

Обробка інформації в складних організаційних системах

УДК 004.89

Д.Б. Куделина¹, И.В. Шостак¹, И.В. Груздо²

¹ *Национальный аэрокосмический университет имени Н.Е. Жуковского «ХАИ», Харьков*

² *Харьковский национальный университет радиоэлектроники, Харьков*

УПРАВЛЕНИЕ ЗНАНИЯМИ РАЗРАБОТЧИКОВ СОФТВЕРНОЙ ФИРМЫ ПО СЕРТИФИКАЦИИ ПРОГРАММНЫХ ПРОДУКТОВ НА ОСНОВЕ ОНТОЛОГИЧЕСКОГО ПОДХОДА

Проведен обзор публикаций, посвященных особенностям управления и извлечения знаний с использованием онтологий. Приведена постановка задачи управления знаниями разработчиков софтверной фирмы на основе прецедентного подхода. Показаны особенности применения методов извлечения знаний, с использованием онтологий, о процессах разработки программных продуктов (ПП), а именно методов кластеризации и аннотации документов. Представлена ролевая модель кластеризации онтологий технической документации к ПП.

Ключевые слова: управление знаниями, софтверная фирма, программный продукт, кластеризация, онтология, аннотирование документов, ролевая кластеризация онтологий, сертификация.

Введение

На данный момент стремительное развитие информационных технологий во всем мире привело к накоплению значительных объемов знаний о процессах разработки программных продуктов (ПП), включая и документальное обеспечение этих процессов. Носителями указанных знаний являются эксперты – в данном случае разработчики программного обеспечения, имеющие достаточно большой профессиональный опыт. Существующий риск того, что специалисты с большими знаниями и опытом, могут покинуть фирму (перейти в другую или выйти на пенсию), определяет необходимость выделения значительных временных и финансовых ресурсов на обучение персонала. К тому же, кроме потери сотрудника, фирма при этом потеряет и знания, которыми сотрудник обладал или которые приобрел ходе своей профессиональной деятельности.

Исходя из приведенных выше соображений, решение проблемы повышения эффективности функционирования софтверных фирм в Украине и других странах, связана, в том числе и с разработкой компьютерных средств формирования комплексного информационного ресурса знаний и опыта сотрудников в виде, пригодном для повторного использования.

Эффективность управления софтверной фирмой определяется стратегией управления и подразумевает: сбор, организацию, использование, накопление информации и знаний. Программное управле-

ние знаниями на уровне ИТ — технологий предполагает решение следующих задач: создания баз данных, баз знаний, хранения, систематизации, анализа информации, организации многопользовательского доступа к сформированному информационному ресурсу [1]. Значительная часть знаний, которые необходимо представить в рассматриваемом информационном ресурсе, изначально сосредоточена в текстовых источниках.

Ввиду большого роста объемов текстовой информации и сложной структурированности естественно-языковых (ЕЯ) текстов, анализ текстов представляет собой актуальную проблему, особенно в последние 15-20 лет, когда наметилась тенденция к информатизации общества.

Особую актуальность в Украине начиная с 2000 годов приобрела проблема обеспечения качества экспертизы ПП. Ведь именно на этапе экспертизы ПП является завершенным и готовым к использованию или даже уже эксплуатируется. Таким образом, целесообразно проводить извлечение знаний из имеющейся текстовой документации к ПП именно на этом этапе. За время существования процедуры сертификации ПП, экспертам пришлось столкнуться с рядом проблем, поскольку данная процедура является слабо формализованным и слабо структурированным видом профессиональной деятельности сертификационных аудиторов (СА). Велик субъективизм и влияние опытности СА на результаты итоговых оценок. В современных условиях решение задач сертификации ПП затруднено, в силу

увеличения объема и сложности ПП с одной стороны и расширения номенклатуры объектов сертификации – с другой. Эти обстоятельства породили ряд проблем в деятельности СА.

Самой существенной проблемой является большая доля рутинного труда, связанного с анализом существенно расширившейся (за счет международных стандартов) нормативной базы (НБ) и необходимостью формирования нормативного профиля (НП) ПП, разработанных для уникальных объектов.

Другая проблема состоит в необходимости статического и динамического анализа крупных программных проектов, содержащих сотни тысяч и миллионы операторов.

Перечисленные выше проблемы могут быть отчасти преодолены за счет применения средств онтологического инжиниринга. Высокая трудоемкость построения, необходимость наличия определенных навыков создания и массовость онтологий для представления знаний в процессе сертификации требуют разработки специальных методов и алгоритмов обработки текстовых документов.

В основе задачи обработки ЕЯ-текстов лежат морфологический и морфемный анализ, синтаксический и семантический анализ, результатами которых являются модели текста, адекватно отражающие его словообразовательные, грамматические и смысловые конструкции. Основные методы анализа текста с использованием онтологического подхода достаточно подробно освещены в трудах И. Шостака, Ю. Бутенко, О. Corby, И.П. Норенкова, В.И. Шабанова, А.М. Андреева, I. Muslea, N. Kushmerick, D. Weld, R. Doorenbos и др.

Цель статьи состоит в изложении методики извлечения и систематизации знаний разработчиков софтверной фирмы на основе использования онтологий на этапе сертификации ПП.

1. Постановка задачи

Для решения задачи располагаем следующими исходными данными: команда проекта, портфель заказов фирмы, опыт выполнения проектов (знания данной команды), документация проекта. В результате будет построена диалоговая система поддержки принятия решения сертификационного аудитора, ядром которой станет онтологическая система.

2. Анализ проблемы извлечения знаний из технической документации при сертификации ПП

2.1. Особенности организации процесса управления корпоративной памятью

Корпоративная память (Corporation Memory – (СМ)), по аналогии с человеческой памятью, предоставляет возможность использования предыдущего опыта с целью избежания ошибок, она фиксирует

информацию из различных источников компании и делает эту информацию доступной специалистам для решения определенных задач проекта [3]. Данная память не позволяет исчезнуть знаниям выходящих специалистов (уход на пенсию, увольнение и др.). При этом она способна хранить большие объемы данных информации и знаний в различных формах, таких как базы данных, документы и базы знаний. При разработке соответствующего механизма для управления корпоративной памятью выделяют следующие этапы:

1. Накопление информации в организации
2. Извлечение – процесс, идентичный извлечению знаний при разработке экспертных систем
3. Структурирование. На этом этапе должны быть выделены основные понятия. Выбрана структура представления информации, обладающая максимальной наглядностью, простотой изменения и дополнения.
4. Формализация. Представление структурированной информации в форматах машинной обработки.
5. Обслуживание – добавление, обновление формализованных данных и знаний, удаление устаревшей информации [3].

В отличие от экспертных систем, цель построения СМ состоит не в поддержке решения конкретной задачи, а в лучшем использовании существующих корпоративных ресурсов, но, тем не менее, СМ зачастую реализуются на основе оболочек экспертных систем (например, KONUS системы).

2.2. Общая характеристика предметной области «Сертификация программных продуктов»

Функциональные характеристики объектов с интенсивным использованием программного обеспечения (ПО) в большей мере зависят от его качества. ПО, являясь эквивалентной формой представления алгоритмов управления, в значительной степени определяет логику, динамику и точность функционирования. Взаимодействие ПО с физическим оборудованием в реальном времени, непосредственно в контурах управления основными режимами работы, предопределяет предельно жесткие требования к качеству отработки и высокую трудоемкость.

Существует критические зависимости качества, сроков, и стоимости системы управления от качества ПО, решение которых состоит в использовании адекватных технологий, охватывающие не только область разработки управляющих алгоритмов и их программных реализаций, но и определяющих возможности достижения требуемых характеристик, качества, сроков и стоимости создания СУ и объекта управления в целом, а также организации эффективного оценивания ПО, сертификационными центрами (СЦ), привлекающими квалифицированных

специалистов, способных быстро и достоверно оценить качество ПО. При этом качество самой оценки определяется квалификацией экспертов и имеющимся в их распоряжении информационными ресурсом – базой нормативных документов (НД).

Ввиду многообразия НД, целесообразно ввести понятие нормативного профиля (НП) – гармонизированной с международными и национальными стандартами совокупности требований, предъявляемых к данному проекту или группе проектов. НП могут быть вновь разрабатываемые государственные или отраслевые стандарты нормативно-методические документы предприятий и общие требования спецификаций ПО. Формирование и верификация НП является одной из задач оценки качества ПО [1]. Процедура оценки программного обеспечения состоит в решении ряда задач, среди которых следует выделить:

- формирование и верификация нормативного профиля,
- реинжиниринг и оценка процесса проектирования на основе НП,
- статический анализ – определение программных метрик, согласно выбранному НП,
- семантический анализ (СА) ПО – контроль семантической эквивалентности программных конструкций,
- интервальный анализ исполняемого модуля, заключающийся в контроле достоверности результатов всех этапов вычислительных процессов,
- определение степени соответствия исходного кода проектной документации и НП.

В современных условиях решение перечисленных выше задач затруднено в силу увеличения объема и сложности ПО с одной стороны и расширением номенклатуры объектов – с другой.

Эти обстоятельства породили ряд проблем в деятельности экспертов. Самой существенной проблемой является большая доля рутинного труда, связанного с анализом существенно расширившейся (за счет международных стандартов) нормативной базы (НБ) и необходимостью формирования НП ПО уникальных объектов. Вторая проблема состоит в необходимости статического и динамического анализа крупных программных проектов, содержащих сотни тысяч и миллионов операторов. Немаловажным является также и возрастание субъективности оценки соответствия исходного кода ПО проектной документации и НП. [2]

3. Общая характеристика методов для извлечения знаний из технической документации по на основе онтологий

3.1. Применение онтологий в задачах извлечения знаний из технической документации

Значительная часть полезных для организации знаний содержится в документальных базах знаний

(БЗ), являющихся базами текстовых документов. Поскольку число документов в БЗ большинства организаций весьма значительно и продолжает расти, ручное управление знаниями оказывается неэффективным. Поэтому в интеллектуальных системах управления знаниями стремятся в максимально возможной степени автоматизировать процедуры управления знаниями (УЗ) [4]. Однако текстовые документы и соответственно знания, содержащиеся в них, обычно являются слабо структурированными, что существенно сокращает возможности автоматических извлечения и обработки знаний в системах УЗ.

Обработка знаний может быть направлена на достижение ряда целей. Цель данной исследовательской работы показать метод извлечения знаний из документации разработанного ПО. Повышению эффективности решения интеллектуальных задач способствует применение онтологий. Модели знаний в онтологиях выражены в виде множеств понятий (концептов, сущностей) и отношений между ними. Поскольку извлечение знаний из текстовых документов и их обработка связаны именно с понятийным составом рассматриваемых приложений, управление документальными БЗ целесообразно выполнять на основе применения онтологий [4].

Данную цель можно достичь, выполняя ряд задач обработки знаний, решаемых в интеллектуальных системах. К ним, в первую очередь, относятся задачи кластеризации, классификации, аннотирования, упорядочения документов, поддержки принятия решений. В силу слабой структурированности знаний успешно решать эти задачи в автоматическом или полуавтоматическом режимах удастся лишь в отдельных частных случаях, поэтому существующие системы КМ, как правило, являются узкоспециализированными. Также онтологическая модель позволяет разработать модель метаданных, что значительно улучшает использование системы широким кругом пользователей с точки зрения организации взаимодействия.

Онтологические модели за время исследований в этой области претерпели значительное развитие. В настоящее время для создания и поддержки онтологий существует целый ряд инструментов, которые помимо общих функций редактирования и просмотра выполняют поддержку документирования онтологий, импорт и экспорт онтологий разных форматов и языков, поддержку графического редактирования, управление библиотеками онтологий и т.д.

Эффективным способом хранения больших и стабильных онтологий является хранение онтологий с помощью систем управления базами данных. В онтологиях знания формализуются в виде описаний предметной области с помощью иерархии классов. Для каждого класса задается свой набор свойств и объектов. Свойства в онтологиях имеют область

определения – класс, для которого задается это свойство, а также область значений. В зависимости от областей значений свойства делятся на два типа: т-свойства (значениями которых являются константы заданного типа данных) и о-свойства (значениями которых являются объекты заданного класса)[4].

3.2. Построение онтологии с использованием кластеризации

Основная идея построения онтологии с использованием кластеризации заключается в следующем: каждый документ представляется виде набора терминов, множество документов разбивается на подмножества документов близкой тематики (кластеры), в результате получают группы терминов одной тематики. Суть данной идеи графически представлена на рис. 1. Это позволяет установить отношения между терминами и концепциями. Каждый термин характеризуется частотой встречаемости (весом). Термины с весом, больше среднего, задают термины онтологии, а в качестве концепции берутся термины с максимальным весом.



Рис. 1. Схема построения онтологии с использованием кластеризации

Кластеризация документов производится на основе иерархического алгоритма «минимальное покрывающее дерево»[12], позволяющего задать таксономию концепций в полученной онтологии. Коллекция документов представляется в виде дерева, разделение кластеров происходит в месте максимального расстояния между ближайшими документами внутри каждого кластера.

Очевидный метод классификации документов предполагает предварительную ручную разработку пользователями-экспертами обучающей выборки, состоящей из специально отобранных документов [5]. Далее выполняется формирование некоторых классификационных признаков и обучение классификатора. На основе выявленных классификационных признаков формируются правила оценки тематической направленности документа, используемые для классификации входящих в систему документов [6]. Преимущества онтологического подхода заключаются в том, что вместо разработки обучающей выборки создаются прикладные онтологии, которые могут использоваться не только для классификации, но и для решения ряда других задач. В системах УЗ

подразумевается предварительная разработка предметных (domain) онтологий, охватывающих области деятельности организации. Если перечни приложений при предметной кластеризации онтологий и документов совпадают, то наличие предметных онтологий позволяет легко осуществить автоматическое распределение документов по рубрикам, т.е. автоматическую кластеризацию документов.

При классификации оценка R_{kl} степени принадлежности k -го документа l -му тематическому кластеру выполняется по n_{kl} – числу появлений концептов l -й предметной онтологии в k -м документе; $k = 1, 2, \dots, q$; $l = 1, 2, \dots, m$; например:

$$R_{kl} = n_{kl} / \sum_{i=1}^m n_{ki}; \quad n_{kl} = \sum_{j=1}^{M_j} N_{jkl},$$

где N_{jkl} – число появлений j -го концепта l -й предметной онтологии в k -м документе; M_j – число концептов в l -м приложении.

Эта оценка может использоваться для классификации документов с введением весов концептов, учитывающих их информативность. Так, в работе [9] классификация осуществляется по суммарному весу концептов предметной онтологии, имеющих в документе. Вес g_j j -го концепта часто определяют следующим образом: $g_j = \ln((q+1)/(q_j+1))$, где q_j – число документов в базе, содержащих j -й дескриптор (термин, обозначающий концепт).

3.3. Аннотирование документов для создания онтологии

Аннотирование является основной задачей извлечения информации (IE – Information Extraction) из текстовых документов. Аннотирование, которое можно трактовать как составление метаданных документов, является основой для решения ряда задач обработки знаний. Подходы к аннотированию документов произвольной тематики и узкой направленности различны. Наиболее сложны задачи автоматического извлечения информации из документов неструктурированных или слабо структурированных, причем трудности IE возрастают по мере расширения тематики исследуемого корпуса документов.

В существующих системах [8,9] попытки извлечения знаний из таких документов основаны на выявлении в текстах паттернов (словосочетаний, предложений), содержащих определенное ключевое слово (обычно глагол) вместе с сопутствующими словами, выполняющими такие роли, как «субъект», «инструмент», «цель» [7]. Одним из условий применения этого подхода для автоматического аннотирования текстов произвольной тематики является их синтаксическая и семантическая корректность. Чаще аннотирование выполняется для структурированных документов конкретной тематики.

В первой группе методов, ориентированных на такое аннотирование, используется поиск и выявление

ние в документах часто встречающихся слов, характеризующих конкретные события, ситуации, факты. К таким словам относятся экземпляры концептов, такие как собственные имена (NE - named entities), названия организаций, географических пунктов, даты, адреса и т.п.

Во второй группе методов извлечение информации заключается в поиске специфических выражений, характерных для определенных предметных областей [8]. Полуавтоматическое аннотирование документов и извлечение нужных данных при этом обычно происходит на основе предварительного обучения системы ИЕ, выполняемого пользователем [9].

Аннотирование с применением онтологий позволяет составлять аннотацию из терминов концептов или значений (экземпляров) концептов, найденных в тексте документа [10, 11]. Здесь по-прежнему популярно аннотирование на основе использования NE. Аннотация представляет собой сформированные высказывания, содержащие NE и выраженные в формате RDF [12]. Аннотации в форме онтологий, что позволяет использовать средства онтологического анализа как для самих документов, так и для аннотаций, рассматриваются в работе [13].

3.4. Метод ролевой кластеризации онтологий

Прецеденты-решения могут быть представлены в той или иной форме, принятой в интеллектуальных системах, например, в виде фреймов, семантических сетей, значений параметров и т.п. [14, 15]. База прецедентов состоит из пар «проблема/решение», причем решение чаще всего представлено фреймом, в слотах которого содержатся значения параметров, характеризующих условия и результаты решения задачи, в частности, результатами могут быть значения предметных переменных. Задача формирования фрейма становится задачей извлечения знаний из тех или иных источников. При извлечении знаний из текстовых документов, а также при описании решений и при сравнительной оценке прецедентов полезно используются онтологии. Автоматическое построение базы прецедентов возможно только для конкретных структурированных приложений. Существуют прагматический и содержательный подходы к управлению знаниями [15].

Прагматический подход характерен для обработки структурированной информации с использованием баз данных. Содержательный подход имеет место при использовании специальных моделей и операций представления и обработки знаний. При работе с неструктурированной информацией, относящейся к сравнительно широким предметным областям, неизвестна сама структура фрейма «проблема/решение», возможности автоматической или полуавтоматической обработки знаний резко сужаются. На основе содержательного подхода удается

формализовать фазу извлечения знаний, фаза собственно принятия решений остается неавтоматизированной. Описываемый далее метод ролевой кластеризации онтологий можно отнести к числу реализаций содержательного подхода:

1. Концепты онтологий распределяются по кластерам в зависимости от их роли в сложных концептах. Сложные концепты – ‘это словосочетания, выражающие отношения «объект-свойство», «объект-действие», «объект-свойство-действие», «средство-действие-объект» и состоящие из простых концептов, выполняющих роли «объект», «свойство», «действие», «средство» [14]. Хотя распределение концептов по кластерам не является формальной процедурой, но выполняется для каждого приложения однократно.

2. Сложные концепты представляются в виде паттернов, слоты которых соответствуют ролям простых концептов.

3. Аннотации документов в соответствии с методом ролевой кластеризации составляются из сложных концептов или из предложений, в состав которых входят сложные концепты.

4. Извлечение сложных концептов из текста заключается в поиске терминов, относящихся к разным кластерам онтологии и расположенных в нужной последовательности близко друг от друга в тексте, например, в пределах одного предложения. Найденные предложения со сложными концептами включаются в аннотацию (с возможной ручной корректировкой), если частота повторения концептов в документе не ниже заданного порога. Возможно совместное применение аннотирования на основе сложных концептов и на основе специфических терминов типа NE.

Таким образом информационный поиск с использованием ролевой кластеризации концептов выполняется по запросам, содержащим сложные концепты. Сложные концепты выделяются в запросе (или запрос конструируется по паттернам сложных концептов), далее запрос сопоставляется с аннотацией документа. Другими словами, вместо поиска по ключевым словам (простым концептам) выполняется поиск по сложным концептам, входящим в индекс, а для индексирования используются результаты аннотирования.

Метод ролевой кластеризации онтологий перспективен для полуавтоматического решения задач управления знаниями, поскольку сложные концепты гораздо более точно выражают семантику документов и запросов, чем совокупности составляющих их простых концептов. Для принятия решений с использованием ролевой кластеризации онтологий применяется метод CBR, в соответствии с которым каждому документу коллекции сопоставлен паттерн проектирования «проблема/решение».

Метаданні документа, включаючи анотацію, складають автоматично формовану ліву частину паттерна, т.е. слоти «проблема». Права частина в разі структурованих додатків містить значення параметрів прецедентів. В разі неструктурованих додатків метод забезпечує пошук документів, стосуються до проблеми, вираженої складними поняттями, і потенційно містять прецеденти рішення проблеми. Остаточне виявлення прецедентів возкладається на користувача.

Висновки

1. Приведено аналіз проблеми витягання знань з технічної документації при сертифікації ПП.

2. Приведено принцип побудови онтології з використанням кластеризації. Основна ідея заключається в наступному: кожен документ представляється в формі набору термінів, множина документів розбивається на підмножини документів близької тематики (кластери), в результаті утворюються групи термінів однієї тематики.

3. Описано методи аналізу текстової документації до ПП, приведено модель онтології, заснована на кластеризації.

Список літератури

1. Шостак І.В. Знаниеориєнтовані методи формування нормативних профілей до системам критичного застосування на основі онтологій / І.В. Шостак, М.А. Бутенко Ю.І. // *Радиоелектроніка і комп'ютерні системи*. – 2010. – № 2. – С. 104-107.
2. Шостак І.В. Текущее состояние бази знань в динамічних експертних системах управління складними об'єктами / І.В. Шостак // *Радиоелектроніка і інформатика*. – 2000. – № 3. – С. 68-71.
3. Ермаков А.Е. Извлечение знань из текста и их обработка: состояние и перспективы / А.Е. Ермаков // *Информационные технологии*. – 2009. – № 7. – С. 50-55.
4. Норенков И.П. Задачи управления знаниями, извлекаемыми из текстовых [Электронный ресурс] / И.П. Но-

виков // *Наука и образование*. – М.: МГТУ им. Н.Э. Баумана. – 2011. – № 9. – Режим доступа: <https://cyberleninka.ru/article/n/zadachi-...-dokumentov>.

5. Шабанов В.И. Метод классификации текстовых документов, основанный на полнотекстовом поиске / В.И. Шабанов, А.М. Андреев // *Труды РОМИП'2003*. – СПб.: НИИ Химии СПбГУ, 2003. – С. 52-71.

6. Oracle Text Application Developer's Guide [Electronic resource]. – Oracle Corporation, 2003. – Access to: <http://www.stanford.edu/dept/itss/docs/oracle/10g/text.101/b10729.pdf>.

7. Muslea I. Extraction Patterns for Information Extraction Tasks: A Survey // *AAAI-99 Workshop on Machine Learning for Information Extraction*, 2013.

8. Li Y., Zhang L., Yu Y. Learning to Generate Semantic Annotation for Domain Specific Sentences. // *K-CAP 2011 Workshop on Knowledge Markup & Semantic Annotation*, 2013.

9. Kushmerick N., Weld D., Doorenbos R. Wrapper Induction for Information Extraction. // *Proc. 15th Int. Joint Conf. AI, 1997* – P. 729-735.

10. Fernandez M., Vallet D., Castells P. Automatic Annotation and Semantic Search from Prot?g? // *8th International Protege Conference, 2005, Madrid, Spain*.

11. Handschuh S., Staab S., Ciravegna F. S-CREAM - Semi-automatic CREATION of Metadata // *Proc. of the European Conference on Knowledge Acquisition and Management EKAW-2002. Madrid: Springer, 2002*.

12. Castells P., Fernandez M., Vallet D. An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval. // *IEEE Transactions on Knowledge and Data Engineering* 19(2), Special Issue on Knowledge and Data Engineering in the Semantic Web Era, February 2007. – P. 261-272.

13. Watson I. Case-Based Reasoning: A Review / I. Watson, F. Marir // *Knowledge Engineering Review*. – 2012. – V. 9, No. 4. – P. 355-381.

14. Kurbalija V. Case-Based Reasoning Framework for Generating Decision Support Systems / V. Kurbalija, Z. Budimac // *Novi Sad J. Math*. – 2008. – V. 38, No. 3. – P. 219-226.

15. Ситников П.В. Построение систем поддержки принятия решений на основе онтологий. – Автореферат диссертации ... канд. техн. наук. – Самара: 2009, 24 с.

Поступила в редакцію 23.02.2016

Рецензент: д-р техн. наук, проф. Ю.В. Стасев, Харківський університет Воздушних Сил ім. І. Кожедуба, Харків.

УПРАВЛІННЯ ЗНАННЯМИ РОЗРОБНИКІВ СОФТВЕРНОЇ ФІРМИ ПО СЕРТИФІКАЦІЇ ПРОГРАМНИХ ПРОДУКТІВ НА ОСНОВІ ОНТОЛОГІЧНОГО ПІДХОДУ

Д.Б. Куделіна, І.В. Шостак, І.В. Груздо

Проведено огляд публікацій, присвячених особливостям управління і витягання знань з використанням онтологій. Наведено постановку задачі управління знаннями розробників софтверної фірми на основі прецедентного підходу. Показано особливості застосування методів витягання знань, з використанням онтологій, про процеси розробки програмних продуктів (ПП), а саме методів кластеризації та анотації документів. Представлено рольову модель кластеризації онтологій технічної документації до ПП.

Ключові слова: управління знаннями, софтверна фірма, програмний продукт, кластеризація, онтологія, анотування документів, рольова кластеризація онтологій, сертифікація.

MANAGEMENT KNOWLEDGES OF SOFTWARE FIRM DEVELOPERS ON THE CERTIFICATION OF SOFTWARE PRODUCTS ON THE BASIS OF ONTOLOGICAL APPROACH

D.B. Kudelina, I.V. Shostak, I.V. Gruzdo

A review of publications devoted to the peculiarities of management and retrieval of knowledge using ontology's. Shows the production of knowledge management tasks developers software company based case approach. The features of the use of knowledge extraction techniques, using ontology's on development software products (SP), namely clustering methods and annotation of documents. Presented role model clustering ontology of technical documentation to the SP.

Keywords: knowledge management, software development company, software, clustering, ontology, annotating documents, role-ontology clustering, certification.