

Агрегация и использование результатов научной деятельности – одно из важнейших направлений, поддерживаемых Министерством образования и науки РФ как в рамках исполнения государственных программ, направленных на создание единой федеральной базы данных, включающей результаты НИОКР и технологических работ, так и в направлении формирования Единого информационного пространства Минобрнауки России.

24 ноября 2014 г. в Вычислительном центре им. А.А. Дородницына Российской академии наук (ВЦ РАН) состоялась II научно-практическая конференция «Технологии создания, агрегации и использования научного и образовательного контента», организованная консорциумом «Контекстум» совместно с ВЦ РАН и Национальной ассоциацией исследовательских и научно-образовательных электронных инфраструктур «e-APENA». Конференция проводилась при поддержке Комиссии по информатизации Федерального агентства научных организаций России и Российского фонда фундаментальных исследований.

Мероприятие было запланировано как очередной шаг в реализации концепции, лежащей в основе создания системы агрегации и использования результатов научной деятельности, разработанной организаторами конференции в рамках выполнения НИР «Разработка научно-методических и нормативно-организационных основ создания системы агрегации и использования результатов научной деятельности государственных академий наук». Концепция всесторонне обсуждалась участниками первой конференции, а её функционал проектировался как развитие самостоятельных направлений и проектов. По итогам первой конференции была принята резолюция с предложениями по дополнению функционала системы информационными сервисами, связанным с большими объёмами данных и обеспечивающими интеграцию с существующими информационными системами.

На II конференции были подняты вопросы агрегации и обработки больших массивов научной информации, накопленных в научных и образовательных учреждениях, и объединения их в общую информационную систему, а также создания сервисов подготовки и оценки научного контента в рамках современной организации научно-исследовательских работ.

Открыл конференцию **Игорь Соколов** (главный учёный секретарь Президиума РАН). Он отметил особое значение и важность формирования системы сбора и анализа результатов научно-исследовательской деятельности с точки зрения не только разработки планов проведения научных исследований, но и технологий и методов, которые позволяют анализировать современное состояние и прогнозировать будущее науки.

Во вступительном слове **Марат Биктимиров** (заместитель директора ВЦ имени А.А. Дородницына РАН) подчеркнул актуальность создания подобных систем.

Новый механизм организации совместной деятельности участников социэкономических и научных систем внутри сообществ учёных на примере системы «СОЦИОНЕТ» (socionet.ru) продемонстрировал в своём

выступлении заместитель директора Центрального экономико-математического института РАН **Сергей Паринов**. Как пример практической реализации «СОЦИОНЕТ» представляет собой CRIS-систему (модель современной научной информационной системы), развиваемую научным сообществом с 2000 г. Главная задача создания таких систем – развитие методов использования результатов исследований, сбора информации о том как, кем и для чего они использованы, а также формирование средств для генерации публичных показателей результативности ученых и исследовательских организаций. Уже сейчас «СОЦИОНЕТ» предоставляет российскому научному сообществу современную инфраструктуру взаимодействия в духе открытой науки, интегрированную с международной Research e-Infrastructure.

Дмитрий Семячкин (Генеральный директор открытой электронной библиотеки «КиберЛенинка») отметил, что ведущие страны мира уделяют особое внимание открытости результатов научных исследований. В России возможны два сценария повышения открытости: консервативный – выпуск печатных версий и размещение журналов в онлайн-доступе и инновационный – отказ от печатных версий и создание журналов открытого доступа. На основании опыта полученного при реализации проекта «КиберЛенинка» был предложен план создания платформы открытого журнала.

В качестве примера модели открытой науки был представлен новый проект «РУАЭСТ» – сводный цифровой российский архив статей периодических изданий. По мнению **Виталия Иванова** (председателя совета директоров консорциума «Контекстум»), в настоящее время открытые научные проекты становятся нестабильными. Главной задачей проекта ставится возвращение в научный оборот значительного массива научной информации, доступ к которому сегодня затруднён или вообще отсутствует.

Принцип нового проекта строится на объединении аналитической росписи научных публикаций, создаваемых библиотечным сообществом, с технологическими возможностями платформы «Контекстум», позволяющими легитимизировать введение в оборот электронных версий публикаций и организовывать доступ к ним.

Сегодня создан массив более чем из 700 наименований журналов, с издателями которых заключены лицензионные договора. Этот ресурс является основой действующего прототипа «РУАЭСТ» (www.ruaest.ru).

Партнёрами проекта выступают консорциум «АРБИКОН» и российские библиотеки, выполняющие аналитическую библиографическую роспись статей российской научной периодики.

Библиотека по естественным наукам РАН координирует функционирование централизованной библиотечной системы (ЦБС), включающей около 100 библиотек академических научных учреждений. Данная библиотечная система в процессе решения задач по информационному сопровождению научных исследований создаёт и агрегирует большие объёмы данных. В докладе **Николая Калёнова** (директора БЕН РАН) был дан обзор экспертной системы централизованного

комплектования БЕН РАН и создания распределенных библиографических баз данных.

Технологии управления жизненным циклом электронных публикаций в информационной системе научного журнала рассматривались **Александром Елизаровым** (Казанский (Приволжский) федеральный университет). Был представлен обзор и дан сравнительный анализ существующих информационных систем создания управления электронными публикациями, а также представлена платформа научных изданий science.tatarstan.ru, которая предоставляет возможности размещения научных журналов. Данная платформа внедрена и функционирует в качестве системы управления научными публикациями Казанского (Приволжского) федерального университета.

Полный цикл редакционной подготовки научной публикации в редакции научного журнала на основе информационной системы Math-Net.Ru был показан **Дмитрием Чебуковым** (Математический институт им. Стеклова РАН). В редакции журнала РАН полный цикл включает загрузку рукописи в электронном виде, её регистрацию с последующей организацией всех этапов продвижения статьи (классификации, рецензирования, авторской доработки, научного редактирования, перевода, редактирования английской версии, публикации в журнале и онлайн-доступа), организацию взаимодействия редакции с автором, рецензентом, куратором статьи, переводчиком, информирование автора о продвижении его статьи в редакции, составление автоматических отчетов для контроля над редакционным портфелем. Система Math-Net.Ru обеспечивает экспорт данных в Google Scholar, Crossref, Zentralblatt MATH и eLibrary.ru, а также имеет версию для мобильных устройств. Разработкой и апробацией системы занимались восемь ведущих математических научных журналов РАН.

В докладе **Ивана Смирнова** (Институт системного анализа РАН) было представлено дальнейшее развитие технологии Exactus Expert. Система предназначена для оценки оригинальности и выявления заимствований в текстах структурированных документов. Отличительными особенностями системы являются полный лингвистический анализ текстов, включая морфологический и синтактико-семантический анализ, нечувствительность к перефразированию и высокая скорость выявления заимствований.

Система поддерживает функции автоматического извлечения метаинформации, выявления в анализируемом научном тексте заимствованных фрагментов и источников заимствования, автоматического распознавания корректности/некорректности выявленных заимствований, определение степени оригинальности научного текста с учетом даты публикации, а также поиском заимствований в интернет.

Алгоритмы Exactus Expert участвовали в международных соревнованиях по поиску заимствований CLEF-2014 и показали высокие результаты по качеству и скорости поиска заимствований. Система внедрена в консорциуме «Контекстум».

Выступление **Павла Арефьева** (Национальный фонд подготовки кадров) было посвящено вопросам публикационной карьеры российского ученого. Основные проблемы, с которыми сталкивается автор, – это несоответствие системы научной коммуникации учёту и оценке научной деятельности, а также к созданию больших авторских научных коллективов, который может привести к потере автора как научного индивида. В основе неосведомлённости исследователя как автора и незначительной доли публикаций российских авторов в международных журналах лежит не только слабое знание языков, но и отсутствие навыков академического письма и работы с библиографическими и информационными ресурсами.

Решение этих проблем представляется в создании открытых ресурсов и архивов авторских публикаций (препринтов) с обеспечением процедуры рецензирования, а также развитие навыков авторов в области академического письма и работы с библиографическими и информационными ресурсами.

В выступлении **Андрея Гвоздева** (зам. генерального директора ЗАО «Эвентос») «Инфраструктура публикации научных данных в открытых машиночитаемых форматах» были подняты вопросы использования семантических веб-технологий в качестве основы для интеграции разнородных приложений, использующих структурированные базы данных. Были затронуты вопросы использования моделей интеграции данных в формате RDF, в том числе и на примере отечественных программных разработок. Особое внимание было уделено задаче связывания данных с информацией из других хранилищ с целью выявления контекста.

Подходы к обеспечению доступности информационных ресурсов через Интернет рассматривались в докладе **Бориса Логинова** (директор Государственной центральной научной медицинской библиотеки). Сложность решения этой задачи состоит в конфликте между социальной ролью библиотек (доступ должен быть бесплатным) и коммерческими интересами правообладателей. Для реализации этой задачи предлагается технология Электронного библиотечного абонемента (ЭБА), позволяющая использовать электронный экземпляр документа с соблюдением условий, установленных правообладателем, включающая ограничение доступа на одновременное использование правомерно приобретённого произведения в цифровом формате.

Технология ЭБА получила реализацию в программном модуле «Электронный библиотечный абонемент» входящим в состав АБИС ОРАС-Global и в настоящее время внедрена в Центральной научной медицинской библиотеке.

В заключение участники конференции высказали свои предложения по решению указанных проблем. Проект интеграции модуля «Электронная библиотека» АБИС «МегаПро» с сервисом поиска текстовых заимствований РУКОНТекст рассматривался в выступлении **Владимира Грибова** (Генеральный директор «МегаПро»). **Игорь Попов** (и.о. директора издательства МГУ) предложил рассмотреть возможность создания национальной системы оперативной публикации первичных результатов

научных исследований. **Александр Антопольский** (заведующий лабораторией Института научно-педагогической информации Российской академии образования) внес предложения по организации научного информационного пространства России в условиях современной политики государства в научно-образовательной сфере.

Определяя вектор дальнейшего развития науковедческих технологий и учета результатов научно-исследовательской деятельности, участники конференции отметили актуальность задачи агрегации и обработки больших массивов информации, накопленной в российских научных и образовательных учреждениях. Без предоставления эффективного доступа к этим материалам невозможна реализация государственных программ в области науки и образования.

Учитывая значимость рассматриваемых на конференции вопросов, было предложено:

- отметить важность создания научных электронных библиотек открытого доступа (Open access) и перевода отечественной научной периодики в открытый доступ, в первую очередь журналов, финансируемых государством, а также создание инфраструктуры публикации научных открытых данных в России;

- считать целесообразным изучение возможности применения технологии Электронного библиотечного абонемента для предоставления открытого доступа к произведениям, защищённым авторским правом и хранящимся в библиотечных фондах, в рамках действующего законодательства;

- отметить важность размещения в открытом доступе и под открытой лицензией результатов научных исследований, выполненных с привлечением государственного финансирования;

- разработать организационно-экономическую модель деятельности информационно-библиотечных и архивных служб в сфере науки и образования, для обеспечения необходимого уровня информационно-библиотечного обслуживания;

- сформировать перечень информационных ресурсов, отвечающий целям и задачам системы НТИ и с учетом реальных потребностей научно-образовательного сообщества;

- рассмотреть возможность передачи платных ресурсов НТИ в коммерческий сектор.

- разработать организационную модель национальной системы научно-образовательной информации, исходя из межведомственного и межрегионального характера информационной деятельности, наличия инфраструктуры, кадров и других параметров.

- организаторам конференции рассмотреть вопрос о проведении III Научно-практической конференции «Технологии создания, агрегации и использования научного и образовательного контента» в 2015 г.