

УДК 658.512:004

ИСТОЧНИКИ ДАННЫХ ДЛЯ РАЗРАБОТКИ СИСТЕМЫ ПОДДЕРЖКИ ТРИЗ НА ОСНОВЕ ОНТОЛОГИЙ

© 2016 М.В.Гришин¹, В.Г. Тронин²

¹ АО «Ульяновское конструкторское бюро приборостроения»

² Ульяновский государственный технический университет

Статья поступила в редакцию 11.03.2016

Статья посвящена формулированию задачи по применению онтологического подхода в системе поддержки решения изобретательских задач на этапе патентного анализа с рассмотрением различных источников данных.

Ключевые слова: ТРИЗ, онтология, поиск, патент, РИНЦ, таксономия

Теория решения изобретательских задач (ТРИЗ), созданная Г.С. Альшуллером, предназначена для повышения эффективности творческой деятельности, изначально основывалась на анализе патентного фонда. ТРИЗ помогает в организации мышления изобретателя при поиске идеи изобретения и делает этот поиск более целенаправленным, продуктивным, способствует нахождению идеи более высокого изобретательского уровня. Были выявлены 40 типовых приемов устранения технических противоречий, затем создан алгоритм решения изобретательских задач (АРИЗ), законы развития технических систем. Созданные АРИЗ и типовые приемы решения изобретательских задач периодически пополняются.

Одним из преимуществ ТРИЗ декларировалась возможность оценки потенциальной эффективности найденного технического решения, а также возможность переноса решений между разными областями техники. Наибольший массив научно-технической информации со сформулированными решениями технических задач накоплен в патентных фондах. Актуальная задача по автоматизации патентного анализа и решения изобретательских задач с учетом отличий в терминологии может быть рассмотрена через онтологический подход.

Анализ проблемной области и постановка задачи. ТРИЗ содержит большое количество структурно-функциональных элементов, каждый из которых может быть реализован в виде отдельного программного модуля. Система поддержки ТРИЗ может содержать в своем составе такие модули как:

- физические эффекты;
- формулирование противоречий и идеального конечного результата;
- типовые приемы устранения технических противоречий;
- алгоритм перебора полевых воздействий МАТХЭМ;
- управление ресурсами;
- приемы снятия психологической инерции (метод ММЧ, синектика, оператор РВС);
- вепольный анализ;
- алгоритм решения изобретательских задач (АРИЗ) [1].

Система поддержки ТРИЗ в полном объеме предполагает обучение методам ТРИЗ, в том числе изучение материалов по определенной теме и решение

контрольных задач (тестирование). Такой способ обучения может быть реализован, например, для освоения физических эффектов, часто применяемых в решении изобретательских задач. Для освоения ТРИЗ на более высоком уровне, с применением вепольного анализа и нескольких приемов при решении одной задачи, требуется более интеллектуальная система обучения.

Основная часть системы поддержки ТРИЗ – это непосредственно формулирование и решение изобретательских задач. В НИОКР в соответствии с требованиями ГОСТ принято проводить поиск и анализ найденных ранее эффективных решений, этот этап называется «патентными исследованиями». Достаточно непростая задача провести выборку патентов с учетом различий в формулировании задач и терминологии. Описание изобретения содержит формулу полезной модели или реферат изобретения. С точки зрения ТРИЗ требуется оценить то, соответствует ли данное изобретение поставленной задаче, насколько его можно применить, возможно ли перенести эффект из другой области техники.

Вывод о том, насколько решение в конкретном патенте эффективно, можно сделать после его анализа используя инструментарий ТРИЗ. Следует разложить решение на веполь и выяснить, какое противоречие разрешалось, какие приемы и эффекты использовались. Форма описания патента унифицирована, однако для проведения патентного исследования необходимо учитывать отличия в терминологии между различными областями знаний и даже внутри одной предметной области. Для решения данных задач, а также для создания базы по использованию определенного успешного технического решения в других области техники может быть применен онтологический подход. Для оценки качества решения также необходимо составление рейтинговой шкалы для каждого изобретательского приема и эффекта (наподобие классификации МАТХЭМ), соотнесение с идеальным конечным результатом. В классическом варианте оценка эффективности патентного решения рассматривалась в рамках функционально-стоимостного анализа, однако для такой оценки требуется подробная информация по стоимости всех составляющих производственного процесса с учетом тиража изделия (сырье, амортизация оборудования, зарплата).

Рассмотрим еще один вариант определения эффективности используемых в патенте решений, а именно, через количественный анализ ссылок на патент и количество патентов по определенной тематике. Подобный анализ используется в аналитических надстройках к политематическим реферативно-

Гришин Максим Вячеславович, кандидат технических наук, инженер-конструктор отдела систем контроля. E-mail: likani7@mail.ru

Тронин Вадим Георгиевич, кандидат технических наук, доцент кафедры «Информационные системы». E-mail: v.tronin@ulstu.ru

библиографическим и наукометрическим базам данных Web of Science и Scopus. По Российским объектам интеллектуальной собственности в качестве источника данных для определения цитируемости патентов можно выбрать Российский индекс научного цитирования (РИНЦ) на основе электронной научной библиотеки eLibrary, для получения подробной информации о патенте (номер, описание, срок действия) - открытые реестры Роспатента. РИНЦ также позволяет провести количественный анализ патентов по определенной тематике, автору, ключевым словам. В качестве источника данных о частоте упоминания патентов может быть применена система поиска научной литературы Google Scholar, ее преимуществом является наличие в структурированном виде информации об объектах интеллектуальной собственности со всего мира. При наличии недостатков базы Google Scholar (эффект Мэтью), она индексирует и обеспечивает бесплатный доступ к максимальному массиву научных публикаций [2].

Для заполнения онтологического словаря необходимо выбрать источник данных по физическим эффектам и явлениям. Перечень физических явлений и эффектов был создан в рамках развития ТРИЗ в 1970-е годы. С появлением новых научных направлений издаются учебные пособия по таким эффектам с описанием иерархии эффектов и взаимосвязей при их применении [6]. В Википедии имеется набор страниц, посвященных физическим эффектам и явлениям, разобранный по областям знаний. Для первичного формирования онтологического словаря по физическим эффектам и явлениям в системе поддержки ТРИЗ может быть использован научный классификатор либо открытый пополняемый интернет-источник.

Следующим шагом должно быть составление онтологического словаря по предметной области. Ульяновский государственный технический университет является обладателем максимального количества объектов интеллектуальной собственности среди юридических лиц РФ [4]. По распределению патентной активности можно выделить несколько наиболее продуктивных научных школ: «Динамика, прочность машин, приборов и аппаратуры», «Разработка научных основ ресурсосберегающих экологизированных технологий изготовления деталей, инструментов и заготовок», «Разработка технологий нанесения ионно-плазменных тонкопленочных нанокристаллических покрытий режущего инструмента и средств технологического обеспечения операций механической обработки». Перечисленные научные школы работают по тематике машиностроения, и в университете имеется собственная база из нескольких сотен патентов, которую можно использовать для создания онтологического словаря по машиностроению. Кроме этого, ТРИЗ первоначально наиболее широко применялся для решения задач именно в машиностроении, поэтому для формирования онтологии предметной области целесообразно использовать термины машиностроительной тематики.

Для формирования онтологии по содержанию патентов может быть применено концептуальное индексирование, при котором текст индексируется не по словам, а по понятиям, обсуждаемым в данном тексте. При концептуальном индексировании выполняется построение концептуального индекса проектного документа (в данном случае описания патента), затем применяется модифицированный метод классификации концептуальных индексов (Fuzzy means,

FSM)). Предлагаемая последовательность индексации описаний патентов:

- загрузка документов;
- анализ структуры документов;
- удаление стоп-слов;
- стемминг (стемминг – процесс нахождения основы слова (термов) для заданного исходного слова);
- подсчет относительной частоты встречаемости термов;
- расчет степени выраженности понятий;
- построение концептуальных индексов для разделов и документов. [5]

В качестве дополнительного источника данных для составления онтологии предлагается использовать информацию из обозначений документов, например на основе лексического анализатора [3]. Патентный анализ на основе онтологии может быть проведен двумя способами:

- предварительный патентный анализ;
- после формулирования задачи в процессе прохождения алгоритма решения изобретательских задач.

Во втором случае мы будем иметь критерии для оценки патентных решений с учетом онтологии выявленного технического (физического) противоречия, веполь и идеального конечного результата, возможных ресурсов и других уточненных условий задачи.

Применение онтологии как инструмента для создания системы поддержки ТРИЗ. На сегодняшний день существует множество определений онтологии, рассматриваемых применительно к определенной области знаний. В самом общем понятии онтология определяет общий словарь для субъектов системы (или пользователей), которым нужно использовать знания в предметной области. С точки зрения информационных технологий (ИТ) онтология включает в себя интерпретируемые формулировки основных понятий предметной области и отношения между ними [10]. В основе любой онтологии лежит семантическая сеть, являющаяся ее концептуальной схемой и определяющая взаимосвязь понятий и терминов предметной области. Исходя из того, что идея о создании системы поддержки ТРИЗ (см. п. 1) базируется на анализе и поиске однотипных понятий в патентах, авторами и выбирается принцип построения поиска на основе онтологий. Обобщенно схема поиска представлена на рис. 1.

Здесь стоит отметить, что каждый словарь Lb_n состоит из множества специфичных терминов $\{t_1, t_2, t_3, \dots, t_i\}$, поэтому для начала необходимо наполнить данные словари понятиями из близких, родственных предметных областей и осуществлять поиск взаимовыгодных артефактов (патентов) [9]. В качестве примера можно рассмотреть отрасль «Машиностроение». В данном аспекте она является материнской. Дочерними областями будут являться «Авиационное», «Железнодорожная техника», «Авто-мобилестроение» и т.п., под который можно выделить группу специализированных словарей основанных по принципу «род-вид», «часть-целое», конкретизация которых зависит от преследуемых целей. Несмотря на то, что требования, предъявляемые к продукции, и области применения и эксплуатации обозначенных выше областей различны, все же у них имеются и точки соприкосновения. Многие виды транспорта, как показывает западная практика, на сегодняшний день продуваются в аэродинамических трубах, некоторые системы (к примеру, системы кондиционирования воздуха) между ними

практически идентичны, необходимость применения новых материалов (к примеру, композитов) становится все шире. К сожалению, НИИ и люди, работающие в определенной области машиностроения, чаще всего свои изобретения акцентируют только под нее, не расширяя спектр возможного применения на «стыках» или в абсолютно иных областях. Ярким примером может служить вычислительная техника, которая до 80-х годов XX века сугубо использовалась только для расчетов, а сейчас без применения компьютеров не может обойтись ни одна сфера деятельности человека.

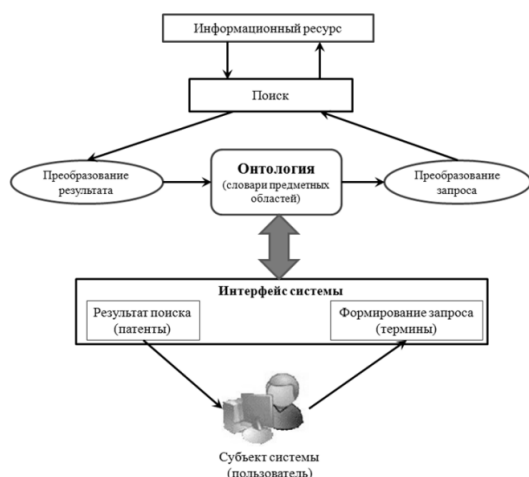


Рис. 1. Схема организации поиска на основе онтологии

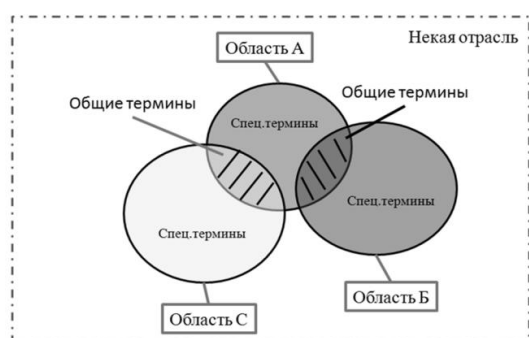


Рис. 2. Взаимосвязь терминов в смежных областях

Развивая идею, следует отметить тот факт, что все отрасли машиностроения используют **базовую онтологию терминов**, что упростит поиск необходимых артефактов для их анализа. Как уже было отмечено ранее, каждый словарь состоит из набора терминов и взаимосвязей между ними. При формализации идейно-семантического метода авторы используют теоретико-множественный подход, как наиболее удачный, поскольку в рамках данной работы приходится работать с большим объемом отдельных определений. Искомые слова и определения планируется брать из производственных стандартов (СТП), директив, технологических инструкций (ТИ), производственных инструкций, ГОСТ, ОСТ и иных производственных документов, которые в общем случае подпадают под общность нормативно-технической документации (НТД), ведь именно здесь как нигде используется схожая база терминов и уже далее осуществлять поиск и выборку. Схематично данную гипотезу можно представить на рис. 2.

Как уже было отмечено ранее каждый $Lb_n = \{t_1, t_2, t_3, \dots, t_i\}$, т.е. отбор артефактов осуществляется по принципу когда некий набор терминов $T_i \in Lb_1$ и $T_i \in Lb_2$, где Lb_1, Lb_2 словари рассматриваемых областей. Процесс систематизации артефактов условно можно связать с предметизацией поиска. В качестве информационной основы предметизации используются термины таксономии и набор логических условий (логиических функций, описывающих связи специализированных терминов по определенной области), с помощью которых осуществляется процесс поиска артефактов по отраслям (или смежным областям) [10]. Для формирования этих логических условий, описывающих содержание публикаций, используются термины таксономии и логические операторы: AND, OR, NOT.

В качестве примера ниже можно рассмотреть простейшее логическое условие, основанное как команда на естественном профессиональном языке IF...THEN...ELSE, являющийся условным оператором «Если...то...иначе». IF (фрезерная обработка OR заготовка OR алюминий OR ЧПУ OR программа OR оператор OR производство OR контроль OR технология OR профиль) AND (самолет OR автомобиль OR поезд); THEN вывод артефактов соответствующих запросу; ELSE уточнить запрос.

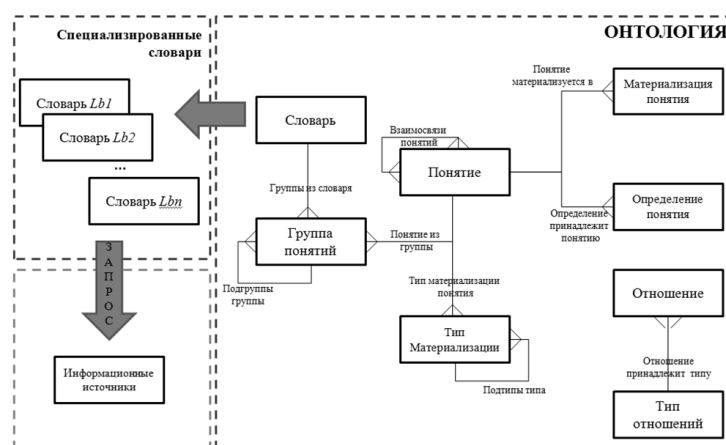


Рис. 3. Представление модели онтологии в рамках разрабатываемой системы

Рассмотрение модели онтологии для решения задач поиска (на примере комплекса средств WIQA). В структуре персональной онтологии O_k

выделяются взаимосвязанные неформальная и формализованная составляющие. Неформальная часть состоит из понятий $\{t_n\}$ онтологии с их толкованиями

(на базе языков L и LP), представленными в статьях «Словаря». В комплексе WIQA [7] для такого отбора используется толковый «Словарь», который строится субъектом системы в процессе создания системы моделей, включающей его персональную онтологию O_k . В этом словаре содержатся и понятия онтологии, требующих их определения. Основным предназначением инструментария WIQA является обслуживание концептуальной активности субъекта в его профессиональной деятельности [8]. Как следствие, положительными эффектами такого подхода является то, что наполнение и использование словарей возможно субъектами в рамках их профессиональной направленности на естественном языке, использующим типовую базовую терминологию. Ориентация в WIQA на разработку проекта привела к решению связать типовую единицу системности профессионального опыта с онтологией проекта, а интеграцию таких онтологий связать с персональной онтологией O_k профессионального опыта субъекта. Это же решение подсказывает отделить построение онтологий от действий по их интеграции, что и привело к нормативной концептуальной структуре «онтологии проекта», представленной на рис. 3.

К числу таких задач принципиальных задач, в которых может быть успешно применена онтология, относится поиск по таксономии. В WIQA решение такой задачи возложено на программного агента, запрограммированного на псевдо-кодovém языке WIQA [8]. Обобщенная структура этого агента приведена на рис. 4. Агент считывает с «доски поиска» описание события, затребовавшего доступ к системе специализированных словарей. Текст этого описания обрабатывается «морфологическим фильтром», который выделяет из описания список потенциальных «ключей доступа». «Онтологический фильтр» исключает из этого списка те лексемы, которые отсутствуют в онтологии.

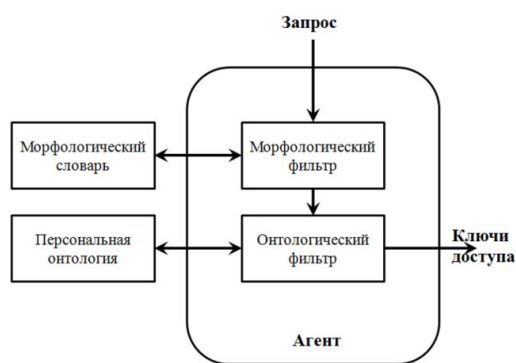


Рис. 4. Обобщенная структура агента, ориентированного на поиск по таксономии

Выводы: предложенная в работе технология, ориентированная на поиск по таксономии при помощи онтологических средств систематизации и ранжирования патентов, позволяет, на наш взгляд, разработать эффективную методику для анализа и применения патентов в смежных областях вследствие создания и использования словарей базовой терминологии. Рассматриваемые в запросе термины семантически связаны между собой определенными отношениями (род-вид, часть-целое, признаки, атрибуты и синонимы), и в этом смысле они должны оказывать большее влияние на качество поиска артефактов системы в отличие от «традиционного» списка ключевых слов.

СПИСОК ЛИТЕРАТУРЫ:

1. Альтиуллер, Г.С. Найти идею. — Новосибирск: Наука, 1986. 298 с.
2. Сафиуллин, А. Сотрудничество и научно-исследовательские сети в экономике знаний / А. Сафиуллин, В. Тронин // Проблемы теории и практики управления. 2015. № 2. С. 104–111.
3. Радионова, Ю.А. Классификация технической документации на основе лексического анализа десятичного номера / Ю.А. Радионова, В.Г. Тронин // Автоматизация процессов управления. 2008. № 3. С. 69–72.
4. Ярушклина, Н.Г. Рейтинги оценки деятельности инженерного вуза / Н.Г. Ярушклина, В.Г. Тронин // Высшее образование в России. 2014. № 5. С. 72–79.
5. Наместников, А.М. Разработка инструментария для интеллектуального анализа технической документации / А.М. Наместников, А.А. Филиппов, Р.А. Субхангулов // Известия Самарского научного центра Российской академии наук. 2011. Т. 13, № 4(4). С. 984–990.
6. Бурковский, В.Л. Физические явления и эффекты в технических системах: учеб. пособие / В.Л. Бурковский, Ю.Н. Глотова, Д.А. Ефремов, А.В. Романов. Воронеж: ГОУВПО «Воронежский государственный технический университет». 2007. 247 с. http://andromanov.narod.ru/Lib/yp_EffFizik.pdf
7. Соснин, П.И. Инструментальные средства для спецификации концептуализаций в проектировании автоматизированных систем / П.И. Соснин, В.А. Маклаев // Онтология проектирования. 2012. № 2. С. 39–52.
8. Соснин, П.И. Персональная онтология профессионального опыта: матер. 4-й Международ. конф. «Открытые семантические технологии проектирования интеллектуальных систем» (OSTIS-2014). — Минск: БГУИР, 2014. С. 147–154.
9. Гришин, М.В. Средства онтологической поддержки процесса проектирования шаблонной оснастки в условиях авиационных производств / М.В. Гришин, С.Н. Ларин, П.И. Соснин // В мире научных открытий. Естественные и технические науки. 2015. № 4 (64). С. 10–43.
10. Вдовичин, В.Т. Технологии систематизации и поиска электронной научной информации с применением онтологий / В.Т. Вдовичин, В.А. Лебедев // Информационные ресурсы России. 2010. № 5. С. 6–10.

DATA SOURCES FOR DEVELOPMENT THE SYSTEM OF SUPPORT THE INVENTIVE TASKS SOLUTION ON THE BASIS OF ONTOLOGIES

© 2016 M.V.Grishin¹, V.G. Tronin²

¹ JSC “Ulyanovsk Instrument Manufacturing Design Bureau”

² Ulyanovsk State Technical University

Article is devoted to formulation the problem of application the ontologic approach in system of support the inventive tasks solution at a stage of the patent analysis with consideration of various data sources.

Key words: TITS, ontology, search, patent, RSCI, taxonomy

Maxim Grishin, Candidate of Technical Sciences, Engineer Designer at the Department of Control Systems. E-mail: likani7@mail.ru; Vadim Tronin, Candidate of Technical Sciences, Associate Professor at the Department “Information Systems”. E-mail: v.tronin@ulstu.ru