



Пост-релиз к конференции SLS 2018

Федеральный Исследовательский Центр «Информатика и управление», совместно с Научно-координационным Центром «Цифровые методы анализа данных», Лабораторией Машинного Интеллекта Физико-технического института провёл 12-ю международную конференцию «Интеллектуализация обработки информации» с 8 по 12 октября 2018 года в г. Гаэта, Италия. В рамках этого форума инициативной группой, в которую кроме организаторов ИОИ 2018 входили филологический факультет РУДН, редакция журнала «Речевые технологии» и Российский центр по науке и культуре в Риме была организована лингвистическая секция «Ситуация, язык, речь. Модели и приложения. («Situation, Language, Speech. Models & Applications» — SLS 2018)».

Заседания секции SLS 2018 были посвящены новому подходу к проблеме машинного понимания смысла текста и спонтанной речи. В фокусе внимания секции были доклады, в которых традиционные задачи распознавания и понимания решались в иной постановке — распознавались не только тематика и семантика текста/речи, но был осуществлён математический, или лингвистический подход к построению модели ситуации, т.е. по тексту и речи выявляется морфология ситуации, этапы её развития, моделируется ситуативная динамика, осуществляется прогноз.

Форуму предшествовали острые дискуссии с идеологами ряда математических и лингвистических конференций, которые прошли на нескольких научных форумах в России и за рубежом. Речь шла о подходах к распознаванию смысла текста и извлечению знаний из данных. В позициях многих авторов (в первую очередь математиков) поражает некоторая излишняя уверенность в эффективности их подходов.

Так математики, штурмующие латентные смыслы текстов, были готовы извлекать знания из данных об экономических, логистических, исторических и любых других процессах, литературных произведениях, данных о здоровье людей и методах их лечения, из историй их болезней, о судебных и следственных делах, о процессах развивающихся в СМИ, о том, что творится в социальных сетях, — и вообще о чём угодно используя чисто математические методы не привлекая в достаточной мере филологические, лингвистические, культурологические и другие необходимые для понимания, как можно было бы ожидать, знания из предметных областей.

С другой стороны, работа исследователей в области компьютерной лингвистики всегда была более основательной, и было ясно, что их проекты обладают более значительным потенциалом, чем то, для чего они проводились — автоматический (машинный) перевод текстов с одного языка на другой, машинное понимание текстов на естественном языке для сравнительно небольшого круга практических задач искусственного интеллекта.

Нельзя сказать, что специалисты, исповедующие сугубо технические методы совсем не осознают некоторую исчерпанность своих подходов. На конференциях лингвистов и математиков в последнее время стало модно выступать с докладами о тональности текста, находить в речи иронию, сарказм, распознавать модальности эмоций и даже юмор. Однако, радикальной постановки задач, резко расширяющих круг возможных применений титанических усилий математиков и лингвистов всего международного экспертного сообщества, в особенности в области корпусной лингвистики, — по сей день не произошло, и это имеет глубокие причины.

Первая из них — изоляционизм, свойственный любой технической предметной области, крепко привязанной к определённому инструментарию. Вторую причину можно увидеть в узкой специализации образования и связанной с ней недостаточностью эрудиции, часто определяющей глубоко мотивированную непроходимость барьера между гуманитарным и техническим знанием. Однако число специалистов, давно преодолевших этот барьер в международном научном сообществе, давно превысило критическую массу, и беременность науки новым направлением уже ощущают многие из них. Кстати, российские журналисты, ведущие исследования в области медиа уже отмечались на международных форумах, и выступали как ключевые спикеры, см. например, <http://internetscienceconference.eu/keynote.html>

Поэтому мы решили, что пора провести установочную конференцию, которая (как мы представили на главной странице сайта нашей конференции — <http://sls2018.info/>) посвящена новому подходу к проблеме машинного понимания смысла текста и спонтанной речи. В центре нашей проблематики оказалось то, что упускали как математики, так и лингвисты — последняя часть семиотической триады — ПРАГМАТИКА. Здесь понадобился весь арсенал теоретических наработок филологов и журналистов, который раньше имел по большей части методическое значение для обучения студентов филологических факультетов. Кроме того, нужен был широкий обзорный охват подходов и инструментов извлечения знаний из данных именно со стороны практиков, занимающихся журналистикой данных.

Нашей задачей на этой конференции было — усадить за общий стол специалистов трёх направлений — лингвистов, математиков и гуманитариев, занятых поиском семантических связей и их коммуникативных значений в тексте. Здесь важно было договориться о расширении взаимодействия специалистов разных направлений исследований в области моделирования ситуации по тексту и речи, которое приведет к созданию принципиально новых систем для широкого круга применений — от университетского образования до автоматизированных динамических новостных информационных систем востребованных медиарынком, или систем машинного анализа и прогнозирования развития ситуаций, в которых заинтересованы госзаказчики.

Традиционным гуманитариям для вхождения в новое междисциплинарное направление нужно было совсем немного — ознакомиться с опытом работ смежников, иметь

представление об инструментарию лингвистов и математиков ведущих исследования в этой области. Нужно было получить представление о корпусных инструментах в грамматических исследованиях русского языка, автоматической разметке семантических ролей в тексте, вероятностном тематическом моделировании. Докладчики готовились с полной ответственностью. Знакомство с инструментариум смежников позволило авторам докладов оценить как производительность применяемого инструментария, так и ограничения предлагаемых методик, узость поставленных задач, и, определив пробелы в существующих разработках, найти свое место в творческом процессе формирования новой научной отрасли, встроиться в тематику междисциплинарных исследований, представленных на конференции, понять ее дух.

В своем докладе (Кальян Виктор Петрович «Морфология ситуации. Структура момента и модели динамики», асс. каф. массовых коммуникаций филфака РУДН, научн. сотр. ФИЦ ИУ РАН) представил опыт разработки системы построения моделей ситуации по тексту/группе текстов с использованием методов математической и лингвистической текстологии. Им была поставлена задача построения языка описания широкого класса ситуаций с выявлением универсальной морфологии ситуаций. Как один из первых подходов к её решению оценивалось создание документальных баз предметных областей с хорошей терминологической оснащённостью. По группе узких специальных текстов, таких, как УПДК, медицинские словари, и подобное, выявлялись именованные сущности, имеющие значения морфем описания нормативных ситуаций в этих областях.

Полученный инвентарь ролей был применён к широкому классу ситуаций. По исходным текстам и экспертным разметкам семантических ролей с использованием средств корпусной лингвистики восстанавливались их свойства, качества, взаимоотношения между сущностями, отслеживались этапы изменения ролей, строились модели ситуации в виде последовательности таблиц и графов отношений, выявлялась динамика изменения ситуации.

Результат был представлен в виде проекта мультимодальной экспертной системы построения динамики ситуационных зависимостей и оценки вариантов развития событий. Описанный в докладе подход дал возможности построения по текстам и речи моделей не только ситуации, но и языка, мышления, восприятия, культуры. Это открыло новые перспективы для решения широкого круга задач — от построения компьютерных систем более точного литературного перевода текста с одного языка на другой, до выявления причин возникновения проблем в описанных в текстах/речи драматических ситуациях.

Большое внимание уделялось возможности описания интенций субъектов. Было построено математическое описание потенциалов движения к цели, оценки сопротивления среды в виде противодействия субъектов, и подсчёта вероятности достижения или не достижения цели, наступления тех или иных событий с помощью метрик и весов потенциалов стремления.

С ключевыми докладами на конференции также выступили Глазков Алексей Владимирович (МПГУ) «Что такое ситуация?», Харламов Александр Александрович (ИВНД и НФ РАН) «Семантика текста как модель ситуации», Кадырова Шуанат Набиевна (Филфак РУДН) «Бедная «русская мысль»: получим ли мы автоматизированное извлечение нарратива? Краткий обзор русскоязычных приложений», Габдрахманова Наиля Талгатовна (МИ РУДН) «Кластеризация документов с помощью нейронных сетей», Смирнов Иван Валентинович (ИСА РАН) «Сетевой дискуссионный текст как источник информации о сценарии», Митрофанова Ирина Игоревна (Филфак РУДН) «Клиповое мышление: реальность и перспективы», Дерябина Светлана Александровна (Филфак РУДН) «Феномен интерязыка машинного текста» и другие.

Была выявлена заинтересованность в совместной работе над проектами по профилю сделанных докладов с такими организациями, как Лаборатория Машинного Интеллекта Физико-Технического Института, корпорация Яндекс, компания Форексис. Большую организационную работу по распространению информации о конференции провел до конференции и продолжает в настоящее время Российский Центр Науки и Культуры в Риме.

В 2019 году в Риме планируется проведение второй конференции «Ситуация, язык, речь. Модели и приложения. («Situation, Language, Speech. Models & Applications» — SLS 2019)».

В этом номере журнала представлены расширенные статьи по некоторым докладам конференции, которые в наибольшей степени соответствуют тематике журнала.