alexander Chuchunkov, St.Petersburg, RU
Mr Anton Churin, Moscow, RU
Mr Artem Churkin, Moscow, RU
Ms Vera Danilova, Bellaterra, Barcelona, ES
Ms Marina Danshina, Moscow, RU
Mr Anton Dergunov, Nizhni Novgorod, RU
Mr Danila Doroshin, Moscow, RU
Mr Mikhail Dykov, Volgograd, RU
Mr Denis Fedorenko, Moscow,, RU
Mr Alexander Fonarev, Moscow, RU
Mr Ilnur Gadelshin, Moscow, RU
Ms Karina Gainutdinova, Moscow, RU
Mr Rinat Gareev, Kazan, RU
Mr Ramil Gataullin, Kazan, RU
Mr Igor Gerbylev, Saint Petersburg, RU
Ms Ekaterina Gladkikh, Moscow, RU
Mr Nikolay Glazyrin, Yekaterinburg, RU
Ms Norma Gomes, Brasília, BR
Mr Dmitry Granovsky, Moscow, RU
Mr Vladimir Gulin, Moscow, RU
Mr Max Ionov, Moscow, RU
Mr Georgy Ivanov, Moscow, RU

Mr Vladimir Ivanov, Kazan, RU
Mr Alexander Kalinin, Krasnoyarsk, RU
Ms Elena Kalinina, Moscow, RU
Mr Nikolay Karpov, Moscow, RU
Ms Maria Karyaeva, Yaroslavl, RU
Mr Andrey Khalliulin, Moscow, RU
Ms Albina Khanbekova, Kazan, RU
Mr Alexander Kirillovich, Kazan, RU
Mr Sergei Koltcov, Moscow, RU
Ms Olessia (Elena) Koltsova, Moscow, RU
Ms Maria Komar, Yaroslavl, RU
Mr Dmitry Kostyrev, Rostov-on-Don, RU
Mr Alexander Kostyuchenko, Moscow, RU
Ms Anna Kozlova, Moscow, RU
Mr Evgeniy Krofto, Moscow, RU
Ms Irina Krylova, St.Petersburg, RU
Mr Andrey Kutuzov, Moscow, RU
Mr Egor Lakomkin, Moscow, RU
Ms Tatiana Lando, Moscow, RU
Mr Kirill Lazarev, St.Petersburg, RU
Ms Galina Lezina, Yekaterinburg, RU
Mr Babak Loni, Delft, NL
Mr Valentin Lubimov, Moscow, RU
Mr Artem Lukanin, Chelyabinsk, RU
Mr Alexander Lukyanov, Moscow, RU
Mr Nikolai Lyfenko, St.Petersburg, RU
Mr Nikolay Lyubimov, Moscow, RU
Mrs Katerina Malahova, St.Petersburg, RU
Mr Valentin Malykh, Moscow, RU
Mr George Mazurkevich, Moscow, RU
Mr Anton Melezhik, Moscow region, RU
Ms Elena Mikhalkova, Tyumen, RU
Mr Timur Mingulov, Moscow, RU
Mr Bulat Muhutdinov, Kazan, RU
Mr Damir Mukhamedshin, Kazan, RU
Mr Andrey Murashev, Moscow, RU
Ms Marina Mytrova, Moscow, RU
Mr Ilia Nekhay, Moscow, RU
Mr Ruslan Nigmatullin, Kazan, RU
Mr Fedor Nikolaev, Kazan, RU
Ms Maria Nikolaeva, Moscow, RU
Mr Michael Nokel, Moscow, RU
Mr Alexander Novikov, Moscow, RU
Ms Natalia Ostapuk, Moscow, RU
Mr Alexander Panchenko, Moscow, RU
Mr Mikhail Payson, Moscow, RU
Mr Sergey Polyashov, Moscow, RU
Ms Svetlana Popova, St Petersburg, RU
Mr Danis Rakhimzyanov, St Petersburg, RU
Ms Elmira Rakhmatulina, Moscow, RU
Mr Ajay Ramaseshan, AALTO, FI
Mr Alexey Raskin, Moscow, RU
Mr Sergey Repalov, Moscow, RU

Ms Alexandra Roshchina, México, Distrito Federal, MX
Ms Kristina Sabirova, Chelyabinsk, RU
Mr Diego Saez-Trumper, Barcelona, ES
Mr Yuriy Savelyev, Moscow, RU
Ms Luiza Sayfullina, AALTO, FI
Mr Alexander Sayko, Moscow, RU
Mr Marsel Sidikov, Kazan, RU
Mr Yury Sokursky, Moscow, RU
Mr Dmitry Soloviev, Moscow, RU
Ms Maria Stepanova, Moscow, RU
Mr Maxim Tkachenko, Moscow, RU
Mr Dmitry Ulyanov, Moscow, RU
Mr Dmirty Ushanov, Moscow, RU
Mr Dmitriy Vasiliev, Moscow, RU
Mr Dmitry Vasilyev, Moscow, RU
Mr Dmitry Vikharev, Moscow, RU
Mr Fedor Vityugin, Moscow, RU
Mr Ramis Yamilov, Kazan, RU
Ms Alexandra Yashina, Kirov, RU
Mr Danila Zaikin, Kazan, RU
Mr Alexey Zakharov, Moscow, RU
Ms Ksenia Zhagorina, Yekaterinburg, RU
Mr Nikita Zhiltsov, Kazan, RU
Mr Maxim Zuev, Moscow region, RU