

Концептуальная модель научной публикации

© В.А. Резниченко

© Г.Ю. Проскудина

© О.М. Овдий

Институт программных систем НАН Украины

Киев

gupros@isofts.kiev.ua

Аннотация

Работа посвящена задаче создания концептуальной модели научных ресурсов для научно-исследовательских и библиотечных информационных систем, а именно модели публикации, как одного из основных ресурсов информационного научного пространства. Также рассмотрены некоторые известные связанные проекты – ЕНИП, CERIF, SPAR, CIDOC CRM, FRBR, FRBRoo.

1 Введение

На сегодняшний день в мире существует большое количество научных информационных систем. С их помощью научные организации представляют данные о публикациях сотрудников, о научных исследованиях и проектах и о полученных результатах.

Как правило, интересы пользователей не обеспечиваются какой-то одной информационной системой, и даже в рамках одной организации может существовать несколько разнородных систем. Для обеспечения их интероперабельности необходимо принять соглашения по стандартизации электронного представления информационных ресурсов и соответствующие модели и средства, поддерживающие интеграцию информационных ресурсов.

В рамках проекта "Модели, методы и средства интеграции гетерогенных научных информационных ресурсов НАН Украины" была разработана концептуальная модель научных ресурсов. В частности, была предложена модель публикации, как одного из основных ресурсов информационного научного пространства.

На начальном этапе были рассмотрены существующие на данный момент модели, описывающие информационные ресурсы в научной и смежных областях, таких как библиотековедение и культурное наследие: ЕНИП [1], CERIF[2-4], SPAR [5], CIDOC CRM [6], FRBR[7], FRBRoo[8]. За основу разработанной модели научной публикации была взята европейская формальная концептуальная модель научных данных CERIF.

Модель описывает такие основные сущности научного информационного пространства, как публикация, персона, организация, проект и событие, а также связи между ними. Наряду с основными сущностями предложен перечень возможных второстепенных. Преимуществом данной модели является поддержка многоязычности и возможность использования различных классификационных схем.

2 Обзор существующих моделей

ЕНИП. В рамках российской инициативы по организации Единого Научного Информационного Пространства (ЕНИП) была разработана модель [1], в которой было выделено четыре основных группы информационных сущностей: участники научной деятельности, научная деятельность, результаты научной деятельности и документы и публикации. К классу Документ относятся разного рода документы и публикации, как печатные, так и цифровые. Класс Публикация в данной модели является подклассом класса Документ. Данный абстрактный класс описывает метаинформацию об официально зарегистрированных печатных изданиях (публикациях). Публикации делятся на 3 группы: издания аналитического уровня, монографического уровня и сводного уровня. При описании конкретных публикаций необходимо указывать конкретные неабстрактные классы, такие как Монография, Многотомник, Выпуск журнала и пр.

CERIF. Появление формальной концептуальной модели научных данных CERIF (Common European Research Information Format), в дословном переводе – Общий европейский исследовательский информационный формат [2-4] стало важным результатом в технологии семантического структурирования контента академических научных сред, и научных электронных библиотек в частности. Модель CERIF предназначена для поддержки управления научно-исследовательской информацией, а также настройки и взаимодействия между исследовательскими информационными системами. Модель имеет следующие основные характеристики: (а) она построена с помощью сущностей или объектов с атрибутами, таких как Публикация, Проект, Персона, Организация, Событие и т.д.; (б) она поддерживает связи многие-ко-многим и рекурсивные связи; с помощью сущностей-связей, которые также обладают определенными свойствами, обеспечивается богатая семантика, включая роль и темпоральность; (с) она поддерживает многоязычность, (г) она является расши-

ряемой, что обеспечивает гарантированную совместимость, по крайней мере, на уровне ядра, но и не исключает более широкую совместимость. Модель предназначена для использования при обмене данными в гетерогенных распределенных средах. Использование так называемого семантического слоя делает модель гибкой и масштабируемой для применения в разнородных средах.

Модель CERIF постоянно развивается, только в 2012 году вышло три ее новых версии 1.3, 1.4 и 1.5. Заявлено о выходе в ближайшее время CERIF 2.0.

CIDOC CRM. Онтология в области документации в сфере культурного наследия CIDOC CRM ("Committee on Documentation" "Conceptual Reference Model") [6] представляет собой формальную онтологию, предназначенную для описания информации в области культурного наследия. Разработка модели началась в 1996 году, а в 2006 она стала стандартом ISO 21127:2006 – "Эталонная онтология для обмена информацией культурного наследия" (A reference ontology for the interchange of cultural heritage information). Модель покрывает контекстной информацией исторического, географического и теоретического характера отдельные экспонаты и музейные коллекции в целом.

Версия 5.0.1 модели CIDOC CRM состоит из 90 классов и 148 свойств (бинарных отношений), связывающих классы между собой и описывающих предметы, понятия, людей, события, место, время и их отношения. Кроме того, классы организованы в иерархии через отношение «является», («is a»). Все классы и свойства модели можно разделить на три группы. Первая группа включает классы и отношения, охватывающие наиболее общие понятия окружающего мира: постоянные и временные сущности, отношения участия, зависимости, совпадения во времени. Вторая группа содержит понятия, частично поддерживающие функции управления: приобретение и учет единиц хранения, передача прав собственности на объекты культуры. К третьей группе относятся классы и свойства, используемые для внутренней организации самой онтологии: средства, необходимые для подключения внешних источников терминов, например, тезаурусов по отраслям культуры.

Обобщая все многообразие музейных коллекций и задач, модель содержит широкий набор универсальных понятий. Важную роль здесь играют временные сущности, так как они связывают объекты (концептуальные или физические) с временным диапазоном, местом и субъектами.

FRBR. Международной федерацией библиотечных ассоциаций и учреждений (International Federation of Library Associations and Institutions, IFLA) разработана модель "Функциональные требования к библиографическим записям" (Functional Requirements for Bibliographic Records, FRBR) как обобщенное представление библиографического универсума [7]. В настоящее время IFLA продолжает контролировать приложения модели FRBR и поддерживает ее использование и развитие.

FRBR включает описание концептуальной модели (сущности, их отношения и атрибуты), предлагает универсальные библиографические записи для всех типов материалов и пользовательских задач, связанных с библиографическими ресурсами, описанными в каталогах, библиографиях и других библиографических инструментах. Модель FRBR различает три группы сущностей:

- для описываемых объектов: Произведение (work), Выражение (expression), Воплощение (manifestation), Экземпляр (item);
- для описателей-субъектов: Персона (person) и Организация (corporate body);
- для описателей-объектов: Концепт, Объект, Событие и Место (concept, object, event, place).

Большое внимание в модели уделено отношениям между сущностями. Помимо логических связей в модели выделена группа так называемых контентных связей (для первой группы сущностей). Они идентифицируют основные типы отношений, которые существуют между экземплярами сущности одного типа (например, сущности Произведения) или между экземплярами разных типов сущностей (например, сущностей Произведение и Воплощение). Например, в группе отношений Произведение-Произведение выделены такие типы отношений: имеет адаптацию (свободный перевод); имеет приложение (сходство, соответствие), имеет продолжение; имеет резюме (обзор, аннотацию); имеет преобразование (стихотворную форму); имеет имитацию (пародию). В группе отношений Выражение-Выражение перечислены следующие типы отношений: имеет сокращение (корректировку, уплотнение); имеет пересмотр (исправленную редакцию, расширенную редакцию); имеет перевод (буквальный перевод) и некоторые другие типы отношений, касающиеся музыкальных произведений. И, наконец, отношения часть/целое и часть в части также представлены в модели FRBR.

FRBRoo. Объектно-ориентированная FRBR [8] моделирует результаты (Произведение, Выражение...) процессов, таких как создание, реализация, планирование, но не сами процессы. FRBRoo, используя подход CIDOC CRM, фокусируется на процессах. Такой подход позволяет принимать во внимание обстоятельства, при которых, например, конкретные произведения были задуманы или реализованы.

В результате в FRBRoo объекты Произведение, Выражение и Воплощение были разбиты на несколько классов со специфическими свойствами. Так, в FRBRoo присутствует класс Произведение, но также декларированы подклассы Индивидуальное Произведение (Individual Work), Составное Произведение (Complex Work), Сопроводительное Произведение (Container Work), Совокупность Произведений (Aggregation Work), Серийное Произведение (Serial Work), Издательское Произведение (Publication Work), Исполнительское Произведение (Performance Work), Записанное Произведение (Recording

Work). Класс Произведение является суперклассом, объединяющим подклассы как частные случаи, каждый из которых имеет свою специфику создания или составления.

FRBRoo следует рассматривать как интерпретацию FRBR, главное новшество FRBRoo – реалистичная, явная модель процесса интеллектуального творчества, которая еще должна получить свое дальнейшее развитие для библиотечарей и ученых.

SPAR. Уже много лет ведутся исследования в области анализа семантики связей между научными материалами. Системным обобщением этих результатов стало появление комплекса онтологий SPAR (Semantic Publishing and Referencing) [5], обеспечивающего достаточно детальную категоризацию отношений, которые могут возникать между научными материалами в электронном виде, и воплощающих их связей. SPAR включает онтологии семантической публикации:

- FaBiO (FRBR-aligned Bibliographic Ontology). Библиографическая онтология совместимая с FRBR, предназначенная для записи и публикации в Semantic Web библиографических научных записей.
- PRO (Publication Roles Ontology). Онтология, характеризующая роли агентов (людей, организаций и вычислительных средств; например, автор, редактор, обозреватель, издатель, библиотекарь) в процессе публикации. Позволяет записывать информацию даты/времени о том, когда исполняются эти роли.
- PSO (Publications Status Ontology). Онтология, характеризующая статус документа и других библиографических сущностей на различных этапах процесса публикации (например, представленные рукописи, отклонение рукописи, прием рукописи, версия записи).
- PWO (Publications Workflow Ontology). Онтология, характеризующая основные этапы в рабочем процессе, связанном с публикацией документа (например, в рассмотрении, в XML формате, дизайн страницы, публикация в Интернете).

и онтологии ссылок:

- CiTO (Citation Typing Ontology). Онтология для описания природы цитат в научно-исследовательских статьях и других академических работах, как из других публикаций, так и из информационных ресурсов Web, и для публикации этих описаний в Semantic Web.
- BiRO (Bibliographic Reference Ontology). Онтология для описания библиографических записей и ссылок, и их сбор в библиографические сборники и списки литературы.
- C4O (Citation Counting and Context Characterization Ontology). Онтология, позволяющая характеризовать библиографические ссылки с точки зрения их количества и контекста.

- DoCO (Document Components Ontology). Онтология, характеризующая составные части библиографического документа. Предоставляет структурированную лексику компонентов документа (например, введение, обсуждение, благодарность, список литературы, рисунки, приложения).

Онтологии FaBiO, CiTO, BiRO и C4O в настоящее время уже доступны для использования, а DoCO, PRO, PSO и PWO находятся в разработке.

3 Основные характеристики разрабатываемой модели

Разрабатываемая модель публикации основана на европейской формальной концептуальной модели научных данных CERIF. Перечислим ее наиболее важные черты.

Базовые сущности. Основу или ядро модели составляют несколько базовых или главных сущностей (объектов) в данной предметной области – единого научного информационного пространства. Это – Публикация, Организация, Персона, Проект и Событие. Для каждой из этих базовых сущностей определяются наборы собственных атрибутов или свойств.

Связи. Экземпляры базовых сущностей могут быть связаны между собой как внутри одной сущности (рекурсивные связи), так и со всеми остальными сущностями связями типа многие-ко-многим. В связи всегда участвует пара сущностей, и эта пара также обладает своими, только ей (этой паре) присущими свойствами. Поэтому в модель вводятся так называемые сущности-связи, например, Организация_Персона, Публикация_Персона, которые также обладают свойствами, например такими свойствами, как роль и темпоральность.

Многоязычность. Модель поддерживает многоязычность, для этого в наборах собственных атрибутов каждой сущности выделяется группа атрибутов, например, название, ключевые слова, аннотация, которые могут иметь несколько значений на разных языках. Такие атрибуты в рамках данной модели также превращаются в отдельные сущности.

Классификация. Как правило, структурированные описания какой-либо сущности начинаются с уникального идентификатора сущности и далее информации о типе данной сущности, которая заносится в поля, названные, например, как Тип, Класс, Категория или Классификация. Все эти поля используются для терминов, которые объявляют, что сущность – член специфического класса или категории сущностей. Представляемая модель не дает строгой привязки к определенным классификациям либо категориям, она только лишь может рекомендовать к использованию те или иные. Для использования классификаций разных типов и их взаимных отображений в данной модели предусматривается набор сущностей под общим названием Классификация. Все базовые сущности модели могут быть

связаны с определенными классификациями отношением типа многие-ко-многим. Для разрешения такой связи для каждой базовой сущности вводится связующая сущность, например Персона_Классификация, Публикация_Классификация. Роль в сущностях-связях (например, Публикация_Персона) также определяется классификационными схемами, для этого в каждой такой сущности предусмотрены атрибуты идентификатора класса и схемы классификации, где определен этот класс, а также атрибуты темпоральности (дата начала, дата окончания) и доли (дня некоторых типов связей). Таким образом, обеспечивается определенная гибкость в выборе классификационных схем и отсутствие строгой привязки к обозначенным форматам.

Сущности второго уровня. Далее определяется набор так называемых сущностей второго уровня – Патент, Продукт, Оснащение, Финансирование, Метрики и т.д. По аналогии с базовыми сущностями, для них также определяются наборы собственных атрибутов, многоязычных атрибутов-сущностей и сущностей-связей. Любая из базовых сущностей может иметь связь с какой-либо из сущностей второго уровня, для них также создаются свои сущности-связи (например, Публикация_Метрики, Публикация_Финансирование).

Наличие семантического слоя. Набор сущностей-связей с привязкой к классификационным схемам и их классам, а также наличие в данной модели фактически вспомогательной или метасущности "Классификация" гарантируют обеспечение богатой семантики, так называемого семантического слоя данных. Использование семантического слоя делает модель гибкой, масштабируемой и применимой в разнородных средах.

Данная модель может быть использована при обмене данными в гетерогенных распределенных средах.

4 Базовые сущности модели и их связи

Ядро модели научного информационного пространства составляют пять базовых сущностей или сущностей первого уровня (рис. 1): Научная публикация – центральная сущность модели, Организация, Персона, Проект и Событие.

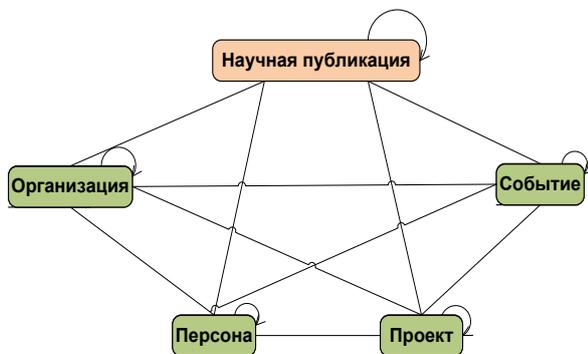


Рис. 1. Базовые сущности модели научной публикации

Экземпляры базовых сущностей могут быть связаны между собой как внутри одной сущности (рекурсивные связи, на рисунке показаны петлей), так и со всеми остальными сущностями отношением типа многие-ко-многим. Поэтому в модель вводятся так называемые сущности-связи (рис. 2), поскольку в связи всегда участвует пара сущностей, и эта пара также обладает своими свойствами.

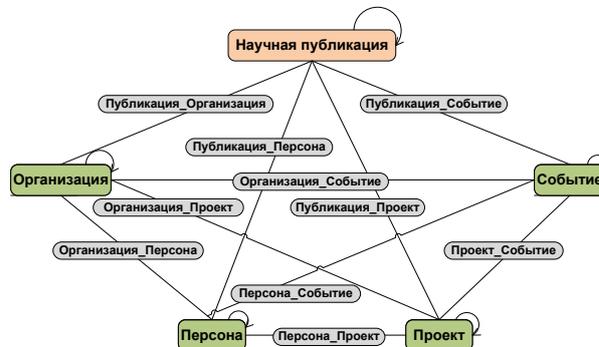


Рис. 2 Сущности-связи, соединяющие базовые сущности

Базовые сущности представляют научные субъекты (персоны и организации) и их основную деятельность (проекты и публикации): сотрудничество одних ученых с другими (Персона_Персона), их участие в проектах (Персона_Проект), аффилирование с организациями (Организация_Персона), публикация статей (Публикация_Персона), проведение мероприятий типа семинаров и конференций (Организация_Событие). Проекты включают людей (Персона_Проект) и организации (Организация_Проект). Научные публикации публикуются организацией (Публикация_Организация) и относятся к проектам (Публикация_Проект), в проектах могут принимать участие целые организации (Организация_Проект), которые могут использовать отдельных людей (Организация_Персона).

Для управления типом этих связей и определения ролей, используются ссылки на классификационные схемы. Они определяются в так называемом семантическом слое. Например, Персона «является автором» Публикации. Для этой связи (Публикация_Персона) могут быть определены и другие роли. Модель семантического слоя учитывает охват любой схемы классификации или структуры.

На рис. 2 показаны сущности-связи для базовых сущностей (кроме рекурсивных, например, Организация_Организация или Публикация_Публикация, они показаны петлей). Базовые сущности могут быть связаны и с другими сущностями (второго уровня) например, Публикация «профинансирована по» Программе финансирования (эта сущность второго уровня в модели кратко названа Финансирование). О сущностях второго уровня, а также о классификации речь пойдет ниже.

5 Классификация

Семантический слой является простым, но мощным инструментом, позволяющим представлять виды связей (типы, роли и т.д.) [3], прикладные представления, предметные классификации (например, терминологии) и другие классификационные схемы (например, классификация языков, стран и валют в соответствии с определенными стандартами), а также установление соответствия между ними.

Все сущности модели, базовые и второго уровня, могут быть классифицированы или типизированы в соответствии с определенными классификациями (рис.3).

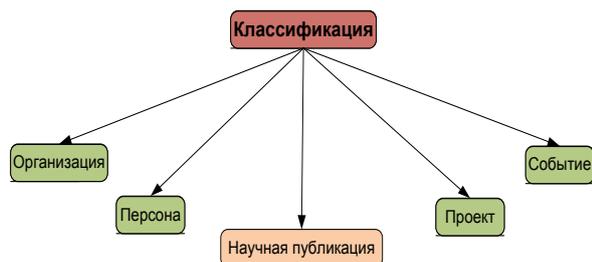


Рис.3 Сущность классификация в базовой модели

Связи сущностей с классификацией также представляются сущностями-связями – Организация_Классификация, Персона_Классификация, Публикация_Классификация, Проект_Классификация и Событие_Классификация. В данных сущностях с помощью атрибутов указывается идентификатор самой сущности и идентификаторы класса и схемы классификации, а также атрибуты темпоральности (дата начала, дата окончания), доля (для некоторых типов связей).

Классификация сущностей строго не привязана к конкретным стандартам, таким образом, могут использоваться различные классификационные схемы. Так, например, по классификации, предложенной в проекте CERIF, сущность Публикация может быть следующих типов: книга, монография, диссертация, учебное издание, журнальная статья, труды конференции и т.д. [3].

По классификации, предложенной ЕНИП [1], все публикации на первоначальном абстрактном уровне классифицируются на издания сводного, монографического и аналитического уровня, а затем конкретизируются как, например, серийное издание, многотомник, труды института, периодическое издание и т.д.

Сущность Персона можно классифицировать, например, по научной степени (доктор наук, кандидат наук), по занимаемой должности (младший научный сотрудник, старший научный сотрудник, ведущий научный сотрудник), по академическим или ученым званиям (профессор, академик, доцент и пр.).

Сущность Событие можно классифицировать как: конференция, семинар, симпозиум, событие культуры, политическое событие, спортивное

событие, выставка, торговая выставка-ярмарка (модель CERIF).

Научные организации (сущность Организация) могут классифицироваться как научно-исследовательские организации, научные организации образовательных учреждений высшего профессионального образования, опытно-конструкторские, проектно-конструкторские, проектно-технологические и иные организации, осуществляющие научную и научно-техническую деятельность.

Сущность Проект можно классифицировать, например, по уровню разрабатываемых проблем (фундаментальные, прикладные); по структуре разрабатываемых проблем (тематические, комплексные) и т.д.

Различные классификации одной и той же сущности могут частично соответствовать друг другу. Так, например, класс Книга по классификации ГОСТ соответствует классу Book по классификации CERIF. В рамках разработки концептуальной модели научных данных CERIF [2], ведутся работы по развитию стандартизированной формальной семантики для отображения отношений между объектами научных информационных систем CRIS (Current Research Information Systems).

6 Сущности второго уровня

Помимо основных сущностей, которые представляют участников научного процесса и их действия, модель охватывает контекст участников и их взаимодействия в более широком диапазоне данной предметной области, так называемыми сущностями второго уровня. На рис. 4 показаны базовые сущности и предложен возможный перечень сущностей второго уровня и их связи, для удобства на рисунке сущности второго уровня продублированы для каждой базовой сущности. Связи между базовыми и сущностями второго уровня в модели представлены сущностями-связями.

Ко второму уровню относятся следующие сущности:

Патент – охраняемый документ, удостоверяющий исключительное право, авторство и приоритет изобретения, полезной модели либо промышленного образца.

Продукт – результаты научной деятельности: интернет-системы, базы данных, экспериментальные данные, математические модели, программные системы, технологии и т.д.;

Финансирование – выделение денежных средств или финансовых ресурсов на выполнение научных работ;

Цитирование – включение в научную работу фрагмента текста другой работы с указанием точных выходных данных позволяющих найти работу и место, из которого был взят фрагмент текста;

Метрики – мера, позволяющая получить численное значение некоторого свойства какого-либо объекта.

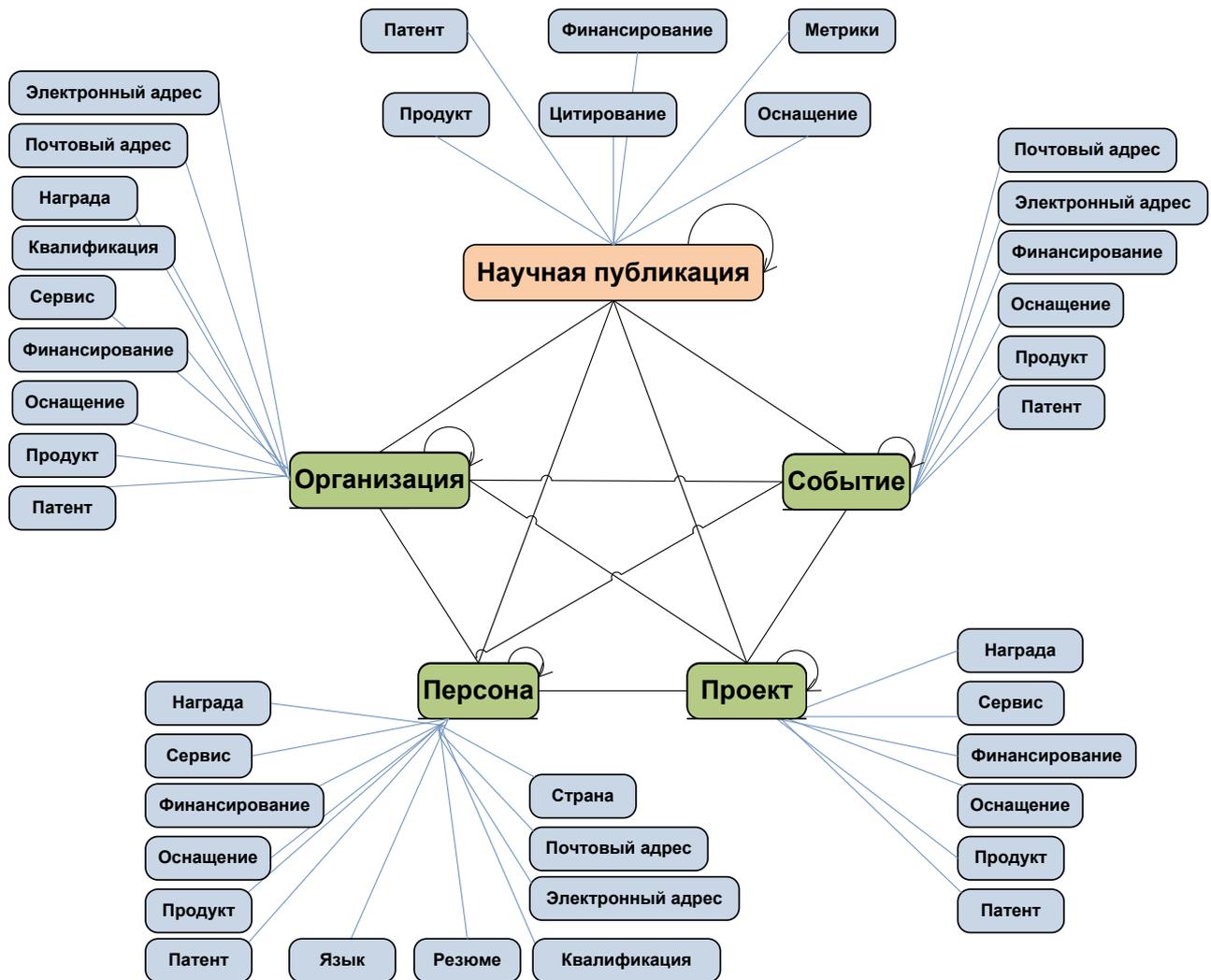


Рис.4 Сущности второго уровня и их связи с базовыми сущностями

Оснащение – совокупность технических средств или приспособлений, которыми оснащено что-либо. В частности, лабораторное оборудование – различные инструменты и снаряжение, используемое учёными, работающими в лаборатории для выполнения экспериментов или осуществления измерений. Лабораторное оборудование подразделяется на общелабораторное, измерительное, специализированное, испытательное и аналитическое.

Сервис – совокупность средств для обслуживания пользователей, процесс оказания услуг.

Награда – что-либо, выдаваемое в качестве поощрения за заслуги, отличия, в знак благодарности за что-либо. К наградам относят: почётные знаки, призы, премии, почётные грамоты, ценные подарки и почётные звания.

Квалификация – опыт и навыки, степень профессиональной подготовленности к выполнению определенного вида работы. Почтовый адрес – текст в

установленной форме, в котором указываются имя и местонахождение получателя.

Страна – гражданство персоны.

Электронный адрес – адреса вида имя_пользователя@имя_домена, используемый в технологии электронной почты.

Язык – язык, которым владеет объект коммуникации.

Резюме – документ, содержащий информацию о навыках, опыте работы, образовании и другой относящейся к делу информации, обычно требуемый при рассмотрении кандидатуры человека для найма на работу.

Сущности второго уровня могут быть классифицированы или типизированы в соответствии с определенными классификациями. Так, например, Финансирование можно классифицировать как [3]: Программа финансирования, Запрос (источник финансирования еще не найден), Тендер, Подарок.

Сущности-связи между базовой и сущностью второго уровня формируются также как и для пары базовых. Ее двойное название всегда указывает на то, между чем и чем создана такая связь, например сущность-связь Персона_Финансирование, подразумевает, что создана связь между базовой сущностью Персона и сущностью второго уровня Финансирование. Соответственно в ее первичный ключ будут входить идентификаторы от каждой из этих сущностей. За идентификаторами далее следуют так называемые семантические идентификаторы класса и классификационной схемы и метки времени (даты начала и окончания). Все вместе идентификаторы, семантические ссылки и метки времени составляют первичный ключ сущности-связи, как это показано на рис. 5.

| ИмяСущности1_ИмяСущности2 |
|----------------------------|
| ИД Сущности1 |
| ИД Сущности2 |
| ИД Класа |
| ИД Классификационной схемы |
| Дата начала |
| Дата окончания |

Рис. 5 Композиция первичных ключей в связующих сущностях

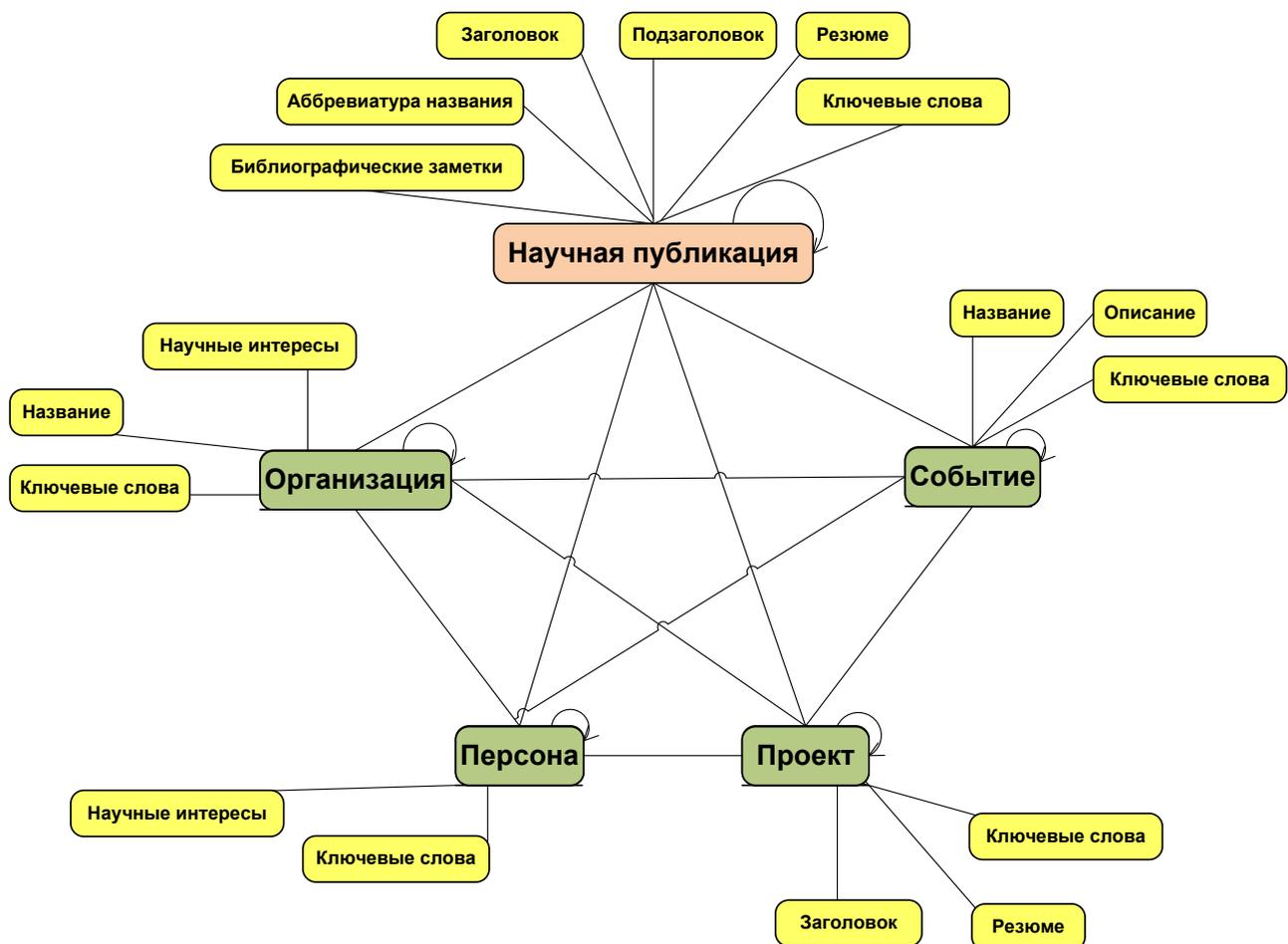


Рис. 6 Поддержка многоязычности для базовых сущностей

7 Многоязычные атрибуты стали сущностями

Много информации в исследовательских средах нуждается в представлении на больше чем одном языке. Как показано на рис.6, модель содержит много сущностей, зависимых от языка, например, Ключевые слова, Резюме, Научные интересы, Названия, Заголовок, Библиографические заметки, Описания и

др. Семантический слой также учитывает многократные языковые представления.

Помимо классификации базовых и сущностей второго уровня, объекты классификации также учитывают языковые ссылки (Код языка) и многие другие классификации. Таким образом, можно поддерживать и сами схемы классификации на нескольких языках. Эта свойство особенно важно для стран, где говорят на нескольких официальных языках. Названия страны, например, можно теперь представить на

разных языках: België (du), Belgien (ge), Belgique (fr), Belgium (en).

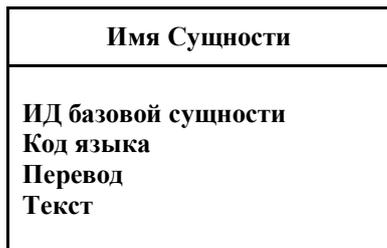


Рис.7 Набор атрибутов многоязычных сущностей

На рис. 7 показано, как идентифицируется каждая зависимая от языка сущность, во-первых, идеентификатором той сущности, с которой она связана (например, ИД Публикации); во-вторых, ссылкой на языковой код (Код языка), в-третьих, типом перевода (Перевод); все эти три поля вместе составляют первичный ключ. Далее за первичным ключом идет – текстовое значение, которое содержит текст на

различных языках (ключевое слово, резюме, описание и т.д.)

8 Описание базовых сущностей модели и их концептуальные схемы

Все сущности представляемой модели имеют набор описательных атрибутов. Те из них, которые требуют многоязычности, представлены как отдельные сущности. На рис. 8 изображены базовые сущности со своими атрибутами, многоязычные атрибуты выделены подчеркиванием. Каждая базовая или сущность второго уровня имеет идентификатор и URI (UniformResourceIdentifier – унифицированный идентификатор ресурса), содержащий ссылку на (постоянный) публичный ресурс, а также прочие описательные атрибуты.

Опишем подробнее сущность Научная публикация (для краткости назовем ее просто Публикация) и все ее связующие сущности.

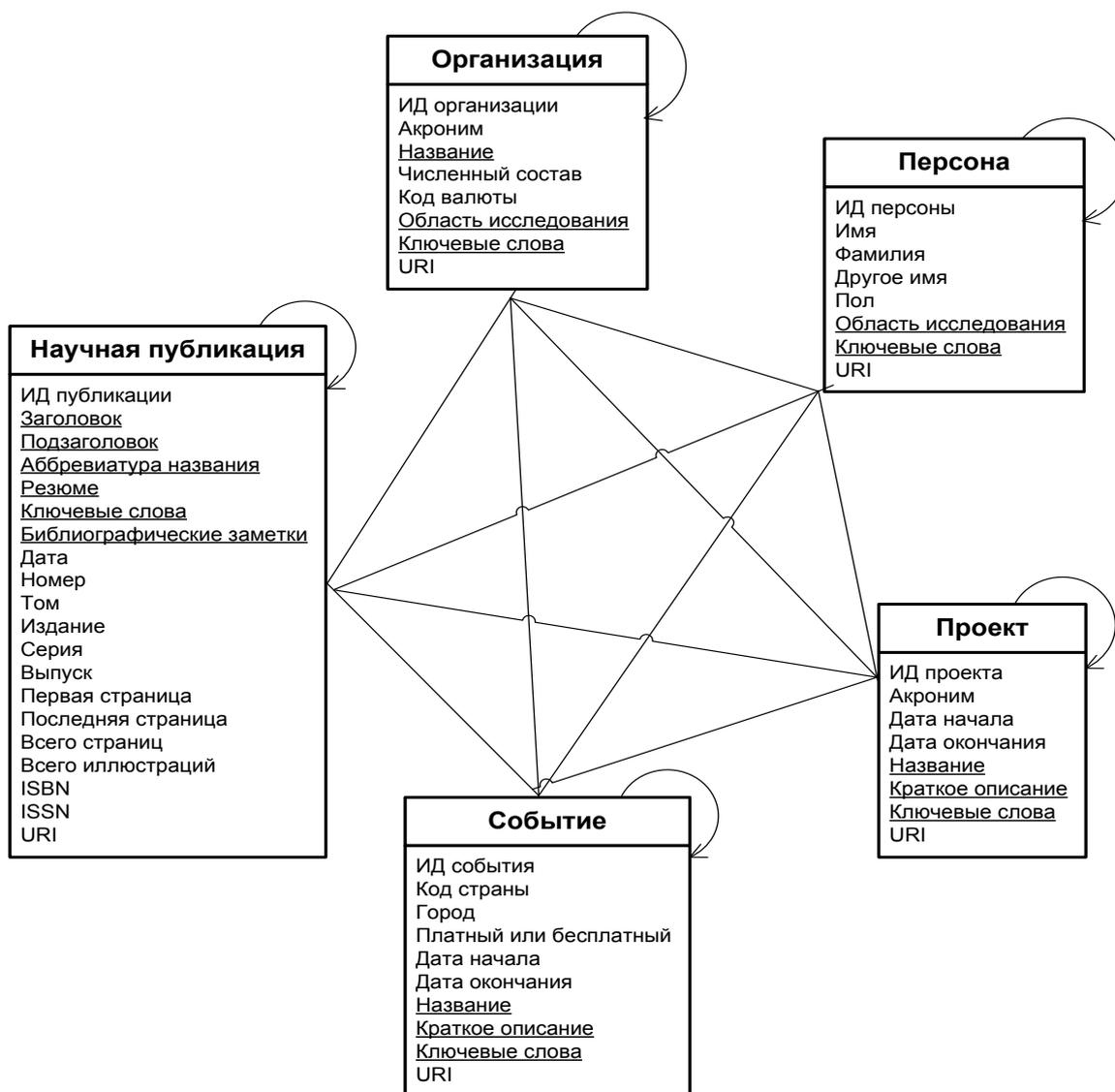


Рис.8 Описательные атрибуты базовых сущностей

9 Публикация

Согласно [10] Издание или Публикация – документ, прошедший редакционно-издательскую обработку, полученный печатью, тиснением или другим способом, содержит информацию, предназначенную для распространения, и соответствует требованиям Государственных стандартов, других нормативных документов по их издательскому оформлению и полиграфическому исполнению. Под документом понимается материальный объект, содержащий информацию, закрепленную созданным человеком способом для передачи во времени и пространстве. Издание, наряду с печатным текстом может содержать аудиозаписи или изображения на других материальных носителях или имеющее сопроводительную вспомогательную информацию, называется комбинированным.

В частности Научное издание или Научная публикация это – издание, содержащее результаты теоретических и (или) экспериментальных исследований, а также научно подготовленные к публикации памятники культуры, исторические документы и литературные тексты.

Для сущности Публикация, с учетом связи между базовыми сущностями, а также того факта, что сущность Публикация сама по себе может иметь одну или несколько классификаций, в модели

научной публикации выделим шесть связующих базовых сущностей:

- Публикация_Классификация;
- Публикация_Публикация;
- Публикация_Персона;
- Публикация_Организация;
- Публикация_Проект;
- Публикация_Событие.

Все сущности-связи содержат атрибуты для определения семантики связей (Класс и Схема классификации) и темпоральности (Даты начала и окончания).

9.1 Семантика сущности-связи Публикация_Классификация

Для классификации сущности Публикация может использоваться одновременно несколько различных классификационных схем. Как было упомянуто выше (см. раздел 4) можно использовать классификации, предложенные в проектах CERIF [3] и ЕНИП [1]. На основании рассмотренных выше классификаций научных публикаций, а также ГОСТ (ДСТУ [9-10]), действующего в настоящее время в Украине, нами предлагается новая классификация (см. рис. 9).

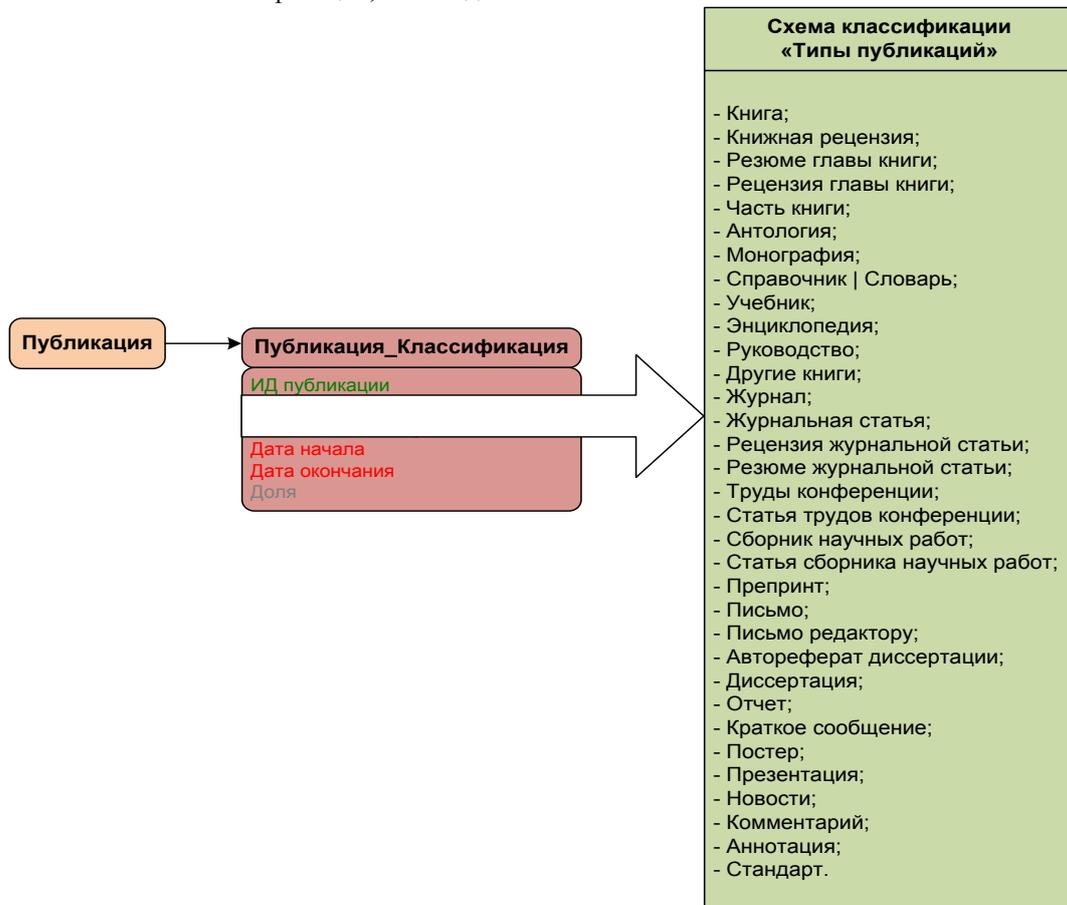


Рис.9 Классификация видов публикаций

На рис. 9 представлена связующая сущность Публикация_Классификация, где перечислены возможные классы предложенной классификации «Типы публикаций».

9.2 Семантика сущности-связи

Публикация_Публикация

Между сущностями Публикация и Публикация в [3] выделено только два отношения:

- Производный – одна публикация происходит из другой публикации.
- Является частью – одна публикация является частью другой публикации.

Этот перечень связей может быть существенным образом расширен. В различных научных дисциплинах были предприняты попытки разработать более подробную категоризацию отношений между научными текстами. Наиболее известными результатами этих попыток являются онтология SWAN (Semantic Web Applications in Neuromedicine) [12], а также CiTO (Citation Typing Ontology) [13], DoCo (Document Components Ontology) [14] и др. В дальнейшем все эти частные разработки были систематизированы, дополнены и объединены в единый комплекс под названием SPAR (Semantic Publishing and Referencing Ontologies) [5], включающий взаимосвязанную совокупность онтологий различного назначения.

В [11] разработана оригинальная категоризация связей Публикация_Публикация исходя из их применения в научной практике. Так, например, была выведена группы связей:

- научного вывода – заимствует основополагающие идеи (obtain background from CiTO), развивает (updates CiTO), подтверждается (cites as evidence CiTO), подтверждает (confirms CiTO), уточняет (qualifies CiTO), исправляет (corrects CiTO);
- использования – содержит утверждения/факты (contains assertion from CiTO), использует

данные из (uses data from CiTO), использует метод из (uses method from CiTO), опровергает (refutes CiTO), совершает плагиат (plagiarizes CiTO);

- мнений и оценок – позитивно оценивает (agrees with CiTO, supports CiTO, credits CiTO, respondsPositively-to SWAN, consistent with SWAN), негативно оценивает (critiques CiTO, disagrees with CiTO, responds negatively to SWAN, inconsistent with SWAN, disputes CiTO, parodies CiTO, ridicules CiTO), нейтрально оценивает (responds neutrally to SWAN);
- иерархические и ассоциативные связи – расширяет (extends CiTO), шире (broader SKOS), сужает (narrower SKOS), родственный (related SKOS, relevantTo SWAN), альтернативен (alternativeTo SWAN);
- между компонентами одного произведения – дублирующая копия, новая редакция, ранняя редакция, аудио/видео версия текстового, текстовая версия аудио/видео, презентация текстового, часть (isPartOf DoCo, paragraph DoCo, part DoCo, section DoCo), включение как части (hasPart, DoCo), абстракт (abstract DoCo), оглавление (table of contents DoCo), предисловие или введение (foreword DoCo, preface DoCo), приложение к (appendix DoCo), библиография (bibliography DoCo), глоссарий (glossary DoCo).

9.3 Семантика сущностей-связей Публикации с базовыми сущностями

Как сказано выше, сущность Публикация связана с другими базовыми сущностями с помощью сущностей-связей, для которых должна быть определена семантика.

На рис. 10 представлен вариант классификации сущности-связи Публикация_Персона:

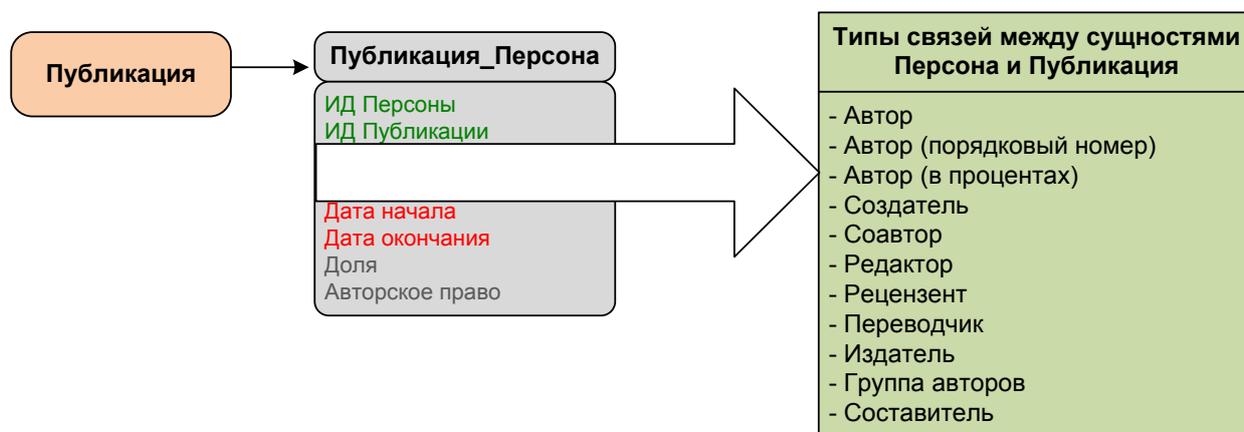


Рис. 10 Семантика (роли) в связи Публикация_Персона

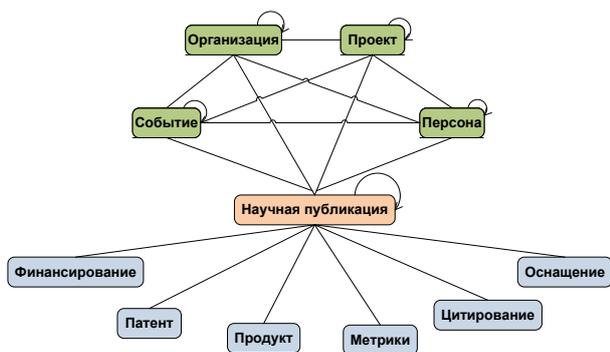


Рис. 11 Сущности первого и второго уровней модели научная публикация и связи между ними

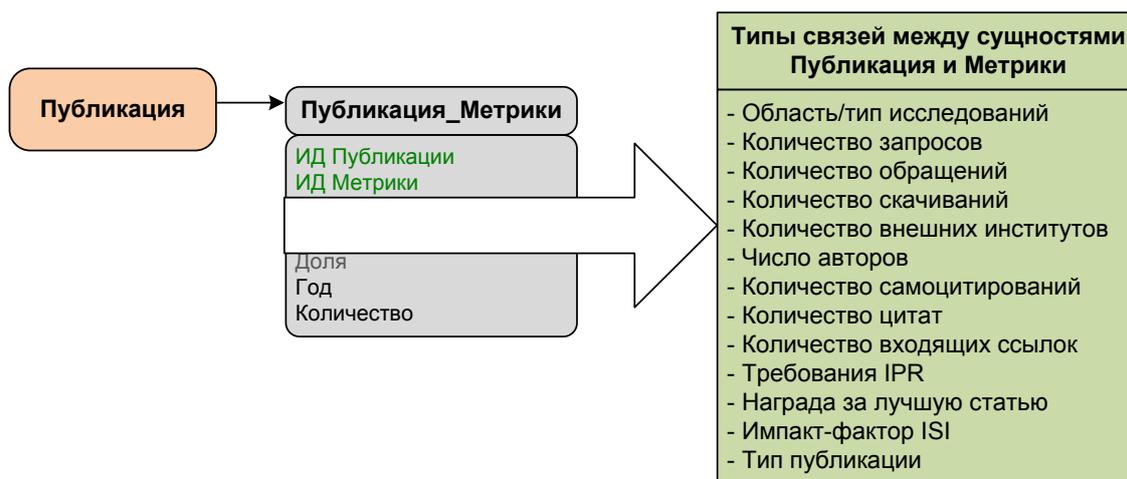


Рис. 12 Семантика (роли) в связи Публикация_Метрики

10 Заключение

В работе была рассмотрена, разработанная в рамках проекта "Модели, методы и средства интеграции гетерогенных научных информационных ресурсов НАН Украины", концептуальная модель научных ресурсов. В частности, была предложена модель публикации, как одного из основных ресурсов информационного научного пространства.

Также были рассмотрены модели, описывающие информационные ресурсы в научной и смежных областях, таких как библиотковедение и культурное наследие: ЕНИП, CERIF, SPAR, CIDOC CRM, FRBR, FRBRoo.

Модель описывает такие основные сущности научного информационного пространства, как публикация, персона, организация, проект и событие, а также связи между ними. Наряду с основными сущностями предложен перечень возможных второстепенных. Преимуществом данной модели является поддержка многоязычности и возможность использования различных классификационных схем.

Предложенная модель научной публикации может найти применение в научно-исследовательских и библиотечных информационных системах, для поддержки стандартизации, интеграции и обмена научными знаниями. Поскольку данная модель

9.4 Связи Публикации с сущностями второго уровня

Публикация связана с некоторым набором сущностей второго уровня (рис. 11): Финансирование, Патент, Продукт, Метрики, Цитирование, Оснащение.

Семантика связей определена в соответствующих сущностях-связях: Публикация_Финансирование, Публикация_Патент, Публикация_Продукт, Публикация_Метрики, Публикация_Цитирование, Публикация_Оснащение.

На рис. 12 приведен пример возможной семантики для сущности-связи Публикация_Метрики [3]:

основана и совместима с широко применяемой в настоящее время европейской формальной концептуальной моделью научных данных CERIF, она обеспечит легкую интеграцию отечественных систем в общее научное пространство.

Литература

- [1] Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН // Вычислительный центр РАН, г. Москва – 2006. – 238с. – Библ. 61 наим.
- [2] CERIF 2008 - 1.2 Full Data Model (FDM). Introduction and Specification http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/CERIF2008_1.2_FDM.pdf
- [3] CERIF 2008 – 1.2 Semantics, euroCRIS. http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_Semantics.pdf
- [4] CERIF 2008 – 1.2 XML Data Exchange Format Specification http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_XML.pdf
- [5] David Shotton. 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/>
- [6] Crofts N., Doerr M., Gill T., Stead S., Stiff M. (editors), Definition of the CIDOC Conceptual Reference Model, January 2008. Version 4.2.4.
- [7] Functional Requirements for Bibliographic Records, Final Report / IFLA Study Group on the Functional Requirements for Bibliographic Records. – München: K.G. Saur, 1998. (UBCIM Publications, New Series; v. 19).
<http://archive.ifla.org/VII/s13/frbr/frbr.htm>
- [8] Doerr M., Leboeuf P. Modelling intellectual processes: The FRBR–CRM harmonization // In Conf. Proc. of ICOM-CIDOC Annual Meeting. Gothenburg, Sweden. – 2006. – p. 10–14.
- [9] Издания. Основные виды. Термины и определения. ГОСТ 7.60-90.
- [10] Видання. Основні види. Терміни та визначення. ДСТУ 3017—95 (Укр.)
- [11] Паринов С.И., Когаловский М.Р. Технология семантического структурирования контента научных электронных библиотек // Труды XIII Всероссийской научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции», Воронеж; 19-22 октября 2011 г. – стр. 197-206 – Воронеж, 2011.
- [12] SWAN (Semantic Web Applications in Neuromedicine) - Scientific Discourse Relationships Ontology Specification.
<http://swan.mindinformatics.org/spec/1.2/discourserelationships.html>
- [13] David Shotton. CiTO, the Citation Typing Ontology. J. of Biomedical Semantics 2010, 1(Suppl 1): S6. <http://www.jbiomedsem.com/content/1/S1/S6>
- [14] David Shotton, Silvio Peroni. DoCO, the Document Components Ontology. 17/02/2011.
<http://purl.org/spar/doco/>

Conceptual model of research publication

Valeriy Reznichenko, Galina Proskudina,
Olga Ovdii

The paper concerns resource conceptual model creation for research and digital libraries information systems, particularly the publication model as one of main resources of a research environment. Also well-known connected projects are discussed such as ЕНИП, CERIF, SPAR, CIDOC CRM, FRBR, FRBRoo.