



Program of

International FRUCT Conference and School on Intelligence, Social Media and Web ISMW-FRUCT 2016

28 August - 4 September 2016



Gaudeamus igitur,
Juvenes dum sumus!
Post jucundam juventutem,
Post molestant senectutem
Nos habebit humus.

Ubi sunt, qui ante nos
In mundo fuere?
Vadite ad superos,
Transite ad inferos,
Ubi jam fuere.

Vita nostra brevis est,
Brevi finietur,
Venit mors velociter,
Rapit nos atrociter,
Nemini parcetur.

Vivat Academia,
Vivant professores!
Vivat membrum quodlibet,
Vivant membra quaelibet!
Semper sint in flore!

Vivant omnes virgines
Faciles, formosae!
Vivant et mulieres,
Tenerae, amabiles,
Bona, laboriosae!

Vivat et respublica,
Et qui illam regit!
Vivat nostra civitas,
Maecenatum caritas,
Quae nos hic protegit

Pereat tristitia,
Pereant dolores,
Pereat diabolus,
Quis antiburschius,
Atque irrisores

Committees of ISMW FRUCT Conference and School

General Chair:

Boris Novikov (Saint Petersburg University, Russia)

Program Committee Co-Chairs:

Paolo Rosso (Technical University of Valencia, Spain)

Mikhail Ageev (Moscow State University, Russia)

School Committee Chair:

Svetlana Popova (Saint Petersburg University and ITMO University, Russia)

Publication Chair:

Sergey Balandin (FRUCT, Finland)

Organizing Committee Co-Chairs:

Svetlana Popova (Saint Petersburg University and ITMO University, Russia)

Elena Mikhailova (Saint Petersburg University, Russia)

Organizing Committee Members:

Dmitry Ustalov (Russian Academy of Science / Ural Federal University)

Alina Bakhitova (Higher School of Economics)

Maxim Buzdalov (ITMO University)

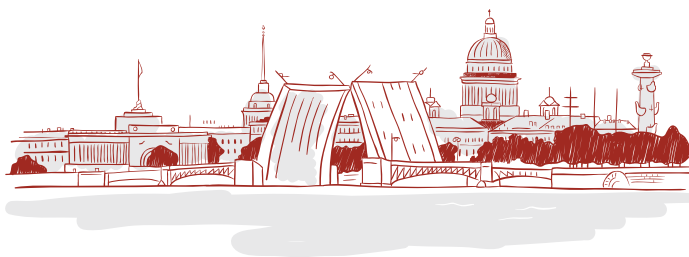
Andrey Filchenkov (ITMO University)

Lidia Perovskaya (ITMO University)

Aleksandr Farseev (National University of Singapore)

Program Committee:

Full list is available: ismw.ru/#callForPapers



Practical Information

Contacts:

Svetlana Popova: +7-921-4388077

Elena Mikhilova: (812) 9979791

Alina Bahitova: +7-918-8496113

office@ismw.ru

The conference and summer school ISMW FRUCT 2016 will be held at **Volkhovskiy Pereulok, 3**, Saint Petersburg.

From **Sportivnaya metro station** we recommend to go by foot, it should take less than 10 minutes. From **Vasileostrovskaya metro station** you can take a trolleybus 6 or buses 6 and 24. You can also reach us on foot. From **Admiralteyskaya metro station** it is faster to take a bus 7, 11 or 24 and get off at the 1-ya liniya/ Bolshoy prospect bus stop.

*A bus ticket costs 30 rubles

*A metro ticket costs 35 rubles

There are plenty of restaurants and cafes near the conference venue where you can have good meals. We listed some of them below.

Marketplace: 7-ya Liniya, 34

Chili Trattoria: Kadetskaya, 9

Dve palochki: Sredniy prospekt V.O, 16

Stolle: 1-ya Liniya V.O. 50

Pub Brugge: nab. Makarova, 22

Villagio: Tuchkov pereulok, 11

Mama Roma: Sredniy prospekt V.O, 6



General Information

Topics of the 2016 ISMW FRUCT conference and school include social media and social networks, artificial intelligence, Web technologies, data mining, knowledge discovery and big data processing.

The event consists of two main parts:

*academic conference

*summer school

The conference and school are intended for interdisciplinary audience with an additional focus on young scientists.

School: August 28 – September 1

Conference: September 2-3

The program of ISMW FRUCT school August 28 – September 1

School objectives: establishment of a scientific platform that will allow to integrate young researchers into the collaboration in the Social Media Analysis field. Participants are invited to attend a course of lectures, visit master classes and to do some research with a group of participants. The research task will bring together students and young researchers from various disciplines to conduct a joint study on the prepared data.

Main tutorials and workshops

Text mining for cybersecurity. Walter Daelemans (University of Antwerp, Tilburg University, University of Nijmegen, co-directing the CLiPS research center)

Authorship analysis. Efstathios Stamatatos (Director of the Artificial Intelligence Lab., University of the Aegean)

Topic Modeling. Olessia Koltsova and Sergei Koltsov, Anna Potapenko, Nadezda Chirkova (Higher School of Economics)

Big Data processing with Spark. Vitaly Khudobakhshov (senior analyst, OK.ru)

Computational experiment in machine learning. Vadim Strijov (Computing Center of the Russian Academy of Sciences)

Introduction to neural networks and deep learning. Eugeny Putin (ITMO University)

Distributional semantics. Andrei Kutuzov (University of Oslo)

Natural language processing. Dmitry Ilvovsky (Higher School of Economics)

Individual and group user profile learning from multiple social networks. Aleksandr Farseev (National University of Singapore)

The role of sociology at Data Science projects. Aleksandr Semenov (Higher School of Economics)

Collection and analysis of network data from Twitter in R. Aleksandr Semenov (Higher School of Economics)

Knowledge extraction from social media services - formats, APIs, tools, and unexpected problems. Alexey Neznanov (Higher School of Economics)

Principles of big data processing: SNA Hackathon 2016 case study. Dmitry Bugaichenko (OK.ru).

The program of ISMW FRUCT conference

September 2-3

Room 403

September 2	
9.30 11.50	Conference Registration
10.00 11.30	School: presentations
11.30 11.45	Tea, coffee
11.45 12.00	Opening session
12.00 13.00	Efstathios Stamatatos «Revealing Authors Behind Texts»
13.00 13.30	Margarita V. Kuznetsova Antiplagiat Research
13.30 14.30	Lunch on your own
14.30 15.10	Olëssia Koltsova «Topic modeling of user-generated content: opportunities and limitations»
15.10 15.40	Alexey Neznanov «Knowledge Extraction from Social Media and other Internet Data Sources: Modern Features and Challenges» (Cases form Criminalistics and Medicine)
15.40 16.40	Session I
16.40 17.00	Coffee break
17.00 19.00	Session II

September 3	
9.30	Conference Registration
10.00 11.40	Session III
11.40 12.00	Coffee break
12.00 13.00	Walter Daelemans «Detecting age, gender, and personality of social media authors»
13.00 13.40	Natalia Loukachevitch «Sentiment analysis: Tasks and Approaches»
13.40 14.00	Coffee break
14.00 15.30	Session: Poster and Demo Conference closing

Keynotes

Walter Daelemans

University of Antwerp, Tilburg University, University of Nijmegen, co-directing the CLiPS research center

Detecting age, gender, and personality of social media authors

What can we learn about the author of a text by analysing his or her language use? We can infer to some extent demographic and psychological properties of the author, and in some cases even identity, not by trying to understand what he or she says, but by analysing how it is said. We will call this computational stylometry. Using patterns of function words, distributions of syntactic structures, punctuation patterns, and more generally anything that can help us analysing the style in which a text was written, a profile of the author (in terms of age, gender and personality) can be made. The approach we take is a kind of reverse sociolinguistics. Where sociolinguists describe how differences in age, gender, education level, social class and related properties are correlated with specific language variation, in computational stylometry we try to predict these author properties using systematic language variation between groups as evidence. The most successful methodology for this is to apply a machine learning methodology. In this talk, I will give an overview of this approach and discuss some achievements and problems with it.

Natalia Loukachevitch

Research Computing Center of Lomonosov Moscow State University

Sentiment analysis: Tasks and Approaches

Web pages and social network sites contain a lot of users' opinions about various subjects. Governments, companies, and ordinary people are interested in extraction and analysis of attitudes contained in these opinions. Thus, automatic sentiment analysis attracts a lot of attention of researchers and practitioners. In my talk I will consider novel approaches to various tasks of automatic sentiment analysis. In addition, I will describe the results of sentiment analysis evaluations for the Russian language and compare them with the state-of-the art approaches.

Olessia Koltsova

Laboratory for Internet Studies, Higher School of Economics

Topic modeling of user-generated content: opportunities and limitations

Efstathios Stamatatos

Director of the Artificial Intelligence Lab., University of the Aegean

Revealing Authors Behind Texts

Authorship attribution is the attempt to identify the authors behind a text based on the assumption that the personal style of authors can be quantified and represented within a computational model. This line of research has a long history and it used to be associated with analysis of literary texts of disputed authorship. Given the vast amount of texts in electronic form in countless media, authorship analysis is now crucial in the framework of forensic and intelligence applications as well as in marketing analysis and social media analytics. Despite the fact that significant progress is reported in recent studies, it is still questionable whether this technology is reliable enough to be used as evidence in court. This talk will present main authorship analysis tasks and review achievements accomplished so far. Moreover, it will highlight limitations of this technology and discuss open problems that need to be faced in the following years.

Alexey Neznanov

Higher School of Economics

Title: "Knowledge Extraction from Social Media and other Internet Data Sources: Modern Features and Challenges". **Subtitle:** "Cases form Criminalistics and Medicine"

We will give practically-oriented overview of some aspects of knowledge extraction, from classical chats and forums, from various modern Social Media Services, and from other communication and collaboration services.

We will discuss the fulfilments of hopes and the actual challenges taking into account Big Data revolution, Open Data movement, and the convergence of technologies. There will be excursions to the application fields where the presenter or their colleagues have personal experience and interesting cases (Medical Informatics, Criminalistics, Education, etc.).

Rita Kuznetsova

«Антиплагиат. Research»

Methods for intrinsic plagiarism detection and author diarization

Session I

Alexander Beloborodov and Pavel Braslavski. Does Everybody Lie? Characterizing Answerers in Health-Related CQA

Maria Khokhlova, Viviana Patti and Paolo Rosso. Distinguishing between Irony and Sarcasm in Social Media Texts: Linguistic Observations

Anastasiya Bodrova and Natalia Grafeeva. Coreference Resolution Using Clusterization

Session II

Dmitry Ustalov and Sergey Igushkin. Sense Inventory Alignment Using Lexical Substitutions and Crowdsourcing

Prateek Mehta and Vasudeva Varma. Improved Topic Models For Social Media via Community Detection Using User Interaction And Content Similarity

Alexandra Vatan, Natalia Gusarova, Natalia Dobrenko and Galina Artemova. An Ontology Approach to Storing Educational Information

Elena Volzhina, Natalia Grafeeva, Elena Mikhailova, Andrei Chudin and Boris Novikov. Discovering Geo-Magnetic Anomalies: a Clustering-based approach

Tatiana Litvinova, Olga Litvinova, Olga Zagorovskaya, Pavel Seredin, Alexandr Sboev and Olga Romanchenko. "Ruspersonality": a Russian Corpus for Authorship Profiling and Deception Detection

Anton Kozmirchuk, Andrei Kokorev, Vyacheslav Nesterov and Elena Mikhailova. PostgreSQL Service with Backup and Recovery for Cloud Foundry

Anton Makarov and Andrey Shabunin. On Design of E-Services for Public Authority in the Russian Federation

Session III

Yelaman Abdullin, Vladimir Ivanov and Radhakrishnan Delhibabu. Deep Recurrent Neural Networks for Bilingual Sentiment Analysis of Short Texts

Polina Panicheva, Olga Bogolyubova and Yanina Ledovaya. Lexical, Semantic and Morphological Correlates of the Dark Triad Personality Traits in Russian Facebook Texts

Oleg Durandin, Nadejda Hilal, Dmitry Strebkov and Nikolai Zolotikh. Using Probability Distribution over Classes in Automatically Obtained Training Corpora

Valerie Mozharova and Natalia Loukachevitch. Two-stage Approach in Russian Named Entity Recognition

Natalia Loukachevitch and Aleksei Alekseev. Use of Neighbor Sentence Co-occurrence to Improve Word Semantic Similarity Detection



Poster and Demo Session

Extended Abstracts

Natalia Archakova, Kirill Boyarsky and Eugeny Kanevsky. Extraction of Low-frequent Terms from Domain-specific Texts by Cluster Semantic Analysis

Samaneh Zolfaghari and Mohammad Reza Keyvanpour. SHARiF: Smart Home Activity Recognition in a Framework

Anna Grabar. Social Infrastructure in Social Networks

Oleg Malafeyev, Nadezhda Redinskikh and Nikolay Rummyantsev. Multi-Agent Interaction in Social Trading Network

Svetlana Bodrunova, Anna Litvinenko and Ivan Blekanov. Defining Influencers in the Russian Twitter: People vs. Institutions in the Discussion on Anti-Migrant Bashings in Biryuliovo

Ekaterina Sharapova. Use the system Avtor.NET to fuzzy duplicates detection

Posters

Liliya Akhtyamova. Extracting important facts from electronic medical records for relevant treatment

Ivan Samborskii, Denis Mehanikov and Aleksandr Farseev. Рекомендация на основе мультимодальных данных из нескольких источников



ANTIPLAGIAT

RESEARCH

AntiPlagiat.Research tackles the most challenging problems in the area of natural language processing and plagiarism detection. Our aim is development of advancing technology and propagation of scientific thought.

We unite young talents from leading institutions: Moscow Phystech (MIPT), Computing Centre of RAS, Moscow State University and others.

Wide adoption of digital documents and word processing software has given rise to unauthorized use of text and cheating, which poses a major problem for educational organizations and independent authors.

We believe that our effort opens the gate to significant improvement in educational quality.

We reveal the truth and show the origins.

AntiPlagiat Research was founded at Anti-Plagiat JSC in 2015 and has already become a remarkable contributor to the Russian research community.

Anti-Plagiat JSC:

- operates on the market of plagiarism detection engines
- tops the list of Russian plagiarism detection software producers
- implements decisions of public authorities in the field of educational quality improvement

Research areas of our interest are:

- information retrieval
- automatic text categorization
- cross-language plagiarism
- machine-generated text detection
- semantic search and topic search

For more detailed information, please refer to:

<http://research.antiplagiat.ru>

CALL FOR PARTICIPATION

19th Conference of Open Innovations

Association FRUCT

Jyväskylä, Finland, 7-11 November 2016



Overview

FRUCT is the largest regional cooperation framework in form of open innovations between academia and industry. FRUCT conferences are attended by the representatives of more than 20 FRUCT member universities from Russia, Finland, Denmark, Italy, Ukraine, industrial experts from Jolla, EMC2, Intel, Nokia, Skolkovo and a number of guests from other companies and universities.

The conference is an R&D forum for the most active students, academic experts, industrial researchers and influential representatives of business and government. The conference invites the world-class academic and industrial researchers to give lectures on the most relevant topics, provides an opportunity for student teams to present progress and results of their R&D projects, meet new interesting people and form new R&D teams. The conference program consists of 3 to 5 intensive (½ or full day) trainings on the most promising technologies, plus 3 days of the main conference.

We warmly welcome all university research teams to participate in the conference, present your research and join the FRUCT Association. IEEE members and representatives of Russian and Finnish universities are entitled to large discounts. Registration to the conference is open at <http://www.fruct.org/conference19>.

Background and motivation

The distinctive feature of modern IT and Telecommunications industries is in dramatic shortening of the period when technology remains commercially viable. On the one hand, this is due to the competition between key market players that are pushing all manufacturers to accelerate innovations; on the other hand, this is due to technological progress speed up caused by the growing expansion of intellectual resource invested into R&D and design activities. This trend is an important call and challenge for the leading educational and research institutions around the globe. In the FRUCT we believe that it is crucial to combine forces of EU and Russia to follow up the competition in adopting university education to the new industrial trends.

The first step is to strength a bridge between Russian and Finnish academic worlds, increase visibility of involved research teams and set direct personal contacts between academic and industrial experts. More information about FRUCT is available at **www.fruct.org**.

Call for papers and presentations

You can select one of the following 3 types of submissions:

- Full papers (min 6 pages and up to 12 pages) - submission deadline is September 30, 2016
- Extended abstracts (min 200 words, max 5 pages) - submission deadline is September 30, 2016
- Poster or demo summary (min 200 words, max 5 pages) - submission deadline is October 31, 2016

All submitted papers will be peer reviewed by the technical committee. Please follow provided paper templates. The list of priority topics is as follows:

- Location Based Services, Navigation, Logistics management, e-Tourism solutions
- Mobile Healthcare, Wellbeing, Automated diagnostics, Fitness, e-Health solutions
- Future services: Proactivity, IoT, Smart Spaces, Context Analysis, Big Data and data mining
- ✓ Energy efficient design of sensors, integration of peripherals
- ✓ Cross-platform software, innovative mobile services, new approaches to application design, innovative UX
- Smart Systems, Inter-device connectivity, embedded networks

The list of additional topics is as follows:

- Mobile device security, management of personal and business privacy
- Modern network architectures, Emerging wireless technologies, Air interfaces and protocols
- Mobile multimedia and video services and solutions

All conference papers and abstracts will be published in FRUCT proceeding (ISSN 2305-7254), all Full papers will be published in IEEE Xplore (Scopus) and selected papers recommended (but not guaranteed) for CPCI indexing (Web of Science). The templates, conference news and other details can be found at

<http://www.fruct.org/conference19>

ISMW FRUCT recommendation

NLPub

NLPub is a catalogue and a community for Russian linguistic resources. NLPub is a nonprofit project, the goal of which is to popularize the natural language processing field in Russia through consolidation of the relevant information. The primary language of NLPub is Russian

<https://nlpub.ru>

ISMW FRUCT Conference and School

Program

Saint-Petersburg, Russia
28 August - 04 September 2016

Printed by ITMO University (Russia)

Approved for publishing on 20.08.2016
Page format 60x84 1/16
Number of copies 200

ITMO university publisher house
197101, Saint Petersburg, Kronverkskiy pr., 49



