

Метрики онлайн-информационных пространств¹

М.Р. Когаловский, С.И. Паринов

(Москва)

e-mail: kogalov@cemi.rssi.ru, sparinov@hse.ru

На примере системы Соционет – первого в нашей стране научно-образовательного информационного пространства в области социально-экономических наук – обсуждается проблематика нового научного направления, порожденного переносом социально-экономической деятельности в среду Интернет-технологий: метрики онлайн-информационных пространств. В статье рассматривается совокупность метрик, основанных на статистических сервисах Соционет.

1. ВВЕДЕНИЕ

Соционет² является крупной распределенной функционирующей в среде Интернет информационной системой, поддерживающей первое и пока единственное в нашей стране глобальное информационное научно-образовательное пространство в области социально-экономических наук. По причинам, описанным ниже, Соционет является пока уникальным примером процессов переноса существенной части профессиональной деятельности (в данном случае – научно-исследовательской) в онлайн-среду и попыток извлечения из этого организационных выгод.

Система была разработана и начала функционировать в 2000–2006 гг. Она развивалась силами сотрудников Института экономики и организации промышленного производства СО РАН (ИЭОПП СО РАН) благодаря грантам Фонда Форда. Начиная с 2007 г. развитие Соционет осуществляется в рамках программы «Открытый доступ к результатам исследований», активными участниками которой являются институты Отделения общественных наук РАН, межинститутская рабочая группа «Соционет» (Паринов, 2006) и Государственный университет – Высшая школа экономики (ГУ–ВШЭ). Координатором этой программы является ЦЭМИ РАН, существенная часть организационной работы выполнена ГУ–ВШЭ. Соционет предоставляет пользователям свободный доступ к ее сервисам и информационным ресурсам, которые включают, в частности, систематические коллекции аннотаций публикаций ведущих отечественных научных экономических журналов – «Экономика и математические методы» и «Вопросы экономики».

Поскольку информационные ресурсы Соционет формируются децентрализованно, эффективность использования этой системы в существенной степени зависит от наличия достаточных обратных связей, сигнализирующих администраторам и пользователям соответствующих информационных ресурсов о происходящих в ней процессах и изменениях. Для

¹ Работа выполнена при финансовой поддержке Российского гуманитарного научного фонда (проект 06-02-12205-в) и Фонда Форда (проект 1055-0505).

² См. домашнюю страницу системы Соционет - <http://socionet.ru/>

обеспечения обратных связей потребовалось разработать системные метрики³ – совокупность показателей, характеризующих свойства информационных ресурсов системы и процессов их использования, а также встроить в Соционет измеряющие их сервисы. Целью данной статьи является обсуждение указанной совокупности показателей и принципов их измерения.

Рассматриваемые метрики имеют явную наукометрическую специфику, но, на наш взгляд, методическая часть этой работы (подходы к сбору онлайн-статистики, характеристики онлайн-поведения действующих лиц и результативности их деятельности) имеет более широкое научное значение, так как предлагает начать обсуждать новую проблематику измерений функционирования социально-экономических систем в онлайн-среде, создаваемой Интернет-технологиями⁴.

В начале статьи приведены общие сведения о системе Соционет и ее месте в активно развиваемой в последние годы международной научно-образовательной информационной сетевой инфраструктуре. В разделе 3 обсуждаются организация информационного пространства системы, представленные в нем информационные объекты и образы действующих лиц, описывающие их метаданные, структура взаимосвязей между информационными ресурсами, используемые в системе методы интеграции информационных ресурсов из множества распределенных источников. В разделе 4 показаны возможности проведения в системе измерений наукометрического характера, а также оценки качества метаданных, определяющих качество представления информационных ресурсов для пользователей. Раздел 5 посвящен обсуждению общих принципов оценки трудозатрат на подготовку информационных ресурсов для системы Соционет, перечислены факторы, влияющие на такие оценки.

2. ОБЩАЯ ХАРАКТЕРИСТИКА СОЦИОНЕТ

2.1. Основные функции системы. Главными функциями системы Соционет являются (подробнее см. (Паринов, 2004)):

- накопление метаданных, описывающих научные публикации и другие научные информационные ресурсы (научные отчеты, рабочие записки, тезисы докладов, статистические таблицы, базы данных, сведения о научных организациях, научная персоналия, материалы научных проектов в социально-экономической области и т.д.);
- формирование развитой структуры связей между объектами метаданных, поступивших в систему из различных источников и описывающих ее информационные ре-

³ Термин «метрика» в области информационных технологий часто используется не в традиционном математическом смысле как некоторого рода расстояние между точками множества, а как количественная характеристика некоторого свойства реальных сущностей, процессов и т.п. Мы используем здесь этот термин именно в таком смысле.

⁴ Другие публикации на эту тему: Алексеев, Киселев, Паринов, 1995; Паринов, 1999, 2002.

сурс, которая обеспечивает расширенные возможности для доступа пользователей к системным ресурсам, а также для проведения наукометрических измерений и оценки показателей качества представления информационных ресурсов в системе;

- предоставление научному сообществу доступа к этим метаданным, а при возможности – и к полным текстам (полному контенту) и/или аннотациям описываемых ими публикаций и других информационных ресурсов;
- использование системы как источника информационных ресурсов для других электронных библиотек и иных информационных систем.

2.2. Соционет и международная научная информационная инфраструктура. Соционет интегрирована с Единым научным информационным пространством РАН (ЕНИП РАН) (Бездушный и др., 2004), является его источником информационных ресурсов. Она интегрирована также в международную научную информационную сетевую инфраструктуру, поскольку разработка Соционет изначально выполнялась в рамках международных инициатив RePEc (Research Papers in Economics) (Крихель, Паринов, 2002; ReDIF, 1996) и Open Archives Initiative (OAI), соглашения и стандарты которых поддерживаются системой. Благодаря этому Соционет может предоставлять своим пользователям доступ к информационным ресурсам других систем, основанных на стандартах OAI, и одновременно – доступ этим системам к своим информационным ресурсам.

В ЦЭМИ РАН, который является теперь базовой организацией Соционет, ведется работа, направленная на поддержку новой инициативы международного научного сообщества – программы «Свободный доступ к результатам исследований» (Берлинская Декларация, 2003; Berlin 3 Open Access, 2005; Budapest Open Access Initiative, 2002). Ученый совет и дирекция института поддерживают присоединение ЦЭМИ РАН к этой программе.

2.3. Источники информационных ресурсов системы. Соционет развивается усилиями научно-образовательного сообщества и используется в его интересах. Координируемое администратором Соционет пополнение информационных ресурсов системы осуществляется сообществом независимых друг от друга владельцев информационных ресурсов (или представителей владельцев, большинство из которых – институты РАН), зарегистрированных и авторизованных системным администратором. Они выполняют свою миссию автономно от других участников этой деятельности.

В последнее время после учреждения деятельности, направленной на развитие и использование Соционет в ООИ РАН, процессы формирования новых информационных ресурсов системы существенно интенсифицировались. В этой работе участвует большая часть институтов отделения общественных наук (на апрель 2007 г. – 22 из 29 институтов).

Важно отметить здесь, что политика представления собственных информационных

ресурсов в системе определяется каждым владельцем самостоятельно и имеет, как правило, институциональную основу. Благодаря этому в рамках системы не возникает вопросов, касающихся защиты авторских прав на публикуемые в Соционет материалы.

2.4. Метрики системы Соционет. Поскольку ресурсы информационного пространства Соционет формируются децентрализованно, для обеспечения их эффективного использования в систему встраивается ряд сервисов, позволяющих измерять и визуализировать различные показатели развития и использования информационных ресурсов. Эти сервисы позволяют, в частности:

- автоматически измерять на основе поддерживаемых в системе метаданных и данных, регистрируемых в системных журналах в процессе ее функционирования, различные полезные наукометрические показатели;
- измерять некоторые показатели качества метаданных, описывающих публикуемые в системе научные информационные ресурсы;
- формировать и предоставлять пользователям динамические ряды указанных показателей.

Важно отметить, что мы имеем здесь дело с электронными информационными ресурсами, которые предоставляются пользователям для свободного доступа в интерактивном (онлайн) режиме с помощью системных механизмов. Поэтому рассматриваемые показатели их качества называют также онлайн-показателями.

К числу таких показателей относятся показатели наукометрического характера, характеризующие на институциональном и персональном уровне деятельность научного сообщества, активность исследований в отдельных научных направлениях и востребованность полученных результатов, степень их влияния на развитие соответствующей ветви науки, притягательность тех или иных направлений исследований. Имеется также возможность измерять показатели качества представления информационных объектов в системе (степень их идентифицированности, доступность, информативность представления и др.), которые можно назвать презентационными показателями.

Некоторые из указанных сервисов уже функционируют в системе Соционет, другие находятся в стадии разработки. Измеряемые статистические показатели доступны ее пользователям. Более того, все исходные данные для расчета показателей также свободно доступны пользователям, и на их основе заинтересованные пользователи могут осуществлять собственный анализ. Описание исходных данных можно найти в работе (Паринов, 2007).

3. ИНФОРМАЦИОННЫЕ РЕСУРСЫ СИСТЕМЫ

Система Соционет функционирует в среде Всемирной паутины (Веб), и ее информа-

ционные ресурсы представляют собой совокупность взаимосвязанных гиперссылками информационных объектов и коллекций информационных объектов (далее для краткости – просто *коллекций*), а также метаданных, которые описывают их свойства в соответствии с принятыми в системе стандартами. Для представления информационных ресурсов в системе используются веб-технологии, в частности, технологии XML.

Управление информационными ресурсами Соционет осуществляется средствами доступного в среде Веб набора сервисов, обеспечивающих их создание авторами либо администраторами информационных ресурсов, поддержку и доступ к ним пользователей, поддержку деятельности персонала системного администратора, измерение онлайн-показателей качества информационных ресурсов системы, а также ее интеграцию в отечественную и международную научно-образовательную информационную сетевую инфраструктуру.

Рассмотрим состав и особенности организации информационных ресурсов системы, а также используемые в Соционет методы интеграции информационных ресурсов из множества доступных ей источников.

3.1. Информационные объекты. Основу научного контента системы Соционет составляют информационные объекты. *Информационным объектом* (или точнее, экземпляром информационного объекта) называют уникальным образом идентифицируемое представление некоторой сущности предметной области системы с помощью информационных ресурсов. Свойства сущностей предметной области представляются *атрибутами* соответствующих информационных объектов. Информационные объекты системы типизируются в соответствии с типизацией представляемых ими сущностей предметной области.

В Соционет предусматривается ряд типов информационных объектов, которые представляют такие сущности, как организации, поставляющие свои информационные ресурсы системе, публикации различного рода и их авторов, а также коллекции публикаций, рецензии, комментарии, проекты, средства программного обеспечения и т.д.

Организации являются, как правило, институтами РАН – местом работы авторов публикаций. Авторы публикаций могут самостоятельно представлять свои публикации в Соционет и воспользоваться для этой цели услугами назначенного организацией администратора ее информационных ресурсов, представляемых в Соционет.

Используемый в данном контексте термин «публикация» обозначает не только публикации в традиционном смысле, т.е. работы, изданные типографским способом. Это может быть любой электронный научный контент в различных возможных формах: электронные версии статей, опубликованных в периодике; полных текстов или фрагментов монографий; полных текстов или тезисов докладов, представленных на научных конференциях; препринтов, аннотаций таких изданий и других печатных материалов. Вместе с тем, это могут быть

электронные версии научных отчетов, авторефератов диссертаций и полных текстов диссертаций, различных рабочих записок и других рабочих материалов, новостных сообщений, презентации докладов; веб-публикации; аудио- или видео-ресурсы, научные веб-сервисы, страницы веб-сайтов, файлы в FTP-архивах и т.п.

Многие включаемые в настоящее время в информационное пространство Соционет публикации уже доступны в среде Веб в веб-версиях научных журналов, на веб-сайтах организаций, в электронных библиотеках и в других сетевых источниках. В противном случае электронные версии публикаций могут создаваться и поддерживаться их авторами либо веб-мастерами организаций на принадлежащих им узлах Интернет. При необходимости система Соционет может предоставлять институтам РАН собственные ресурсы – пространство памяти на ее серверах для поддержки веб-сайтов институтов, отдельных файлов различных форматов, содержащих их информационные объекты – тексты научных работ, аналитические обзоры, фотографии, статистические таблицы и др., а также предоставлять доступ к системным сервисам, обеспечивающим создание этих информационных объектов и/или описывающих их метаданных.

Авторы или организации, обладающие авторскими правами на публикацию, самостоятельно принимают решение о представлении ее электронной версии в Соционет, поддерживая при положительном решении ее доступность в Веб средствами собственных ресурсов или помещая ее на серверы Соционет. Важно заметить, что управление полнотекстовыми публикациями и другим научным контентом (их создание, предоставление к ним доступа, актуализация, обеспечение надежного хранения и т.п.) не является функцией системы Соционет. Эти ресурсы полностью управляются их авторами или ответственным за эту функцию персоналом институтов РАН. В силу этого персонал Соционет не несет ответственности за соблюдении авторских прав на представленные в системе публикации.

Информационные объекты каждого типа обладают атрибутами, которые характеризуют их свойства. Набор атрибутов информационного объекта зависит от его типа. Например, к числу атрибутов публикации относится гиперссылка в формате URL, идентифицирующая местоположение ее полного текста в Интернет. К числу атрибутов организации относится URL домашней страницы ее веб-сайта. Важным обязательным атрибутом информационного объекта каждого типа является его уникальный в рамках системы идентификатор, который назначается механизмами Соционет каждому представляемому в ней информационному объекту.

Информационные объекты некоторых типов, таких как организации, авторы, коллекции, представляются описывающими их метаданными – наборами значений их атрибутов. Информационные объекты–публикации наряду с этим располагают еще полным текстом

(полным контентом). Однако полный текст оказывается доступным пользователям системы не для каждой публикации, хотя это весьма желательно. Как указывалось ранее, автор публикации или организация могут проводить такую информационную политику, в соответствии с которой, полные тексты некоторых принадлежащих им публикаций не предоставляются для свободного доступа. Электронная версия полного ее контента может вообще не существовать. В таких случаях система позволяет пользователю, по крайней мере, выявить факт существования публикации и получить полную библиографическую ссылку, аннотацию и/или некоторые другие характеристики данной публикации.

3.2. Коллекции информационных объектов. Администратор множества представляемых в Соционет информационных объектов (публикаций, авторов, новостей и др.) объединяет их в *коллекции* на основе какого-либо принятого им критерия: общности авторства публикаций; их хронологической или тематической общности; общности места работы авторов, составляющих коллекцию публикаций, и т.п. Автор публикаций, самостоятельно создающий и помещающий коллекцию публикаций в Соционет, или его доверенное лицо, выполняющее эту функцию, должны обладать правами администратора этой коллекции.

Администратор учреждается для каждой коллекции, и она связывается с организацией, под эгидой которой эта коллекция создается. Для соответствующей организации, в свою очередь, устанавливается симметричная связь с данной коллекцией. Совокупность коллекций организации называется ее *архивом*.

Некоторые коллекции могут создаваться непосредственно администратором системы Соционет. К числу таких коллекций относится, в частности, коллекция организаций, информационные ресурсы которых уже представлены или планируется представлять в Соционет.

При создании коллекции система назначает ей, как и другим представляемым в ней информационным объектам, уникальный в рамках ее информационного пространства идентификатор. Этот идентификатор используется в ссылках на данную коллекцию из различных информационных объектов. Кроме того, он используется в качестве *префикса* к назначаемым создателем коллекции локальным уникальным внутри коллекции идентификаторам составляющих эту коллекцию информационных объектов. Вместе с таким префиксом полный идентификатор информационного объекта становится уникальным в системе.

3.3. Связи между информационными объектами. Как уже указывалось, информационные объекты и коллекции информационных объектов информационного пространства Соционет связаны между собой бинарными ориентированными связями вида «1:1» или «1:N». Действительно, организация связана с ее сотрудниками – авторами публикаций и с коллекциями публикаций, представленных в Соционет от ее имени. Организация связана с ее руководителями, как правило, авторами представленных в системе публикаций. Автор свя-

зан с его публикациями и с организацией – его местом работы. Публикация связана с ее авторами. Коллекция связана с входящими в ее состав публикациями. Связи могут существовать между версиями информационных объектов в системе – текущей и созданными ранее. В виде связей представляются библиографические ссылки одной публикации на другие.

Полная структура связей между информационными объектами в Соционет и процедура их создания детально описаны в (Создание сети связей между материалами в Соционет, 2006).

Связи в информационном пространстве Соционет реализуются в виде гиперссылок Веб. Составной частью URL целевого информационного объекта или коллекции в гиперссылке является при этом их уникальный системный идентификатор, о принципах назначения которого говорилось выше.

Поддержка связей между информационными объектами играет важную роль в Соционет. С помощью связей осуществляется верификация целостности структуры информационных ресурсов, принадлежащих отдельным институтам РАН. Связи между информационными объектами используются для навигации по структуре информационного пространства и для доступа к образующим его информационным объектам, в частности, и к полным текстам публикаций (полному контенту информационных объектов), если их электронные версии существуют и доступны в среде системы.

Ценным побочным эффектом поддержки связей между информационными объектами в Соционет является возможность формирования на основе текущего состояния ее информационного пространства онлайн-показателей, о которых шла речь выше. Благодаря поддержке связей между информационными объектами в Соционет можно измерять, в частности, различные агрегированные показатели.

Связи между информационными объектами и/или коллекциями выполняют и другую функцию. Они позволяют выстраивать в информационном пространстве системы полезные структуры информационных объектов, неоднородных в том смысле, что они принадлежат различным коллекциям, различным архивам и т.п.

3.4. Метаданные информационных объектов и коллекций. Каждый информационный объект и каждая коллекция информационных объектов снабжается в системе метаданными, описывающими их свойства. Эти метаданные в системной документации называются *карточкой информационного объекта* или *коллекции*. Карточка информационного объекта (коллекции) представляет собой совокупность значений атрибутов, принадлежащих некоторому набору. Для каждого типа информационных объектов и для коллекций предусматривается свой специфический набор атрибутов и собственный формат представления их значений в карточках, т.е. соответствующий этому типу информационных объектов тип карточки. Для

конкретного типа информационных объектов одни атрибуты являются обязательными, другие – факультативными.

Важно отметить, что карточки являются формой материализации описываемых ими информационных объектов и коллекций. Иначе говоря, карточки являются их фактическими представителями в системе. Идентификатор объекта или коллекции – это одновременно идентификатор соответствующей карточки.

Наряду с описанием свойств информационных объектов и коллекций их карточки выполняют еще одну важную задачу – они служат для явного представления связей между информационными объектами и/или коллекциями, которые обсуждались выше. Ссылки других информационных объектов на данный объект представляются в виде ссылок из их карточек на его карточку. При наличии доступного системе полного контента описываемого информационного объекта ссылка, по которой может быть осуществлен доступ к полному контенту, содержится в его карточке. Поэтому структура взаимосвязей информационных объектов и коллекций в информационном пространстве системы – это структура взаимосвязей описывающих их карточек.

Карточки информационных объектов имеют внутреннее системное и внешнее представление. Внутреннее представление карточек базируется на языке XML, в соответствии с инициативой OAI, стандарты и технологии которой поддерживает система Соционет. Вместе с тем, для сохранения преемственности и интероперабельности с системой RePEc, которая, как и Соционет, создавалась ранее объявления инициативы OAI, Соционет «понимает» также и формат метаданных ReDIF (Research Documents Information Format) (ReDIF version 1, 1996) системы RePEc.

Международная система RePEc поддерживает в настоящее время описания и доступ почти к полумиллиону англоязычных публикаций по экономике. Система Соционет создавалась в содружестве с разработчиками RePEc и является для нее одним из открытых архивов – источников информационных ресурсов. В свою очередь информационные ресурсы RePEc как множество ReDIF-архивов зарегистрированы в Соционет и доступны его пользователям. Именно этими обстоятельствами объясняется решение сохранить интероперабельность Соционет с RePEc.

Внешнее представление карточек информационных объектов и коллекций предназначено для пользователей, администраторов коллекций и системного администратора. Оно определяется принятым в системе форматом визуализации карточек каждого типа.

В системе Соционет предусматриваются, в частности, следующие типы карточек:

- карточки профиля организации;
- карточки коллекций;

- карточки профилей авторов публикаций;
- карточки публикаций.

Для каждого типа информационных объектов-публикаций (статья, монография, рабочий материал, научный отчет, автореферат диссертации, комментарий, новость и т.д.) предусматривается собственный тип карточек. Важным отличием карточек публикаций традиционного характера от карточек других типов информационных объектов является обязательное наличие в них библиографического описания публикации. Карточки организаций и карточки авторов называются в системной документации, соответственно, профилями организаций и профилями авторов публикаций. Поэтому коллекция информационных объектов представляется в системе как коллекция их карточек. Навигация по структуре информационного пространства – это навигация по связям, определенным в карточках. Поиск информационных объектов – это поиск их карточек.

Состав атрибутов для всех типов карточек детально описан в (Личная зона Соционет, 2001).

3.5. Интеграция информационных ресурсов в Соционет. Как уже указывалось, система Соционет использует принципы и стандарты технологии ОАИ и расширяет традиционные функциональные возможности базирующихся на них открытых архивов за счет развитых средств управления содержанием архива, а также средств создания связей между информационными объектами как внутри архива, так и с объектами из других архивов.

Инициатива ОАИ – подход к созданию распределенных электронных *репозиториев* информационных ресурсов, предложенный в Корнельском университете (США) (Open Archives Initiative Protocol, 2000). Ключевая концепция этого подхода – *открытый архив*. Открытый архив – это распределенная совокупность доступных в среде Интернет уникальным образом идентифицируемых информационных объектов, для которой поддерживается репозиторий описывающих их стандартизованных метаданных, представляемых средствами языка разметки XML. Репозитории метаданных открытых архивов должны быть доступны для авторизованных сборщиков метаданных, функционирующих в соответствии с разработанным в рамках инициативы ОАИ протоколом «The Open Archives Initiative Protocol for Metadata Harvesting» (ОАИ-РМН) (Open Archives Initiative Protocol, 2002). Для этого каждый репозиторий располагает механизмом доступа, способным выполнять по запросам сборщиков метаданных набор операций, определенных указанным протоколом.

Благодаря стандартизации представления метаданных в репозиториях и стандартизации протокола ОАИ-РМН, определяющего взаимодействия открытых архивов, обеспечивается интероперабельность открытых архивов. Согласно протоколу ОАИ-РМН для заданного открытого архива может осуществляться в соответствии с установленным регламентом сбор

метаданных из репозитория метаданных зарегистрированных в нем других открытых архивов, а также синхронизация состояния репозитория метаданных рассматриваемого открытого архива с изменениями состояний репозитория метаданных зарегистрированных в нем архивов-источников метаданных.

В результате каждый открытый архив выступает в качестве системы интеграции данных, основанной на материализованной интеграции метаданных из зарегистрированных в нем открытых архивов-источников. Описываемые метаданными информационные объекты-публикации при этом никак не затрагиваются, поскольку это система виртуальной интеграции. В интегрированном репозитории метаданных обеспечивается их уникальная идентификация. Однако их полные тексты сохраняются в прежних местах хранения и становятся доступными с помощью пользовательских сервисов интегрированного открытого архива по ссылке на него в описывающей этот объект карточке.

Организация системы Соционет следует описанным принципам. Интегрируемые открытые архивы-источники ее информационных ресурсов распределены и доступны в среде Веб. Некоторые локальные репозитории метаданных могут размещаться на серверах системы Соционет в так называемых Личных зонах. Личная зона – ресурс системной памяти, выделяемой авторизованному пользователю системы (автору публикации, помещаемой в систему или администратору создаваемых коллекций).

После создания коллекции информационных объектов в локальном репозитории администратор этой коллекции включает так называемый *режим трансляции*, после чего эта коллекция становится доступной для сборщика метаданных и включается в процедуру сбора метаданных, осуществляемую сервисами системы Соционет. В соответствии с установленным регламентом, системные сервисы осуществляют синхронизацию состояния метаданных в централизованном глобальном системном репозитории с состоянием их во всех транслируемых коллекциях локальных репозитория метаданных интегрируемых источников. Эта процедура осуществляется, как уже отмечалось, в соответствии с протоколом OAI-PMH.

Система Соционет с ее централизованным интегрированным репозиторием метаданных, в свою очередь, может служить источником информационных ресурсов для других систем, представляющих собой открытые архивы OAI. Она зарегистрирована, в частности, как источник данных в системе RePEc.

Являясь системой интеграции информационных ресурсов и одновременно открытым архивом OAI, Соционет выполняет следующие функции:

- регистрация открытых архивов – источников информационных ресурсов для Соционет, в качестве которых могут выступать, в частности, и локальные репозитории метаданных, базирующиеся на серверах системы Соционет (в Личных зонах);

- поддержка создания коллекций информационных объектов и описывающих их метаданных в Личных зонах администраторов коллекций;
- обеспечение уникальной идентификации информационных объектов в интегрированном репозитории метаданных системы;
- сбор метаданных из зарегистрированных в Соционет открытых архивов, синхронизация состояний репозитория метаданных системы и репозитория метаданных открытых архивов – источников информационных ресурсов;
- обеспечение для пользователей Соционет навигации по информационному пространству системы, просмотра метаданных информационных объектов и доступа к их полному контенту, если он доступен;
- поиск информационных объектов по заданным критериям с помощью поисковых сервисов Соционет.

Оценивая Соционет как систему интеграции информационных ресурсов множества открытых архивов, следует отметить, что она обладает вместе с тем дополнительными возможностями интеграции информационных ресурсов по сравнению с другими системами, основанными на технологиях ОАИ. Эти возможности обеспечиваются естественным образом как побочный эффект «обитания» Соционет в среде Веб. Способ интеграции информационных ресурсов, который имеется в виду, можно назвать *навигационной интеграцией*. Информационные объекты Соционет могут быть связаны с информационными объектами, доступными в Веб, но не описанными в Соционет и тем самым «не принадлежащими» ресурсам системы. Например, в карточке публикации может быть гиперссылка на страницу Веб, содержащую рецензию на эту публикацию. В свою очередь, на этой странице могут присутствовать ссылки на другие страницы Веб. Такие внешние (по отношению к ресурсам Соционет) связи позволяют пользователю осуществлять навигацию не только по структуре информационного пространства Соционет, но и «бесшовно» выходить за его пределы в среду Веб и возвращаться обратно.

4. МЕТРИКИ СОЦИОНЕТ

Как отмечалось ранее (см. п. 2.4), для обеспечения эффективного функционирования системы Соционет в нее необходимо встроить *сигнальную систему*, которая должна измерять ряд показателей, характеризующих свойства информационных ресурсов, процессы их создания и использования, в частности, и динамику этих процессов. В настоящее время ведутся работы по созданию сигнальной системы. Некоторые ее функции уже реализованы и доступны для использования (Статистика Соционет, 2007).

Автоматические измерения показателей сигнальной системы возможны благодаря

поддержке в Соционет метаданных, описывающих представленные в ней информационные объекты и развитую структуру связей между ними, а также журнализации доступа пользователей к информационным ресурсам системы.

Источниками исходных данных для измерений в системе Соционет являются ее репозиторий метаданных (характеристики информационных объектов, описываемых в карточках, связи между информационными объектами, даты помещения в репозиторий новых карточек, т.е. включения в архивы системы новых коллекций и включения в коллекции новых информационных объектов), а также журнал доступа пользователей к информационным ресурсам системы. При использовании данных журнала в нем обнаруживаются и исключаются искажающие смысл измеряемых показателей сведения о доступах индексаторов поисковых машин Веб к информационным ресурсам системы, сведения об обнаруживаемых многократных обращениях одного и того же пользователя к одному и тому же информационному объекту системы.

Рассмотрим подробнее предлагаемый состав измеряемых показателей. Заметим, что поскольку события включения новых информационных объектов в состав информационных ресурсов системы, а также операции доступа пользователей к информационным ресурсам регистрируются в системе ассоциированными с отметками времени, то для большинства рассматриваемых показателей могут быть построены динамические ряды.

4.1. Наукометрические показатели. В связи с массовым характером размещения научных статей и материалов в Веб получают распространение различные подходы к построению онлайн-наукометрических показателей. В западной научной литературе (см. например (Harnad, 2006, 2007; Shadbolt et al., 2006)) рассматриваются онлайн-показатели, построенные на основе числа просмотров аннотаций статей, скачивания из Веб их полных текстов, а также данных об их цитировании.

К числу наукометрических показателей, которые могут измеряться в системе Соционет, относятся показатели *востребованности публикаций* пользователями, *влияния* статей и материалов *на другие научные тексты* (электронное цитирование), показатели *притягательности* отдельных областей науки, а также показатели *активности публикации результатов* исследований.

Показатель востребованности публикации измеряется как сумма числа уникальных (по отношению к обращающимся пользователям) доступов к ее аннотации, а также и к ее полному тексту. Когда множество связей в системе Соционет, представляющих ссылки одних публикаций на другие, станет достаточно насыщенным и представительным, будет целесообразно оценивать этот показатель с учетом числа ссылок на данную публикацию в других публикациях без учета самоцитирования. Для этого можно, например, добавлять число ссы-

лок на данную публикацию к указанной выше сумме.

Оперируя показателем востребованности публикации, оцененным средствами системы Соционет, следует, конечно, иметь в виду относительный характер такой оценки. Эта оценка не учитывает внимания научного сообщества к версиям тех же публикаций, представленным в типографских изданиях, в других электронных источниках, ссылок на них из публикаций, не представленных в Соционет. Кроме того, в некоторых публикациях, представленных в Соционет, могут присутствовать ссылки на другие публикации, представленные в системе, но не описанные в их карточках.

Показатель востребованности может оцениваться не только для отдельной публикации, но и как агрегированный показатель для коллекций, для множества публикаций отдельного автора, публикаций сотрудников организации или группы организаций, публикаций в конкретном периодическом издании. Для сравнительной оценки востребованности публикаций организаций или различных групп организаций, наряду с абсолютными, могут измеряться и относительные значения этих показателей, равные их абсолютным значениям, нормированным по числу сотрудников этих организаций или групп организаций, публикации которых представлены в системе.

Показатель влияния представленной в системе публикации на другие публикации, имеющиеся в Соционет, может измеряться как значение функции от следующих аргументов:

- число уникальных (относительно пользователей) просмотров аннотации или полного текста публикации;
- число цитирований публикации в других публикациях, представленных в системе, исключая самоцитирование;
- показатели влияния авторов, которые цитируют данную публикацию в своих публикациях;
- импакт-фактор издания, в котором была опубликована данная работа.

Показатель влияния может измеряться для отдельной публикации, а также как агрегированный показатель для отдельных ученых, для организаций или групп организаций, для направления науки, идентифицируемого рубриками классификатора, по отношению к другим научным направлениям. Интересная возможность – вычленение тематических направлений активных исследований – открывается за счет использования кластеризации корпуса публикаций, посвященных данному научному направлению, на основе использования связей между ними по цитированию.

Соционет располагает техническими возможностями измерения показателя влияния. Однако пользоваться ими имеет смысл лишь после насыщения системы сведениями о публикациях с полноценным определением в их карточках библиографических ссылок на другие

публикации, представленные в Соционет.

Показатель притягательности данной области науки (идентифицируется рубриками классификатора) может измеряться как значение функции от следующих аргументов:

- число связанных с ней публикаций, представленных в системе;
- число доступов к публикациям в данной области;
- число цитирований работ, посвященных данной области, исключая самоцитирование.

Показатель активности (или интенсивности публикаций) отдельного ученого за рассматриваемый интервал времени может быть измерен числом его публикаций, пополнивших информационные ресурсы системы в данном интервале времени. Этот показатель может оцениваться и как агрегированный для организаций или групп организаций. Отнесенный к отдельным коллекциям или архивам, поддерживаемым системой, этот показатель характеризует одновременно динамику их пополнения. Для сопоставления организаций или групп организаций по показателю активности целесообразно использовать относительный показатель, равный абсолютному, нормированному по числу сотрудников организации, сведения о которых имеются в Соционет.

Представляют интерес *динамические ряды показателей активности*. Среди них весьма важным является показатель активности всего научного сообщества в данной области науки или в отдельных ее направлениях.

Показатели активности публикации результатов исследований имеют двойственный характер и могут интерпретироваться иным образом. Характеризуя активность исследователей и организаций, они, вместе с тем, одновременно характеризуют активность авторов и/или администраторов коллекций (чаще всего ответственных лиц, назначенных для выполнения этой функции руководством организаций), направленную на рост объема информационных ресурсов системы. Поэтому показатели активности, наряду с некоторыми другими рассматриваемыми здесь показателями, могут использоваться для ориентировочной оценки трудозатрат авторов публикаций или администраторов коллекций, связанных с представлением предлагаемых им ресурсов в Соционет.

Показатели активности предполагается использовать и по отношению к пользователям системы на основе статистики их запросов, распределенных во времени, которые регистрируются в системном журнале. Здесь также имеют смысл динамические ряды этих показателей. В Соционет отсутствует традиционный для многих аналогичных систем механизм регистрации пользователей, являющийся необходимым условием предоставления доступа к ресурсам системы. Соционет предоставляет пользователям свободный доступ к ее ресурсам без предварительной регистрации. Поэтому при формировании статистики по пользователям

их идентифицируют в системе по IP-адресам их компьютеров. Распознавание «мигрирующих» пользователей, т.е. пользователей, осуществляющих доступ к системе с разных компьютеров, или пользователей, которым коммуникационный провайдер предоставляет динамические IP-адреса (они различаются при разных подключениях), пока не реализовано.

Более подробное обсуждение проблемы измерения наукометрических показателей в Соционет, а также используемых для этой цели алгоритмов, можно найти в (Паринов, 2007). Системный сервис, измеряющий ряд описанных наукометрических показателей и генерирующий их динамические ряды, в настоящее время доступен пользователям (Статистика Соционет, 2007).

4.2. Показатели качества метаданных. Степень близости значений измеряемых в системе наукометрических показателей, рассмотренных в предыдущем разделе, к реальным значениям представляемых ими характеристик в существенной степени зависит от качества метаданных, описывающих информационные объекты и коллекции. Действительно, если например, в карточке, представленной в системе публикации, содержащей важные новые научные результаты в некоторой области исследований, будет отсутствовать аннотация, соответствующие рубрики классификатора и другие характеризующие содержание публикации атрибуты, то такое описание будет мало информативно и не будет способствовать высокой востребованности этой публикации, ее влиянию на развитие исследований в данной области.

В Соционет имеется возможность измерять, наряду с показателями научного качества публикаций, также и показатели качества их представления в системе, характеризующие качество метаданных, описывающих эти публикации. Такие показатели мы будем называть *презентационными*.

Рассматриваемые далее показатели качества метаданных относятся к каждой отдельной карточке и не имеют количественных значений. Эти показатели взаимосвязаны – улучшение одних способствует улучшению других. Наряду с этими качественными показателями к данной группе относятся и некоторые количественные статистические показатели.

Следует отметить, что презентационные показатели характеризуют не только качество метаданных, но и качество работы администраторов коллекций и архивов. Рассмотрим презентационные показатели подробнее.

Высокая *информативность метаданных*, описывающих публикацию, позволяет пользователю осуществлять предварительную оценку ее релевантности его информационным потребностям, не обращаясь к полному тексту публикации. Это позволяет исключить потери времени, связанные с извлечением из системы полного текста публикации и ознакомлением с ним в ситуациях, когда эта публикация, по оценке пользователя, не представляет для него интереса. Действительно, важную информацию о содержании публикации для

пользователя несут содержащаяся в карточке ее аннотация, соответствующие данной публикации рубрики классификатора, ключевые слова. Для оценки степени релевантности описываемой публикации используются также содержащиеся в карточке публикации ее библиографическое описание, сведения об организации – месте работы автора и другие сведения. Повышение информативности метаданных публикаций способствует повышению их доступности в системе и точности их поиска системными сервисами.

Повышению информативности метаданных способствует по возможности максимально полное заполнение карточек информационных объектов и коллекций. *Полнота метаданных*, представленных в карточке, является важным фактором обеспечения их информативности и тем самым доступности описываемой публикации, обеспечения измерения статистических оценок информационных ресурсов системы, о которых говорилось выше. Так, полное библиографическое описание публикации часто позволяет пользователю получить возможность познакомиться с ее полным текстом, даже если он не доступен в среде системы Соционет, – в типографском издании, а иногда и в электронном виде в других электронных библиотеках или на веб-сайтах, где она опубликована. Более полное указание связей данного информационного объекта с другими информационными объектами, представленными в системе, способствует повышению доступности данного объекта навигационным путем. Кроме того, от полноты определения указанных связей зависит степень точности измерения некоторых рассмотренных наукометрических показателей.

Достоверность метаданных существенно влияет на эффективность использования системы. Например, неточности библиографической ссылки в карточке публикации приводят к потере времени пользователем, пытающимся получить доступ к полному тексту этой публикации, если такая возможность отсутствует в Соционет.

Еще один презентационный показатель – *доступность описываемых информационных объектов*. В системе Соционет предусматривается три способа доступа к ее информационным ресурсам:

- с помощью непосредственного просмотра каталогов системы различных уровней, например, каталога нужной коллекции;
- путем навигации по структуре связей между карточками и между карточками и публикациями;
- с помощью поисковых сервисов системы;
- комбинированный способ доступа.

Первый способ не требует комментариев. Относительно *навигационной доступности* ясно, что чем полнее определены связи между информационными ресурсами системы, тем выше их навигационная доступность. Уровень *поисковой доступности* информационных

ресурсов в системе определяется степенью полноты определения в соответствующих им карточках значений их *ключевых атрибутов*. Например, возможности доступа с помощью поисковых сервисов открываются благодаря указанию в карточках, соответствующих описываемым публикациям, рубрик классификатора, названий организаций – мест работы авторов и других ключевых атрибутов. *Комбинированный способ доступа* сочетает использование поискового сервиса и навигационных механизмов. Уровень интенсивности использования этого способа зависит от обеспеченного уровня как поисковой, так и навигационной доступности ресурсов.

Показатель степени идентифицированности информационного объекта в Соционет характеризует, в какой мере полно в его карточке определены значения первичных и вторичных ключей данного информационного объекта – автора, названия, организации, рубрик классификатора и т.д. Значения этих ключей используются поисковыми сервисами системы и для поддержки различного рода связей между информационными объектами, а, следовательно, и для навигационного доступа. Поэтому степень идентифицированности информационного объекта непосредственно влияет на его навигационную и поисковую доступность.

Группа *статистических показателей качества метаданных* включает следующие показатели:

- долю полнотекстовых публикаций;
- долю рубрицированных публикаций;
- долю аннотированных публикаций;
- долю публикаций, снабженных ключевыми словами в их карточках;
- среднюю связность информационных объектов (среднее число связей на один информационный объект).

Перечисленные показатели могут оцениваться для отдельного информационного объекта, для коллекций, архивов и для репозитория метаданных системы в целом. Для всех этих случаев могут также строиться динамические ряды.

4.3. Соответствие метаданных стандартам. Заполняемые в карточки значения атрибутов должны соответствовать установленным в системе стандартам, касающимся формата и возможных значений этих атрибутов. Для значений некоторых используемых атрибутов существуют общепринятые официальные стандарты представления их значений. Это касается, например, библиографических описаний публикаций (Берлинская Декларация, 2003). Следование стандартам представления значений атрибутов метаданных в карточках способствует полноте и точности метаданных и обеспечивает корректное выполнение функций системы.

4.4. Перспективы развития измерений в системе. Адекватность отображения реальной ситуации в исследовательском сообществе с помощью измеряемых в системе Социо-

нет наукометрических показателей в существенной степени зависит от представительности того множества публикаций, которым располагает система, по отношению к их полному реальному множеству. Оптимизм авторов в этой области связан с рядом объявленных в последние годы международных инициатив, направленных на развитие открытого доступа к результатам научных исследований, на рост числа публикаций, предоставляемых для открытого доступа в разнообразных научных электронных библиотеках, созданных при поддержке РФФИ, на веб-сайтах исследовательских учреждений и в других репозиториях. Активную поддержку научной общественности встретили, в частности, Инициатива открытых архивов OAI (2000) (Open Archives Initiative, 2000), Будапештская инициатива открытого доступа (2001) (Budapest Open, 2002), Берлинская декларация об открытом доступе к научным и гуманитарным знаниям (2003) (Берлинская декларация, 2003; Berlin 3, 2005).

Реализация этих инициатив создаст представительную базу для измерений в Соционет и других научных системах, а также реальные возможности использовать измеряемые показатели для наукометрических целей, повышения эффективности научных исследований, обеспечения высокого качества онлайн-информационных ресурсов, представляющих результаты научных исследований, и повышение их доступности.

5. ОЦЕНКА ТРУДОЗАТРАТ НА ФОРМИРОВАНИЕ ИНФОРМАЦИОННЫХ РЕСУРСОВ

В деятельности, направленной на развитие информационных ресурсов системы Соционет, участвует в настоящее время большая группа специалистов. В 2006 году в работу по созданию архивов и коллекций информационных ресурсов системы, как уже отмечалось, были вовлечены представители более двадцати институтов Отделения общественных наук РАН.

Для эффективной организации развития информационного пространства, наряду с рассмотренными показателями качества информационных ресурсов Соционет, необходимо располагать также критериями оценки трудоемкости работ, необходимых для осуществления публикации информационных объектов в системе их автором либо администратором коллекций.

По нашему мнению, в качестве факторов такой оценки можно использовать:

- число коллекций, созданных за указанный промежуток времени данным лицом или группой лиц;
- число информационных объектов, созданных за указанный промежуток времени данным лицом или группой лиц;
- средняя связность созданных информационных объектов, равная частному от

деления числа определенных в рассматриваемом промежутке времени связей между информационными объектами рассматриваемого множества (отдельной коллекции, совокупности коллекций организации, архива системы и т.п.) на число этих объектов;

- доля аннотированных информационных объектов в рассматриваемом множестве объектов, созданном в данном промежутке времени;
- доля информационных объектов, снабженных ключевыми словами в их описаниях, в общем числе объектов, созданных за указанный период времени;
- доля рубрицированных по тематическому классификатору предметной области информационных объектов в общем числе объектов, созданных за указанный период времени.

Затраты времени, необходимые на создание заполненных карточек, описывающих создаваемые коллекции и информационные объекты, могут быть хронометрированы и учитываться в качестве весовых коэффициентов при общей оценке трудоемкости выполненных работ.

ЗАКЛЮЧЕНИЕ

Онлайновые информационные пространства, примером которых является Соционет, пополняются и развиваются в результате децентрализованных усилий большого числа людей. Такого рода системы не могут полноценно функционировать без набора метрик и встроенных сервисов, осуществляющих измерения в системе. Подобная сигнальная система обеспечивает автоматическое измерение и предоставление администраторам и пользователям значений системных метрик, которые измеряют и визуализируют два вида показателей: а) показателей использования информационных ресурсов (включая показатели востребованности, влияния и т.п.), информирующих пользователей, авторов и администраторов о реакции профессионального сообщества на их работу; б) показателей качества метаданных информационных ресурсов (презентационные показатели), которые необходимы в первую очередь разработчикам (администраторам) информационных ресурсов для определения возможных причин низкого интереса к их работе, а также для определения способов повышения качества метаданных.

СПИСОК ЛИТЕРАТУРЫ

1. **Алексеев А.В., Киселев А.В., Паринов С.И.** (1995): Анализ деловой активности в регионах России в 1994 г. Вопросы экономики, 3.
2. **Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М.** (2004): Возможности технологий ИСИР в поддержке Единого Научного Информационного Пространства РАН //Сб. докладов Шестой Всероссийской конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2004, Пущино.
3. Берлинская Декларация (2003): Берлинская Декларация об открытом доступе к научному и гуманитарному знанию (перевод Э.М. Мирского).
<http://informika.ru/text/magaz/newpaper/messedu/2003/cour0311/200.htm>
4. ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание.
<http://www.bookresearch.ru/download/71-2003.pdf>
5. **Крихель Т., Паринов С.** (2002): База данных RePEc и ее российский партнер система Соционет // *Электронные библиотеки*. Т. 1. Вып. 2.
<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2002/part2/КР>
6. Личная зона Соционет (2001): типы данных/коллекции электронных материалов.
<http://swb.socionet.ru/datatype.htm>.
7. **Паринов С.И.** (1999): Трафик сети Интернет как источник экономических данных», РАРДЭС. Май. <http://rvles.ieie.nsc.ru/parinov/net-data/intro.htm>
8. **Паринов С.И.** (2002): К теории сетевой экономики, Новосибирск, ИЭОПП СО РАН, <http://socionet.ru/publication.xml?h=repec:rus:ieieli:97421>
9. **Паринов С.И.** (2006): Программа «Открытый доступ к результатам исследований» в Отделении общественных наук РАН». Соционет, электронный депонент.
<http://socionet.ru/publication.xml?h=repec:rus:mqijxk:12&type=paper>
10. **Паринов С.И.** (2007): Онлайн-показатели результатов научно-исследовательской деятельности. <http://socionet.ru/publication.xml?h=repec:rus:mqijxk:16>
11. **Паринов С. И.** (2004): О тенденциях и ориентирах развития открытых электронных библиотек (из опыта создания и эксплуатации системы Соционет) // *Электронные библиотеки*. Вып.6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part6/sp>
12. Создание сети связей между материалами в Соционет (2006).
http://socionet.ru/docs/instr1/instruction_09.htm
13. Статистика Соционет (2007). <http://socionet.ru/stats.xml>
14. Berlin 3 (2005): Open Access: Progress in Implementing the Berlin Declaration on Open

Access to Knowledge in the Sciences and Humanities. Feb 28th – Mar 1st. University of Southampton, UK. <http://www.eprints.org/events/berlin3/outcomes.html>

15. Budapest Open Access Initiative (2002). <http://www.soros.org/openaccess/read.shtml>

16. **Harnad St.** (2006): Online, Continuous, Metrics-Based Research Assessment. Future UK Research Assessment Exercise (RAE) to be Metrics-Based.

<http://eprints.ecs.soton.ac.uk/12130/01/rae-metric.html>

17. **Harnad St.** (2007): Open Access Scientometrics and the UK Research Assessment Exercise. Preprint of Invited Keynote Address to 11th Annual Meeting of the International Society for Scientometrics and Informetrics. Madrid, Spain. 25–27 June.

<http://eprints.ecs.soton.ac.uk/13804/03/oa-scientometrics.pdf>

18. Open Archives Initiative (2000): <http://www.openarchives.org/>

19. ReDIF version 1 (1996). http://openlib.org/acmes/root/docu/redif_1.html

20. RePEc (1997). <http://repec.org/>

21. **Shadbolt N., Brody T., Carr L., Harnad St.** (2006): The Open Research Web: A Preview of the Optimal and the Inevitable. – <http://eprints.ecs.soton.ac.uk/12453/>

22. The Open Archives (2002): The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Protocol Version 2.0 of 2002-06-14. Document Version 2003/02/21T00:00:00Z. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>