

Предисловие

В 2016 году Международная конференция «Аналитика и управление данными в областях с интенсивным использованием данных» (“Data Analytics and Management in Data Intensive Domains” – DAMDID/RCDL’2016) проводилась с 11 по 14 октября в доме отдыха Ершово (Московская область, Одинцовский район).

Традиционно конференция «Аналитика и управление данными в областях с интенсивным использованием данных» представляет собой мультидисциплинарный форум исследователей и практиков из разнообразных областей деятельности людей, содействующий сотрудничеству и обмену идеями в сфере анализа и управления данными в областях исследований, движимых интенсивным использованием данных (ОИИД). Подходы к анализу данных и управлению данными, развиваемые в конкретных ОИИД X-информатики (таких как X=астро, био, гео, нейро, медицина, физика, химия, и пр.), социальных наук, а также различных ОИИД информатики, промышленности, новых технологий, финансов и бизнеса составляют предметную область конференции.

Программа конференции 2016 года отражает наряду с традиционной для управления данными тематикой движение в направлении науки о данных (data science) и аналитики с интенсивным использованием данных. Три приглашенных пленарных доклада конференции акцентируют ключевые проблемы развития методов и средств аналитики и управления данными в ОИИД. В докладе Андрея Ржецкого (профессора генетической медицины Чикагского университета), открывающем конференцию, рассматриваются подходы к изучению механизмов рака, развиваемые консорциумом UChicago, работающим под руководством Андрея Ржецкого по программе DARPA Big Mechanism (<http://www.darpa.mil/program/big-mechanism>). Эти подходы сфокусированы на развитии современных методов анализа данных в ОИИД, включая когнитивные методы, основанные на понимании естественного языка, подобные методам системы Ватсон IBM, с упором на формирование гипотез на основе обнаруживаемых в текстах причинных отношений, моделирование механизмов рака для автоматического предсказания терапевтических решений, применение методов организации экспериментов на основе их роботизации. Второй день открывает доклад Софии Ананиаду (директора Национального исследовательского центра Соединенного королевства по извлечению информации из текстов (NaCTeM) при Манчестерском университете, входящего составной частью в консорциум UChicago, финансируемого DARPA), в котором рассматриваются проблемы автоматической реконструкции траекторных моделей как результата обнаружения отношений между понятиями различной природы в текстах. Наконец, программа третьего дня начинается докладом Димитриоса Тзовараса (профессора и руководителя Института информационных технологий Центра исследований и технологий Hellas (Эллада) в Салониках), в котором дан аналитический обзор Европейской стратегии в области исследовательских инфраструктур.

Программный комитет конференции рассмотрел 57 заявок. Из них: 27 приняты как полные статьи, 16 – как краткие, 3 – как постеры, 2 – как демо, 9 – отклонены. 43 доклада (полные и краткие) представлены в 13 сессиях, таких как семантическое моделирование в DID, методы анализа данных, управление знаниями, системы управления обучением, семантический поиск и навигация, анализ паттернов в рекомендательных системах, исследовательские инфраструктуры (в астрономии, астрофизике, материаловедении, для мониторинга Земли), извлечение информации из текстов, интеграция и совместное использование данных, системы анализа текстов. Большинство докладов посвящены результатам исследований, выполняемых в исследовательских организациях, расположенных в различных местах на территории России, включая: Дубну, Екатеринбург, Звенигород, Иркутск, Казань, Москву, Новосибирск, Обнинск, Пущино, Томск, Тулу, Санкт Петербург, Ярославль.

Кроме того, в состав конференции включены следующие ассоциированные мероприятия: тьюториал, подготовленный профессором Вереной Кантере и ее аспирантом Максимом Филатовым в Женевском университете и содержащий анализ эффективности средств поддержки аналитики данных в многомашинных средах; диссертационный семинар, содержащий 5 докладов; а также открытый симпозиум «Здоровьесбережение на основе интенсивного использования данных», включающий три круглых стола, на которых обсуждаются бизнес-модели, мобильные технологии, вопросы диагностики в персонализированной медицине, а также две сессии, тексты докладов на которых, посвященные рассмотрению роли информатики в Здоровьесбережении, включены в сборник трудов конференции. Симпозиум подготовлен по инициативе члена-корреспондента РАН Андрея В. Лисицы (Научно-исследовательский институт биомедицинской химии (ИБМХ)).

Председатели Программного и Организационного комитетов конференции выражают благодарность Николаю Скворцову за взаимодействие при помощи систем СМТ с авторами присланных работ и с членами ПК – рецензентами докладов, а также за подготовку верстки сборника трудов конференции в процессе издания его печатной версии. Председатели ПК также выражают благодарность членам Программного комитета за выполненную ими работу по рецензированию и отбору докладов, а также Дмитрию Брюхову за поддержку актуального содержания сайта конференции на всех этапах подготовки DAMDID/RCDL'2016.

Дом отдыха Ершово, в котором проводилась конференция, расположен в уникальном историческом уголке Подмосковья в усадьбе династии Олсуфьевых. Усадьба Ершово находится в 50 км от Москвы в окрестностях Звенигорода, одном из красивейших мест Подмосковья, где неповторимая природа, живописный ландшафт и чистый воздух создают особую атмосферу для проведения научных форумов. Обладая современной инфраструктурой, удобной логистикой (порядка одного часа до Москвы по железной и автодорогам), историческими и природными достопримечательностями, Ершово способствовало проведению очередной научной конференции DAMDID/RCDL на высоком уровне.

Председатели Организационного комитета и Программного комитета конференции выражают благодарность авторам поданных на конференцию заявок, а также Российскому Фонду Фундаментальных Исследований и Национальному исследовательскому ядерному университету МИФИ за финансовую поддержку конференции.

**Сопредседатели
Программного комитета**

Калиниченко Леонид Андреевич
(ФИЦ ИУ РАН)
Кузнецов Сергей Олегович
(НИУ ВШЭ)
Манолопулос Янис
(Университет Аристотеля, Салоники)

**Сопредседатель
Организационного комитета**

Захаров Виктор Николаевич
(ФИЦ ИУ РАН)

Preface

In 2016 the International Conference “Data Analytics and Management in Data Intensive Domains” (DAMDID/RCDL’2016) was held on October 11 – 14 in the Holiday Center, Ershovo (Moscow region).

By tradition the “Data Analytics and Management in Data Intensive Domains” conference (DAMDID) is planned as a multidisciplinary forum of researchers and practitioners from various domains of science and research, promoting cooperation and exchange of ideas in the area of data analysis and management in domains driven by data intensive research. Approaches to data analysis and management being developed in specific data intensive domains (DID) of X-informatics (such as X = astro, bio, chemo, geo, medicine, neuro, physics, etc.), social sciences, as well as in various branches of informatics, industry, new technologies, finance and business are expected to contribute to the conference content.

The program of the DAMDID/RCDL’2016 conference alongside with the traditional data management topics reflects a rapid move into the direction of data science and data intensive analytics. Three invited plenary talks of the conference emphasize the key problems of methods and facilities of data analytics and management in DID. In the keynote of Andrey Rzhetsky (Professor of genetic medicine of the Chicago University) that opens the conference, the approaches for studying the mechanisms of cancer are presented, which are being developed by the UChicago consortium in which Andrey Rzhetsky acts as the PI according to the DARPA Big Mechanism program (<http://www.darpa.mil/program/big-mechanism>). These approaches are focused on the development of the contemporary methods of data analysis in DID including cognitive methods based on the understanding of the natural language resembling those of IBM Watson center, but focused on the hypothesis formation based on the causal relationships detected in the texts, cancer mechanism modeling to automatically predict therapeutic clues, application of robot scientist methods for the organization of experiments. The invited talk of Sophia Ananiadou (Director of the National Centre for Text Mining (NaCTeM) at the Manchester University) opens the second day of the conference. NaCTeM is included as a part of the UChicago consortium supported by DARPA. In this talk the approaches for automatic reconstruction of the pathway models are considered. These approaches are based on the discovery of the relationships of various kinds between the concepts of arbitrary nature in the texts. Finally, the program of the third day is opened with the keynote of Dimitrios Tzovaras (Director at the Information Technologies Institute of the Centre for Research and Technology Hellas in Thessaloniki) in which the analytical survey of the European strategy in the area of research infrastructures is presented.

The conference Program Committee has reviewed 57 submissions and accepted of them 27 as full papers, 16 as short papers, 3 as posters, 2 as demos, whereas 9 submissions were rejected. According to the conference program, these 43 oral presentations (of the full and short papers) are structured into 13 sessions including Semantic Modeling in DID, Data Analysis Methods, Knowledge Management, Learning Management, Semantic Search and Navigation, Pattern Analysis in Recommender Systems, Research Data Infrastructures (in Astronomy, Astrophysics, Material Sciences, Earth Monitoring), Data Extraction from Texts, Data Integration and Sharing, Text Analysis Systems. Most of the presentations are dedicated to the results of researches conducted in the research organizations located on the territory of the Russian Federation including Dubna, Ekaterinburg, Irkutsk, Kazan, Moscow, Novosibirsk, Obninsk, Puschino, Tomsk, Tula, Saint Petersburg, Yaroslavl, Zvenigorod,

Besides the above, the conference contains also the following satellite events: a tutorial “Data Analytics in Multi-Engine Environments” prepared by Dr. Verena Kantere and Ph.D. student Maxim Filatov from the University of Geneva; a Ph.D. Workshop containing 5 presentations; an open workshop “Data Intensive Healthcare”, including three round tables with the discussions of business models, mobile technologies, diagnostics and coaching in the personalized medicine, as

well as two sessions entitled as “Informatics in Healthcare” (the papers reflecting the presentations at these sessions are included into the conference proceedings). This workshop is prepared according to the initiative of Andrey Lisitsa (corresponding member of RAS, Research Institute of the Biomedical Chemistry, Moscow).

The chairs of the Program Committee and Organizing Committee of DAMDID/RCDL’2016 express their gratitude to Nikolay Skvortsov for the effective interactions by the CMT system with the authors of submissions and with the PC members reviewing the submissions, as well as for preparing a layout of the conference proceedings during the process of its publishing. The chairs of PC also express their gratitude to the PC members for carrying out the reviewing of the submissions and selection of the papers for presentation, as well as to Dmitry Briukhov for keeping of the up-to-date content of the conference site at all stages of the conference preparation.

The Ershovo Holiday Center in which the conference was held is located in the historic Ershovo mansion placed at the unique historical corner in the vicinity of Moscow, in the neighborhood of Zvenigorod. Providing an up-to-date infrastructure, convenient logistics (50 km from Moscow, about one hour to Moscow by railway or highway), historical and natural attractions, the venue contributed to the organization of the DAMDID/RCDL conference on a high standard.

The chairs of the Organizing Committee and Program Committee of DAMDID/RCDL’2016 express their gratitude to the authors of the submissions as well as to the Russian Foundation for Basic Research and the National Research Nuclear University MEPhI for the financial support to the Conference.

Co-chairs of the Program committee

Leonid A. Kalinichenko
(FRC CSC RAS)
Sergey O. Kuznetsov
(NRU HSE)
Yannis Manolopoulos
(Aristotle University, Thessaloniki)

Co-chair of the Organizing committee

Victor N. Zakharov
(FRC CSC RAS)