



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 681.3.

РАЗРАБОТКА ОБЪЕКТИВНЫХ МЕТОДОВ ОЦЕНКИ КОМПЕТЕНТНОСТИ УЧАСТНИКОВ НАУЧНЫХ ПРОЕКТОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО АНАЛИЗА

Рогушина Ю.В. *, Гладун А.Я. **

**Институт программных систем НАН Украины, Киев, Украина*

ladamandraka2010@gmail.com

***Международный научно-учебный центр информационных технологий и систем НАН Украины,
Киев, Украина*

glanat@yahoo.com

Предложены объективные методы оценки компетентности участников научных и научно-образовательных проектов. Представленные методы базируются на семантическом сопоставлении описания проекта и документов, характеризующих компетентность его участником в выбранной сфере. Кроме того, предлагается использовать знания, извлекаемые из открытой среды Web – Wiki-справочники, наукометрические базы, официальные сайты организаций и сопровождающие их метаданные, онтологии предметных областей. Разработана специализированная онтология научной деятельности, позволяющая стандартизировать терминологическую базу описания квалификации участников проектов.

Ключевые слова: онтология, тезаурус, компетентность.

Введение

В настоящее время в различных сферах научной и технической деятельности предлагается большое количество разнообразных проектов – научных исследований, издания научных трудов, проведения междисциплинарных исследований, проведения научных исследований отдельными учеными и т.д.

При их сравнении и анализе таких проектов, как правило, оценивается научный уровень самого проекта, возможности его выполнения участниками и целесообразность реализации, прогнозируется научная и экономическая эффективность предполагаемых результатов [Васенкова и др., 2011]. Проекты могут отличаться как по сложности, так и по источникам финансирования – государственным, из международных фондов, от корпораций и частных фондов.

В общем случае целесообразность финансирования проекта сводится к двум факторам – к актуальности и полезности решаемой проблемы и к прогнозу того, что данный коллектив участников способен решить поставленную проблему за указанное время.

1. Компетенции и компетентность исследователей

Правилам подготовки различных научных проектов посвящено много методических разработок [Зінченко и др., 2015] а анализ их актуальности и прогнозирование эффективности является задачей экспертов соответствующих предметных областей (ПрО) [Петровский и др., 2009].

Эксперты, оценивающие проекты, должны обладать достаточно глубокими знаниями в той сфере, к которой относятся все проекты в целом, но могут не являться экспертами во всех анализируемых подобластях каждого отдельного проекта. В то же время требуется, чтобы авторы предлагаемых проектов обладали высокой квалификацией именно в той предметной области (ПрО), к которой относится конкретный проект, а не только в данном научном направлении. Поэтому возникает необходимость в разработке средств автоматизированной оценки компетентности потенциальных участников проекта на основе сопоставления материалов, которые описывают их знания, квалификацию и навыки, с описанием самого проекта. Необходимо, чтобы в таком

сопоставлении использовались знания о ПрО проекта.

Важным фактором, влияющим на эффективность выполнения научно-исследовательской работы, является компетентность исполнителей в соответствующей проекту ПрО. При этом следует различать общую квалификацию исследователей и их компетентность в тех вопросах, которые непосредственно связаны с выполнением проекта, включая как теоретические знания в этой сфере, так и практические навыки, и опыт, необходимый для успешной реализации проекта.

Следует различать компетенцию и компетентность. *Компетенция* – это понятие, в общем случае не связанное с конкретным лицом; *компетентность* – это отношения между лицом и компетенцией, которое означает, что определенный человек владеет данной компетенцией. В данном случае компетенция – это свойство проекта, а компетентность – свойство исследователя.

В общем смысле *компетенция* – способность успешно действовать на основе имеющихся знаний и практического опыта при решении задач. Элементы компетенции – знания и навыки, жизненный опыт, способности, черты характера, интеллект и т.д., объединенные в разных конфигурациях для решения человеком конкретных проблем.

В области управления персоналом под компетенцией обычно понимают формально описанные требования к личностным, профессиональным и другим качествам сотрудника. Однако в области научных исследований понятие «компетенция» является нечетким и сильно зависимым от специфики конкретной ПрО.

В то же время результаты работы научных сотрудников обычно довольно четко формализованы и открыты для анализа: это научные публикации, патенты, отчеты, описания прикладных разработок и прочие материалы, представленные в виде естественных языковых текстов с элементами структурированных данных (таблиц и графиков) и мультимедиа. В большинстве случаев все эти материалы представлены в электронной форме, которая предполагает создание средств их автоматизированного анализа. Значительная часть научных публикаций доступна через Web.

Компетентность – уровень достижений (опыта, знаний, привычек) лица в определенной ПрО. Компетентность может быть определена на основе анализа деятельности специалиста, его осведомленности в достижениях науки и техники, его понимания исследуемых проблем и возможных путей их решения. Для количественной оценки уровня компетентности используется *коэффициент компетентности*.

Оценки, получаемые на основе анализа формальных характеристик участников проекта (такие как дипломы, количество публикаций в

указанной ПрО, стаж работы и т.п.), являются слишком приблизительными и не учитывают семантику решаемой задачи.

Поэтому разработка объективных методов оценки компетентности претендентов на основе онтологического анализа с целью прогнозирования успешности реализации научных проектов представляется важной научной проблемой, решение которой должно базироваться на использовании и обработке знаний о той ПрО, для которой планируется выполнять исследования.

2. Постановка задачи

Для обеспечения объективных методов оценки компетентности участников проекта на поддержку и участие в разработке научных и научно-образовательных проектов необходимо разработать средства сопоставления описания научно-технического проекта и сведений о его участниках, характеризующих их знания и опыт в ПрО проекта, учитывающее знания об этой ПрО.

Такое сопоставление сводится к анализу двух множеств естественно-языковых документов и связанных с ними метаданных и знаний. Рассматриваемый подход позволит интегрировать персонифицированные онтологические знания об экспертах в интересующей нас ПрО и логические методы их обработки, используя онтологический анализ и методы семантического поиска.

3. Источники сведений о ПрО и специалистах

Предполагается, что основным источником сведений о проекте является его естественно-языковое описание (заявка, техническое задание), а также дополнительные внешние информационные ресурсы – онтологии, Wiki-страницы и т.д., содержащие структурированные и семантически значимые сведения о ПрО.

Более сложно проанализировать информация об участниках проекта. Часть сведений о них четко формализована и может быть однозначно оценена без учета семантики ПрО и дополнительных знаний о проекте. В частности, это уровень образования участников, их стаж работы в соответствующей области, наличие ранее выполненных проектов. К сожалению, этих сведений недостаточно для того, чтобы определить их компетентность для конкретного проекта, особенно для новых и динамично изменяющихся ПрО. Проблема заключается в том, что научные сотрудники, получившие одинаковое образование и даже защитившие диссертации по одной специальности, в процессе своих исследований специализируются в различных вопросах.

Одним из наиболее объективных критериев оценки сферы компетентности научных сотрудников является анализ их публикаций,

представленных в Web, – научных статей, докладов, отчетов и презентаций.

Пертигентность документа (публикации) проекту зависит от таких параметров, как количество упоминаний основных терминов проекта в документе и количество основных терминов, упомянутых в проекте. Существует много работ по автоматическому определению компетентностей на основе документов.

Однако различные информационные источники в Web имеют различную оценку качества представленной в них информации.

Кроме того, важно учитывать оценку деятельности исследователей научным сообществом – через наличие ссылок на их работы, а также упоминанием их работ в учебной и справочной литературе. Кроме того, источники сведений об исследователях могут быть импортированы из баз знаний интеллектуальных приложений, обеспечивающих персонализированное информационное обслуживание пользователей – например, из семантических поисковых и рекомендующих систем [Рогущина, 2015].

Таким образом, в качестве источников сведений о сфере компетентности авторов проекта предлагается рассматривать следующие ИР:

- Тексты статей (публикаций, тезисов, монографий, учебников, технических отчетов, авторских свидетельств и прочих объектов интеллектуальной собственности, представленных в виде естественно-языковых документов), опубликованные в Web, с учетом рейтингов наукометрических баз данных;
- Тексты статей (публикаций, тезисов, монографий, учебников, технических отчетов, авторских свидетельств и прочих объектов интеллектуальной собственности, представленных в виде естественно-языковых документов), представленные самими авторами (например, тексты статей из платных журналов, промежуточные отчеты и т.п.);
- Wiki-страницы авторов проекта, содержащиеся в различных Википедиях и Wiki-справочниках, которые обеспечивают структурированное представление информации;
- Wiki-страницы организаций, с которыми связаны авторы проекта;
- описание специальностей авторов проекта (паспорт специальности для научной степени, диплома);
- отчеты о ранее реализованных проектах и исследованиях разных типов, с учетом успешности их выполнения.

Одним из наиболее значимых факторов, определяющих квалификацию научного сотрудника, являются его диплом и – при наличии ученой степени – специальность, по которой защищена диссертация. Каждая специальность имеет свой паспорт – документ, закрепляющий область

исследований, за которые может быть присуждена учёная степень кандидата или доктора наук по соответствующей специальности. В паспорте специальности заложено также определение специальности, науки, в которые входит данная специальность.

В значительной мере компетентность сотрудника определяется местом его работы. Поэтому целесообразно учитывать сведения об организациях (коллективах), в которых работают авторы проекта, и сопоставлять их с ПрО проекта. Источником информации об организациях может служить Web – значительная часть предприятий и учреждений имеет собственные сайты и порталы. Если сведения представлены на официальном сайте организации, то они должны соответствовать действительности.

Иногда полезнее использовать уже структурированные знания об организации. *Организационная онтология* – это онтология, которая отображает знание об организационной и функциональной структуре определенного субъекта экономической деятельности, то есть его основные компоненты и связи между ними. Она содержит информацию о работниках предприятия, иерархию производственных отношений между ними; ресурсы, которые используются на предприятии в процессе производства; продукцию, создание которой есть следствием функционирования предприятия, и структурные единицы предприятия и связи между ними.

4. Описание наукометрических БД

Эффективность научной деятельности может оцениваться с использованием как качественных, так и количественных показателей. В основе качественных оценок лежат выводы экспертов. Однако субъективность подобных оценок значительно снижает достоверность получаемых результатов, а отсутствие количественного выражения затрудняет их использование.

Наукометрические показатели удобны для оценки фундаментальных исследований, востребованность которых оценивают через отзывы научного сообщества на публикации с результатами исследований. Возросший интерес к наукометрическим показателям вызван в первую очередь возможностью автоматизации процесса оценивания результатов научной деятельности [Штовба и др., 2013].

Сейчас при формировании рейтинга исследователя используют различные параметры – количество публикаций – суммарное или по отдельным типам (монографии, статьи, тезисы, публикации в изданиях, входящих в список ВАК, проиндексированные в Web of Science, Scopus или Google Scholar и т.п.) и ссылок на них. Иногда учитывают объем публикаций. На их основе формируются интегральные критерии, в которых баллы за публикацию определяются ее типом и

импакт-фактором издания, в котором она опубликована. Несколько сложнее рассчитывается рейтинг автора при наличии соавторов в публикациях. Обычно считается, что вклад всех соавторов одинаков, и оценку публикацию делят на число соавторов.

Сейчас в мире издается около 25 тыс. реферируемых журналов, в которых ежегодно публикуется около 2,5 млн. статей. Эффективность научной деятельности может оцениваться с использованием как качественных, так и количественных показателей. В [Бурков и др., 2013] анализируются наиболее эффективные и распространённые характеристики продуктивности научной деятельности, в частности, индекс Хирша и импакт-фактор.

Чтобы выявить ученых, пишущих много и качественно, в 2005 г. Х. Хирш предложил новый показатель – индекс Хирша (h-индекс) [Hirsch, 2005] который представляет собой максимальное целое число h , указывающее, что автор опубликовал h статей, каждая из которых процитирована хотя бы h раз. Простота расчетов и нечувствительность к типовым приемам искусственного улучшения вышерассмотренных показателей, мгновенно сделали индекс Хирша популярным наукометрическим индикатором.

Индекс Хирша может вычисляться с использованием как бесплатных общедоступных наукометрических баз данных в Интернете, (например, Google Scholar, Elibrary.ru, ADS NASA), так и баз данных с платной подпиской (например, Scopus или ISI Web of Science); однако платные базы данных часто тоже приводят h -индекс учёных в свободном доступе. Следует отметить, что индекс Хирша имеет разные значения одного и того же ученого в зависимости от проиндексированных ИР. Индекс Хирша дает более объективные результаты при отбрасывании ссылок авторов на собственные статьи. Например, в рейтинге ученых Украины согласно индексу Хирша выполняется подсчет по базе данных Scopus с отбрасыванием самоцитирования.

Следует отметить, что индекс Хирша хорошо работает лишь при сравнении ученых, работающих в одной области исследований из-за традиций, связанных с цитированием, в различных областях наук.

Импакт-фактор показывает, сколько раз в среднем цитируется каждая опубликованная в журнале статья в течение x последующих лет после выхода. Это численный показатель важности научного журнала, который ежегодно рассчитывается Институтом научной информации (Institute for Scientific Information, ISI) и публикуется в журнале Journal Citation Report. Импакт-фактор позволяет по формальным признакам сравнивать разные журналы и исследовательские группы.

Как правило, расчет импакт-фактора основан на трехлетнем периоде. Импакт-фактор журнала A за год A вычисляется по формуле

$$\text{Imp}(A, x) = \frac{\text{Cit}(A, x - 2, x) + \text{Cit}(A, x - 1, x)}{\text{Pub}(A, x - 2) + \text{Pub}(A, x - 1)}, \quad \text{где}$$

$\text{Cit}(A, y, z)$ – количество цитирований в течение года z статей, опубликованных в журнале A в году y в публикациях журналов, которые отслеживаются Институтом научной информации (СИИ), а $\text{Pub}(A, y)$ – количество публикаций в журнале A за год y .

Индекс цитирования – это принятая в научном мире мера значимости научной работы какого-либо ученого или научного коллектива, которая представляет собой суммарное количество ссылок в проиндексированных работах на рассматриваемые публикации. Он отражает реакцию научного сообщества на приведенные в ней результатами исследований, т.е. уровень их востребованности другими учеными. Как правило, слабые и вторичные работы не цитируют. Цитируемость зависит не только от уровня научных результатов, но и от других факторов, например, своевременности. Для более точных оценок при определении индекса цитируемости не учитывают самоцитирование или цитирование соавторами, а также повторные цитирования одной работы одним и тем же ученым.

Для получения этих оценок используют наукометрические базы данных (НМБД) – библиографические и реферативные БД с инструментами для отслеживания цитируемости статей, опубликованных в научных изданиях.

Наиболее известна НМБД Scopus издательской корпорации Elsevier – библиографическая и реферативная база данных, индексирующая 18 000 научных изданий. Издания, которые индексируются этой НМБД, должны удовлетворять ряду условий – иметь англоязычное название и публиковать англоязычные версии аннотаций всех статей на своем Web-сайте; публиковать новые выпуски не реже одного раза в год; обладать авторитетностью (с учетом научного авторитета членов редколлегии), популярностью и доступностью и предусматривать контроль за качеством публикаций (например, научное рецензирование).

Наукометрический аппарат Scopus обеспечивает статистику цитируемости публикаций ученых и учреждений. В Scopus не используется понятие импакт-фактора, но широко применяется индекс Хирша. База данных доступна в условиях подписки через веб-интерфейс (<http://www.scopus.com>). Кроме того, авторы могут без регистрации просматривать свою страницу <http://www.scopus.com/search/form/authorFreeLookup.uri>

Не менее популярная НМБД *Web of Science* (WoS) компании Thomson Reuters объединяет реферативные базы данных публикаций в научных

журналах и патентов, в том числе базы, учитывающие взаимное цитирование публикаций.

Она содержит ссылки на полные тексты в первоисточниках и списки всех библиографических ссылок, встречающихся в каждой публикации, что позволяет в краткие сроки получить самую полную библиографию по интересующей теме. Одним из ключевых понятий ее наукометрического аппарата платформы является импакт-фактор научного издания. Доступна только для подписчиков.

Менее известна международная НМБД *Index Copernicus* (Польша) (<http://www.indexcopernicus.com>), которая включает индексирование, ранжирование и реферирование журналов, а также является платформой для научного сотрудничества и выполнения совместных научных проектов. База данных имеет несколько инструментов для оценки производительности, позволяющие отслеживать влияние научных работ и публикаций отдельных ученых или научных учреждений. В дополнение к оценке производительности, индекс Коперникус предлагает также традиционные реферирования и индексирования научных публикаций. Для использования требует регистрации.

Наиболее полно представленные в Web научные работы проиндексированы в *Google Scholar* (<http://scholar.google.com/>) – свободно доступной поисковой системе, которая индексирует полный текст научных публикаций всех форматов и дисциплин. Система обеспечивает поиск и формирование корректных ссылок в различных форматах. При поиске можно учитывать время создания и частоту цитирования документов. Основным наукометрическим показателем, который генерирует эта НМБД – индекс Хирша (как общий, так и за последние пять лет).

Существуют и национальные НМБД, ориентированные на индексацию и оценку публикаций на языках, отличных от английского. Например, Web-сайт «Украинский индекс научного цитирования» (<http://uincit.uran.ua>) предназначен для сбора, обработки и предоставления доступа к данным по показателям активности индивидуальных и коллективных субъектов научной деятельности Украины.

Предполагается, что сайт позволит просмотреть показатели публикационной активности как отдельных ученых, так и научных учреждений Украины, узнать показатели цитируемости их работ и ключевые наукометрические показатели.

Информация о публикации и показатели их цитируемости получают сайтом с внешних наукометрических источников, среди которых: международный реестр ученых ORCID, наукометрические платформы Web of Science и SciVerse Scopus, научно-издательская инфраструктура «Научная периодика Украины». На

данный момент сайт работает в тестовом режиме и не обеспечивает заявленные возможности.

При определении рейтинга ученых для определения возможного участия в Международных инвестиционных проектах, например проектах Рамочной Программы ЕС (Horizon-2020, TEMPUS, Erasmus, TESIS, DAAD, EUREKA, CRDF и др.) одним из важных аспектов выбора участников для международных проектов является, кроме его рейтинга научных трудов, его участие и опыт в предыдущих международных проектах, что является гарантией своевременного выполнения запланированных рабочих пакетов – WP [Бедоев, 2015]. Эту информацию также можно найти в Интернете на специализированных сайтах Еврокомиссии.

5. Использование семантических Wiki-ресурсов

В различных Wiki-проектах широко используются категории и семантические свойства. Чтобы связать страницу с определённой категорией, достаточно добавить в ее код имя категории в двойных квадратных скобках. Для добавления семантических свойств, задающих связи между страницами (и соответствующими понятиями ПрО) используется следующая конструкция: [[имя свойства::имя страницы]], которая обеспечивает формирование семантических троек “текущая страница”(свойство)”указанная страница”. Например, конструкция [[столица::Украина]] на странице “Киев” указывает, что Киев является страницей Украины.

В различных Wiki-справочниках представлены страницы организаций и – реже – их сотрудников. Если авторы проекта указывают на свою связь с такими страницами, то анализ их семантических свойств позволяет извлечь из них знания об этих объектах и – при наличии дополнительных навыков – пополнить соответствующую онтологию, классами которой являются категории, экземплярами – страницы, а отношениями – семантические свойства страниц. В частности, можно выявить связи с понятиями онтологии научной деятельности и онтологии ПрО проекта. Например, Wiki-страница позволяет установить, что организация, в которой работают авторы, относится к Национальной академии наук, а в разделе “Сфера исследований” есть понятия, описываемые в проекте. Наличие таких соответствий является положительным фактором.

6. Роль онтологии в оценке компетенций

Значительно повысить эффективность анализа компетентности авторов проекта позволит обработка информации на семантическом уровне. При этом полезно применять как специальные знания о предметной области, к которой относится проект, так и знания о типах и структуре документов, связанных с научной деятельностью.

Именно онтологии являются на сегодня широко распространенным средством представления знаний о ПрО, позволяющим анализировать и сопоставлять компетенции экспертов и разработчиков в новых исследовательских областях [Rogushina et al., 2012].

Чтобы использовать знания ПрО, необходимо использовать онтологию этой ПрО. Поиск (или разработка) такой онтологии является задачей авторов проекта, так как именно они должны определить, в какой сфере они планируют вести исследования и добиваться новых результатов на основе имеющихся у них знаний. Если же авторы не предоставляют такую онтологию, то используется более общая онтология той области знаний, к которой относится рассматриваемый проект. Например, авторы проекта «Разработка интеллектуальной системы информационного и когнитивного сопровождения функционирования Национальной рамки квалификаций» могут предоставить онтологию «Компьютерно-ориентированные учебные среды» (рис.1), но, если они это не сделали, то будет использована онтология соответствующего научного направления «Информационные и коммуникационные технологии».

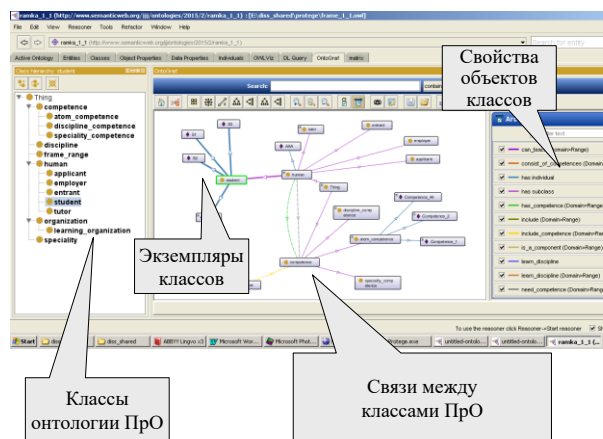


Рисунок 1 – Пример онтологической модели ПрО

Использование более общих онтологий приводит к двум проблемам: 1) необходимо обрабатывать значительно больший объем терминов (на этапе построения тезауруса проекта); 2) в проекте могут встречаться более общие, не существенные для сути проекта термины, обработка которых несколько понизит точность результирующей оценки.

Кроме того, сам факт наличия у авторов проекта онтологии, соответствующей ПрО, как правило, свидетельствует о более глубоких познаниях в выбранной области (особенно если речь идет об информационных технологиях).

Наряду с онтологиями ПрО, целесообразно применять общую онтологию научной деятельности, которая позволяет однозначно устанавливать терминологию, связанную с рейтингом публикаций, научными степенями и учеными званиями, типами организаций и т.д. Такая онтология специально разрабатывается для целей

определения компетентности авторов проекта на основе организационных онтологий научных организаций, классификатора УДК и паспортов специальностей ВАК. В такой онтологии отражаются, к примеру, такие отношения между классами, как «быть соавтором», «работать в организации», «быть автором публикации», «иметь ученую степень по специальности» и свойства «являться публикацией ВАК», «иметь индекс Хирша» (рис.2).

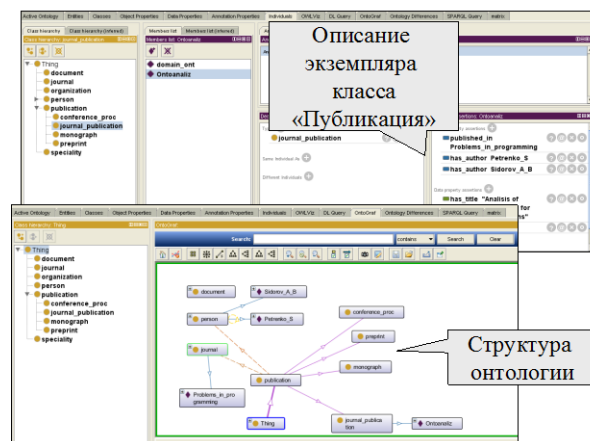


Рисунок 2. Онтология научной деятельности, используемая при анализе компетентности авторов проекта.

Эта онтология предлагается авторам проекта в качестве образца для описания и классификации предлагаемых документов, которые должны удостоверить уровень их компетентности в научной деятельности в целом и в предметной области проекта в частности.

7. Тезаурус и онтология ПрО как средство сопоставления естественных языковых документов

Для сопоставления проекта и компетенций участников проекта предлагается построить тезаурус проекта и тезаурусы всех ИР, описывающих участников проекта.

Тезаурус естественного языка ИР представляет собой проекцию онтологии, анализируемой ПрО на рассматриваемый документ. Более подробно алгоритм построения таких тезаурусов рассматривается в [Гладун и др., 2010].

Тезаурус участников определяется как объединение тезаурусов отдельных ИР. При этом следует учитывать вес отдельных ИР, в котором учитываются как значимость самого документа для описания компетенции, так и уровень доверия к самому ИР. Например, больший вес имеет автореферат диссертации, чем описание диплома, а при оценке публикаций полезно учитывать рейтинг различных журналов и конференций в наукометрических базах данных.

Можно говорить о том, что формируется корпус текстов, описывающий авторов проекта. Корпус —

не просто большое собрание текстов. Это собрание сбалансированное (тексты разного типа отобраны в определенной пропорции). Предполагается, что каждый из авторов проекта сам формирует набор документов, которые наиболее пертинентны предложенному проекту. Например, если у автора есть n научных публикаций, то он сам отбирает m из них, которые связаны с той проблемой, которой посвящен проект, причем не обязательно это должны быть статьи с наибольшим индексом цитирования или входящие в Scopus. Однако автор должен стремиться к тому, чтобы для всех понятий онтологии ПрО, для которых обнаружены лингвистические аналоги в тексте проекта, были найдены такие соответствия и в его работах (вес каждого отдельного сопоставления определяется весом ИР, определяемым как функция от статуса документа и его рейтинга).

Статус документа характеризует уровень его документальной подтвержденности, а рейтинг – его оценку наукометрическими БД.

Вначале строится тезаурус проекта Th_{proj} , который зависит от описания проекта и от выбранной онтологии ПрО. Он представляет собой множество пар (t_i, q_i) , где $t_i \in T$, T – множество терминов онтологии ПрО $O_{ПрО} = \langle T, R, A \rangle$, а q_i – количество найденных соответствий, определяющее вес этого термина (если определенный термин встретился в описании проекта 10 раз, то предполагается, что он более важен, чем тот, который встретился 2 раза). Для каждого термина онтологии ПрО предпринимается попытка найти фрагменты текста

Общая оценка компетентности коллектива авторов определяется как массив $\left(t_i, \sum_{j=1}^m R_{ИР_{j_i}} * v_{ИР_{j_i}} \right)$, где $t_i \in T$, T – множество терминов онтологии ПрО $O_{ПрО} = \langle T, R, A \rangle$, $R_{ИР_{j_i}}$ – количество найденных соответствий с этим термином в j -м ИР, а $v_{ИР_{j_i}}$ – вес j -го ИР.

Чтобы определить вес j -го ИР $v_{ИР_{j_i}}$, учитывается: 1) для научных публикаций – является ли статья публикацией ВАК (+5), материалами конференции (+1), опубликована в зарубежном издании (+3), проиндексирована Google Scholar (+2) и Scopus (+10); 2) для паспорта специальности, диплома – +5; 3) для описания профиля деятельности организации – +3; 4) для описаний ранее выполненных проектов – +3; 5) для автореферата диссертации – +3. Прочие документы учитываются с весом 1.

Важно, что в этом массиве присутствуют не все термины онтологии ПрО (которая в целом может быть значительно шире, чем непосредственно

рассматривается в проекте), а лишь те, для которых найдены соответствия в проекте.

Очевидно, что различные ИР имеют разный вес при оценке компетенций специалистов. Например, публикация в реферируемом журнале более показательна, чем выложенный в Web технический отчет. Поэтому при оценке публикаций целесообразно учитывать импакт-фактор опубликовавшего его издания.

В таком случае общая оценка компетентности авторов проекта оценивается следующим образом:

$$C = \sum_{i=1}^n q_i * \left(\sum_{j=1}^m R_{ИР_{j_i}} * v_{ИР_{j_i}} * Imp(ИР_{j_i}) \right) \quad (1),$$

где $Imp(ИР_{j_i})$ – импакт-фактор издания, опубликовавшего ИР j_i .

В дальнейшем целесообразно ввести различные нормализующие коэффициенты, позволяющие уменьшить воздействие документов большого объема, слабо насыщенных терминами ПрО. Однако это требует более детального исследования состава документов, предоставляемых авторами на экспертизу, и в значительной области зависит от специфики проводимого конкурса.

Основным недостатком оценки (1) является то, что она не использует семантику ПрО, т.е. в ней не учитываются связи между терминами онтологиями (классами и экземплярами), представленные в онтологии ПрО. Для более корректного учёта этих связей предлагается использовать следующую оценку:

$$C = \sum_{i=1}^n q_i * \left(\sum_{j=1}^m R_{ИР_{j_i}} * v_{ИР_{j_i}} \right) * s_i \quad (2),$$

где параметр s_i определяет значимость i -го термина онтологии через количество его отношений с другими терминами онтологии, также вошедшими в тезаурус проекта, с учетом семантического расстояния между ними.

Чтобы, кроме специализации в выбранной ПрО, учитывать и общую квалификацию каждого из авторов проекта, целесообразно учитывать в оценке и их рейтинг исследователей, полученный из наукометрических баз данных. В частности, предлагается использовать сведения из Google Scholar и Scopus, так как эта информация является открытой и доступна всем пользователям Web. Кроме того, это позволяет дифференцировать квалификацию отдельных авторов, а не просто суммировать их результаты. При том, что оценка качества формирования коллектива авторов не является задачей экспертов, однако наличие в нем людей с резко различающейся квалификацией может вызывать вопросы. Предлагается

использовать следующий критерий оценки x -го автора проекта:

$$C_x = \sum_{i=1}^n q_i * \left(\sum_{j=1}^m R_{IP-x_j} * v_{IP-x_j} \right) * s_i * h_x \quad (3),$$

где учитываются не все ИР, а только непосредственно связанные с ним, а h_x – *сев*vf индекса Хирша ученого из Google Scholar и Scopus.

Общая оценка квалификации участников проекта может оцениваться как сумма оценок участников либо как их нормализованная сумма. Первый подход предпочтительнее, так как знания и опыт каждого могут использоваться независимо от количества участников. Поэтому нормализованная оценка может применяться только как дополнительная.

Заклучение

Предложен новый подход к проблеме объективного оценивания компетентности в контексте новых информационно-коммуникационных технологий, характеризующихся высокой динамикой развития научных исследований, использования и обработки информационных ресурсов об экспертах, интересующей нас ПрО. Целью разработки представленных методов является объективизация процесса оценки квалификации потенциальных участников на основе учет большого количества доступных знаний об их квалификации в сфере разработки.

Предложенный подход позволяет на основе описания проекта и естественной языковой информации об участниках проекта (их публикаций, дипломов, описаний ранее выполненных проектов, сведений об их организациях и т.д.) строить тезаурус проекта и тезаурусы участников, которые базируются на онтологии предметной области проекта, сопоставлять эти тезаурусы и определять на основе этого уровень компетентности участников.

Библиографический список

- [Hirsch, 2005] Hirsch, J. E. An index to quantify an individual's scientific research output / J.E.Hirsch // Proc. of the National academy of Sciences of the United States of America, 2005, 102(46), P.16569-16572.
- [Rogushina et al., 2012] Rogushina, J., Ontology-based competency analyses in new research domains / J.Rogushina., A. Gladun // Journal of Computing and Information Technology. V.20, N. 4, 2012, P.277-293].
- [Бедоев, 2015] Бедоев, Д.Г. Инвестиционное сотрудничество РФ и ЕС в сфере информационных технологий: проблемы и перспективы / Д. Г. Бедоев // Молодой ученый, №14, 2015, С. 226-230.
- [Бурков и др., 2013] Бурков, В.Н. Параметры цитируемости научных публикаций в наукометрических базах данных / В.Н.Бурков, А. А. Белошицкий, В. Д. Гогунский // Управління розвитком складних систем, 15, 2013, С.134-139.

[Васенкова и др., 2011] Васенкова, Е.И. Сравнительный анализ эффективности научных проектов / Е.И. Васенкова, Г.М. Казляк // XX международная научно-практическая конференция «Управление в социальных и экономических системах, 2011, С.101-103. – <http://www.bsu.by/Cache/pdf/239773.pdf>.

[Гладун и др., 2010] Гладун, А.Я. Тезаурус предметной области – инструмент представления знаний при проблемно-ориентированном поиске в Web / А.Я. Гладун, Рогушина Ю.В. // Праці Міжнародної науково-технічної конференції «Штучний інтелект. Інтелектуальні системи», 2010, Т. 2, С.78-90.

[Зінченко и др., 2015] Зінченко, О.С. Грантрайтинг: метод. рек. для органів публічної влади щодо написання проектних заявок / О.С. Зінченко, О.В. Кулініч, П.Ю. Куліш, за заг. ред. О.В. Кулініча. – Харків, 2015. — 80 с.

[Петровский и др., 2009] Петровский, А.Б. Построение интегральных показателей оценки результативности научных проектов / А.Б. Петровский, Г.В. Ройзензон, И.П. Тихонов // Intelligent Support of Decision Making / Ed. by K. Markov, A. Voloshyn, K. Ivanova, I. Mitov, –No. 10, Sofia: FOI ITHEA, 2009, P. 59-66.

[Рогушина, 2015] Рогушина, Ю.В. Разработка средств персонификации интеллектуальных Web-приложений / Ю.В. Рогушина // Материалы V международной научно-технической конференции «Открытые семантические технологии проектирования интеллектуальных систем» OSTIS-2015, Минск, БГУИР, 2015, С.265-270.

[Штовба и др., 2013] Штовба, С.Д. Обзор наукометрических показателей для оценки публи-кационной деятельности ученого / С. Д. Штовба, Е.В. Штовба // Управление большими системами: сборник трудов, (44). – <http://belmapo.by/assets/files/naukometricheskie-pokazateli-dlya-ocenki-deyatelnosti-uchenogo.pdf>.

DEVELOPMENT OF OBJECTIVE ONTOLOGY BASED METHODS FOR COMPETENCE EVALUATION OF SCIENTIFIC PROJECTS AUTHORS

Rogushina J., Gladun A.

Institute of software systems of National Academy of Sciences Ukraine, Kiev, Ukraine

ladamandraka2010@gmail.com

International Research and Training Center of Information Technology and Systems of National Academy of Sciences Ukraine, Kiev, Ukraine

glanat@yahoo.com

Objective methods of competence evaluation of scientific and educational projects researchers are proposed. These methods are based on the semantic mapping of project description and documents characterizing the competence of its researchers in selected subject domain. In addition, it is proposed to use the knowledge acquired from the Web open environment – Wiki-encyclopedias, scientometric database, the official sites of the organizations and their metadata, domain ontologies. Specialized ontology of scientific activity which allows to standardize the terminological base for researcher qualifications describing is designed. In addition, it is assumed ontology projects.

Keywords: ontology, thesaurus, competence.