

В СПбПУ прошел научно-образовательный интенсив «Digital linguistics: подготовка лингвистов нового поколения»



19 мая в Политехе состоялось долгожданное событие: научно-образовательный интенсив «Digital linguistics: подготовка лингвистов нового поколения», организованный Высшей школой лингводидактики и перевода Гуманитарного института.

Его актуальность обусловлена тем, что с 2022 года Высшая школа начинает подготовку бакалавров по направлению 45.03.04 «Интеллектуальные системы в гуманитарной сфере» (программа 45.03.04_01 «Цифровые технологии и иностранные языки») и подготовку магистров по направлениям 45.04.04 «Интеллектуальные системы в гуманитарной среде» (программа 45.04.04_01 «Цифровая лингвистика» (международная образовательная программа) «Digital Linguistics» (International Educational Program).

Основными докладчиками выступили к. фил. наук, доцент кафедры «Математическая лингвистика» СПбГУ, исполнитель и руководитель ряда грантов РФФИ, РФФИ, СПбГУ Ольга Александровна Митрофанова и к. фил. наук, имеющая степень магистра лингвистики дублинского университета (Ирландия), имеющая опыт сотрудничества

как с коммерческими компаниями, так и с разными учебными заведениями и исследовательскими организациями в России и за рубежом Полина Вадимовна Паничева, обе выпускницы кафедры «Математическая лингвистика» СПбГУ.



Гости от первого лица рассказали о своем опыте работы в и взаимодействия с коммерческими компаниями, обратив внимание, что сейчас нужно бизнесу от компьютерных/цифровых лингвистов, какие задачи призваны решать специалисты, имеющие эту квалификацию, занимая должности под самыми разными названиями. Так, специалисты, имеющие квалификацию «Цифровой лингвист», могут занимать должности компьютерный лингвист, продуктовый лингвист, специалист по диалоговым системам, лексикограф; инженер по автоматической обработке текста на естественном языке (NLP), лингвист-разработчик, архитектор систем NLP, специалист по анализу данных (data scientist), прикладной исследователь NLP, исследователь, инженер-исследователь, научный сотрудник, руководитель лаборатории и др.

Задачи, которые решают специалисты этого профиля, можно разделить на 2 группы.

Первая – автоматическая обработка текста на естественном языке (NLP): анализ тональности, извлечение фактов, автоматический перевод, информационный поиск, классификация текстов, автоматическое исправление ошибок; профилирование авторов текстов, оценка рисков неблагоприятных черт и состояний авторов, выявление экстремистских и человеконенавистнических высказываний, медицинские

приложения.

Вторая группа включает в себя лингвистические задачи: лексикографические исследования, создание и пополнение онтологий и терминологических словарей, сценариев для чат-ботов, разметка текстов и руководство, оценка; постановка задачи для задач 1-й группы.

Особый акцент П.В. Паничева сделала на том, чего хотят работодатели. Кандидаты на занятие вакансии должны иметь опыт программирования, знать алгоритмы и структуры данных, иметь опыт работы с текстом, словарями, онтологиями; иметь опыт участия в завершенных проектах, практической направленности, и решения релевантных задач, опыт работы с системой контроля версий, например GitHub, опыт применения нейросетевых моделей языка.

По мнению докладчика вооружить выпускника требуемыми навыками и компетенциями можно только при условии, что все изучаемые дисциплины будут содержать практические задачи, которые помогут студенту понять значимость (практическую, коммерческую, фундаментальную, академическую) конкретной информации, которую он получает при изучении дисциплины лингвистического, математического блока или блока программирования, которые в свою очередь получат развитие/будут закреплены в рамках НИР и производственной практики.



О том, как на практике реализована модель подготовки компьютерных лингвистов на кафедре Математической лингвистики СПбПУ, которая в этом году отмечает 60-летие со дня своего основания, рассказала О. А. Митрофанова. Первые выпускники кафедры осваивали программу 2-х факультетов СПбГУ: филологического и математико-механического. Эта особенность в дальнейшем выразилась в интеграции лингвистического и инженерного подходов, ориентации на практические задачи: тезаурусное моделирование, машинный перевод, информационный поиск, электронные библиотеки и т.д., в участии студентов в хозяйственных работах с внешними заказчиками. Сегодня в учебных планах большое внимание уделяется привлечению студентов к НИР по грантам и организации практик в компаниях-партнерах СПбГУ. В подготовке бакалавров и магистров время между основными блоками дисциплин лингвистического, математического и связанного с изучением языков программирования и освоением самых современных программных продуктов распределено примерно одинаково. Программа выстроена так, что помогает студентам «лингвистического склада» подтянуть программирование, а программистам развить лингвистический потенциал. Всячески поощряется взаимодействие и взаимопомощь студентов при решении учебных задач. Докладчик привела конкретные примеры гармонизации учебных планов и курсов на всех этапах подготовки специалистов, ориентации студентов на постоянное саморазвитие с помощью доступных онлайн курсов.



После выступлений докладчиков состоялось оживленное обсуждение, в котором приняли участие профессора ВШЛиП В.Е. Чернявская, Н.В. Попова, А.В. Рубцова, доцент Н.Б. Смольская. В завершении дискуссии директор ГИ Н.В. Чичерина предложила продолжить обсуждение актуальных вопросов подготовки цифровых лингвистов в разных форматах.