

УДК 004.9
ГРНТИ 20.23.25

Научные электронные информационные ресурсы ИПР РАН и их востребованность научным сообществом

М.Р. Коголовский, к.т.н., доцент
e-mail: kagalovsky@ipr-ras.ru

Аннотация

Обсуждаются этапы и направления развития научных электронных информационных ресурсов ИПР РАН. Важную роль сыграло создание электронной библиотеки института и формирование на ее основе открытого архива научных публикаций с использованием технологии Инициативы открытых архивов. Кратко рассматриваются ключевые принципы этой технологии, а также система Соционет, средствами которой поддерживается открытый архив. Рассматривается также возможность использования на контенте архива реализованного в системе подхода, обеспечивающего семантическое его обогащение пользователями системы в онлайн-режиме путем создания семантических связей научного характера между его информационными объектами. Семантика создаваемых связей определяется средствами разработанной таксономии. Обсуждаются новые формы научной деятельности, обеспечиваемые благодаря использованию предложенного подхода, а также новые возможности для наукометрических исследований на контенте открытого архива института и всего множества открытых архивов, поддерживаемых системой. Приводятся показатели востребованности информационных ресурсов открытых архивов ИПР РАН и других академических институтов научным сообществом пользователей системы Соционет.

Статья подготовлена при поддержке Российского гуманитарного научного фонда – проект 14-02-12010-в.

Ключевые слова: электронная библиотека, веб-сайт, открытый архив, система Соционет, семантическая связь, таксономия связей, наукометрия, научная коммуникация, востребованность научных публикаций

Введение

Работы по обеспечению электронного представительства ИПР РАН в глобальной сети начались в 1996 г. с создания официального веб-сайта ИПР РАН. При этом в качестве основной задачи рассматривалось представление на сайте результатов научных исследований института. Имелось в виду предоставлять в открытый доступ научному сообществу не только библиографические описания материалов, представляющих результаты выполненных исследований, но и полные тексты работ или, по крайней мере, их аннотации. На сайте начали публиковаться электронные версии опубликованных статей и фрагментов монографий, научные отчеты по выполненным сотрудниками института исследованиям, авторефераты защищенных диссертаций, тезисы и полные тексты докладов на научных конференциях, а также работы, не опубликованные типографским образом (веб-публикации). В состав библиотеки включались не только текущие, но и ретроспективные публикации. Фактически начала формироваться научная электронная библиотека института. Ее развитие осуществлялось в несколько этапов.

В 2003 г. библиотека как составная часть сайта института была выделена в автономный его модуль, обеспечивающий доступ только к публикациям, представленным полными текстами или, по крайней мере, их аннотациями.

Библиотечный модуль, по сути, является многоаспектным каталогом таких ресурсов. Сами эти ресурсы по-прежнему остаются физически составной частью сайта института. Они остаются доступными для пользователей сайта, как и ранее, традиционным образом через каталог научных отчетов сайта, через общий каталог публикаций сотрудников и с их личных страниц на сайте. Вместе с тем, полнотекстовые публикации доступны также и через каталоги электронной библиотеки. Такая архитектура сайта и его электронной библиотеки поддерживается до настоящего времени. Она обеспечивает множественность путей доступа к ресурсам сайта, представляющим результаты научной деятельности института, и повышает уровень их доступности для пользователей. Пользователям при этом предоставляется открытый доступ ко всем информационным ресурсам, представленным на сайте и в электронной библиотеке.

Следующий этап развития электронной библиотеки института начался в 2006 г. Его целью была интеграция ее информационных ресурсов в среду крупного отечественного онлайн-научно-образовательного информационного пространства Соционет [7, 8]. Эта система основана на технологии открытых архивов и содержит большой объем информационных ресурсов по социально-экономической тематике. Соционет функционирует уже более 15 лет и приобрела в последние годы де-факто институциональный статус в Отделении общественных наук РАН. Информационное пространство Соционет содержит также публикации ряда образовательных учреждений и других организаций, импортирует метаданные из репозитория метаданных более 1500 источников, включая сотни отечественных и зарубежных открытых архивов. Система предоставляет также метаданные собственного репозитория метаданных для импорта другим системам. В настоящее время в Соционет доступно около 2.5 млн. информационных объектов, содержатся описания более 8 миллионов семантических связей между ними. Система Соционет стала также полигоном для проведения исследований в области инновационных технологий электронных библиотек. Постоянно проводятся работы по расширению разнообразия типов информационных ресурсов, представляемых в этой системе, и развитию функциональности механизмов управления ими.

В 2007 г. в среде системы Соционет на основе электронной библиотеки института был создан открытый архив [6, 14], который был зарегистрирован в ряде международных реестров репозитория метаданных открытых архивов. Благодаря использованию технологии открытых архивов (Open Archives Initiative, OAI) [17, 26] публикации сотрудников и другие научные ресурсы института по-прежнему поддерживаются на веб-сайте института, а в среде Соционет создан репозиторий метаданных этих ресурсов, который и позволяет пользователям системы получать доступ к ним. Интеграция ресурсов электронной библиотеки института в систему Соционет позволила отказаться от развития функциональности поддерживаемой на сайте версии электронной библиотеки, поскольку широкий спектр возможностей для пользователей предоставляют сервисы этой системы. В системе обеспечивается не только доступ к ресурсам библиотеки, но и целый ряд других

функций, в частности, регистрация доступов пользователей к ресурсам и генерация соответствующих статистических показателей, а также средства новых семантических технологий.

Базирование механизмов системы на технологии открытых архивов обеспечивает также импорт этих информационных ресурсов в другие открытые архивы. Начиная с 2012 г. коллекции монографий и научных отчетов института, представленные в электронной библиотеке, импортируются в международную систему RePEc (Research Papers in Economics) [20]. Эта система также основана на технологии открытых архивов. Импорт информационных ресурсов из других открытых архивов означает импорт метаданных, описывающих эти ресурсы, в репозиторий метаданных импортирующей системы из репозитория метаданных открытых архивов - источников. Сервисы, созданные для системы RePEc, генерируют ежемесячную статистику просмотров описаний и доступов к отчетам и монографиям, описания которых импортированы из Соционет в эту систему.

Параллельно с указанными процессами с 2009 г. на полигоне Соционет начались исследования, разработки инструментария и эксперименты по семантическому обогащению поддерживаемых в ней информационных ресурсов, а также по созданию сервисов системы, которые позволяют обеспечивать на этой основе новые функциональные возможности для пользователей [1-4]. Семантическое обогащение контента осуществляется путем создания семантических связей между информационными ресурсами, представленными в открытых архивах, которые поддерживаются в системе. Семантика связей отображается классами создаваемых связей. Допустимые классы определяются таксономией семантических связей, специально разработанной для системы Соционет [5]. Эта таксономия определяет разнообразные отношения между информационными объектами, характерные для научной и научно-организационной деятельности. Функциональные возможности системы Соционет, основанные на использовании семантических связей, могут применяться и к информационным ресурсам научного открытого архива ИПР РАН.

Отметим, наконец, что в 2015 г. комплекс научных электронных информационных ресурсов института пополнился еще одним важным ресурсом. Институт начал выпускать

научный электронный журнал «Проблемы рыночной экономики». Высокий научный уровень публикаций в журнале призван обеспечивать авторитетный международный редакционный совет, включающий ряд ведущих ученых, работающих по тематике журнала. Издание журнала обеспечивает еще один, более статусный способ оперативной публикации результатов научных исследований по актуальным проблемам рыночной экономики в нашей стране, но уже не только для сотрудников ИПР РАН.

В остальной части статьи рассматриваются состояние контента научного открытого архива ИПР РАН и ключевые идеи технологии Инициативы открытых архивов, на которой основана поддерживаемая в Соционет его версия, реализованный в системе подход, обеспечивающий семантическое обогащение контента архива пользователями системы в онлайн-режиме путем создания семантических связей научного характера между его информационными объектами на основе разработанной таксономии семантических связей. Обсуждаются новые формы научной деятельности, а также возможности более информативных наукометрических исследований на контенте архива, которые обеспечиваются благодаря использованию предложенного подхода. Приводятся оценки востребованности информационных ресурсов открытых архивов ИПР РАН и других институтов ООН РАН научным сообществом пользователей системы Соционет, которая их поддерживает.

Состояние научных информационных ресурсов института

В настоящее время научные информационные ресурсы ИПР РАН включают, как уже отмечалось, электронные версии полных текстов или аннотаций статей, опубликованных в периодических изданиях, полных текстов и фрагментов изданных монографий, коллекцию аннотаций научных отчетов по результатам исследований, выполненных в институте со дня его основания, авторефераты защищенных диссертаций и полные тексты некоторых диссертаций, тезисы и полные тексты докладов на научных конференциях, а также работы, не опубликованные типографским образом (веб-публикации). Среди электронных публикаций только текущие, но ретроспективные работы сотрудников.

Все эти ресурсы и описывающие их метаданные поддерживаются на веб-сайте инсти-

тута. Существует несколько возможностей доступа к представленным публикациям. Прежде всего, возможен обычный текстовый поиск на сайте. Лучшие возможности предоставляются сервисами системы Соционет. Более удобен доступ к ресурсам конкретного сотрудника – автора публикаций возможен из его личной страницы, на которой приводятся библиографические описания его публикаций и ссылки либо на их версии на сайте, либо версии на сайте издания, где она опубликована. Возможен также доступ через сводный каталог публикаций всех сотрудников института.

Библиотечный модуль веб-сайта института обеспечивает дополнительные возможности доступа. Он включает ряд каталогов, через которые возможен доступ. Поддерживаются, в частности, предметный и авторский указатели. Авторский указатель ссылается на персональные страницы авторов ресурсов, представленных в электронной библиотеке. Такая страница содержит список публикаций соответствующего автора со ссылками на эти публикации. Справочный аппарат библиотеки включает также каталоги представленных в библиотеке монографий, статей, научных отчетов, депонированных в ВИНТИ работ, препринтов, трудов конференций, проведенных институтом, веб-публикаций и авторефератов диссертаций со ссылками на тексты соответствующих ресурсов. Кроме того, имеются тематические указатели публикаций по направлениям исследования института.

Представление электронной библиотеки в среде системы Соционет в форме открытого архива также включает коллекции описателей публикаций архива отдельно по типу публикаций – монографии, статьи, научные отчеты, веб-публикации и т.д. Каждый описатель ссылается на файл соответствующей публикации, хранимый на веб-сайте института. Кроме того, имеется авторский указатель, который для каждого автора ссылается на его описатель – профиль автора. Профиль автора содержит сведения об авторе, ссылку на его личную страницу на сайте института и список публикаций этого автора в открытом архиве со ссылками на описатели этих публикаций. Имеется также описатель (профиль) института, который помимо характеристик института в целом содержит список всех авторов – сотрудников института со ссылками на профили авторов, а также список всех публикаций всех сотрудников института, сгруппированных по

годам их издания или независимо от годов издания. Имеются также тематические указатели публикаций по направлениям исследований института.

Контенты электронной библиотеки на сайте института и открытого архива в среде системы Соционет постоянно актуализируются.

Технология открытых архивов

Как уже указывалось, научная электронная библиотека ИПР РАН представлена помимо веб-сайта института также и в среде системы Соционет в виде открытого архива.

Термин *открытый архив* был предложен авторами *Инициативы открытых архивов* (Open Archives Initiative, OAI) [17], разработавшими технологию интеграции информационных ресурсов из источников, доступных в среде Веба. Предполагается, что единицы информационных ресурсов (информационные объекты) интегрируемых источников являются файлами различных форматов, доступны в Вебе с помощью обычных браузеров, поддерживаются и управляются их владельцами. *Открытым архивом* создатели технологии OAI называют виртуальный набор информационных объектов, организованных в виде коллекций. Каждой коллекции соответствует материализованная коллекция метаобъектов. Каждый метаобъект представляет собой совокупность метаданных, описывающих соответствующий информационный объект, включающих, в частности, его адрес в Вебе. Совокупность метаданных открытого архива называется его *репозиторием метаданных*.

Основная задача технологии открытых архивов заключается в обеспечении возможности интеграции коллекций информационных объектов, принадлежащих различным открытым архивам. Для решения этой задачи открытые архивы должны быть *интероперабельными*. Каждый из них должен иметь возможность импортировать в соответствии с заданным регламентом нужные для него коллекции информационных объектов из других зарегистрированных в нем открытых архивов и обеспечивать их синхронизированное состояние с источниками импорта.

Следует заметить, что используемый в этой технологии термин *открытый архив* вовсе не предполагает свободного доступа к содержащимся в нем информационным объектам. Предоставление пользователям полномочий доступа и введение или отсутствие ограничений доступа относятся полностью к ком-

петенции владельца ресурсов архива и определяются проводимой им политикой доступа. Открытость архива следует понимать *в технологическом смысле*, означающем следование в используемой технологии открытым стандартам, т.е. широко распространенным стандартам де-факто. Именно это обстоятельство обеспечивает интероперабельность открытых архивов. В технологии OAI таким стандартом служит протокол *сборщика метаданных* OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) [26]. Более точно, роль этого протокола следует рассматривать как обеспечение *интероперабельности репозитория метаданных* открытых архивов и при необходимости взаимного импорта содержащихся в них метаданных.

Интеграция данных открытых архивов не сводится к созданию физически единого хранилища информационных объектов интегрируемых архивов, включающего копии импортируемых информационных объектов из интегрируемых источников. Информационные объекты импортируются виртуально. Коллекции информационных объектов открытых архивов – источников интеграции хранятся и поддерживаются их владельцами, доступны в среде Веб и не перемещаются в импортирующий архив. Интеграция информационных ресурсов обеспечивается путем физического импорта *лишь метаданных* этих информационных объектов из репозитория метаданных открытого архива, содержащего эти объекты.

Каждый открытый архив обычно снабжается некоторой операционной онлайн-оболочкой, предоставляющей конечным пользователям интерфейс для доступа к ресурсам его репозитория метаданных, поиска и просмотра в нем описаний требуемых информационных объектов и доступа к ним путем навигации по гиперссылкам, содержащимся в описывающих их метаданных. Предоставляемые через такие интерфейсы функции различных оболочек могут различаться.

Вместе с тем, в соответствии с протоколом OAI-PMH, репозиторий метаданных каждого открытого архива должен также обладать *механизмом сборщика метаданных* (Metadata Harvester) из других открытых архивов и *сервером с программным интерфейсом* (API) для сборщиков метаданных других открытых архивов, которые могут импортировать метаданные из репозитория метаданных данного архива. Указанный сервер позволяет запрашивающему сборщику метаданных выпол-

нять запросы шести видов, определенных протоколом OAI-PMH. Сборщик метаданных импортирующего открытого архива является клиентским приложением этого сервера, которое может издавать такие запросы и активизируется в соответствии с регламентом, установленным администратором импортирующего открытого архива. Поэтому интеграция информационных ресурсов на основе технологии OAI осуществляется в автоматическом режиме.

Таким образом, важным фактором обеспечения интеграции источников информационных ресурсов в технологии OAI является предложенная архитектура поддержки интероперабельности распределенного множества репозиторий метаданных. Стандартизация интерфейсов репозиторий метаданных обеспечивает их *гомогенизацию* и тем самым интероперабельность. Однако для обеспечения интероперабельности репозиторий метаданных различных открытых архивов необходима также *стандартизация представления метаданных*, в них содержащихся. Технология OAI допускает использование в каждом репозитории метаданных одного или нескольких форматов их представления. Можно, например, использовать набор элементов Дублинского ядра [15], стандарт MARC [16] и др. При этом используются XML-представление метаописаний информационных ресурсов и XML-схемы для верификации таких описаний. Сборщик метаданных может запросить у репозитория метаданных – источника импортируемых метаданных, какие форматы представления метаданных в нем используются. Запрашивая собственно метаданные, он может указать, в каком формате их представить в результате обработки его запроса.

Технология OAI имеет в настоящее время множество конкретных реализаций и используется, в частности, в системах Соционет [7, 8] и RePEc [20]. Созданы и поддерживаются реестры существующих открытых архивов, например, OpenDOAR [25], ROAR [19] и др.

Семантическое обогащение контента открытого архива

В последние годы в области научных электронных библиотек активно развиваются исследования и разработки, направленные на обогащение их контента семантической информацией. Благодаря этому может быть повышена информативность, а, следовательно, и

ценность представленных в этих системах информационных ресурсов. Помимо традиционного доступа непосредственно к информационным объектам библиотеки, становится возможным решение новых важных задач, открываются новые возможности для использования систем этого класса.

В частности, активно развивается подход к семантическому обогащению информационных ресурсов научных электронных библиотек путем декларации явным образом семантических связей между информационными объектами контента библиотеки. Такие связи могут нести различную информацию. Они могут представлять отношения научного характера между публикациями, в частности, мотивы цитирования, отношения, характеризующие влияние одной публикации на другую или использование в ней результатов, обсуждаемых в другой, оценку результатов одной публикации в тексте другой, структурные связи между компонентами или различными версиями публикации, отношения авторства между персонами и публикациями, вклад каждого из авторов в создании коллективных публикаций и др. Такой подход, обеспечивающий семантическое обогащение информационных ресурсов, реализован в системе Соционет, и его механизмы применимы к ресурсам открытого архива ИПР РАН, поддерживаемого этой системой.

В Соционет допускаются только бинарные ориентированные семантические связи между ее информационными объектами. Информационный объект-участник связи, из которого исходит связь, называется ее *исходным* информационным объектом. Второй участник, на который направлена связь, называется *целевым* объектом этой связи. Участниками семантических связей могут быть научные публикации, наборы научных данных, профили (метаданные) их авторов и других зарегистрированных пользователей системы, профили организаций, в которых они работают. Связываемыми информационными объектами могут быть также представленные в электронной форме такие продукты научной деятельности, как программы научных исследований, научные отчеты, рецензии, описания проектов, онтологии различных исследуемых предметных областей, средства программного обеспечения и их описания, спецификации стандартов метаданных и др.

Семантические связи декларируются децентрализованно в онлайн-режиме любыми

ми авторизованными пользователями системы. Исключение составляют связи цитирования, семантика которых может декларироваться только пользователями-авторами цитирующих публикаций. При создании связей пользователи могут руководствоваться различными мотивами. Авторы информационных объектов могут получать оповещения о возникновении новых связей, затрагивающих их объекты, и об изменениях свойств существующих связей с их участием, реагировать на такие события созданием новых связей.

В результате такой совместной деятельности независимых друг от друга пользователей и авторов над контентом системы динамически формируется семантическая сеть [1, 2], поддержка и использование которой обеспечивают новые функциональные возможности системы.

Семантические связи представляются в системе Соционет как самостоятельные информационные объекты. Автономное представление семантических связей позволяет использовать объекты-связи как участников других связей. Кроме того, благодаря этому можно формировать автономные репозитории семантических связей, которые могут реализовываться как самостоятельные источники информации и могут интегрироваться с другими подобными репозиториями, базирующимися на контенте иных открытых архивов, основанных на единых с ним стандартах представления метаданных.

Сочетание указанных возможностей позволяет использовать систему с семантически структурированным контентом как полигон для новых видов научной деятельности, как информационный источник для новых форм наукометрии и исследования корпуса научных знаний, а также для решения ряда других задач. Эти возможности и разработанные механизмы для их практического применения подробно обсуждаются в работах [4, 18].

Определение семантики связей

Декларация семантики связей, создаваемых пользователями в системе Соционет, осуществляется на основе *таксономии семантических связей*. При разработке этой таксономии были приняты во внимание и использованы результаты ряда известных проектов последних лет. Наиболее продвинутые работы в этой области были выполнены специалистами в области биомедицины [21-23] и консорциума W3C [24]. Значительное влияние

в рассматриваемой области имеет разработанная в рамках проекта CERIF организацией euroCRIS (<http://www.eurocris.org/>) унифицированная концептуальная схема научно-исследовательской среды, называемая ее авторами *полной моделью данных (Full Data Model)* [11] и онтология [12], определяющая систему терминов для обозначения сущностей этой модели и отношения между ними. Следует также упомянуть таксономию видов вклада авторов в создание коллективных публикаций [10] и основанный на ней проект открытого стандарта для выражения существенных для исследования ролей его участников [13].

В таксономии связей, разработанной для системы Соционет [5], с точностью до имен классов и структурирования использованы классы связей, определенные в указанных источниках, включен также ряд дополнений, например, классы связей «пользователь-публикация», характеризующих рекомендации и информацию для автора публикации. Эта таксономия имеет *двухуровневую организацию*. Верхний ее уровень включает одиннадцать классов семантических связей, для каждого из которых определено множество его подклассов.

Семантика связей может быть весьма разнообразной и зависит от видов связываемых информационных объектов. Например, связь между профилем пользователя и научной публикацией может представлять *оценку* этой публикации в целом или рассмотренных в ней отдельных результатов исследований - своего рода *рецензию* публикации. Связь между двумя публикациями может констатировать *использование* в одной из них метода, подхода или данных, опубликованных в другой, либо указывать на факт плагиата. Поскольку в предлагаемом в этой работе подходе предполагается, что семантические связи могут представляться как самостоятельные информационные объекты контента архива, связь между профилем пользователя системы и некоторой другой связью может выражать мнение этого пользователя о факте или мнении, выраженном этой другой связью. Связь между профилем персоны и коллективной публикацией может описывать *вид вклада* в ее создание данной персоны как соавтора.

Другие примеры касаются семантики связей цитирования между научными публикациями. Факты существования таких связей, представленные в пристатейных списках ис-

пользуемых источников, составляют источник данных для традиционной практики наукометрических измерений. Слабой стороной наукометрии, основанной на таком источнике, является отсутствие явным образом определенной семантики связей цитирования. Такие связи естественно называть «немыми» [1], поскольку они сами по себе не несут какой-либо информации, характеризующей, например, мнение автора цитирующей работы о цитируемом источнике или мотив цитирования. В связи с этим возможны такие парадоксальные ситуации, когда благодаря высокому количеству цитирований высоко ценится статья, содержащая грубые ошибки и/или принципиальные заблуждения, касающиеся обсуждаемой проблемы, и в связи с этим вызывающая активный отклик научного сообщества. Во избежание таких ситуаций в предлагаемом подходе предусматривается наделение связей цитирования семантикой. Их семантику на основе используемой онтологии определяют пользователи системы - авторы цитирующих публикаций.

Таксономия связей, используемая в системе Соционет, реализована в виде набора контролируемых словарей имен классов (подклассов) связей. *Контролируемым словарем* называют набор наперед заданных элементов, из которого пользователь системы может выбрать нужный ему для ввода значения в какое-либо поле экранной формы используемой системы. В нашем случае каждому классу верхнего уровня таксономии системы Соционет соответствует свой контролируемый словарь, а его подклассам – элементы этого словаря, представленные в виде имен этих подклассов.

Подробнее таксономия семантических связей системы Соционет обсуждается в работе [5]. В Соционет имеются средства создания, модификации и использования контролируемых словарей классов семантических связей для декларации семантических связей в системе и изменения их свойств.

Новые формы научной деятельности в среде открытого архива

На полигоне открытого архива ИПР РАН, поддерживаемого в системе Соционет в режиме своеобразной социальной сети, может обеспечиваться деятельность пользователей по созданию семантических связей. Благодаря этому становятся возможными некоторые новые формы научной деятельности, результаты

которой будут открытыми для всего научного сообщества. Это, естественно, обуславливает более высокий уровень ответственности ее участников. В частности, в такой среде возможны новые формы научных коммуникаций [4].

Важно здесь отметить, что, по существу, создание в онлайн-режиме семантической связи некоторого класса воплощает в среде системы некоторый *акт коммуникации* автора создаваемой связи как со всем сообществом пользователей системы, так и персональной коммуникации с автором целевого, а иногда и исходного объектов связи. Содержание передаваемого сообщения представляется при этом классом/подклассом связи, а также, возможно, и комментарием автора связи.

Коммуникация с автором целевого объекта имеет место, когда создается семантическая связь между профилем автора этой связи и некоторой публикацией. Например, это может быть связь какого-либо оценочного класса, выражающая одобрение целевой публикации связи, негативное к ней отношение и др. Другой случай – коммуникация между автором создаваемой связи и авторами исходной и целевой публикаций этой связи. Такая коммуникация возникает, например, при создании связи между двумя публикациями некоторым пользователем системы. Важно при этом, что авторы публикаций – участников создаваемой связи – в обоих случаях уведомляются о создании связи с участием их публикаций. Они могут не только ознакомиться с параметрами связи, но и отреагировать на неё. Конечно, персональная коммуникация в виде уведомления авторов публикаций - участников созданной связи - возможна в том случае, если в их профилях, созданных в системе при их регистрации, указан адрес электронной почты. Акт коммуникации с авторами публикаций-участников связи порождается также и в случае, когда эта связь уже существует и изменяются значения некоторых ее атрибутов, прежде всего атрибута, указывающего класс связи или содержание комментария.

Рассмотрим наиболее характерные разновидности научных коммуникаций, реализованные в системе Соционет.

Публикация результатов исследований и предоставление их научному сообществу без посредничества издателей. Благодаря развитию веб-технологий появились новые возможности для публикации результатов выполненных исследований в научных элек-

тронных библиотеках с помощью специалистов или самостоятельно (*самоархивирование*) без традиционного посредничества издателей. При этом радикальным образом сокращается время поступления сообщения ее автора адресатам в виде его публикации и многократно расширяется круг потенциальных адресатов этого сообщения - специалистов, заинтересованных в ознакомлении с этой публикацией. Развитие такой формы коммуникации стимулируется *инициативами открытого доступа* к результатам научных исследований, активно поддерживаемыми мировым научным сообществом. В системе Соционет имеется механизм, называемый *персональным роботом*, который отслеживает появление в ее контенте публикаций по тематике, интересующей данного пользователя. В случае появления публикаций такого рода этот пользователь получает уведомление от системы.

Сообщение пользователя системы оценочного мнения о некоторой публикации ее автору. В случае, когда пользователь системы создает связь одного из оценочных классов, которые предусмотрены в таксономии связей, между своим профилем, хранимым в системе, и некоторой публикацией (ее метаобъектом), то он тем самым осуществляет коммуникацию с ее автором. Посредством этой коммуникации создатель связи сообщает автору публикации свое мнение о его работе. Такая коммуникация близка по смыслу традиционному рецензированию. Смысл рецензии здесь может выражаться не только в виде структурированных данных - именем класса связи (позитивная оценка, негативная оценка и т.п.), но и неструктурированными данными - текстовым комментарием в метаобъекте-описателе связи. При создании связи система активизирует сервис уведомления, направляющий автору публикации сообщение по электронной почте о появлении новой связи, участником которой является его публикация. В сообщении указывается ссылка на ее метаобъект. Работа механизма уведомления возможна, только если участники коммуникации имеют в Соционет персональные профили с указанными в них адресами их электронной почты. В метаобъекте связи автор публикации может узнать о классе связи, ознакомиться с комментарием, авторством и временем создания связи.

Сообщение автору публикации о возможном изменении оценочного мнения некоторого пользователя о его работе. Зарегистрирован-

ный пользователь системы обладает полномочиями на изменение описания созданной им семантической связи. Изменение, в частности, может быть вызвано изменением мнения автора связи о ее целевой публикации. Оно может осуществляться заменой ранее указанного класса/подкласса связи или изменением содержания комментария в метаобъекте данной связи. После сохранения модифицированного метаобъекта связи сервис уведомления, как и в предыдущем случае, отправляет автору целевой публикации сообщение по электронной почте.

Сообщение о мнении пользователя системы относительно научного отношения между публикациями. Зарегистрированный пользователь может создать связь между двумя публикациями из открытого архива. Например, по содержанию исходной публикации он может утверждать, что в ней высказано некоторое оценочное мнение о целевой публикации. Между такими публикациями может быть создана связь независимо от того, имеется ли в исходной публикации связи явное цитирование целевой публикации. Уведомление о создании связи направляется авторам обеих публикаций с указанием, как и ранее, уникального идентификатора созданной связи.

Сообщение об изменении мнения пользователя системы о научном соотношении между некоторыми публикациями. Такое сообщение порождается в случае, когда пользователь системы, например, осознал ошибочность своего мнения и вносит какие-либо изменения в метаобъект созданной им связи. После запоминания модифицированного метаобъекта, как и в предыдущем случае, авторам обеих связанных публикаций-участниц обновленной связи направляется уведомление электронной почте.

Проведение научных дискуссий. Благодаря представлению автономных семантических связей в системе Соционет как обычных информационных объектов, они могут сами становиться участниками других связей. Используя эту возможность, можно поддерживать в системе дискуссионные форумы.

Пусть, например, пользователь системы выразил мнение о некоторой представленной в ней публикации, создав для этого связь какого-либо оценочного класса между своим профилем и метаобъектом указанной публикации. Уведомленный системным сервисом автор этой публикации или какой-либо иной

пользователь системы может, в свою очередь, выразить мнение как о рассматриваемой публикации, так и о мнении, высказанном первым пользователем. В последнем случае он создает связь между своим профилем и метаобъектом связи, созданной первым пользователем, и т.д. Во всех случаях создания новых связей при этом сервис уведомления оповещает авторов затрагиваемых публикаций и связей. В такой дискуссионный процесс могут вовлекаться и авторы оцениваемых публикаций, и создатели связей.

Более информативная наукометрия

Поддержка семантических связей между информационными объектами открытого архива в среде системы Соционет позволяет осуществлять наряду с традиционными наукометрическими измерениями, основанными на «немых» ссылках цитирования, не несущих информации о его мотивах, также и более информативные измерения, а также проводить исследования корпуса научных знаний, представленных в системе.

Прежде всего, система генерирует статистические данные, характер которых разнообразен и определяется семантикой (таксономическими классами) связей между цитирующими и цитируемыми публикациями. Например, можно получить для конкретной публикации количество входящих или исходящих из нее связей (т.е. связей, в которых данная публикация участвует как целевая или, соответственно, как исходная), относящихся к некоторому классу связей заданной таксономии связей. Получаемая статистика может показывать, в частности, сколько имеется позитивных или негативных ссылок на данную работу, в каком количестве работ используются методы, предложенные в ней, или содержащиеся в ней научные данные и т.д.

Помимо результатов статистических запросов предусматривается возможность получения перечня конкретных публикаций, представленных в системе и связанных с заданной публикацией как с исходной или с целевой в связях цитирования заданных классов. Например, можно выяснить, на результаты каких публикаций опирается данная конкретная работа или, наоборот, в каких публикациях получены результаты, основанные на данной работе.

Еще одна важная группа возможностей связана с использованием полного графа связей цитирования. Исследуя этот граф, можно

решать различные задачи, связанные как с анализом его топологии и вычленением подграфов с дугами разных классов таксономии связей цитирования, имеющих ясную семантическую интерпретацию, так и с визуализацией подграфов. Например, можно с помощью средства фильтрации классов связей вычленять и визуализировать из многослойной структуры полного графа связей слой, соответствующий классу связей, указывающему на использование одних публикаций как основополагающих для других публикаций. Этот слой будет в общем случае представлять множество графов, каждый из которых соответствует одной из публикаций, на результатах которой базируются какие-либо иные публикации. Можно также получить подграф, образованный связями цитирования, характеризующими цитируемые публикации как развивающие результаты, обсуждаемые в цитируемой публикации и принадлежащие ее автору. Полученный подграф будет характеризовать логику развития данной области науки, конечно, если контент системы содержит достаточно представительное множество публикации, относящихся к рассматриваемой области.

Структура графа семантических связей может использоваться также для семантической навигации пользователя в структуре контента открытого архива.

Указанные возможности системы Соционет обсуждаются подробнее в работах [6, 18].

Востребованность информационных ресурсов ИПР РАН

Для наукометрических измерений на открытом архиве института, поддерживаемого в системе Соционет, могут использоваться ее специальные сервисы. Имеется две группы таких сервисов. Первые из них генерируют статистические показатели доступов пользователей к информационным ресурсам на основе ежедневной регистрации их обращений к отдельным информационным объектам [9]. Здесь различаются показатели просмотров и скачиваний. Просмотром считается доступ к описателю публикации – к ее метаданным. Скачиванием считается скачивание из архива полного текста публикации. Оба показателя характеризуют интерес пользователя к данной публикации, однако более естественно оценивать действительную ее востребованность количеством скачиваний полного текста публикации. Такой показатель используется в ряде

электронных библиотек как характеристика востребованности.

Статистические показатели доступов к публикациям (просмотров метаданных и скачиваний) сервисы Соционет могут агрегировать по авторам, по организациям, в которых работают авторы, а также по рубрикам поддерживаемых в системе классификаторов научной информации (ГРНТИ и JEL). Эти показатели могут генерироваться для заданных пользователем периодов времени.

Вторая группа наукометрических сервисов системы генерирует статистику существующих на данный момент семантических связей между информационными объектами системы. Эти сервисы не имеют специального пользовательского интерфейса. Они выдают данные, касающиеся того или иного информационного объекта, при обращении к этому объекту. Генерируемая ими статистика дифференцируется по классам связей и по авторам связанных информационных объектов. Ее генерация основана, как уже отмечалось, на поддерживаемой в системе таксономии семантических связей. Статистика семантических связей характеризует степень влияния одних публикаций на другие, различные варианты использования результатов, опубликованных в одной работе, в другой, оценки публикаций пользователями системы или авторами других публикаций в их работах и др. Хотя сервисы этой второй группы уже разработаны и функционируют в системе, анализ генерируемых ими результатов пока не имеет практического смысла, поскольку в системе еще не сформирован достаточно представительный массив семантических связей.

В прилагаемой к статье таблице приведены оценки востребованности (скачиваний и просмотров) публикаций сотрудников первых десяти институтов Отделения общественных наук из их списка, упорядоченного по убыванию среднегодового за последние пять лет показателя скачиваний. Анализ приведенной статистики показывает, что несмотря на небольшую численность научного персонала ИПР РАН и связанное с этим относительно небольшое количество публикаций в открытом архиве института по сравнению с архивами других значительно более крупных институтов Отделения, ИПР РАН устойчиво на протяжении последних пяти лет занимает в рейтинге востребованности его научной продукции наиболее высокое место, даже по сравнению с такими крупными научными

учреждениями, как ЦЭМИ РАН, Институт экономики и организации промышленного производства СО РАН и др. Возможно, это связано со значительным количеством публикаций сотрудников ИПР РАН, посвященных наиболее актуальным проблемам отечественной экономической науки. Нужно отметить также, что институты ООИ РАН проявляют разную степень активности в представлении публикаций своих сотрудников в системе Соционет. Это, естественно, также сказывается на показателях востребованности их информационных ресурсов сообществом пользователей системы.

Заключение

Научные электронные информационные ресурсы ИПР РАН активно используются сообществом пользователей системы Соционет, а также веб-сайта института и представленной на нем научной электронной библиотеки. Они постоянно пополняются и доступны по всем трем указанным каналам. Вместе с тем, продолжаются работы по развитию семантической оболочки контента открытого архива института в среде системы Соционет в виде структуры семантических связей ее информационных объектов, в результате которых существенно повышается их информационная значимость. Развиваемая уникальная система функциональных сервисов Соционет позволяет пользователям системы получать ранее недоступную информацию, проводить качественно новые наукометрические измерения, поддерживать в оперативном режиме некоторые важные формы научной деятельности.

Литература

1. Когаловский М.Р., Паринов С.И. Семантическое структурирование контента научных электронных библиотек на основе онтологий. В кн.: "Современные технологии интеграции информационных ресурсов: сборник научных трудов". - Санкт-Петербург: Президентская библиотека им. Б.Н. Ельцина, 2011. - С. 26-45.
2. Когаловский М.Р., Паринов С.И. Классификация и использование семантических связей между информационными объектами в научных электронных библиотеках // Информатика и ее применения. - 2012. - Т. 6. Вып. 3. - С. 32-42.
3. Когаловский М.Р., Паринов С.И. Технологии социальной сети для создания семанти-

- ческих связей информационных объектов в научной электронной библиотеке //Программирование. МАИК/Наука «Интерпериодика». 2014. Т. 40. № 6.
4. Коголовский М.Р., Паринов С.И. Научные коммуникации на базе электронных библиотек с онлайн-декларацией семантических связей. В сб.: «Электронные библиотеки: перспективные методы и технологии, электронные коллекции: XVI Всероссийская научная конференция RCDL-2014 (Дубна, 13-16 октября): труды конференции» /Сост. Л.А. Калмыкова, М.Р. Коголовский. – Дубна: ОИЯИ, 2014. – С. 115-123.
 5. Коголовский М.Р., Паринов С.И. Таксономия семантических связей информационных объектов контента научной электронной библиотеки //НТИ. Серия 2. Информационные процессы и системы. - 2015. - No. 9. – С. 15-23.
 6. Коголовский М.Р., Паринов С.И., Ильменская Е.М. Открытый архив публикаций по рыночной экономике и наукометрические измерения. Труды XII Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции – RCDL-2010. Казань, 13-17 октября 2010 г.». – Казань: Казанский университет, 2010. – С. 155-161.
 7. Паринов С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов //Российский научный электронный журнал «Электронные библиотеки». – 2003. – Том 6. – Вып. 1. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP>
 8. Система Соционет. <http://socionet.ru>
 9. Статистика Соционет. <http://socionet.ru/stats.xml>
 10. Allen L., Brand A., Scott J., Altman M. and Hlava M. Credit where credit is due. Nature, Vol. 508, April 2014. http://www.nature.com/polopoly_fs/1.15033!/menu/main/topColumns/topLeftColumn/pdf/508312a.pdf
 11. CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS, 2012. http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf
 12. CERIF 1.3 Semantics: Research Vocabulary. CERIF Task Group, euroCRIS, 2012. http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf
 13. CRediT. An open standard for expressing roles intrinsic to research. <http://credit.casrai.org/proposed-taxonomy/>
 14. Institutional Open Archive Web Page. http://mei.socionet.ru/oai/eoorg_inst-1/oai.xml
 15. ISO 15836:2003 (E). Information and Documentation – The Dublin Core metadata element set.
 16. MARC XML Architecture. MARC 21 XML Schema. <http://www.loc.gov/standards/marcxml/>
 17. Open Archives Initiative (2000): <http://www.openarchives.org/>
 18. Parinov S., Kogalovsky M. Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. Scientometrics. Volume 98, Issue 2 (2014), Page 927-943.
 19. Registry of Open Access Repositories/ <http://roar.eprints.org/>
 20. RePEc. <http://repec.org/>
 21. Semantic Web Applications in Neuromedicine (SWAN) Ontology. W3C Interest Group Note, 20 October 2009. <http://www.w3.org/TR/2009/NOTE-hcls-swan-20091020/>
 22. Shotton D. Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/>
 23. Shotton D. and Peroni S. Semantic annotation of publication entities using the SPAR (Semantic Publishing and Referencing) Ontologies /Beyond the PDF Workshop, La Jolla, 19 January 2011. http://imageweb.zoo.ox.ac.uk/pub/2010/Publications/Shotton&Peroni_semantic_annotation_of_publication_entities.pdf
 24. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 Aug. 2009. <http://www.w3.org/TR/skos-reference/>
 25. The Directory of Open Access Repositories OpenDOAR. - OpenDOAR. <http://www.opendoar.org/index.html>
 26. The Open Archives Initiative Protocol for Metadata Harvesting. <http://www.openarchives.org/OAI/openarchivesprotocol.html>

Scientific digital information resources of MEI RAS and demand for them by scientific community

Mikhail R. Kogalovsky, PhD (Tech.), Associate Professor
e-mail: kogalovsky@ipr-ras.ru

Abstract

The article describes stages and development directions of the MEI RAS scientific digital information resources. Among them the important role was played by creation of the institute digital library and formation on its basis of open archive of scientific publications based on the Open Archive Initiative technology. The key principles of this technology as well as Socionet system, which means support open archive, are briefly considered. The possibility of use on content of the open archive the approach implemented in Socionet for providing his semantic enrichment by the system users in online mode by creation of semantic scientific linkages between his information objects also is considered. Semantics of the linkages is defined by the developed taxonomy of semantic links. New kinds of scientific activity as well as new possibility for scientometric research on the institute open archive and all set of open archives supported by the system are described. In the article, indicators are represented which characterize demand of the information resources from the open archives of MEI RAS and other RAS institutes by scientific community of the Socionet users.

Keywords: digital library, web site, open archive, Socionet, semantic linkage, linkage taxonomy, scientometrics, scientific communication, publication demand

Об авторе

Коголовский Михаил Рувимович, кандидат технических наук, доцент, зав. лабораторией Института проблем рынка РАН, Москва.

ПРИЛОЖЕНИЕ

Востребованность в Соционет публикаций институтов РАН общественно научного профиля в период 2011-2015 гг.*

Название института	2011	2012	2013	2014	2015	Среднегодовые за 5 лет
Институт проблем рынка РАН	1559/9877	1221/6257	922/5006	528/4075	452/3887	936/5820
Институт экономики и организации промышленного производства СО РАН	945/11454	918/8104	570/7339	423/5168	369/4177	645/7248
Институт США и Канады РАН	956/5615	526/3691	401/3211	396/3582	344/3092	524/3838
Центральный экономико-математический институт РАН	1001/13436	623/8799	378/8468	277/7122	236/5919	503/8749
Институт народно-хозяйственного прогнозирования РАН	803/3053	533/1959	450/1681	283/1006	128/581	439/1656
Институт философии РАН	866/11022	358/4001	271/2406	145/1803	108/1746	350/4196
Институт экономики РАН	418/3357	243/1981	155/1786	97/1501	215/3118	226/2349
Институт Европы РАН	368/1669	113/438	89/278	31/160	16/258	123/561
Институт проблем региональной экономики	228/459	190/408	126/303	42/191	18/80	121/288
Институт социально-политических исследований РАН	127/509	70/275	40/224	31/184	22/116	58/262

* В числителе указано число скачиваний публикаций, в знаменателе – число просмотров их метаданных.