



ВОЛОХОВ Вадим Маркович - доктор физико-математических наук, заведующий отделом вычислительных и информационных ресурсов Учреждения Российской академии наук «Институт проблем химической физики» РАН
Адрес: 142432, Московская область, Ногинский район, г. Черноголовка, проспект академика Семенова, 1
e-mail: vmt@icp.ac.ru



ПРОХОРОВ Андрей Иванович - научный сотрудник Учреждения Российской академии наук «Институт проблем химической физики» РАН
Адрес: 142432, Московская область, Ногинский район, г. Черноголовка, проспект академика Семенова, 1
e-mail: aipro@icp.ac.ru



АМОСОВА Елена Сергеевна - младший научный сотрудник Учреждения Российской академии наук «Институт проблем химической физики» РАН
Адрес: 142432, Московская область, Ногинский район, г. Черноголовка, проспект академика Семенова, 1
e-mail: alene2000@mail.ru

Онтологическая модель предметной области информационной системы поддержки инновационных разработок институтов РАН

Введение

В последнее время использование онтологий для моделирования предметных областей автоматизированных информационных систем получает все более широкое распространение [1, 2]. Наиболее часто такой подход применяется для интеллектуальных систем [3], в частности, предназначенных для функционирования в сети Интернет. Это связано с тем, что онтологическая модель позволяет разработать модель метаданных, что значительно улучшает использование системы широким кругом пользователей с точки зрения организации взаимодействия.

Онтология - это структура, описывающая значения элементов некоторой системы, попытка структурировать окружающий мир, описать какую-то конкретную предметную область в виде понятий и правил, утверждений об этих понятиях, с помощью которых можно формировать отношения, классы, функции и пр. Онтологии предметных областей ограничиваются описанием мира в рамках конкретной предметной области.

Задача построения онтологической модели предметной области информационной системы для поддержки коммерциализации результатов инновационных разработок в научных исследованиях является актуальной и сложной научно-практической задачей. Сложность поставленной задачи определяется, в частности, наличием множества межпредметных и междисциплинарных связей и различными целями конечных пользователей системы: ученых, экспертов, бизнесменов, политиков, сотрудников общественных и коммерческих организаций.

Целью настоящей работы является разработка и создание онтологической модели предметной области информационной системы для поддержки коммерциализации результатов научных исследований.

Краткий обзор стандарта CERIF 2008

Во всех странах мира проводится множество различных научных исследований, причем схема проведения научных исследований в разных странах похожа. Как правило, сначала осуществляется стратегическое планирование, потом объявляется программа исследований, проходит поиск предложений, подходящие предложения принимаются к работе, результаты исследований отслеживаются, анализируются и впоследствии используются в тех или иных целях.

Исследования в одной и той же области знаний могут производиться одновременно в нескольких научных организациях, в том числе и в одной стране. Кроме того, в век глобализации исследовательские организации одной страны могут опираться в своей работе на результаты, полученные в других странах. Поэтому важно обеспечить обмен полной и достоверной информацией, наборами данных между различными странами и фондами на всех этапах проведения исследований, начиная с этапа подачи заявки и заканчивая этапом публикации рецензии на инновационную разработку. Проблема стандартизации данных научных исследований возникла еще в 80-е годы прошлого века, и в качестве решения этой проблемы сначала появлялись варианты обобщения схем баз данных для хранения результатов научных исследований, на основе которых позднее возник стандарт CERIF (Common European Research Information Format - общий европейский формат для исследовательской информации) [4].

Моделированием предметной области научных исследований на основе этого стандарта в последние 14 лет в Евросоюзе активно занимается организация euroCRIS. Основные свойства этого стандарта:

1) стандарт поддерживает концепцию объектов или сущностей с атрибутами: например, таких как проект, человек, организация;

2) стандарт поддерживает n:m отношения между объектами, используя «связывающие отношения», и таким образом обеспечивает богатую семантику, включающую роли и временные характеристики;

3) стандарт полностью интернационален с точки зрения языкового или знакового набора;

4) стандарт расширяет без повреждения основной модели данных, что предоставляет возможность оперирования на основном уровне, не препятствуя еще более широкому взаимодействию.

Основные объекты в стандарте CERIF - это Person, OrganisationUnit и Project (Человек, Организация и Проект), каждый из которых рекурсивно связан сам с собой и поддерживает отношения с другими объектами. Стандарт описывает множество дополнительных объектов, с помощью которых полностью описываются исследовательские проекты, их участники, результаты их совместной работы и пр. Семантика данных задается на специальном семантическом уровне, в таблицах, описывающих возможные роли и взаимодействия между отдельными объектами.

Взаимоотношения между проектом, человеком, организацией показываются в стандарте CERIF при помощи специальных связей, и их считают одной из сильных сторон модели CERIF. Связь всегда соединяет два объекта. Все связи строятся по одинаковой схеме: они наследуют названия и идентификаторы от объектов-родителей и дополнительно обладают атрибутами даты начала и конца действия связи, в каждой связи отражается семантика через ссылку на семантический слой CERIF посредством специальных идентификаторов. Таким образом, все возможные взаимоотношения между проектами, людьми и организациями задаются с помощью этих связей, а характер взаимоотношений подчиненности (кто чей автор, кто чей подданный, что часть чего и т.д.) показывается благодаря семантическому слою, в котором все эти роли расписаны.

Для отображения результатов научной деятельности в стандарте CERIF предусмотрены специальные объекты: ResultPublication, ResultPatent, ResultProduct (Публикация, Патент, Продукт). Помимо основных и результирующих объектов в CERIF также используется множество так называемых объектов второго уровня, таких как: FundProg - программа финансирования, Event - событие, Prize - вознаграждение, Facil - средства обслуживания, Equip - оборудование и т.д. Объекты второго уровня позволяют представить контекст исследования посредством связей с основными и результирующими объектами.

Модель CERIF поддерживает возможность многоязычности для имен, названий, описаний, ключевых слов, обобщений и даже для семантики. Используемый язык хранится в атрибуте LangCode с максимум пятизначными значениями (например, en, de, fr, si, en-uk, en-us, fr-fr, fr-be, fr-nl). Атрибут Trans предоставляет информацию о типе перевода: o=original (язык оригинала), h=human (перевод человеком) или m=machine (машинный перевод). Помимо основных, результирующих и объектов второго уровня многоязычность поддерживают также и классификаторы на семантическом уровне CERIF. Таким образом, становится возможным поддерживать классификационные схемы на различных языках.

Стандарт CERIF рекомендован к использованию в системах CRIS (Current Research Information Systems - информационные системы по актуальным научным исследованиям) [5], которые собирают воедино всю информацию, лежащую в основе научных исследований. Использование подобных систем значительно облегчает взаимодействие инвесторов и исследователей. Исследовательские группы получают легкий доступ к информации, необходимой для разработки инновационных идей, руково-

дители и управляющий персонал получают возможность проще отслеживать и оценивать текущую исследовательскую деятельность, инвесторы и исследовательские советы могут оптимизировать процесс финансирования инновационных проектов.

Реальный пример использования стандарта - это портал IST World, построенный на основе стандарта CERIF. Он предоставляет информацию об экспертах, исследовательских группах, центрах и компаниях, задействованных в создании технологий для растущего информационного сообщества. Главный акцент сервиса - экспертиза и опыт основных участников процесса в европейских странах. Репозиторий содержит информацию по проектам пятой, шестой и седьмой рамочных программ Европейской Комиссии, а также информацию, связанную с этими научно-исследовательскими проектами, собранную в Болгарии, Кипре, Чехии, Эстонии, Венгрии, Латвии, Литве, Мальте, Польше, Румынии, России, Сербии, Словении, Словакии и Турции.

В России единой системы по текущим научным исследованиям не существует. Все попытки создания таких систем проходят разобщенно в рамках различных программ и проектов. В Черноголовке в рамках Российской академии наук по гранту НААВ реализуется проект, целью которого является создание и разработка информационной системы для поддержки коммерциализации результатов интеллектуальной деятельности для предоставления заинтересованным юридическим и физическим лицам данных об инновационных разработках институтов РАН с возможной их последующей коммерциализацией. В этой системе под инновационными разработками понимаются информационные образы объектов интеллектуальной собственности, технических решений, а также технологические запросы, идеи и иные нематериальные активы, полученные в результате научно-технической деятельности.

Анализируя стандарт CERIF, обнаруживаем, что он не охватывает предметные области, связанные с работой экспертов и подготовкой инновационной разработки к процессу коммерциализации. Поэтому авторами было предложено расширение модели, предлагаемой этим стандартом на указанные выше предметные области.

Инновационный процесс со структурной точки зрения представляет собой комплекс последовательно взаимосвязанных действий по созданию, освоению и распространению инновации. Инновационный процесс предполагает эволюционное изменение состояния инновационного продукта, его превращение из идеи в товар, а также мониторинг его дальнейшей рыночной судьбы.

Модель предметной области для поддержки инновационных разработок

Предметная область информационной системы для поддержки коммерциализации результатов научных исследований является суммой объединения совокупности нескольких предметных областей, а именно предметной области научных исследований, предметной области возможным областям внедрения и предметной области экспертов по коммерциализации инновационных разработок. При этом последнее слагаемое должно помогать решать следующую задачу: динамически формировать пути взаимодействия в отношении «многие-ко-многим» между первыми двумя слагаемыми.

Онтология области научно-исследовательской деятельности представляет собой структуру системы, отображающей процесс научной деятельности. Научные исследования возможны только при наличии полной и достоверной информации и наборов данных: начиная с этапа подачи заявки и заканчивая этапом публикации рецензии на разработку. Информационные системы по текущим исследованиям должны собирать воедино всю информацию,

лежащую в основе научных исследований. Подобные системы могут использоваться широким кругом лиц: от исследователей до инвесторов. Научно-исследовательские организации могут размещать через интернет информацию о своих инновационных разработках и выполнять поиск предложений потенциальных инвесторов и заказчиков, потенциальные инвесторы и заказчики могут размещать заказы на выполнение НИОКР и предложения об инвестициях в сфере высоких технологий и выполнять поиск инновационных разработок.

В предметной области по научным исследованиям можно выделить следующие основные классы (рис. 1).

Проект содержит информацию о проектах, исследованиях, результатом которых будут инновационные разработки в том или ином виде, а также об их сроках. Проекты могут быть связаны с другими проектами, связаны с людьми, организациями, патентами, публикациями, продуктами и др. объектами системы.

Организация содержит информацию об организациях, имеющих отношение к проектам. Содержит описание организации: валюту расчетов, численность сотрудников, оборот и т.д. Организации также могут быть взаимосвязаны и связаны с другими объектами системы.

Человек содержит информацию о людях, задействованных в научных проектах. Люди также могут быть взаимосвязаны и связаны с другими объектами.

Дополнительный объект **Имена** содержит информацию о различных вариантах написания имени одного человека, в том числе и на разных языках.

Публикация содержит информацию о результатах исследований в виде публикаций. Содержит выходные данные о публикации: о дате выхода, издании, серии, страницах, ISBN, ISSN, краткое содержание, комментарии и пр. Публикации могут быть взаимосвязаны и связаны с другими результатами исследований, а так-

же с другими объектами системы: проектом, организациями, людьми и т.д.

Патент содержит информацию о патентах, выданных на результаты исследований. Содержит сведения о стране выдачи патента, дате регистрации и краткое содержание. Патенты могут быть связаны с публикациями, проектами, организациями и людьми.

Продукт содержит информацию о продуктах, полученных в результате исследований, т.е. об инновационных разработках, а также описание продукта. Продукты могут быть связаны с публикациями, проектами, людьми, организациями.

Дополнительные объекты, предусмотренные стандартом CERIF, также оказываются задействованными в подсистеме: **Язык** служит для отображения информации о языке представления данных в системе, **Адрес** - для отображения информации о физических адресах людей и организаций, **ЭлектронныйАдрес** - для отображения информации об электронных адресах людей и организаций, **Страна** - для отображения информации о странах, **Валюта** - для информации о валютах, **ПрограммаФинансирования** - для информации о программе, в рамках которой выполняется проект, и т.д.

При помощи объектов семантического уровня **Класс** и **КлассификационнаяСхема** характеризуются типы отношений, формы заявлений, классификация субъектов. Например, для обозначения типов публикаций или видов продуктов и т.д.

В предметной области возможных областей внедрения можно выделить следующие классы (рис. 2).

Организация содержит информацию об организациях, заинтересованных в инвестировании в инновационные разработки, в проведении НИОКР. Содержит описание организации: валюту расчетов, численность сотрудников, оборот и т.д.

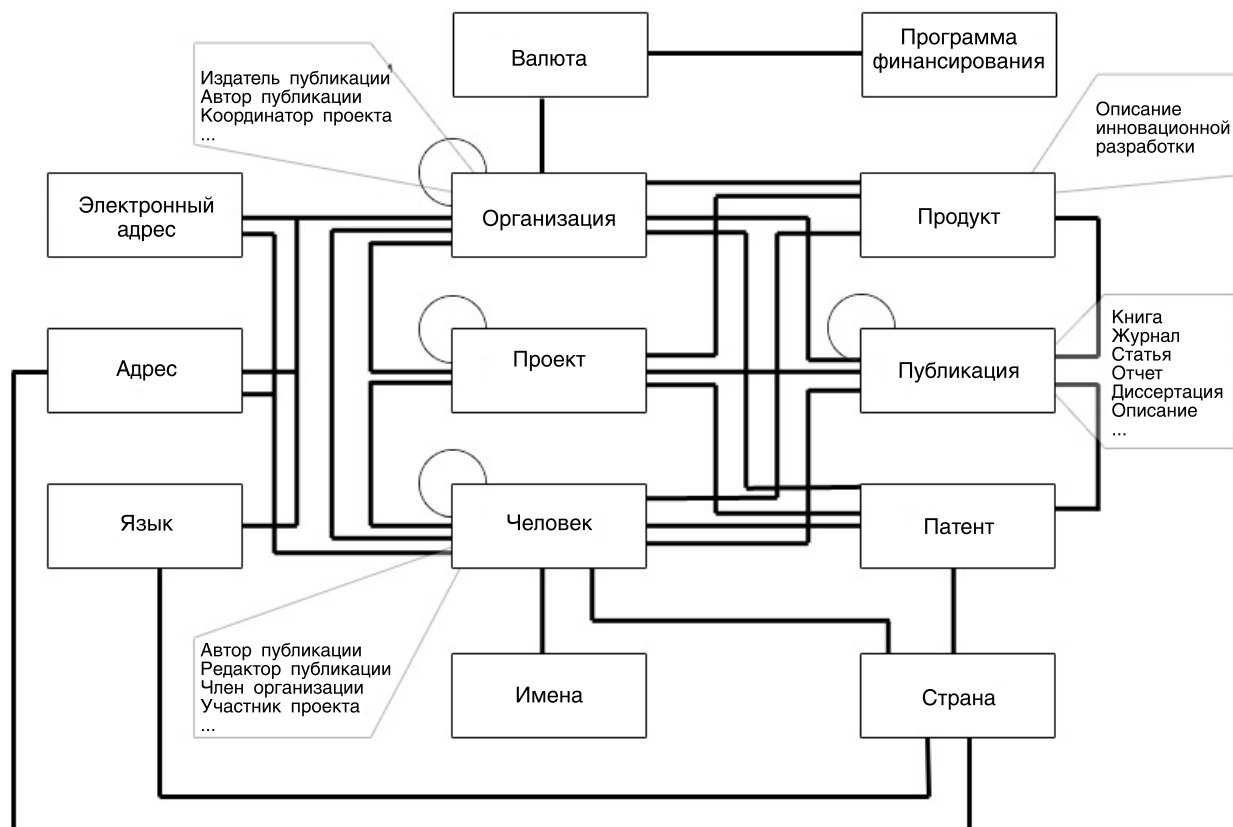


Рис. 1. Основные классы предметной области по научным исследованиям

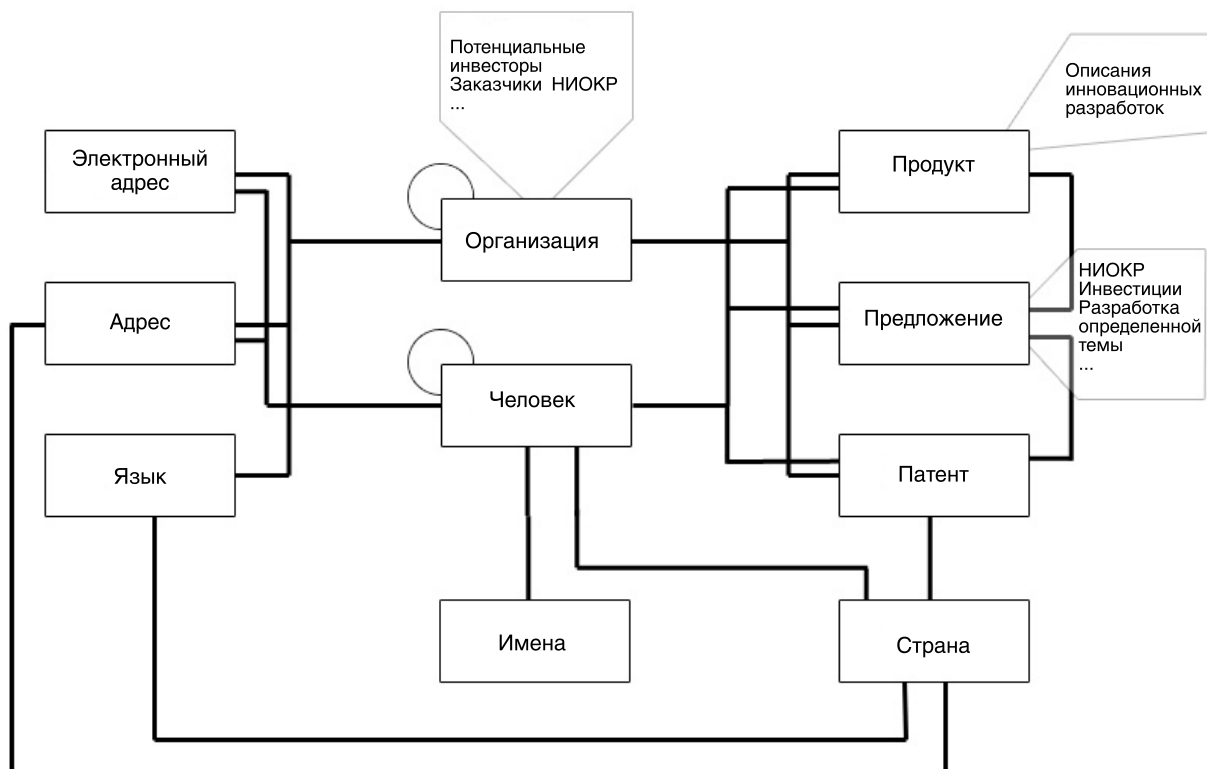


Рис. 2. Основные классы предметной области по возможным областям внедрения

Человек содержит информацию о людях, занятых в организациях, или об индивидуальных потенциальных инвесторах. Люди могут быть взаимосвязаны и связаны с другими объектами.

Для данной предметной области также применим объект **Имена**, который содержит информацию о различных вариантах написания имени одного человека.

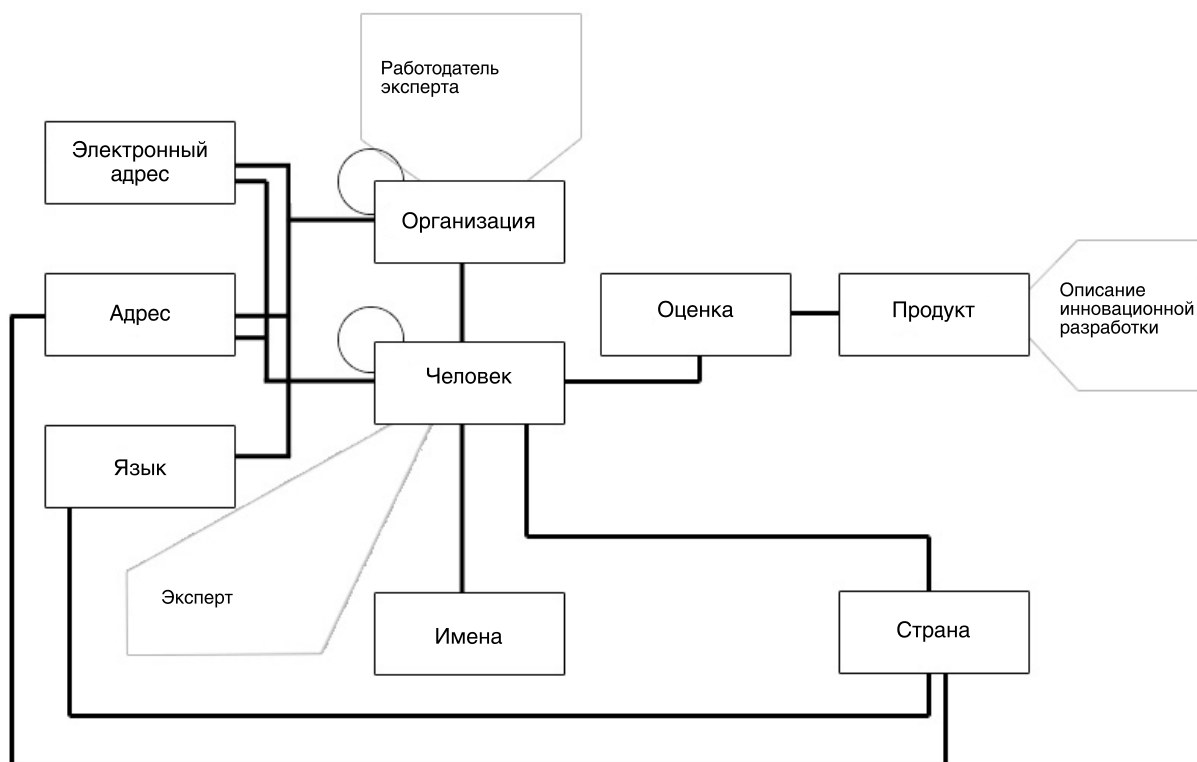


Рис. 3. Основные классы предметной области экспертов

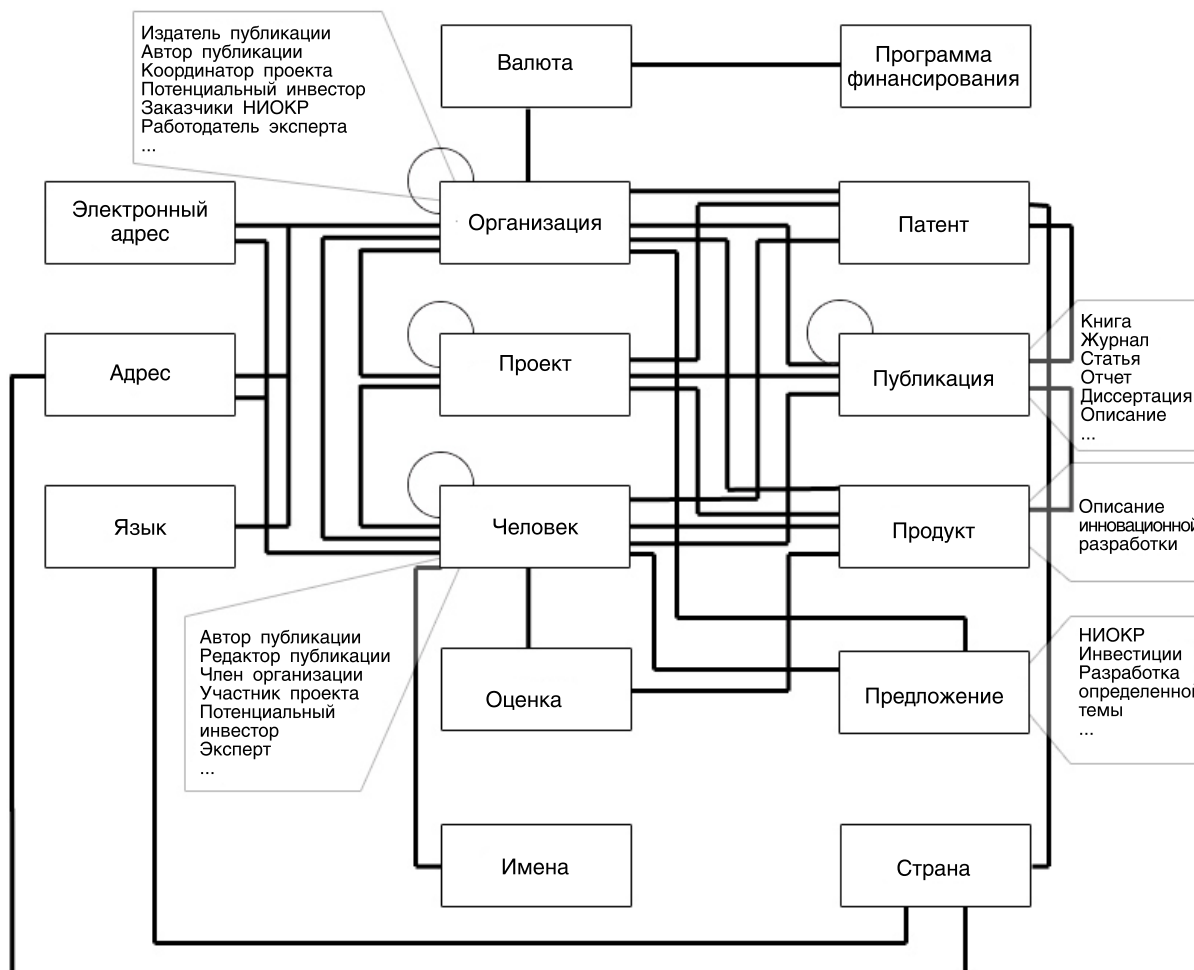


Рис. 4. Предметная область информационной системы для поддержки коммерциализации результатов научных исследований

Предложение содержит информацию о предложениях от потенциальных инвесторов на проведение НИОКР, на инвестиции, на разработку определенной темы. Содержит описания предложений, а также информацию об их сроках. Предложения могут быть взаимосвязаны, а также связаны с людьми, организациями и др. объектами системы.

Патент содержит информацию о патентах на разработки, в которые организация хочет инвестировать.

Продукт содержит информацию об интересных инвесторам продуктах.

По аналогии с предметной областью научных исследований в предметной области возможных областей внедрения можно выделить дополнительные объекты: **Язык, Адрес, ЭлектронныйАдрес, Страна, Валюта** и др.

Для характеристики типов отношений между объектами и для классификации самих объектов также можно использовать объекты семантического уровня **Класс** и **КлассификационнаяСхема**.

В предметной области по экспертной оценке возможности коммерциализации инновационных разработок можно выделить следующие классы (рис. 3).

Человек содержит информацию об экспертах, проводящих оценку и анализ инновационных разработок и выносящих решение о возможности их коммерциализации. Тот же дополнительный объект **Имена** содержит информацию о различных вариантах написания имени одного человека.

Организация содержит информацию об организациях, в которых заняты эксперты.

Продукт содержит информацию о научно-технических разработках, оценкой которых занимаются эксперты.

Отдельно можно выделить объект **Оценка** для хранения заключений экспертов о возможности коммерциализации разработок.

По аналогии с предметными областями научных исследований и возможных областей внедрения в предметной области экспертов можно выделить дополнительные объекты: **Язык, Адрес, ЭлектронныйАдрес, Страна, Валюта** и др.

Общая структура, объединяющая все три подсистемы, в полной мере отражает процесс проведения научных исследований и оценки возможности их коммерциализации (рис. 4).

Основные принципы построения информационной системы и ее пользователи

В информационной системе для поддержки коммерциализации результатов научных исследований РАН можно выделить три подсистемы: подсистема научных исследований, проводимых в институтах РАН (подсистема институтов), подсистема возможных областей внедрения (подсистема потенциальных инвесторов) и подсистема экспертной оценки возможности коммерциализации инновационных разработок (подсистема экспертов). Соответственно, в каждой подсистеме можно выделить три

группы пользователей - группа владельцев интеллектуальной собственности (исследователей), группа экспертов и группа инвесторов.

В информационной системе каждый пользователь - владелец объекта интеллектуальной собственности (исследователь) - независимо от степени завершенности своей разработки (патент, решение, идея и т. п.) может представить информацию об ОИС, о своих научно-технических разработках в виде совокупного информационного образа инновационной разработки, в которую могут быть включены резюме, технологическое предложение, информация о владельце и пр. Кроме того, он может добавить сведения о патентной защищенности своих разработок, а также разместить дополнительную информацию о них. Потенциальные инвесторы, заказчики НИОКР или их представители могут разместить в системе свои предложения об инвестициях, информацию о своих потребностях (интересах) и заказы на проведение НИОКР, на экспертную оценку инновационной разработки, проводить поиск инновационных разработок, ознакомиться с уже имеющимися экспертными оценками разработок. В системе может быть предусмотрена отдельная виртуальная площадка для экспертов, которые могут разрабатывать опросный лист (оформить технологический аудит), проводить анализ бизнес-идей и оценивать инвестиционную привлекательность инновационных разработок. Каждый пользователь информационной системы в зависимости от своих интересов име-

ет возможность проводить поиск информационных объектов и сопутствующей информации, их отбор, анализ, чтобы впоследствии принять решение о целесообразности дальнейших контактов.

Незарегистрированный в системе пользователь, используя возможности гостевого входа, может также принять заочное участие в работе информационной системы. Ознакомившись с открытыми к просмотру резюме инновационных разработок, предложениями инвесторов, оценками экспертов, он может решить, имеются ли в системе интересные его разработки или предложения для исследований, понять, по каким критериям идет оценка инвестиционной привлекательности экспертами, а затем принять решение о регистрации и дальнейшей работе в информационной системе по поддержке коммерциализации научных исследований.

Заключение

Авторы считают, что в данной работе новыми являются следующие положения и результаты: онтологическая модель предметной области информационной системы поддержки жизненного цикла инновационных разработок института РАН.

Разработанная модель позволяет разработать программную архитектуру такой системы, разработать метаданные и построить совокупность взаимосвязанных тезаурусов для поддержки семантики запросов конечных пользователей.

Литература:

1. Лалшин В.А. *Онтологии в компьютерных системах*. - М.: Научный мир, 2010. - 222 с.
2. Gruber T.R. *The role of common ontology in achieving sharable, reusable knowledge bases // Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference*. J.A. Allen, R. Fikes, E. Sandewell - eds. Morgan Kaufmann, 1991, P. 601-602.
3. Константинова Н.С. *Онтологии как системы хранения знаний [Электронный ресурс] / Н.С. Константинова, О.А. Митрофанова*. - Режим доступа: [http://www.sci-innov.ru/icatalog_new/index.php?action=send_att&entry_id=68352&fname=68352e2-st08_\(Митрофанова_О.А.\).pdf](http://www.sci-innov.ru/icatalog_new/index.php?action=send_att&entry_id=68352&fname=68352e2-st08_(Митрофанова_О.А.).pdf).

4. CERIF 2008 - 1.2 Full Data Model (FDM). *Introduction and Specification [Электронный ресурс]*. - Режим доступа: http://www.eurocris.org/Uploads/Web%20pages/CERIF2008/Release_1.2/CERIF2008_1.2_FDM.pdf.
5. Кулагин М.В., Лопатенко А.С. *Научные информационные системы и электронные библиотеки. Потребность в интеграции // Сборник трудов Третьей Всероссийской конференции по электронным библиотекам. RCDL2001 Петрозаводск, 11-13 сентября 2001 г.* - С. 14-19.

НАША ИНФОРМАЦИЯ

10 мая 2011 г. в рамках Международной выставки «Связь-Экспокомм-2011» (г. Москва) прошел День массовых коммуникаций, темой которого стали развитие масс-медиа в Рунете и перспективы развития отечественного сегмента сети.

Особое внимание было уделено актуальной проблеме авторского права на «круглом столе» «Защита авторских прав в эпоху интернета: щедры ли разбогатеет».

Дискуссия развивалась вокруг основных тенденций, существующих в Рунете на сегодняшний день: пиратство и нелегальный доступ к «бесплатному» контенту, с одной стороны, и готовность многих пользователей вносить умеренную плату при условии явной выгоды от покупки - с другой.

«Законодательство не совсем адаптировано к тому, что мы называем Интернет-эпохой», - заявил во вступительном слове председатель Комиссии РАЭК по правовым вопросам Михаил Якушев (Mail.ru Group). Директор Ассоциации Интернет-издателей Иван Засурский (зав. кафедрой новых медиа и теории коммуникации факультета журналистики МГУ) напомнил о состоявшейся 29 апреля встрече представителей Интернет-отрасли и Дмитрия Медведева. В ходе нее Президент и рунетчики активно обсуждали лицензию Creative Commons, которая может стать альтернативой решениям, существующим в отечественной практике.

О том, как проходит законодательная работа сейчас, рассказали Анатолий Семенов (координатор эксперт-

ной рабочей группы по интеллектуальным правам подкомитета по технологическому развитию Комитета ГД ФС РФ по информационной политике, информационным технологиям и связи) и генеральный директор Wikimedia.ru Станислав Козловский. Докладчики представили проекты изменений, которые необходимо внести в существующие законы, в том числе и в 4-ю часть Гражданского кодекса РФ.

Начальник отдела маркетинга компании ГК ФИЛАНКО Матвей Алексеев представил точку зрения хостинг-провайдеров на существующие проблемы. Он предложил сформировать комиссию, которая бы руководствовалась кодексом профессиональной деятельности, разработанным РАЭК.

Директор РАЭК Сергей Плуготаренко прокомментировал существующую ситуацию следующим образом: «Разрешение различных споров - весьма сложный процесс. Членами РАЭК была проделана огромная работа по формированию единой позиции отрасли. Комиссия РАЭК по правовым вопросам оптимизировала работу по аудиту текущего законодательства и разработке новых законодательных инициатив в Интернет-отрасли».

Участники выразили надежду на продолжение конструктивного диалога между отраслью и властью, а также на дальнейшую плодотворную законодательную работу в сфере авторского права. РАЭК еще раз призвала всех заинтересованных представителей Рунета к подготовке итогового документа для оптимизации законодательного регулирования отрасли.