

Таксономия семантических связей информационных объектов системы Соционет

М.Р. Когаловский, ИПР РАН

С.И. Паринов, ЦЭМИ РАН

Депонирована в Соционет 31.10.2014

Новая версия от 07.11.2014

Обсуждается таксономия семантических связей, используемая в системе Соционет, рассматривается ее структура и обосновывается состав предусмотренных в ней классов, рассматривается организация таксономии, представленной в системе в виде набора контролируемых словарей классов связей. Описаны способы создания связей в системе и возможности, которые обеспечивает для пользователей система с семантически структурированным контентом. Работа поддержана грантом РГНФ 14-02-12010-в.

Ключевые слова: *таксономия, семантическая связь, контролируемый словарь*

1. Введение

В последние годы в области научных электронных библиотек активно развиваются исследования и разработки, направленные на обогащение их контента семантической информацией. Благодаря этому может быть повышена информативность, а, следовательно, и ценность представленных в этих системах информационных ресурсов. Помимо традиционного доступа непосредственно к информационным объектам библиотеки, становится возможным решение новых важных задач, открываются новые возможности для использования систем этого класса.

Один из активно развивающихся подходов к семантическому обогащению информационных ресурсов научных электронных библиотек, на котором основаны и работы авторов данной статьи, заключается в декларации явным образом семантических связей между информационными объектами контента библиотеки. Такие связи могут нести различную информацию. Они могут представлять отношения научного характера между публикациями, в частности, мотивы связей цитирования, отношения, характеризующие влияние одной публикации на другую или использование в ней обсуждаемых в другой результатов, оценку результатов одной публикации в тексте другой, структурные связи между компонентами или различными версиями научного произведения, отношения авторства между персонами и публикациями, вклад каждого из авторов в создании коллективных публикаций и др.

Исследования, обсуждаемые в данной статье, используют в качестве полигона крупную отечественную электронную библиотеку Соционет [1], функционирующую уже около 15 лет. Соционет основана на технологии открытых архивов, импортирует метаданные из репозитория метаданных более 1500 источников, включая сотни отечественных и зарубежных открытых архивов. Предоставляет метаданные собственного репозитория метаданных для импорта другим системам. В настоящее время в Соционет доступно около 2.5 млн. информационным объектам, содержатся описания более 8 миллионов семантических связей.

В Соционет допускаются только *бинарные ориентированные семантические связи* между ее информационными объектами. Участник связей, из которого исходит связь, называется *исходным* ее информационным объектом. Второй участник, на который направлена связь, называется *целевым* объектом этой связи.

Семантические связи представляются в системе Соционет как самостоятельные информационные объекты [2]. Автономное представление семантических связей

позволяет использовать объекты-связи как участников других связей. Кроме того, благодаря этому можно формировать автономные репозитории семантических связей, которые могут реализовываться как самостоятельные источники информации и могут интегрироваться с другими подобными репозиториями, базирующимися на контенте иных электронных библиотек, основанными на единых с ним стандартах представления метаданных.

В системе Соционет семантические связи, кроме связей цитирования и некоторых других, декларируются децентрализованно любыми авторизованными ее пользователями в онлайн-режиме. Семантика связей цитирования может декларироваться только пользователями-авторами цитирующих публикаций. В результате этой деятельности над контентом системы динамически формируется семантическая сеть, поддержка и использование которой обеспечивают новые возможности системы [3, 4].

Сочетание указанных возможностей позволяет использовать систему с семантически структурированным контентом как полигон для новых видов научной деятельности, как информационный источник для новых форм наукометрии и исследования корпуса научных знаний, а также решения ряда других задач. Эти возможности и разработанные механизмы для их практического применения подробно обсуждаются в наших работах [2, 3, 5-9].

2. Средства определения семантики декларируемых связей

Декларация семантики связей, создаваемых пользователями в системе Соционет, осуществляется на основе *таксономии связей*. При разработке этой таксономии были приняты во внимание и использованы результаты ряда известных проектов последних лет. Наиболее продвинутые работы в этой области были выполнены специалистами в области биомедицины.

Одну из ранних фундаментальных разработок в этой области представляет собой модульный комплекс онтологий *SPAR (The Semantic Publishing and Referencing Ontologies)* [10, 11], созданный сотрудниками Оксфордского и Болонского университетов. Этот комплекс включает восемь независимых онтологий, позволяющих описывать семантику библиографических объектов, а также их отношений, в частности библиографических объектов, библиографических записей и источников в пристатейных списках, связи цитирования, контексты цитирования и их связи с релевантными разделами цитируемых публикаций. Кроме того, онтологии комплекса SPAR могут использоваться для описания семантики компонентов документов, ролей и состояний публикаций, потоков работ в издательских процессах.

В таксономии связей системы Соционет используются также фрагменты другого комплекса онтологий - *SWAN (Semantic Web Applications in Neuromedicine)* [12], созданного специалистами в области нейромедицины в Главном госпитале Массачусетса и Медицинской школе Гарварда. Комплекс SWAN предназначен для обеспечения комфортной среды в Семантическом Вебе, названной авторами *социально-технической экосистемой*. Эта среда позволяет создавать и сохранять семантический контекст научных коммуникаций, обеспечивает доступ к нему, его интеграцию, а также обмен неструктурированной и слабоструктурированной цифровой научной информацией. В онтологиях SWAN классы связей более агрегированы по сравнению с классами SPAR.

Другой проект, результаты которого были использованы в таксономии связей системы Соционет, это разработанные консорциумом W3C рекомендации *SKOS (Simple Knowledge Organization System)* [13]. Разработчики спецификации SKOS создавали ее для использования в системах организации знаний - тезаурусах, схемах классификации,

таксономиях и рубрикаторах (*Subject Heading Systems*), поддерживаемых в среде Семантического Веба. Рекомендация SKOS определяет концептуальную схему, называемую *общей моделью данных*, которая позволяет совместно использовать и связывать различные системы организации знаний в среде Веба. При использовании такой унифицированной концептуальной схемы упрощается интеграция существующих систем организации знаний в среде Семантического Веба.

Основную часть информационных объектов в контенте системы Соционет составляют научные публикации. Среди них значительна доля коллективных работ. В таком контексте представляет интерес поддержка семантических связей, характеризующих вклад каждого автора в создание коллективных публикаций. Для этого таксономия связей должна включать классы связей с такой семантикой. В таксономии Соционет в качестве прототипа соответствующего ее фрагмента использована работа [14] и основанный на ней проект открытого стандарта для выражения существенных для исследования ролей его участников [15].

Нужно, наконец, упомянуть также еще один важный источник, использованный при разработке таксономии связей для системы Соционет, - модель научных данных CERIF, – развиваемая euroCRIS (<http://www.eurocris.org/>). Один из главных результатов развития CERIF – унифицированная концептуальная схема научно-исследовательской среды, называемая ее авторами *полной моделью данных (Full Data Model)* [16]. Эта модель рассматривается как единая основа разработок информационных систем (*Current Research Information Systems, CRIS*) для поддержки научно-организационной деятельности в разных странах и в различных научных организациях. Использование стандартизированной концептуальной схемы обеспечивает интероперабельность таких систем. В последнее время в развитии CERIF уделяется серьезное внимание проблеме спецификации семантики полной модели данных. В частности, предложена онтология [17], определяющая систему терминов для обозначения сущностей этой модели и отношения между ними.

В таксономии связей, разработанной для системы Соционет, с точностью до имен классов и структурирования использованы классы связей, определенные в указанных источниках, включен также ряд дополнений, например, классы связей «пользователь-публикация», характеризующих рекомендации и информацию для автора публикации.

Таксономия связей, разработанная для системы Соционет, имеет *двухуровневую организацию*. Верхний ее уровень включает одиннадцать классов семантических связей, для каждого из которых определено множество его подклассов.

Подробнее таксономия семантических связей системы Соционет обсуждается в следующем разделе.

3. Таксономия семантических связей системы Соционет

Работа с собственными публикациями и публикациями других исследователей является одним из важнейших элементов научной деятельности. В процессе этой работы возможны различные реакции. Если эта деятельность осуществляется в среде развитой онлайн-информационно-коммуникационной инфраструктуры, такие реакции могут реализоваться оперативно и достаточно комфортно. Их результаты в кратчайшие сроки становятся доступными научному сообществу. Такого рода условия для реализации указанных реакций предоставляет система Соционет.

Возможные реакции определяются, в частности, желанием исследователя определить и предоставить научному сообществу:

- собственные мнения о соотношении научных результатов, обсуждаемых в некоторых публикациях, с результатами его собственных работ или работ иных авторов

- оценки доступных в среде системы публикаций
- информацию о взаимосвязях между версиями своих публикаций и/или их компонентами
- свои мнения об оценках других исследователей относительно представленных в системе публикаций
- рекомендации авторам прочитанных публикаций, представленных в системе, направленные на развитие и/или улучшение этих публикаций
- оценки своего вклада в создание коллективных публикаций
- свои комментарии, аннотирующие фрагменты собственных публикаций и публикаций других авторов.

Перечисленные продукты научной деятельности используются при подготовке публикаций, отражающих собственные результаты исследований ученых, тематических научных обзоров, научных отчетов, рецензий и отзывов на результаты исследований других ученых.

Именно исходя из необходимости обеспечения средств для представления в системе результатов перечисленных возможных реакций исследователя – пользователя системы, была определена таксономия семантических связей информационных объектов для системы Соционет.

Таксономия связей, используемая в системе Соционет, реализована в виде набора *контролируемых словарей* имен классов (подклассов) связей. *Контролируемым словарем* называется набор наперед заданных элементов, из которого пользователь системы может выбрать нужный ему для ввода значения в какое-либо поле экранной формы используемой системы.

В нашем случае каждому классу верхнего уровня таксономии системы Соционет соответствует свой контролируемый словарь, а его подклассам – элементы этого словаря, представленные в виде имен этих подклассов.

Для поддержки рассмотренных выше возможных реакций пользователей системы таксономия семантических связей системы Соционет должна включать следующий набор контролируемых словарей:

- Словарь развития и дополнения результатов публикаций
- Словарь оценочных связей
- Словарь связей между версиями и компонентами публикаций
- Словарь мнений об оценках публикаций
- Словарь рекомендаций авторам публикаций
- Словарь видов вклада в создание коллективных публикаций.

При аннотировании абстрактов публикаций, представленных в Соционет, пользователь часто ограничивается комментарием аннотируемого фрагмента абстракта. Поскольку аннотации представляются в системе также в виде семантических связей между профилем пользователя системы, осуществляющим аннотирование, и данной публикацией, в таком случае создается связь с неопределенным классом таксономии. Для создания таких связей используется специальный словарь таксономии с единственным классом связей «Семантика не определена».

Кроме указанных словарей таксономия семантических связей включает также словари, необходимые для представления в системе административных связей между сотрудниками и организациями, а также между сотрудниками организации, связей авторства между сотрудниками/организациями и их публикациями. Эти связи поддерживаются в системе для отображения некоторой научно-организационной информации, а также для генерации наукометрических показателей.

Все указанные контролируемые словари более подробно рассматриваются ниже.

Важно отметить, что связи каждого класса (подкласса) таксономии могут быть определены только на определенном множестве пар типов исходного и целевого объекта связи. При создании конкретной связи после указания исходного и целевого объектов система использует *матрицу допустимости* классов связей и позволяет пользователю делать выбор возможного класса создаваемой связи только из некоторого допустимого набора контролируемых словарей, соответствующего паре типов связываемых информационных объектов.

В Соционет имеются средства создания, модификации и использования контролируемых словарей классов семантических связей для декларации семантических связей в системе и изменения их свойств. Каждый словарь представлен в системе как коллекция информационных объектов типа *metrics*, каждый из которых представляет некоторый подкласс связей класса, которому соответствует данный словарь.

4. Контролируемые словари классов семантических связей Соционет

Контролируемые словари, представляющие таксономию системы Соционет, предназначены для классификаций семантических связей (отношений), которые пользователь устанавливает между заданными информационными объектами.

Для дальнейшего обсуждения необходимо уточнить некоторые используемые термины. Прежде всего, отметим, что в описании приведенных ниже контролируемых словарей термин *публикация* обозначает информационный объект системы такой, как книга, статья, научный отчет, тезисы доклада, автореферат или полный текст диссертации, препринт и т.п. Эти объекты могут быть следующих типов: *book, article, paper, news, comment, thesis, artifact, metrics*. Термин *персона* означает информационный объект, представляющий некоторое лицо (пользователя системы, администратора, автора публикации) и относящийся к типу *person*. Наконец, термин *организация* означает информационный объект такой, как вуз, научно-исследовательский институт или другое учреждение. Объекты, представляющие организации в Соционет, имеют тип *institution*.

Приведенное ниже описание каждого контролируемого словаря включает его имя на русском и английском языке (имя соответствующего класса таксономии), описание семантики подклассов связей, представленных в словаре, область его возможного применения, а также набор имен входящих в него подклассов связей, каждое из которых также представлено на русском и английском языке.

4.1. Словарь развития и дополнений //Development and Additions

Семантика: Этот словарь включает классы связей, описывающих варианты использования результатов *целевой публикации* в исходной, в частности, для научного вывода новых результатов.

Применение: Классы словаря могут использоваться только для связей вида «публикация-публикация».

Классы словаря:

- Детализирует идею/метод/модель из //Details idea/method/model from
- Обобщает идею/метод/модель из //Generalize idea/method/model from
- Развивает результаты из //Develops results from
- Уточняет результаты из //Refines results from
- Исправляет ошибки в //Corrects errors in
- Опровергает результаты из //Refutes results from
- Анализирует результаты из //Analyzes results from

- Интерпретирует результаты из //Interprets results from
- Иллюстрирует результаты из //Illustrates results from
- Реализует идею/метод/модель из //Implements idea/method/model from
- Использует метод из //Uses method from
- Использует модель из //Uses model from
- Использует данные из //Uses data/statistics from
- Использует статистику из //Uses statistics from
- Использует программное обеспечение из //Uses software from

4.2. Словарь оценок //Estimation linkages

Семантика: Словарь включает классы связей, описывающих варианты оценки результатов *целевой публикации*, отражающей мнение о ней пользователя системы.

Применение: Классы словаря могут использоваться только для связей вида «персона-публикация».

Классы словаря:

- Наилучшая, наиболее релевантная работа по теме //Best, most relevant on the subject
- Подозрения на плагиат //Suspected plagiarism
- Негативно оценивает целевой //Responds negatively to
- Ненаучный подход к получению результата //Unscientific approach
- Новаторский результат //Innovative result
- Очень интересный результат //Very interesting result
- Поворотный пункт для развития науки //Turning point for the science development
- Позитивно оценивает целевой //Responds positively to
- Результат с возможным опасным влиянием //Potentially dangerous effect
- Результат, основанный на заблуждении //Result based on confusion

4.3. Словарь связей между фрагментами и полными текстами публикаций //Linkages between fragments and full text publications

Семантика: Словарь включает классы связей, характеризующих отношения информационных объектов - исходных участков связей (абстрактов, оглавлений, предисловий и других компонентов публикаций, их версий, презентаций, родственных публикаций и др.) к целевым публикациям.

Применение: Классы словаря могут использоваться только для связей вида «публикация-публикация».

Классы словаря:

- Авторская версия (рукопись) для //Author's version (manuscript) of
- Версия с открытым доступом для //Open access version of
- Версия с небольшими изменениями для //Version with slight changes of
- Версия с несущественными изменениями для //Version with minor changes of
- Существенно переработанная версия для //Substantially revised version of
- Идентичная копия для //Duplicated copy of
- Презентация к //Presentation of
- Раздел, часть, глава из //Part of

- Абстракт (аннотация) из //Abstract of
- Новая редакция для //Revised or new version of
- Оглавление из //Table of contents
- Предисловие (введение) из //Foreword of
- Список литературы из //References of

4.4. Словарь мнений об оценках и о семантике других связей **//Linkages of opinions about estimates**

Семантика: Этот словарь включает классы связей, характеризующих мнение персон о семантике оценочных и других связей, выступающих в качестве целевых в данных связях.

Применение: Классы словаря могут использоваться только для связей вида «персона-связь».

Классы словаря:

- Согласие с оценкой/семантикой связи // Agreement with
- Несогласие с оценкой/семантикой связи //Disagreement with
- Безразличие к оценке/ семантике связи //Indifference to
- Ошибочность оценки //Estimation as erroneous
- Завышенность оценки //Estimation as overrated
- Заниженность оценки //Estimation as underrated

4.5. Словарь рекомендаций и информации для авторов **//Recommendations and Information for authors**

Семантика: Словарь включает классы связей публикации автора с рассматриваемой публикацией, представляющих полезные для ее улучшения рекомендации или информацию для ее автора, основанные на содержании выбранной пользователем собственной публикации.

Применение: Классы словаря могут использоваться только для связей вида «публикация-публикация».

Классы словаря:

- У меня обсуждается близкая проблема //Near problem is discussed in my publication
- У меня получены те же результаты //I received the same results
- У меня анализируются ваши результаты //Your results are analysed in my publication
- У меня иллюстрируются ваши данные/метод/модель/результаты //Your data/method/model/results are illustrated in my publication
- У меня интерпретируются ваши результаты //Your results are interpreted in my publication
- У меня детализируются ваши идея/метод/модель/результаты //Your idea/method/model/results are itemized in my publication
- У меня обобщаются ваши идея/метод/модель/результаты //Your idea/method/model/results are generalized in my publication
- У меня реализуются ваши идея/метод/модель //Your idea/method/model are implemented in my work
- У меня лучше данные/метод/модель //My data/method/model are better

- Вы можете применить использованное мною ПО //You can apply software that I used
- У меня указаны и исправлены ваши ошибки //Your errors are detected and corrected in my publication
- У меня опровергаются ваши результаты //Your results are refuted in my publication

4.6. Словарь вкладов авторов в подготовку коллективных публикаций //Linkage of author contributions

Семантика: Словарь включает классы связей, характеризующих вклад персоны-соавтора в подготовку *коллективной* публикации - целевого объекта связи. При этом персона-соавтор выступает в качестве исходного участника связи.

Применение: Классы словаря могут использоваться только для связей вида «персона-публикация».

Классы словаря:

- Администрирование проекта //Project administration
- Научное руководство //Supervision
- Компьютерные работы //Computation
- Проведение экспериментов //Experiments
- Проведение исследования //Investigation
- Концептуализация //Conceptualization
- Курирование данных //Data curation
- Разработка методологии //Methodology
- Обеспечение финансирования //Funding acquisition
- Визуализация результатов //Visualization
- Подготовка начального варианта текста //Writing - original draft
- Критический анализ/доработка текста //Writing – review&edition
- Обеспечение ресурсами //Resources
- Формальный анализ данных //Formal analysis

4.7. Словарь ролей организации по отношению к изданию //Linkages of organizations with respect to publications

Семантика: Словарь включает классы связей, характеризующих роли организаций (исходных участников связей) по отношению к издаваемым публикациям (их целевым участникам).

Применение: Классы словаря могут использоваться только для связей вида «организация-публикация».

Классы словаря:

- Издатель //Publisher of
- Организация-автор //Organization as author of
- Владелец прав //IPR Claim of

4.8. Словарь административных отношений между персонами //Linkages of administrative relationships between persons

Семантика: Словарь включает классы связей, характеризующих административные отношения в организации между персонами – исходными участниками связей и персонами – целевыми их участниками.

Применение: Классы словаря могут использоваться только для связей вида «персона-персона».

Классы словаря:

- Начальник //Supervisor of
- Руководитель-наставник //Mentor of
- Менеджер //Manager of
- Коллега, сослуживец //Colleague of

4.9. Словарь статусов персон в организациях
//Linkages of person status in organizations

Семантика: Словарь включает классы связей, характеризующих статус персон – исходных участников связей в организациях – целевых участниках связей.

Применение: Классы словаря могут использоваться только для связей вида «персона-организация».

Классы словаря:

- Директор/начальник/руководитель организации //Director/principal/head of
- Ректор вуза //High school rector
- Заместитель директора организации //Deputy director of
- Проректор вуза //Pro-rector of
- Руководитель подразделения организации //Head of division
- Декан факультета //Dean of
- Сотрудник/член организации //Employee/member of

4.10. Словарь ролей персон в издании публикаций
//Linkages of person roles in publication production

Семантика: Словарь включает классы связей, характеризующих роли персон – исходных участников связей по отношению к издаваемым публикациям - целевым их участникам

Применение: Классы словаря могут использоваться только для связей вида «персона-публикация».

Классы словаря:

- Автор/соавтор публикации //Author/coauthor of
- Переводчик публикации //Translator of
- Редактор публикации //Editor of
- Рецензент публикации //Reviewer of

4.11. Словарь связей без явно определенной семантики
//Linkages without explicitly defined semantics

Семантика: Коллекция содержит контролируемый словарь классов связей между информационными объектами системы, семантика которых определяется по умолчанию и «встроена» в системные механизмы. Словарь включает единственный класс. Описание связей, относящихся к этому классу, может в дальнейшем обогащаться путем спецификации явного определения их семантики.

Применение: Класс этого словаря используется в системе Соционет для связей вида «персона-публикация», «персона-организация» и некоторых других.

Классы словаря:

- Семантика явным образом не определена //No explicitly defined semantics.

Приведенный выше набор контролируемых словарей семантических связей, поддерживаемых в системе Соционет, может легко пополняться новыми словарями и их состав может корректироваться.

5. Создание семантических связей

Возможность создания семантических связей в системе Соционет доступна только для авторизованных ее пользователей. Это связано с тем, что в описании создаваемой связи должен содержаться уникальный системный идентификатор пользователя – ее создателя. Поэтому при создании связи некоторым пользователем система должна его идентифицировать. Это возможно только при условии, если пользователь был зарегистрирован в системе и авторизовался в начале своего сеанса работы. Кроме того, идентификация пользователя необходима системе и в некоторых других случаях при создании связей. Например, для определения функций, доступных пользователю, в некоторых случаях система должна установить, является ли он автором рассматриваемой публикации (см. ниже). Но это невозможно, если пользователь не авторизовался в начале своего сеанса работы.

В системе Соционет в настоящее время реализовано несколько способов создания семантических связей. Один из них требует использования специального «универсального» сервиса создания связей. Второй способ предназначен для использования пользователями-авторами представленных в системе публикаций. Третий способ служит для создания семантических связей путем семантического «обогащения» связей цитирования, представленных в списках использованных источников публикаций. Он также может использоваться только авторами цитирующих публикаций.

Описание каждой семантической связи, созданной пользователем системы, запоминается в указанной им коллекции.

Рассмотрим кратко реализованные в системе способы создания семантических связей.

Использование «универсального» сервиса системы. С помощью этого сервиса любой авторизованный пользователь системы может в онлайн-режиме создать семантическую связь любого возможного в системе класса между представленными в системе или «внешними» публикациями, за исключением семантических связей цитирования. Для создания связи пользователь выбирает публикации, которые станут участниками создаваемой связи. В зависимости от типов представляющих их информационных объектов ему доступны для выбора некоторые контролируемые словари классов связей. Пользователь выбирает подходящий для его цели словарь и один из входящих в этот словарь класс (подкласс) связей, тем самым определяя семантику создаваемой связи. Кроме того, пользователь – создатель связи может задать некоторый комментарий. Наконец, пользователь задает также локальный идентификатор создаваемой связи – идентификатор, уникальный в рамках той коллекции информационных объектов, в которую он намерен включить эту связь.

В результате для создаваемой связи создается ее описатель [2, 5], включающий: глобальный (в рамках всей системы) уникальный идентификатор, генерируемый системой на основе заданного пользователем локального идентификатора, глобальные системные идентификаторы публикаций – участников связи, идентификатор класса таксономии, к которому относится создаваемая связь, если необходимо, комментарий создателя связи. В описателе указывается также идентификатор автора связи. В случае использования в качестве участника связи «внешней» публикации, она должна быть

доступна в Вебе и в качестве ее глобального идентификатора используется URI этой публикации. Комментарий в описателе связи может использоваться для неформального описания некоторых ее свойств или причин, побудивших автора связи ее создать.

Использование специального интерфейса на странице описания публикации.

Этот способ более дружелюбен для пользователя системы по сравнению с использованием «универсального» сервиса. Для того, чтобы воспользоваться этим способом, нужно вызвать на экран описание публикации, которая станет одним из участников создаваемой связи. Доступ к рассматриваемому интерфейсу осуществляется кликами на активных областях («виртуальных кнопках») этого экрана и возможен как для заявленного автора данной публикации, так и для ее читателя (см. раздел 4.1).

При использовании рассматриваемого интерфейса поведение системы и доступные пользователю возможности зависят от того, какую роль по отношению к данной публикации играет пользователь. Когда пользователь системы просматривает описание некоторой публикации, система рассматривает его как ее автора или как читателя. Пользователь квалифицируется как ее автор в случае, если в системе установлена связь между профилем этого пользователя и рассматриваемой публикацией (пользователь «заявлен» как ее автор). В остальных случаях она рассматривает его как читателя.

Возможности пользователя-автора. Если пользователь выступает в роли автора, исходным объектом создаваемой связи может являться его персональный профиль, а целевым объектом – его публикация. При этом автор может указать свой вклад в подготовку этой публикации (используя словарь 4.6), если у него есть соавторы. Он может также *аннотировать* абстракт своей публикации (используя словарь 4.11) с целью актуализации или уточнения его содержания.

Кроме этого, автор может установить семантическую связь между парой своих публикаций. В этом случае он имеет возможность использовать словарь связей *дополнения и развития* (словарь 4.1) для того, чтобы указать подходящее отношение между изложенными в них результатами исследований, либо указать характер отношений *между различными версиями* (словарь 4.3) или компонентами одной и той же своей публикации.

Возможности пользователя-читателя. Если пользователь выступает в роли «читатель», то в ряде случаев он может установить связь от своего персонального профиля (исходный объект) к публикациям других авторов (целевые объекты). Таким способом «читатель» может, просматривая описание выбранной публикации, *аннотировать* фрагменты ее абстракта и при желании дать при этом *профессиональную оценку* такой публикации (словарь 4.2), которая станет целевой в создаваемой семантической связи.

Кроме того, читатель может создать связь от выбранной своей публикации (исходный объект) к чужой (целевой объект). Таким способом могут быть созданы, например, связи *рекомендательно-информационного характера* (словарь 4.5), с помощью которых автор исходной публикации сообщает автору целевой публикации о возможно полезном ее содержании. Могут быть также установлены связи *дополнения и развития* (словарь 4.1) от выбранной публикации читателя к публикации, описание которой просматривается на экране.

Семантическое обогащение связей цитирования. Третий способ создания семантических связей в системе Соционет относится к связям между данной публикацией как исходной и публикациями, указанными в списке литературы. По существу, речь идет о семантическом обогащении «немых» связей цитирования, уже описанных ранее в описании публикации. «Немыми» мы называем традиционные связи цитирования, поскольку они не несут какой-либо информации о мотивах, побудивших

автора цитировать ту или иную публикацию. Для таких связей в результате определяется явно указанная семантика на основе таксономии связей, поддерживаемой в системе. Такие связи может создавать только пользователь-автор цитирующей публикации после того, как он авторизовался при входе в систему.

Для создания таких связей пользователь должен вызвать на экран описание цитирующей публикации. Если этот пользователь был заявлен как ее автор (см. выше), то система предоставит ему возможность определения семантики связей цитирования. У каждой из публикаций списка использованных источников цитирующей публикации на экране появляется виртуальная кнопка, кликнув на которую пользователь получает доступ к словарю связей *дополнения и развития* (словарь 4.1). Выбрав подходящий класс для создаваемой связи, пользователь тем самым определяет характер использования цитируемого источника в этой своей работе и порождает соответствующую связь, которая запоминается в указанной им коллекции информационных объектов системы.

6. Использование семантических связей в Соционет

Поддерживаемые в системе семантические связи между ее информационными объектами обеспечивают ряд новых возможностей по сравнению с обеспечиваемыми в традиционных электронных библиотеках, в том числе, позволяют осуществлять некоторые новые формы научной деятельности. Ограничимся здесь кратким обсуждением наиболее важных из этих возможностей.

Семантическое структурирование корпуса научных знаний в библиотеке. Прежде всего, явная декларация в системе семантических связей между публикациями позволяет семантически структурировать корпус научных знаний, содержащихся в системе [3]. Формирование семантических связей осуществляется децентрализованно авторизованными пользователями в онлайн-режиме, и в этой деятельности система поддерживает взаимодействие пользователей, обеспечивая им возможности для реагирования на создание связей другими пользователями. В результате этой деятельности над контентом системы формируется многослойная семантическая сеть, своеобразный его семантический ореол [18]. Каждый ее слой соответствует одному из классов связей, определенных поддерживаемой в системе таксономии семантических связей.

Визуализация связей и семантическая навигация. При обращении пользователя к какой-либо из имеющихся в системе публикаций он имеет возможность визуально анализировать ближайшую окрестность узла семантической сети, соответствующего данной публикации, т.е. данный узел и все непосредственно связанные с ним узлы с указанием классов связей. При просмотре описания публикации при желании можно видеть все входящие и исходящие связи данного узла. Это позволяет получать некоторую информацию о месте данной публикации в корпусе научных знаний, о соотношении данной публикации с другими представленными в системе. Вместе с тем, визуализация связей каждого узла семантической сети дает возможность пользователю осуществлять семантическую навигацию по сети, пошагово перемещаясь от одной публикации к другой.

Пользуясь визуализированным представлением окрестности каждого неизолированного узла семантической сети, можно легко найти иные версии соответствующей этому узлу публикации, ее разделы или презентации, если они также представлены в системе. Вместе с тем, отфильтровывая слой сети, соответствующий интересующему пользователя классу связей, например, связей, определяющих характер использования результатов, представленных в одних публикациях, в других, и стартуя с некоторого узла сети, можно навигацией по полученной подсети проследить

эволюцию научных результатов в некоторой области науки, выявить относящиеся к ней основополагающие работы. Конечно же, все это возможно при условии достаточно представительного контента системы как относительно состава представленных в нем публикаций, так и в уровне определенности их семантических связей.

Используя подсеть оценочных связей, можно найти работы либо мнения конкретных пользователей системы, в которых данная публикация подвергается критике или, наоборот, получает положительные оценки.

Наукометрические исследования. Другое важное направление использования семантических связей между публикациями в системе – использование построенной над ее контентом семантической сети как источника данных для наукометрических исследований, более информативного по сравнению с традиционно используемыми на практике [8]. Традиционные наукометрические исследования обычно основаны на использовании связей цитирования в пристатейных списках публикаций в периодике. Однако эти связи не несут никакой информации о мотивах цитирования в отличие от семантически обогащенных связей цитирования в системе Соционет, предоставляющих информацию о характере использования цитируемого источника в цитирующей публикации.

Наукометрические сервисы системы генерируют статистические данные на основе связей цитирования дифференцировано по классам связей. Такие данные позволяют более глубоко осмысливать и оценивать степень влияния тех или иных публикаций на развитие научных исследований, изучать содержательные наукометрические «портреты» как отдельных ученых, так и исследовательских организаций в целом, получать более осмысленные оценки их научной продуктивности. Использование в системе семантических связей, отражающих вклад автора в создание коллективной публикации, позволяет также проводить наукометрические измерения, связанные с персональными оценками этого аспекта научной деятельности для отдельных исследователей.

Научные коммуникации. Нужно, наконец, упомянуть еще одно достаточно существенное направление использования семантических связей в системе [19]. С помощью семантических связей авторизованные пользователи могут с использованием структурированных и/или неструктурированных данных публиковать в системе свои профессиональные оценки публикаций других авторов, предлагать им рекомендации по развитию конкретных их работ, комментировать абстракты своих и чужих публикаций. Наряду с этим, имеется возможность публиковать мнения и оценки относительно оценок, высказанных другими пользователями. Таким образом, на полигоне системы могут поддерживаться альтернативные формы рецензирования научных публикаций и научные дискуссии. Авторы публикаций и других научных информационных объектов, которые становятся участниками создаваемых связей, информируются об этом по электронной почте сервисом оповещения, имеющимся в системе Соционет. Получение оповещения стимулирует получателя реагировать на него и тем самым поддерживать такую дискуссию. Все используемые в описанных процессах семантические связи включены в пространство наукометрических измерений в системе с учетом их классов в таксономии.

7. Заключение

Предполагается продолжить работы по развитию функционирующих в рамках системы Соционет механизмов создания и использования семантических связей для обеспечения более комфортных условий работы пользователей и расширения возможностей системы по формированию и использованию ее контента, по анализу динамически формируемой над ним семантической сети связей информационных

объектов. Будет уточняться и развиваться также используемая в системе таксономия семантических связей.

Литература

[1]. Паринов С.И., Ляпунов В.М., Пузырев Р.Л. Система Соционет как платформа для разработки научных информационных ресурсов и онлайн-сервисов //Российский научный электронный журнал «Электронные библиотеки». 2003. Том 6. Вып. 1. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP>

[2]. Когаловский М.Р., Паринов С.И. Технологии социальной сети для создания семантических связей информационных объектов в научной электронной библиотеке //Программирование. МАИК/Наука «Интерпериодика». 2014. Т. 40, № 5 (в печати).

[3]. Когаловский М.Р., Паринов С.И. Семантическое структурирование контента научных электронных библиотек на основе онтологий. В кн.: "Современные технологии интеграции информационных ресурсов: сборник научных трудов». Санкт-Петербург: Президентская библиотека им. Б.Н. Ельцина, 2011. С. 26-45.

[4]. Паринов С.И., Когаловский М.Р. Технология семантического структурирования контента научных электронных библиотек. Труды XIII Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции – RCDL-2011. Воронеж, 19-22 октября 2011 г.». г. Воронеж: Воронежский государственный университет, 2011. С. 94-103.

[5]. Когаловский М.Р., Паринов С.И. Новый источник данных для наукометрических исследований. Электронные библиотеки: перспективные методы и технологии, электронные коллекции. XV Всероссийская научная конференция RCDL'2013. Ярославль, Россия, 14-17 октября 2013 г. Труды конференции. С. 107-117.

[6]. Parinov S. Open Repository of Semantic Linkages. In: Proceedings of 11th International Conference on Current Research Information Systems e-Infrastructure for Research and Innovations (CRIS 2012), Prague, 2012. <http://socionet.ru/publication.xml?h=repec:rus:mqijxk:29>

[7]. Parinov S. Towards a Semantic Segment of a Research e-Infrastructure: necessary information objects, tools and services. Metadata and Semantics Research, Communications in Computer and Information Science. J. M. Doderо, M. Palomo-Duarte, P. Karampipерis, Eds. Springer, vol. 343, 2012, pp. 133-145. <http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30>

[8]. Parinov S., Kogalovsky M. Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. Scientometrics, Volume 98, Issue 2 (2014), pp. 927-943. DOI 10.1007/s11192-013-1108-3

[9]. Parinov Sergey. Semantic enrichment of research outputs metadata: new CRIS facilities for authors. Submitted to MTSR 2014, 8th Metadata and Semantics Research Conference, 27-29 November 2014, Karlsruhe, Germany.

[10]. Shotton D. Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/>

[11]. Shotton D., Peroni S. Semantic annotation of publication entities using the SPAR (Semantic Publishing and Referencing) Ontologies /Beyond the PDF Workshop, La Jolla, 19 January 2011. http://imageweb.zoo.ox.ac.uk/pub/2010/Publications/Shotton&Peroni_semantic_annotation_of_publication_entities.pdf

[12]. Semantic Web Applications in Neuromedicine (SWAN) Ontology. W3C Interest Group Note, 20 October 2009. <http://www.w3.org/TR/2009/NOTE-hcls-swan-20091020/>

- [13]. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 August 2009. <http://www.w3.org/TR/skos-reference/>
- [14]. Liz Allen, Amy Brand, Jo Scott, Micah Altman and Marjorie Hlava. Credit where credit is due. *Nature*, Vol. 508, April 2014. http://www.nature.com/polopoly_fs/1.15033!/menu/main/topColumns/topLeftColumn/pdf/508312a.pdf
- [15]. [15] CRediT. An open standard for expressing roles intrinsic to research. <http://credit.casrai.org/proposed-taxonomy/>
- [16]. CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS, 2012. http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf
- [17]. CERIF 1.3 Semantics: Research Vocabulary. CERIF Task Group, euroCRIS, 2012. http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf
- [18]. Dix A., Levialdi S. & Malizia A. Semantic halo for collaboration tagging systems. In the Social Navigation and Community-Based Adaptation Technologies Workshop-June 20th, 2006.
- [19]. Kogalovsky M.R., Parinov S.I. Social Network Technologies for Semantic Linking of Information Objects in Scientific Digital Library //Programming and Computer Software, 2014, Vol. 40, No. 6, pp. 313–320.