

Таксономия семантических связей информационных объектов контента научной электронной библиотеки

М.Р. Когаловский, С.И. Паринов

Научно-техническая информация. Серия 2. - 2015. - № 9. - С. 15-23.

Авторская редакция.

Обсуждается таксономия семантических связей, позволяющая авторизованным пользователям научной электронной библиотеки в онлайн-режиме декларировать различные виды отношений между информационными объектами ее контента. В статье рассматривается организация таксономии и ее представление в виде набора контролируемых словарей классов связей. Описывается реализация таксономии в среде крупной научной информационной системы Соционет, а также ряд новых возможностей для сообщества пользователей системы, обеспечиваемых благодаря семантическому обогащению ее контента путем спецификации семантических связей. Работа поддержана грантом РГНФ 14-02-12010-в.

Ключевые слова: *научная электронная библиотека, контент библиотеки, система Соционет, семантическая связь, таксономия связей, контролируемый словарь, новые функции электронной библиотеки*

1. Введение

В последние годы в области научных электронных библиотек активно развиваются исследования и разработки, направленные на семантическое обогащение их контента. Благодаря этому может быть повышена информативность, а, следовательно, и ценность информационных ресурсов, представленных в этих системах. Помимо традиционного доступа пользователей непосредственно к информационным объектам библиотеки, открываются новые возможности для практического применения семантически обогащаемых электронных библиотек и решения ряда новых нетрадиционных задач.

Один из активно развивающихся подходов к семантическому обогащению информационных ресурсов научных электронных библиотек, на котором основаны и работы авторов данной статьи, заключается в предоставлении пользователям возможностей декларации *семантических связей* между информационными объектами контента библиотеки и визуализация созданных связей. Такие связи могут нести различную информацию, определенную явным образом. Они могут представлять разнообразные отношения научного характера между публикациями, в частности, констатировать факт цитирования с указанием его мотивов, характеризовать влияние одной публикации на другую или использование в ней результатов, обсуждаемых в другой публикации, определять структурные связи между компонентами или различными версиями научного произведения, представлять отношения авторства между персонами и публикациями, декларировать вклад соавторов в создание коллективной публикации и др. Могут представляться также отношения между пользователями системы (представленными их персональными профилями) и публикациями библиотеки. Например, пользователи библиотеки могут декларировать оценки представленных в ней публикаций, выражать рекомендации их авторам или публиковать свое мнение относительно информации, содержащейся в созданных семантических связях.

Семантические связи, кроме связей цитирования и некоторых других, в предлагаемом подходе декларированы *децентрализованно авторизованными пользователями* электронной библиотеки в онлайн-режиме, имеющими при этом дело со своеобразной социальной сетью. Семантика связей цитирования может декларироваться только пользователями-авторами цитирующих публикаций. Возможности пользователей декларировать новые семантические связи в онлайн-режиме в процессе их работы с контентом библиотеки приводят к формированию динамически обогащаемого *«семантического ореола»* библиотеки - его *многослойной семантической структуры*, слои которой соответствуют различным классам декларированных семантических связей [1, 2, 3].

Научная электронная библиотека с семантически структурированным контентом может служить основой для поддержки новых видов научных коммуникаций [4], информационным источником для качественно новых наукометрических измерений и исследований на представленном в библиотеке корпусе научных знаний, а также для ряда других целей. Эти возможности и разработанные механизмы для их практического применения подробно обсуждаются в работах [1, 3, 5, 6-10].

Описываемая в данной статье таксономия семантических связей является одним из ключевых элементов, предложенного и реализованного авторами подхода к созданию семантически обогащаемых научных электронных библиотек.

В принятом авторами подходе допускаются только *бинарные ориентированные семантические связи* между ее информационными объектами. Объект - участник связи, из которого исходит связь, называется ее *исходным* информационным объектом. Второй участник, на который направлена связь, называется *целевым* объектом этой связи.

Семантические связи в предлагаемом подходе представляются в электронной библиотеке как *самостоятельные информационные объекты* [1, 5]. 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 как с другими подобными репозиториями, так и с различными электронными библиотеками, если они используют единый с данной библиотекой стандарт представления метаданных и общий с нею метод идентификации информационных объектов.

В качестве полигона для реализации рассматриваемого подхода используется крупная отечественная *научная информационная система Соционет* [11], функционирующая уже полтора десятилетия. Соционет основана на технологии открытых архивов [12-13], импортирует метаданные из репозитория более 1700 организаций, включая сотни отечественных и зарубежных открытых архивов, предоставляет метаданные собственного репозитория для импорта другим системам и т.д. В настоящее время в Соционет доступно около 2.5 млн. информационных объектов, содержатся описания более 8 миллионов семантических связей.

В остальной части статьи обсуждаются известные средства определения семантики декларируемых связей (разд. 2), на которые опирается предлагаемый подход, дается общая характеристика предлагаемой таксономии (разд. 3) и ее представления в виде контролируемых словарей (разд. 4), описывается техника создания семантических связей в среде системы Соционет (разд. 5). Наконец, кратко обсуждаются возможности, которые обеспечиваются в системе для пользователей благодаря поддержке семантической структуры связей и принятому способу их представления в системе (разд. 6).

2. Средства определения семантики декларируемых связей

Декларация семантики связей в созданной авторами технологии осуществляется пользователями электронной библиотеки на основе *таксономии связей*. При разработке 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 представляет собой модульный комплекс онтологий *SPAR (The Semantic Publishing and Referencing Ontologies)* [14, 15], 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.

Другой комплекс онтологий - *SWAN (Semantic Web Applications in Neuromedicine)* [16] создан 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.

В данном контексте заслуживают внимания разработанные консорциумом W3C рекомендации *SKOS (Simple Knowledge Organization System)* [17]. Разработчики 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.

В научных электронных библиотеках значительна доля коллективных работ. Поэтому представляет интерес поддержка семантических связей между авторами и публикациями, характеризующих вклад каждого автора в создание коллективных публикаций. Для этого таксономия связей должна включать классы связей с такой семантикой. Группой редакторов медицинских журналов совместно с Гарвардским университетом и медицинским благотворительным фондом Wellcome Trust из Лондона предложен *проект CRediT* [18, 19] открытого стандарта классификатора существенных для коллективно выполняемого исследования ролей его участников.

Нужно, наконец, упомянуть также модель научных данных *CERIF*, – развиваемую европейской организацией 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)* [20]. 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* [21], определяющая систему терминов для обозначения сущностей этой модели и отношения между ними.

В таксономии семантических связей, реализованной в системе Соционет, использованы указанные таксономии или их фрагменты. Включен также ряд дополнений, например, классы связей «пользователь-публикация», характеризующих рекомендации и информацию для автора публикации, или «пользователь-связь», характеризующие оценки семантики установленных связей пользователями системы. Подробнее эта таксономия 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*, которые мотивируются, 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
- свои оценки содержания публикаций, доступных в среде системы

- 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 или публикаций других авторов.

Оперативно и достаточно комфортно «материализовать» результаты таких реакций позволяет научная электронная библиотека, предоставляющая пользователям механизмы создания, публикации и использования семантических связей между информационными объектами ее контента. Каждая из указанных реакций *реализуется путем декларации семантической связи* с соответствующей семантикой. Результаты такой деятельности становятся доступными научному сообществу пользователей библиотеки. Именно исходя из необходимости обеспечения средств для представления результатов перечисленных возможных реакций исследователя – пользователя системы, была определена описываемая таксономия семантических связей информационных объектов.

Таксономия семантических связей, реализованная в системе Соционет, имеет двухуровневую организацию и организована в виде набора контролируемых словарей. *Контролируемым словарем* называется некоторый упорядоченный набор допустимых значений, предоставляемый пользователю компьютерной системы для выбора конкретного значения, вводимого в окно экранной формы, с которым ассоциирован данный словарь. В нашем случае каждому классу верхнего уровня таксономии соответствует некоторый контролируемый словарь, а классам второго уровня – значения из этого словаря, представляющие имена классов, являющихся подклассами соответствующего класса верхнего уровня.

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

- Словарь видов развития и дополнений результатов исследований
- Словарь видов профессиональных оценок публикаций
- Словарь отношений между компонентами и версиями публикаций
- Словарь видов мнений о существующих связях
- Словарь видов близости публикаций
- Словарь видов вкладов авторов в создание коллективных публикаций.

В системе используется также вспомогательный словарь связей без семантики, определенной в явном виде.

При аннотировании абстрактов публикаций, представленных в Соционет, пользователь ограничивается *только комментарием* к аннотируемому фрагменту 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, рассматриваемая таксономия 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, которые в данной статье не рассматриваются. Эти связи поддерживаются в системе для отображения научно-организационной структуры организации и генерации некоторых связанных с нею наукометрических показателей.

Важно отметить, что связи каждого класса таксономии могут быть определены только *на определенном множестве пар типов* исходного и целевого объекта связи. При создании конкретной связи после того, как пользователь укажет исходный и целевой ее объекты, система использует *матрицу допустимости* классов связей, определяющую для каждой пары типов связываемых объектов возможные классы семантических связей между ними и предоставляет пользователю возможность выбора нужного класса связи из соответствующих контролируемых словарей.

В Соционет имеются средства создания, модификации и использования контролируемых словарей классов семантических связей для декларации семантических связей в системе, их публикации 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 связей в системе Соционет, предназначены для *классификаций* семантических связей (отношений), которые пользователь устанавливает между выбранными им информационными объектами контента системы. Каждый из словарей определяет набор классов, являющихся подклассами соответствующего этому словарю класса семантических связей таксономии. Классы связей представляются в словарях своими именами, отражающими семантику связей, относящихся к этому классу.

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* (книга), *article* (статья), *paper* (отчет, рабочая записка, тезисы доклада и др.), *news* (новость), *comment* (комментарий), *thesis* (диссертация), *artifact* (артефакт), *metrics* (класс связей).

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*.

В соответствии с технологией открытых архивов, публикации, персоны, организации и другие информационные объекты представляются в системе Соционет их *описаниями*, содержащими для публикаций и других «материализованных» объектов ссылки на них в среде Веба, где они хранятся. Описания организаций и персон называются их *профилями*. Семантические связи между информационными объектами представляются в системе связями между их описаниями.

Приведенное ниже описание каждого контролируемого словаря, представляющего используемую в системе Соционет таксономию семантических связей, включает его название (имя соответствующего класса верхнего уровня таксономии), общее описание семантики принадлежащих ему классов связей второго уровня и набор их имен, область его возможного применения.

4.1. Словарь видов развития и дополнений результатов исследований

Семантика: Этот словарь включает классы связей, описывающих варианты использования результатов *целевых публикаций* в исходных, в частности, для научного вывода новых результатов.

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

Классы связей:

- Детализирует идею/метод/модель из
- Обобщает идею/метод/модель из

- Развивает результаты из
- Уточняет результаты из
- Исправляет ошибки в
- Опровергает результаты из
- Анализирует результаты из
- Интерпретирует результаты из
- Иллюстрирует результаты из
- Реализует идею/метод/модель из
- Использует метод из
- Использует модель из
- Использует данные из
- Использует статистику из
- Использует программное обеспечение из

4.2. Словарь видов профессиональных оценок публикаций

Семантика: Словарь включает классы связей, описывающих варианты оценки *целевых публикаций*, которые отражают мнение о них пользователей системы, профили которых служат исходными объектами таких связей.

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

Классы связей:

- Рассматривает как наилучшую, наиболее релевантную работу по теме
- Подозревает плагиат
- Негативно оценивает целевую
- Считает, что использован ненаучный подход к получению результата
- Рассматривает как новаторский результат
- Рассматривает как очень интересный результат
- Считает поворотным пунктом для развития науки
- Позитивно оценивает целевую
- Рассматривает как результат с возможным опасным влиянием
- Полагает, что результат основан на заблуждении.

4.3. Словарь отношений между компонентами и версиями публикаций

Семантика: Словарь включает классы связей, характеризующих отношения публикаций - исходных участников связей (абстрактов, оглавлений, предисловий и других компонентов публикаций, их версий, презентаций, родственных публикаций и др.) с целевыми публикациями.

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

Классы связей:

- Авторская версия (рукопись) для
- Версия с открытым доступом для
- Версия с небольшими изменениями для
- Версия с несущественными изменениями для
- Существенно переработанная версия для
- Идентичная копия для
- Презентация к
- Раздел, часть, глава из
- Абстракт (аннотация) для
- Новая редакция для
- Оглавление из

- Предисловие (введение) из
- Список литературы из

4.4. Словарь видов мнений о существующих связях

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

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

Классы связей:

- Согласен с семантикой связи
- Не согласен с семантикой связи
- Считает неверно понятой оцениваемую/используемую работу
- Считает ошибочной оценку, представленную связью
- Считает завышенной оценку, представленную связью
- Считает заниженной оценку, представленную связью.

4.5. Словарь видов близости публикаций

Семантика: Этот словарь включает классы связей публикаций создателя связи - с рассматриваемой публикацией. Семантика создаваемых связей предоставляет авторам целевых публикаций полезные для их улучшения сведения, основанные на содержании выбранной создателем связи собственной публикации (исходной публикации связи). Связь данного класса позволяет независимо от связей цитирования целевой публикации с определенными для них мотивами цитирования сообщить ее автору сведения о развитии представленных в ней результатов или о результатах, близких к ним, обсуждаемых в исходной публикации связи.

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

Классы связей:

- У меня обсуждается близкая проблема
- У меня получены те же результаты
- У меня анализируются ваши результаты
- У меня иллюстрируются ваши данные/метод/модель/результаты
- У меня интерпретируются ваши результаты
- У меня детализируются ваши идея/метод/модель/результаты
- У меня обобщаются ваши идея/метод/модель/результаты
- У меня реализуются ваши идея/метод/модель
- У меня лучше данные/метод/модель
- В вашем случае применимо использованные мною программные средства
- У меня указаны и исправлены ваши ошибки
- У меня опровергаются ваши результаты.

4.6. Словарь видов вкладов авторов в создание коллективных публикаций

Семантика: Словарь включает классы связей, характеризующих вклад персон-соавторов в подготовку коллективных публикаций - целевых объектов связи. При этом персона-соавтор выступает в качестве исходного участника связи.

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

Классы связей:

- Администрирование проекта
- Научное руководство
- Компьютерные работы
- Проведение экспериментов

- Проведение исследования
- Концептуализация проблемы
- Курирование данных
- Разработка методологии
- Обеспечение финансирования
- Визуализация результатов
- Подготовка начального варианта текста
- Критический анализ/доработка текста
- Обеспечение ресурсами
- Формальный анализ данных.

4.7. Словарь связей, семантика которых не представлена в явном виде

Семантика: Это - особый словарь, содержащий классы связей между информационными объектами системы, семантика которых явным образом не представлена, а определяется по умолчанию и «встроена» в системные механизмы. В настоящее время этот словарь включает единственный класс. Описания связей, относящихся к этому классу, могут в дальнейшем обогащаться путем отнесения их к одному из классов связей описанных выше словарей и тем самым явного определения их семантики.

Применение: В системе Соционет класс из этого словаря используется для связей вида «персона-публикация» со встроенной семантикой авторства, «персона-организация» с семантикой, характеризующей персон как сотрудников данных организаций, связей вида «публикация-публикация» для представления «немых» связей цитирования и для некоторых других связей.

Классы связей:

- Семантика явным образом не определена.

4.8. Пример семантической связи

Предлагаемый ниже пример семантической связи между вымышленными публикациями использует класс семантической связи из контролируемого словаря разд. 3.1:

- *Исходная публикация связи:* Петров П.П. Оценка уровня инфляции в российской экономике. – М.: Экономика, 2015. – 200 с.
- *Класс связи:* Анализирует результаты из
- *Целевая публикация:* Васильев В.В. Промышленность и сельское хозяйство России в 2010 – 2014 гг. //Экономическая теория. - 2014. - № 4. – С. 12-25.

Приведенный выше набор контролируемых словарей семантических связей, поддерживаемых в системе Соционет, может пополняться новыми словарями, словари могут дополняться новыми классами, и имена классов в словарях могут изменяться.

5. Создание семантических связей

Возможность создания семантических связей в системе Соционет доступна только для ее *авторизованных пользователей*. Это связано, прежде всего, 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.

Идентификация создателя семантической связи необходима и потому, что факт существования связи и ее семантика становятся в системе *публично доступными*, и отсутствие

анонимности автора повышает его ответственность перед научным сообществом за публикуемые мнение или оценку, представляемые создаваемой им связью.

В системе Соционет в настоящее время реализовано несколько способов создания семантических связей. Один из них основан на использовании 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 публикаций, представленных в списках использованных источников. Он также может использоваться только авторами цитирующих публикаций.

Описание каждой семантической связи, созданной пользователем системы, запоминается в указанной им его коллекции информационных объектов в среде хранения данных системы. Коллекции информационных объектов в системе представляют собой совокупность описаний однотипных информационных объектов.

Рассмотрим кратко указанные способы создания семантических связей.

Использование «универсального» сервиса создания связей. С помощью этого сервиса авторизованный пользователь системы может в онлайн-режиме создать семантическую связь любого допустимого класса между представленными в системе или «внешними» (не содержащимися в ее контенте, но доступными в Вебе) публикациями, за исключением семантических связей цитирования. Создание связей цитирования возможно только для автора цитирующей публикации. Для создания связи пользователь выбирает публикации, которые станут ее участниками. 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 связей, тем самым определяя семантику создаваемой связи. Кроме того, пользователь – создатель связи может задать некоторый комментарий, входящий в описание связи. 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 *описание* [1, 6], включающее: глобальный (в рамках всей системы) уникальный идентификатор, генерируемый системой на основе заданного пользователем локального идентификатора; глобальные системные идентификаторы публикаций – участников связи; идентификатор словаря (класса) таксономии, к которому относится создаваемая связь; если необходимо, комментарий создателя связи. В описании указывается также идентификатор автора связи. В случае использования «внешней» публикации в качестве участника связи, она должна быть доступна в Вебе, и в качестве ее глобального идентификатора используется URI этой публикации. Комментарий в описании 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 прост в использовании для пользователя системы по сравнению с использованием «универсального» сервиса. Для того чтобы им воспользоваться, нужно вызвать на экран описание публикации, которая станет *целевым участником* создаваемой связи. Доступ к рассматриваемому интерфейсу осуществляется кликами на активных областях («виртуальных кнопках») этого экрана и возможен как для автора данной публикации, так и для ее читателя.

При использовании рассматриваемого интерфейса поведение системы и доступные пользователю возможности зависят от того, какую роль по отношению к данной публикации играет пользователь. Когда авторизованный пользователь просматривает описание некоторой публикации, система опознает его как ее автора или как читателя. Пользователь квалифицируется как ее *автор* в случае, если в идентификатор этого пользователя указан как значение соответствующего атрибута в описании этой публикации. В остальных случаях она рассматривает его как читателя.

Если пользователь выступает в роли автора, исходным объектом создаваемой связи может являться его персональный профиль, а целевым объектом – его публикация (ее описание). При этом автор может указать свой *вклад в подготовку этой публикации* (используя словарь из разд. 4.6), если она – коллективная. Он может также *аннотировать* абстракт своей публикации (используя словарь из разд. 4.7) с целью актуализации или уточнения его содержания.

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

Пользователь, выступающий в роли *читателя*, может устанавливать связи от представляющего его персонального профиля (исходный объект) к публикациям других авторов (целевые объекты). Таким способом, просматривая описание выбранной публикации, он может *аннотировать* фрагменты ее абстракта, при желании дать *профессиональную оценку* такой публикации (разд. 4.2), которая станет целевой в создаваемой семантической связи.

Пользователь-читатель может также создать связь от выбранной своей публикации (исходный объект) к чужой публикации (целевой объект). Это могут быть, например, связи *информационного характера* (разд. 4.5), с помощью которых автор исходной публикации сообщает автору целевой публикации *о близости их содержания*. Могут быть также установлены связи классов из словаря *видов развития и дополнений* (разд. 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 пристатейном 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 объявлен как ее автор, 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 *видов развития и дополнений* (разд. 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 Соционет семантические связи между ее информационными объектами обеспечивают ряд новых возможностей по сравнению с обеспечиваемыми в традиционных электронных библиотеках. 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 [2, 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. Стимулирует ответные реакции пользователей на создание семантических связей, участниками которых являются их публикации, предусмотренный в системе сервис оповещения о таких событиях.

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 информацию о месте данной публикации в корпусе научных знаний, о соотношении данной публикации с другими представленными в системе, об оценках его работ пользователями системы. Вместе с тем, визуализация связей каждого узла семантической сети дает возможность пользователю осуществлять *семантическую навигацию* по сети, пошагово перемещаясь от одной публикации к другой по связям интересующих его классов.

Пользуясь визуализированным представлением окрестности каждого неизолированного узла семантической сети, можно легко найти иные версии публикации, соответствующей этому узлу, ее разделы или презентации, если они также представлены в системе. Вместе с тем, можно отфильтровать слой сети, соответствующий интересующему пользователя классу связей, например, связей, определяющих характер использования результатов, представленных в одних публикациях, в других. Далее, стартуя от текущего 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 полезно при условии достаточно представительного контента системы как относительно состава представленных в нем публикаций, так и семантических связей между ними.

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

Наукометрические исследования. Другое важное направление использования семантических связей между информационными объектами системы – это использование семантической сети, построенной на ее контенте, как *источника данных* для наукометрических исследований, более информативного по сравнению с традиционно используемыми на практике [9]. Традиционные наукометрические исследования обычно основаны на использовании связей цитирования в пристатейных списках публикаций из периодики. Однако эти связи не несут никакой информации о мотивах цитирования в отличие от семантически обогащенных связей цитирования в системе Соционет. Кроме того, в наукометрии могут быть использованы и поддерживаемые в системе связи, отличные от связей цитирования, имеющие различную семантику.

Наукометрические сервисы системы Соционет генерируют статистические данные дифференцировано по классам связей, в том числе, и для связей цитирования. Такие данные позволяют более глубоко анализировать и оценивать научное качество представленных в системе публикаций, степень их влияния 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 [4]. Создавая семантические связи, авторизованные пользователи могут с использованием структурированных (имен классов связей) и/или неструктурированных данных (комментариев в описаниях создаваемых связей) 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. Для этого нужно создавать связи собственного профиля со связями, выражающими оцениваемые оценки. Таким образом, на полигоне системы могут поддерживаться альтернативные формы рецензирования научных публикаций, научные дискуссии и обмен научной информацией, ассоциированной с представленными в системе публикациями. Авторы публикаций и других научных информационных объектов, которые становятся участниками создаваемых связей, информируются об этом по электронной почте сервисом уведомления, имеющимся в системе Соционет. Получение оповещения стимулирует получателя реагировать на него, создавая соответствующую новую связь, и тем самым поддерживать такую дискуссию. Все созданные в описанных процессах семантические связи включаются в пространство наукометрических измерений в системе с учетом их классов в таксономии.

7. Заключение

Описанная в статье таксономия семантических связей реализована в системе Соционет. Созданы механизмы формирования, поддержки и использования контролируемых словарей, а также декларируемых с их помощью семантических связей. Эти механизмы позволяют при необходимости расширять и корректировать поддерживаемую в системе таксономию семантических связей, не затрагивая программного обеспечения системы. Разработаны средства контекстной визуализации семантических связей для узлов семантической сети, а также ряд других пользовательских сервисов системы. Все эти средства находятся в стадии опытной эксплуатации.

Литература

- [1]. 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. - № 6.
- [2]. 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.
- [3]. 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.
- [4]. Когаловский М.Р., Паринов С.И. Научные коммуникации в среде семантически обогащаемых электронных библиотек // Программная инженерия. – 2015. - № 4. – С. 31-38 (в печати).
- [5]. 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.
- [6]. Когаловский М.Р., Паринов С.И. Новый источник данных для наукометрических исследований. Труды XV Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции – RCDL-2013. Ярославль, Россия, 14-17 октября 2013 г.». г. Ярославль: Ярославский университет, 2013. – С. 107-117.
- [7]. 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. [Электрон. ресурс]. – 2012. – URL: <http://socionet.ru/publication.xml?h=repec:rus:mqijxk:29> (дата обращения: 03.03.2015).

[8]. 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. Dodero, M. Palomo-Duarte, P. Karampiperis, Eds. Springer. Vol. 343, 2012, pp. 133-145. [Электрон. ресурс]. – 2012. – URL: <http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30> (дата обращения: 03.03.2015).

[9]. Parinov S., Kogalovsky M. Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. *Scientometric*. Vol. 98, Issue 2 (2014), pp. 927-943.

[10]. 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.

[11]. 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. [Электрон. ресурс]. – 2003. – URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal /2003 /part1/PLP> (дата обращения: 03.03.2015).

[12]. Open Archives Initiative. [Электрон. ресурс]. - URL: <http://www.openarchives.org/> (дата обращения: 03.03.2015).

[13]. The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Document Version 2015-01-08. [Электрон. ресурс]. – 2015. - URL: <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm> (дата обращения: 03.03.2015).

[14]. Shotton D. Open Citations and Related Work. Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. [Электрон. ресурс]. – 2010. - URL: <http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/> (дата обращения: 03.03.2015).

[15]. 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. [Электрон. ресурс]. – 2011. - URL: <http://speroni.web.cs.unibo.it/publications /shotton-2010-semantic-annotation-publication.pdf> (дата обращения: 03.03.2015).

[16]. Semantic Web Applications in Neuromedicine (SWAN) Ontology. W3C Interest Group Note, 20 October 2009. [Электрон. ресурс]. – 2009. - URL: <http://www.w3.org/TR/hcls-swan/> (дата обращения: 03.03.2015).

[17]. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 August 2009. [Электрон. ресурс]. – 2009. - URL: <http://www.w3.org/TR/skos-reference/> (дата обращения: 03.03.2015).

[18]. Liz Allen, Amy Brand, Jo Scott, Micah Altman and Marjorie Hlava. Credit where credit is due. *Nature/ International weekly journal of science*. Vol. 508, Issue 7496, April 2014. [Электрон. ресурс]. – 2014. - URL: http://www.nature.com/polopoly_fs/1.15033!/menu/main/topColumns/topLeftColumn/pdf/508312a.pdf (дата обращения: 03.03.2015).

[19]. CRediT. An open standard for expressing roles intrinsic to research. *Taxonomy*. [Электрон. ресурс]. – 2014. – URL: <http://credit.casrai.org/proposed-taxonomy/> (дата обращения: 03.03.2015).

[20]. CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS, 2012. [Электрон. ресурс]. – 2012. - URL: http://www.eurocris.org/Uploads/Web%20pages/ CERIF-1.3/Specifications/CERIF1.3_FDM.pdf (дата обращения: 03.03.2015).

[21]. CERIF 1.3 Semantics: Research Vocabulary. CERIF Task Group, euroCRIS, 2012. [Электрон. ресурс]. – 2012. - URL: http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf (дата обращения: 03.03.2015).

[22]. 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.