

Информационные ресурсы, наукометрические показатели и показатели качества метаданных системы Соционет*

© М.Р. Когаловский

Институт проблем рынка РАН
kogalov@cemi.rssi.ru

С.И. Паринов

Центральный экономико-математический
институт РАН
sparinov@hse.ru

Аннотация

Система Соционет является первой и в настоящее время единственной в нашей стране глобальной информационной научно-образовательной социальной сетью в области общественных наук. Система предоставляет свободный доступ пользователей к ее информационным ресурсам, интегрирована в международную научную информационную сетевую инфраструктуру, является одним из источников информационных ресурсов для разрабатываемого Единого научного информационного пространства РАН. Соционет разрабатывалась и развивается в рамках международных инициатив RePEc (Research Papers in Economics) и OAI (Open Archives Initiative), поддерживает их соглашения и стандарты. Наряду с этим в системе поддерживается сложная структура связей между информационными ресурсами. Благодаря более развитой архитектуре информационных ресурсов система обеспечивает новые функциональные возможности для доступа пользователей, появляются возможности для расширенных наукометрических измерений и оценки показателей качества поддерживаемых метаданных. В предлагаемой работе приводятся общие характеристики Соционет, рассматриваются организация и методы интеграции ее информационных ресурсов, обсуждаются возможности наукометрических измерений в системе, а также показатели качества системных метаданных.

1 Система Соционет и ее функции

Система Соционет является первой и пока единственной в нашей стране глобальной информацион-

Труды 9^{ой} Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2007, Переславль Залесский, Россия, 2007.

ной научно-образовательной социальной сетью в области общественных наук (<http://socionet.ru/>). Ее ресурсы свободно доступны пользователям Интернет с 2000 г. Все эти годы система функционировала и развивалась в ИЭОПП СО РАН благодаря грантам Фонда Форда. Название системы можно рассматривать как отражение ее связи с концепцией профессиональной социальной сети. Напомним, что социальная сеть рассматривается как социальная структура, состоящая из множества узлов - индивидумов или организаций, которые находятся между собой в отношениях одного или более типов (см., например [10]).

Соционет интегрирована с Единым научным информационным пространством РАН (ЕНИП РАН) [11], являясь для него источником информационных ресурсов. Она интегрирована также в международную научную информационную сетевую инфраструктуру благодаря тому, что ее разработка выполнялась в рамках международных инициатив RePEc (Research Papers in Economics) [7, 14] и Open Archives Initiative (OAI) [5], соглашения и стандарты которых поддерживаются системой.

Соционет развивается благодаря усилиям 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 – как правило, институтов РАН), которые зарегистрированы и авторизованы системным администратором. Они выполняют свою миссию автономно от других участников этой деятельности. Пользователям системы предоставляется свободный доступ к ее информационным ресурсам и сервисам. Для этого достаточно иметь доступ к Интернет.

Важно отметить здесь, что политика представления собственных информационных ресурсов в системе определяется каждым из их владельцев самостоятельно и имеет, как правило, институциональную основу.

Указанный выше статус системы Соционет определяет ее задачи, главными из которых являются:

- накопление метаданных, описывающих научные публикации и другие научные информацион-

ные ресурсы (научные отчеты о проведенных исследованиях, рабочие записки, тезисы докладов, статистические таблицы или базы данных, материалы научных проектов в социально-экономической области общественных наук, сведения о научных организациях, научная персоналия и т.д.)

- формирование развитой структуры связей между объектами метаданных, поступивших в систему из различных источников и описывающих ее информационные ресурсы, 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.

При этом архитектура информационных ресурсов Соционет обеспечивает возможность функционирования ряда системных сервисов, которые позволяют:

- автоматически измерять на основе поддерживаемых в системе метаданных и данных, регистрируемых в системных журналах в процессе ее функционирования (лог-файлы) различные полезные наукометрические показатели

- измерять некоторые показатели качества метаданных, описывающих публикуемые в системе научные информационные ресурсы

- формировать и предоставлять пользователям динамические ряды указанных показателей.

Измеряемые наукометрические показатели позволяют получать некоторые оценки качества представленного в системе научного контента (актуальность его тематики, востребованность представляемых идей и результатов научным сообществом, степень его влияния на развитие соответствующей ветви науки и т.п.). В то же время показатели качества метаданных характеризуют уровень обеспечения в системе возможностей эффективного восприятия пользователями представленного в ней научного контента.

Некоторые из указанных сервисов уже функционируют в системе Соционет, другие находятся в стадии разработки.

Важно отметить, что мы имеем здесь дело с электронными информационными ресурсами, предоставленными пользователям для свободного доступа в интерактивном (on-line, онлайн) режиме с помощью системных механизмов. Поэтому измеряемые показатели их качества принято называть *онлайн-показателями*. Их оценка в автоматизированном режиме с помощью специальных сервисов системы Соционет не вызывает принципиальных трудностей. Однако проблематичными являются

попытки практического использования некоторых из этих показателей в научно-организационной деятельности в связи с факторами психологического и социального характера.

Обсуждение онлайн-показателей качества информационных ресурсов системы Соционет является целью данного документа. Поскольку предлагаемые здесь показатели могут оцениваться автоматизированным образом благодаря особенностям информационного пространства системы и архитектуры ее информационных ресурсов, необходимо, прежде всего, кратко рассмотреть эти свойства системы. Более детальное описание системы Соционет можно найти в [16].

Остальная часть данной работы построена следующим образом. В разделе 2 обсуждаются организация информационного пространства системы, представленные в нем информационные объекты и описывающие их метаданные, структура взаимосвязей между информационными ресурсами, используемые в системе методы интеграции информационных ресурсов из множества распределенных источников. В разделе 3 обсуждаются возможности проведения в системе измерений наукометрического характера, а также измерений качества метаданных, определяющих, в свою очередь, качество представления информационных ресурсов для пользователей. Поэтому показатели качества метаданных мы называем презентационными. Наконец, в заключении подводятся итоги работы и характеризуются перспективы развития наукометрического аппарата системы, связанные с выдвигаемыми в последние годы международным научным сообществом инициативами, направленными на обеспечение открытого доступа к результатам исследований.

2 Информационное пространство системы

Информационное пространство системы Соционет поддерживается в среде Всемирной паутины и реализовано с использованием веб-технологий, в частности, технологий XML. Оно включает совокупность взаимосвязанных информационных объектов и коллекций информационных объектов, метаданных, описывающих их свойства, а также набор сервисов, которые обеспечивают их создание авторами либо администраторами информационных ресурсов, поддержку и доступ к ним пользователей, поддержку деятельности персонала системного администратора, измерение показателей качества информационных ресурсов системы, интеграцию Соционет с отечественной и международной научной информационной сетевой инфраструктурой.

Рассмотрим состав и особенности организации информационных ресурсов системы, представляющих собой, как указывалось выше, совокупность взаимосвязанных информационных объектов и коллекций информационных объектов, а также описывающих их свойства метаданных.

2.1 Информационные объекты

Основу научного контента системы Соционет составляют информационные объекты. *Информационным объектом* (или точнее, экземпляром информационного объекта) называют представление идентифицируемой сущности предметной области информационной системы с помощью ее информационных ресурсов. Свойства сущностей предметной области представляются атрибутами соответствующих информационных объектов.

В Соционет предусматривается несколько типов информационных объектов, которые представляют организации, поставляющие свои информационные ресурсы системе, публикации различного рода и их авторов, а также коллекции публикаций.

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

Используемый в данном контексте термин *публикация* обозначает любой электронный научный контент в различных возможных формах – электронные версии статей, опубликованных в периодике, полных текстов или фрагментов опубликованных монографий, полные тексты или тезисы докладов, опубликованных в трудах конференций, препринты, аннотации 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, а также презентации докладов, веб-публикации, аудио или видеоресурсы, научные веб-сервисы, страницы веб-сайтов, доступные в Веб системы баз данных, файлы FTP-архивов и т.п.

Многие включаемые в настоящее время в информационное пространство системы Соционет публикации доступны в среде Веб. Для полных текстов таких публикаций существуют их электронные версии, представленные в веб-версиях научных журналов, на веб-сайтах организаций, в электронных библиотеках и в других источниках, доступных в среде Веб. Электронные версии иных публикаций могут создаваться и поддерживаться их авторами либо веб-мастерами сайтов или администраторами репозитория информационных ресурсов 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 этих институтов, отдельных файлов различных форматов, содержащих такие их информационные объекты, как аннотации или полные тексты научных работ, фотографии авторов публикаций или руководителей организаций и др., а также предоставлять доступ к системным сервисам,

обеспечивающим создание этих информационных объектов и/или описывающих их метаданных.

Авторы или организации – обладатели авторских прав на публикацию, самостоятельно принимают решение о публикации 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 ответственными за эту функцию персоналом институтов РАН. В силу этого персонал Соционет не несет ответственности за соблюдение 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 относятся гиперссылка в формате URL, идентифицирующая местоположение ее полного текста в Интернет, или соответствующие ей рубрики классификатора. К числу атрибутов организации относятся URL домашней страницы ее веб-сайта. Важным обязательным атрибутом информационного объекта каждого типа является его системный идентификатор. Технологии Соционет позволяют назначить каждому представленному в ней информационному объекту уникальный идентификатор, используемый в системе.

Информационные объекты некоторых типов, таких как организации, авторы, коллекции, полностью представляются описывающими их метаданными – наборами значений их атрибутов. Информационные объекты–публикации наряду с этим располагают еще полным текстом (полным контентом). Однако не для каждой публикации полный ее текст оказывается доступным пользователям системы, хотя это весьма желательно. Как уже указывалось, автор публикации или организация могут проводить такую информационную политику относительно своих информационных ресурсов, в соответствии с которой полные тексты некоторых принадлежащих им публикаций не предоставляются для свободного доступа. Электронная версия полного ее контента может вообще не существовать. Однако система позволяет пользователю, по крайней мере, выявить факт ее существования и получить полную библиографическую ссылку, аннотацию и/или некоторые другие характеристики данной публикации.

Одним из типов информационных объектов, поддерживаемых в системе, являются научные классификаторы. Коллекции объектов этого типа по функциональным возможностям не отличаются от других коллекций. Вместе с тем, классификаторы служат в Соционет для рубрикации информационных объектов других типов, и соответствующие рубрики могут использоваться для их поиска. В Со-

ционет поддерживаются международный классификатор JEL (*Journal of Economic Literature Classification System*) для рубрикации экономических информационных объектов и ГРНТИ для объектов в других областях общественных наук.

2.2 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, хронологической или тематической их общности, общности места работы авторов представленных в системе публикаций и т.п. Автор публикаций, самостоятельно создающий и помещающий коллекцию публикаций в Соционет, или его доверенное лицо, выполняющие 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 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 как префикс 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 информационном пространстве системы.

2.3 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, каждая 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 публикаций. Каждый автор связан 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 объек-

тов в системе - текущей и созданными ранее. С помощью связей поддерживаются библиографические ссылки одной публикации на другие.

Полная структура связей между информационными объектами в Соционет и процедура их создания детально описаны в материале [18].

Поддержка связей между информационными объектами играет важную роль в системе 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 ее информационным объектам, в частности, и к полным текстам публикаций (полному контенту информационных объектов), если их электронные версии существуют и доступны в среде системы. Кроме того, ценным побочным эффектом поддержки связей между информационными объектами в системе является возможность формирования на основе текущего состояния ее информационного пространства полезных статистических показателей различного рода, в том числе наукометрического характера, характеризующих качество научного контента информационного пространства системы и описывающих его метаданных. Благодаря поддержке связей между информационными объектами в Соционет можно измерять агрегированные наукометрические показатели (см. разд. 3.1).

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.

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.

2.4 Метаданные информационных объектов и коллекций

Каждый информационный объект и каждая коллекция информационных объектов представляется в системе Соционет их описателем, который в системной документации называется карточкой этого объекта или коллекции. *Карточка информационного объекта* – это визуализированная совокупность значений его атрибутов из некоторого набора метаданных, описывающих объекты данного типа. Для каждого типа информационных объектов предусматривается специфический набор атрибутов метаданных и собственный формат представления их значений в карточках, т.е. соответствующий этому типу информационных объектов *тип карточки метаданных*. Некоторые атрибуты метаданных являются обязательными для конкретного типа инфор-

мационных объектов, другие – факультативными.

Важно здесь отметить, что карточки метаданных являются формой материализации описываемых ими информационных объектов. Иначе говоря, карточка метаданных информационного объекта является фактически его *представителем* в системе. Идентификатор объекта – это одновременно идентификатор его карточки. Ссылки других информационных объектов на данный объект представляются в виде ссылок на его карточку. При наличии доступного системе полного контента описываемого информационного объекта ссылка, по которой может быть осуществлен доступ к полному контенту, имеется 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 метаданных в карточках соответствует стандарту языка разметки 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) [6] системы RePEc.

Международная система RePEc поддерживает в настоящее время описание 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 организаций зарегистрированы в Соционет и доступны его пользователям. Именно этими обстоятельствами объясняется решение о сохранении интероперабельности Соционет с RePEc.

Внешнее представление карточек информационных объектов для пользователей, администраторов коллекций и системного администратора определяется принятым разработчиками системы форматом визуализации карточек для каждого типа информационных объектов.

Наряду с описанием свойств соответствующих информационных объектов метаданных карточки выполняют еще одну важную задачу – они служат для явного представления связей между информационными объектами и/или коллекциями информационных объектов, которые обсуждались выше.

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

- карточка профиля организации
- карточка коллекции
- карточка профиля автора публикаций
- карточки публикаций – свой тип карточек для каждого типа информационных объектов публика-

ций (статья, монография, рабочий материал, научный отчет, автореферат диссертации, комментарий, новость...).

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 и в терминологии пользовательского интерфейса системы, соответственно, профилями организаций и профилями авторов публикаций.

В соответствии с указанным, коллекция информационных объектов представляется в системе как коллекция их карточек. Навигация по структуре информационного пространства – это навигация по связям, определенным в карточках метаданных. Поиск информационных объектов – это поиск их карточек.

Состав атрибутов всех типов карточек описан в документации системы Соционет [15], и поэтому он здесь не рассматривается.

2.5 Интеграция информационных ресурсов в системе

Как уже указывалось ранее, система Соционет использует принципы и стандарты технологии OAI 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 OAI – это подход к созданию распределенных электронных репозиториев информационных ресурсов, предложенный в Корнельском университете (США). Ключевой для этого подхода является концепция *открытого архива*. 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 OAI протоколом “The Open Archives Initiative Protocol for Metadata Harvesting” (OAI-PMH) [9]. Для этого каждый репозиторий располагает механизмом доступа, способным выполнять по запросам сборщиков метаданных набор операций, определенных протоколом OAI-PMH.

Благодаря стандартизации представления метаданных в репозиториях и стандартизации протокола OAI-PMH, определяющего взаимодействия открытых архивов, обеспечивается интероперабельность открытых архивов. Согласно протоколу OAI-PMH, для заданного открытого архива может осуществляться сбор метаданных из репозиториев метаданных зарегистрированных в нем других открытых

архивов в соответствии с установленным регламентом, а также синхронизация состояния репозитория метаданных рассматриваемого открытого архива с изменениями состояний репозитория зарегистрированных в нем источников-источников метаданных.

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

Организация Соционет следует описанным принципам. Интегрируемые открытые архивы - источники ее информационных ресурсов - распределены в Веб и доступны в этой среде. Некоторые локальные репозитории метаданных могут размещаться на серверах самой системы в так называемых Личных зонах.

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 используется, в частности, как источник данных системой RePEc.

В качестве системы интеграции информационных ресурсов и открытого архива OAI система Соционет выполняет следующие функции:

- регистрация открытых архивов - источников информационных ресурсов для системы 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 Личной зоне администратора коллекции
- обеспечение уникальной идентификации информационных объектов в интегрированном репо-

зитории метаданных системы

- сбор метаданных из зарегистрированных открытых архивов и 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 поисковыми сервисами Соционет.

Оценивая Соционет как систему интеграции информационных ресурсов множества открытых архивов, нужно отметить, что она обладает дополнительными возможностями интеграции информационных ресурсов по сравнению с другими системами, основанными на технологиях OAI. 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 метаданных публикации может быть гиперссылка на страницу Веб, содержащую рецензию на эту публикацию. В свою очередь, на этой странице могут иметься 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.

3. Измерения в системе Соционет

Благодаря поддержке в системе Соционет метаданных, описывающих представленные в ней информационные объекты и развитую структуру связей между ними, а также журнализации доступа пользователей к информационным ресурсам системы, имеется возможность проводить в системе автоматические измерения ряда показателей, характеризующих свойства информационных объектов, отношений между ними, отношений между пользователями и информационными объектами.

К числу таких показателей относятся показатели наукометрического характера, характеризующие на институциональном и персональном уровне деятельность научного сообщества, объединяемого системой как социальной сетью, активность исследований в отдельных научных направлениях и востребованность полученных результатов, притягательность тех или иных направлений исследований. Наконец, имеется возможность измерять показатели качества представления информационных объектов в системе (их идентифицируемость и доступность,

информативность представления и др.), которые можно назвать презентационными показателями информационных ресурсов.

Межинститутской рабочей группой Соционет, учрежденной в Отделении общественных наук РАН, разрабатывается система измеряемых показателей. Как указывалось выше, источниками исходных данных для измерений в системе Соционет являются ее репозиторий метаданных (характеристики информационных объектов, описываемых в карточках метаданных, 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 индексов поисковых машин Веб к информационным ресурсам системы, обнаруживаемые многократные обращения одного и того же пользователя к одним и тем же информационным ресурсам системы.

Рассмотрим предлагаемый состав измеряемых показателей. Некоторые из них в настоящее время уже измеряются в системе, для измерения других разрабатываются системные сервисы. 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, для большинства рассматриваемых показателей могут быть построены динамические ряды.

3.1 Наукометрические показатели

В связи с массовым характером размещения научных статей и материалов в Веб получают все большее распространение различные подходы к построению онлайн-наукометрических показателей. В западной научной литературе (см. например, [3, 4, 8]) рассматриваются онлайн-показатели, построенные на основе количества просмотров аннотаций статей, загрузок их полных текстов, а также данных об их цитировании.

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

Показатели востребованности. Показатель востребованности статей измеряется как сумма количества уникальных (по отношению к обращающимся пользователям) доступов к аннотации статьи, а также и к ее полному тексту. После того, как множество связей в системе Соционет, представляющих ссылки одних статей на другие, станет достаточно насыщенным и представительным, будет

целесообразно оценивать этот показатель с учетом количества ссылок на данную статью в других статьях и материалах без учета самоцитирования. Для этого можно, например, добавлять количество ссылок на данную статью к указанной выше сумме.

Оперируя показателем востребованности статьи, оцененным средствами системы Соционет, следует, конечно, иметь в виду относительный характер такой оценки. Эта оценка не учитывает внимания научного сообщества, проявляемого 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 иметься ссылки на другие публикации, представленные в системе, но они не описаны в их карточках метаданных.

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 абсолютным их значениям, нормированным по числу сотрудников этих организаций или групп организаций, публикации которых представлены в системе.

Показатели влияния. Показатель влияния представленной в системе публикации на другие опубликованные результаты исследований может измеряться как функция от следующих аргументов:

- количества уникальных (относительно пользователей) просмотров аннотации или полного текста публикации
- количества цитирований публикации в других публикациях, представленных в системе (в карточках, описывающих публикации, предусматривается задание списка ссылок на цитируемые в них другие публикации), исключая самоцитирование
- показателей влияния авторов, которые цитируют данную публикацию в своих публикациях
- импакт-фактора издания, в котором была опубликована работа.

Показатели влияния могут измеряться для отдельной публикации, а также как агрегированные показатели 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 статей, посвященных данному научному направлению, на основе использования связей между ними по цитированию.

Соционет располагает техническими возможностями измерения показателей влияния. Однако они

могут быть полезными для практического использования лишь после насыщения системы информационными ресурсами и полноценного определения библиографических ссылок в описаниях публикаций на другие публикации, представленные в Соционет.

Показатель притягательности. *Показатель притягательности* данной области науки (идентифицируется рубриками классификатора) для исследований измеряется как функция от следующих аргументов:

- количество связанных с ней публикаций, помещенных в систему
- количество доступов к публикациям в данной области
- количество цитирований публикаций, относящихся к данной области, исключая самоцитирование.

Показатели активности. *Показатель активности* (или *интенсивности публикаций*) отдельного ученого в рассматриваемом интервале времени измеряется как количество его публикаций, пополнивших информационные ресурсы системы в данном интервале времени. Этот показатель можно агрегировать для организаций (учитываются все публикации сотрудников организации, представленные в системе) или для групп организаций. Отнесенный к отдельным коллекциям или архивам, поддерживаемым системой, он характеризует динамику пополнения их информационных ресурсов.

Для сравнительного 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 также 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.

Показатели активности предполагается использовать и по отношению к пользователям системы на основе статистики их запросов, распределенных во времени. Здесь также имеют смысл динамические ряды измерений этих показателей.

Более подробное обсуждение проблемы измерения наукометрических показателей в Соционет, а также используемых для этой цели алгоритмов, можно найти в рабочей записке [7].

3.2. Показатели качества метаданных

Степень близости значений измеряемых в системе наукометрических показателей, обсуждаемых в предыдущем разделе, к реальным значениям представляемых ими характеристик, в существенной степени зависит от *качества метаданных*, их описывающих. Действительно, если, например, в карточке представленной в системе публикации фундаментального характера, содержащей весьма важные новые научные результаты в некоторой области исследований, будут отсутствовать аннотация, соответствующие рубрики классификатора и другие характеризующие содержание публикации атрибуты, то такое описание будет мало информативно. Оно не будет способствовать высокой востребованности этой публикации и тем самым 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 метаданных, и они не имеют количественных значений. Эти показатели взаимосвязаны – улучшение одних способствует улучшению других. Наряду с этими качественными показателями к данной группе относятся некоторые количественные статистические показатели.

Нужно отметить, что презентационные показатели характеризуют не только качество метаданных, но и, с другой стороны, качество работы администраторов коллекций и архивов.

Рассмотрим презентационные показатели подробнее.

Информативность метаданных. Высокая информативность метаданных, описывающих публикацию, позволяет пользователю осуществлять предварительную оценку ее релевантности собственным информационным потребностям, не обращаясь к полному тексту публикации. Повышению информативности метаданных способствует по возможности максимально полное заполнение значений атрибутов информационных объектов в их карточки метаданных.

При работе с системой пользователь осуществляет предварительную оценку релевантности рассматриваемой публикации его информационным потребностям на основе тех метаданных, которые содержатся в описывающей ее карточке. Это позволяет исключить потери времени, связанные с необходимостью извлечения из системы полного текста публикации и ознакомления 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 метаданных 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 повышению точности их поиска системными сервисами.

Полнота метаданных. 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 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 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 наукометрических показателей, рассмотренных ранее.

Точность метаданных. Недостаточная точность метаданных не позволяет эффективно их использовать, зачастую вводит в заблуждение пользователя.

Доступность описываемых информационных объектов. В системе Соционет предусматривается три способа доступа к ее информационным ресурсам:

- прямой просмотр каталогов системы различных уровней, например, каталога нужной коллекции
- навигация по структуре связей между карточками метаданных 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. Поисковая доступность информационных ресурсов в системе определяется степенью полноты определения в карточках метаданных 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 системы. Степень интенсивности использования этого способа доступа к информационным ресурсам

системы зависит от уровня как поисковой, так и навигационной доступности ресурсов.

Показатель степени идентифицированности информационных объектов. Степень идентифицированности информационных объектов в системе Соционет зависит от того, в какой мере полно в их карточках метаданных 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 связей различного рода между информационными объектами. Поэтому степень идентифицированности информационного объекта непосредственно влияет на его навигационную и поисковую доступность.

Соответствие метаданных стандартам. Заполняемые в карточки метаданных 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 [13]. Следование стандартам представления значений атрибутов метаданных в карточках метаданных способствует полноте и точности метаданных, обеспечивает корректное выполнение функций системы.

Статистические показатели качества метаданных. Эта группа включает следующие показатели:

- доля полнотекстовых публикаций
- доля рубрицированных публикаций
- доля аннотированных публикаций
- доля публикаций, снабженных ключевыми словами в карточках метаданных
- связность информационных объектов - среднее количество 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.

Заключение

Адекватность отображения реальной ситуации в исследовательском сообществе с помощью измеряемых в системе Соционет наукометрических показателей, конечно же, в существенной степени зависит от представительности того множества публикаций, которым располагает система по отношению к полному их реальному множеству. Оптимистические надежды здесь связаны с рядом объявленных в последние годы международных инициатив, направленных на развитие открытого доступа к результатам исследований, на рост количества 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) [5], Будапештская инициатива открытого доступа (2001) [2], Берлинская декларация об открытом доступе к научным и гуманитарным знаниям (2003) [1, 12].

Реализация указанных инициатив создаст реальные возможности использования показателей, измеряемых в Соционет и в других функционирующих в среде Веб научных системах, для наукометрических целей, для стимулирования более высокой эффективности научных исследований, обеспечения высокого качества онлайн-информационных ресурсов, представляющих результаты научных исследований, и улучшения их доступности.

Литература

1. Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Feb 28th - Mar 1st, 2005, University of Southampton, UK. <http://www.eprints.org/events/berlin3/outcomes.html>
2. Budapest Open Access Initiative. <http://www.soros.org/openaccess/read.shtml>
3. Harnad Stevan. Online, Continuous, Metrics-Based Research Assessment. Future UK Research Assessment Exercise (RAE) to be Metrics-Based. - 2006. <http://eprints.ecs.soton.ac.uk/12130/01/rae-metric.html>
4. Harnad Stevan. 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, 2007. <http://eprints.ecs.soton.ac.uk/13804/03/oa-scientometrics.pdf>
5. Open Archives Initiative, 2000. <http://www.openarchives.org/>
6. ReDIF version 1. http://openlib.org/acmes/root/docu/redif_1.html
7. RePEc. <http://repec.org/>
8. Shadbolt, N., Brody, T., Carr, L. and Harnad, S. The Open Research Web: A Preview of the Optimal and the Inevitable. - 2006. <http://eprints.ecs.soton.ac.uk/12453/>
9. 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>
10. Wikipedia. The Free Encyclopedia. <http://en.wikipedia.org/wiki/>
11. Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М. Возможности технологий ИСИР в поддержке Единого информационного пространства РАН //Сб. докладов. 6-й Всероссийской научной конф. "Электронные библиотеки: перспективные методы и тех-

нологии, электронные коллекции" - RCDL'2004, Пушкино, 2004.

12. 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>

13. ГОСТ 7.1-2003. Библиографическая запись. Библиографическое описание. <http://www.bookresearch.ru/download/71-2003.pdf>

14. Крихель Т., Паринов С. База данных RePEc и ее российский партнер система Соционет. Электронные библиотеки. – 2002. – Т. 1, вып. 2. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2002/part2/КР>

15. Личная зона Соционет: типы данных / коллекции электронных материалов. <http://swb.socionet.ru/datatype.htm>

16. Паринов С.И. О тенденциях и ориентирах развития открытых электронных библиотек (из опыта создания и эксплуатации системы Соционет). Электронные библиотеки. - 2004. - Вып.6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2004/part6/sp>

17. Паринов С.И. Онлайн-показатели результатов научно-исследовательской деятельности. <http://socionet.ru/publication.xml?h=репес:rus:mqijxk:16>

18. Создание сети связей между материалами в Соционет. http://socionet.ru/docs/instr1/instruction_09.htm

Socionet Information Resources, Scientometrics and Metadata Quality Indicators

M.R. Kogalovsky, S.I. Parinov

The Socionet system is the first Russian professional social network for education and research areas. It is also a platform to build e-Social Sciences infrastructure at national level. Socionet information resources are integrated into national Common Scientific Information Space of Russian Academy of Sciences and have relations with international e-Science online infrastructure. The Socionet system used some ideas, methods and tools of RePEc (Research Papers in Economics) project and OAI (Open Archives Initiative). An added value of the Socionet is advanced architecture of relations among different types of information objects, which allows better navigation, research performance indexes, scientometric analysis and metadata quality indicators. The proposed article describes general features of the Socionet and its methods of information resources integration. The article also discusses possible implementations of scientometrics based on Socionet statistics and a problem of metadata quality measurements.

* Работа поддерживается грантами РГНФ 06-02-12205-в и Фонда Форда № 1055-0505.