инжиниринг онтологий

УДК 004.822:004.89

DOI: 10.18287/2223-9537-2021-11-4-500-520

Автоматизация разработки онтологий научных предметных областей на основе паттернов онтологического проектирования

Ю.А. Загорулько, Е.А. Сидорова, Г.Б. Загорулько, И.Р. Ахмадеева, А.С. Серый

Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия

Аннотация

В настоящее время онтологии признаны наиболее эффективным средством формализации и систематизации знаний и данных в научных предметных областях (НПрО). Однако разработка онтологии является сложным и трудоёмким процессом. Практика показала, что при разработке онтологий НПрО, особенно эффективно применение паттернов онтологического проектирования. Это связано с тем, что в онтологии НПрО, как правило, содержится большое количество типовых фрагментов, которые хорошо описываются паттернами онтологического проектирования. Благодаря тому, что использование паттернов онтологического проектирования значительно облегчает разработку онтологии НПрО, к ней можно привлечь экспертов в моделируемой НПрО, не владеющих навыками онтологического моделирования. Для получения онтологии НПрО, необходимо обработать огромное количество публикаций, релевантных моделируемой НПрО. Облегчить и ускорить процесс пополнения онтологии информацией из таких источников можно за счёт использования лексико-синтаксических паттернов онтологического проектирования. В статье представлен подход к автоматизированному построению онтологий НПрО на основе системы разнородных паттернов ОП. Система включает паттерны ОП, предназначенные для разработчиков онтологий, и автоматически построенные на их основе лексико-синтаксические паттерны, с помощью которых может выполняться автоматическое пополнение онтологии информацией, извлекаемой из текстов на естественном языке.

Ключевые слова: научная предметная область, паттерн. онтологическое проектирование, паттерны содержания, автоматическая генерация паттернов, пополнение онтологии.

Цитирование: Загорулько, Ю.А. Автоматизация разработки онтологий научных предметных областей на основе паттернов онтологического проектирования / Ю.А. Загорулько, Е.А. Сидорова, Г.Б. Загорулько, И.Р. Ахмадеева, А.С. Серый // Онтология проектирования. − 2021. − Т.11, №4(42). − С.500-520. − DOI: 10.18287/2223-9537-2021-11-4-500-520.

Введение

В настоящее время онтологии признаны наиболее эффективным средством формализации и систематизации знаний и данных в научных предметных областях (НПрО), под которыми принято понимать предметные области (ПрО), охватывающие некоторую научную дисциплину или область научных знаний во всех её аспектах, включая характерные для неё объекты и предметы исследования, применяемые методы, выполняемую научную деятельность и полученные результаты.

Разработка онтологии ПрО является сложным и трудоёмким процессом. Для его упрощения и облегчения предложены и применяются различные методы и подходы [1, 2]. Можно отметить три основных подхода:

- 1. Разработка онтологии «с нуля» [3-5].
- 2. Создание онтологии из «готовых блоков» [6-8].

3. Автоматическое построение онтологии [9].

Первый подход является самым трудоёмким и требует привлечения опытных специалистов в области онтологического инжиниринга.

Второй подход менее трудоёмкий, так как предполагает использование заранее созданных базовых онтологий и/или их фрагментов, которые можно специализировать под конкретную ПрО. Он позволяет вовлечь в процесс построения онтологий специалистов в ПрО, для которых создаются онтологии.

Третий подход наименее трудоёмкий, но не позволяет построить онтологии высокого качества.

В данной статье описан подход, реализующий автоматическое пополнение онтологии, построенной в рамках второго подхода. Его особенностью является то, что на первом этапе инженерами знаний и экспертами в НПрО выполняется разработка и начальное наполнение онтологии с использованием базовых онтологий и их фрагментов (паттернов онтологического проектирования, ОП), а на втором этапе - автоматическое пополнение онтологии НПрО онтологическими сущностями, извлечёнными из тематического корпуса текстов.

Первый этап предложенного подхода опирается на использование паттернов ОП (Ontology Design Patterns или ODPs) [7, 10, 11], представляющих собой документально зафиксированные описания проверенных на практике решений типовых проблем онтологического моделирования.

Опыт авторов показывает, что применение паттернов ОП особенно эффективно при разработке онтологий НПрО [12, 13]. Это связано с тем, что в онтологии НПрО, как правило, содержится большое количество типовых фрагментов, которые хорошо описываются паттернами ОП. Благодаря этому к разработке онтологии НПрО можно привлечь экспертов в моделируемой НПрО, не владеющих навыками онтологического моделирования, что позволяет ускорить разработку онтологии НПрО.

Для того, чтобы получить онтологию НПрО требуется обработать огромное количество публикаций, релевантных моделируемой области, и извлечь из них онтологические сущности, дополняющие онтологию, построенную на первом этапе. Для пополнения онтологии на основе извлечения информации из текстов на естественном языке разрабатываются различные методы и подходы [14-17]. Особенностью рассматриваемого в статье подхода является автоматическое пополнение онтологии НПрО на основе анализа тематического корпуса текстов с использованием лексико-синтаксических паттернов (ЛСП), построенных на основе паттернов ОП, используемых на первом этапе, словаря общенаучной лексики и текущей версии онтологии.

1 Онтология паттернов ОП

Онтология паттернов ОП (см. рисунок 1) включает систематизацию паттернов ОП, описание их свойств и отношений между ними, методов и областей их применения, публикаций, информационных ресурсов и др.

Систематизация паттернов ОП может выполняться по следующим основаниям: по типам решаемых проблем онтологического моделирования, по назначению (по типам решаемых прикладных задач), по областям использования, т.е. ПрО.

В качестве базовой принята систематизация паттернов по типам решаемых проблем онтологического моделирования. При разработке такой систематизации использована классификация, предложенная в исследовательском проекте NeOn [18]. Паттерны ОП разделяются на шесть основных типов: Структурные паттерны (Structural ODPs), Паттерны соответствия (Correspondence ODPs), Паттерны содержания (Content ODPs), Паттерны логического выво-

да (Reasoning ODPs), Паттерны представления (Presentation ODPs) и ЛСП (Lexico-Syntactic ODPs). Структурные паттерны подразделяются на Архитектурные паттерны (Architectural ODPs) и Структурные логические паттерны (Structural Logical ODPs).

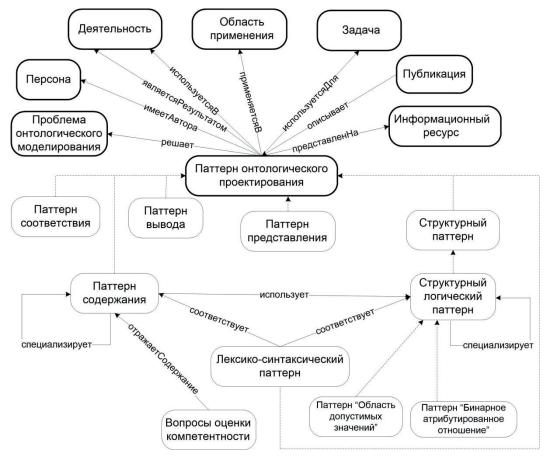


Рисунок 1 – Онтология паттернов онтологического проектирования

Ядро онтологии ОП составляют класс *Паттерн ОП*, задающий основные свойства паттернов ОП, и его подклассы, используемые для представления типов паттернов ОП. Такими классами являются *Структурный логический паттерн*, *Паттерн содержания*, *Паттерн представления* и др.

Описание свойств паттернов ОП выполняется на основе формата, предложенного на портале ассоциации *ODPA* [19]. В описание паттерна включаются сведения об авторе и области применения, текстовое описание, графическое представление, ссылки на другие паттерны, набор сценариев и примеров использования. Паттерн содержания дополяется набором вопросов оценки компетентности (*Competency questions*) [20].

Формат описания паттернов дополнен элементами, служащими для описания контекста разработки и использования паттернов ОП. Для этих целей в онтологию паттернов ОП включены следующие классы: Область применения (ПрО), Деятельность, Задача, Публикация, Персона, Организация, Информационный ресурс и др. Для связывания паттернов с экземплярами этих классов в онтологию паттернов ОП включены отношения, позволяющие связать паттерны ОП с ПрО, персонами, организациями и проектами, в которых они используются, а также с публикациями и информационными ресурсами, где они описываются.

Между паттернами содержания, а также между структурными логическими паттернами, определено отношение *«специализирует»*, которое связывает *«исходные»* паттерны с построенными на их основе паттернами, настроенными на моделируемую НПрО.

Наиболее полно в онтологии описаны паттерны, реализованные в предлагаемой системе автоматизированного построения онтологий НПрО [12]: структурные логические паттерны, паттерны содержания, паттерны представления и лексико-синтаксические паттерны.

Необходимость в использовании структурных логических паттернов возникла из-за недостатка в языке *OWL* [21] выразительных средств для представления сложных сущностей и конструкций, актуальных при построении онтологий НПрО, в частности, многоместных и атрибутированных отношений (бинарных отношений с атрибутами), а также областей допустимых значений, определяемых разработчиком онтологии.

Паттерн «Область допустимых значений» предназначен для задания возможных значений какого-либо свойства класса, когда всё множество таких значений (как правило, строковых) известно заранее и поэтому может быть зафиксировано на этапе разработки.

Паттерны содержания предназначены для поддержки единообразного и непротиворечивого представления используемых в НПрО понятий и их свойств. Такие паттерны были разработаны для понятий, характерных для большинства НПрО: Объект исследования, Предмет исследования, Метод, Задача, Раздел науки, Научный результат, Деятельность, Проект, Персона, Организация, Публикация, Информационный ресурс и др. Для каждого из этих паттернов определён набор вопросов проверки компетентности. С помощью этих вопросов выявлены обязательный и факультативный составы онтологических элементов паттерна и описаны требования к ним, которые представлены в виде аксиом и ограничений.

Для каждого паттерна, представляющего понятие НПрО, составлен набор ключевых атрибутов, однозначно определяющих конкретный экземпляр понятия.

В качестве примера рассмотрен паттерн для представления понятия "Метод" (рисунок 2). Элементы описания этого паттерна представлены обязательными классами онтологии Задача, Раздел науки, Персона и Организация, необязательными (факультативными) классами Деятельность, Научный результат и др., а также отношениями "используется в", "реализуется в", "решает", "имеет автора" и др. У паттерна, представляющего понятие "Метод", есть один ключевой атрибут "Название".

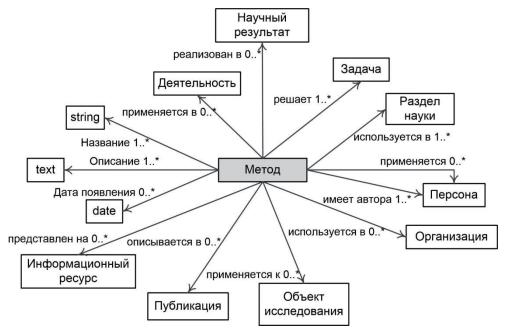


Рисунок 2 – Паттерн для представления понятия "Метод"

Примеры вопросов оценки компетентности, представляющих содержание паттерна "Метод":

- Как называется метод?
- Кто является автором метода?
- Когда был предложен метод?
- К каким объектам исследования применяется метод?
- Какие задачи решаются с помощью метода?
- В какой деятельности используется метод?
- В каких научных результатах реализован метод?
- Кто применяет метод?
- В каких организациях используется метод?

2 Архитектура системы автоматизированного построения онтологий НПрО

Система автоматизированного построения онтологий (САПО) НПрО на основе разнородных паттернов ОП состоит из следующих компонентов (рисунок 3): онтологии паттернов ОП, репозитория паттернов ОП, репозитория базовых онтологий, словаря общенаучной лексики, редактора онтологий, редактора данных, информационно-аналитического интернетресурса и подсистемы автоматического пополнения онтологии на основе ЛСП.

Репозиторий паттернов ОП строится на основе онтологии паттернов ОП и включает реализации паттернов ОП. При этом паттерны содержания, паттерны представления и структурные логические паттерны реализованы средствами языка OWL, в то время как ЛСП представляются на специализированном языке описания шаблонов [22].

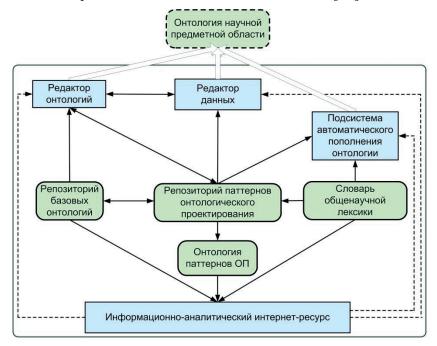


Рисунок 3 – Архитектура системы автоматизированного построения онтологий

САПО поддерживает метод построения онтологии НПрО на основе базовых онтологий, содержащих наиболее общие понятия, характерные для большинства НПрО. В связи с этим в систему включён репозиторий базовых онтологий, содержащий следующие онтологии: онтологию научного знания, онтологию научной деятельности, базовую онтологию задач и методов, базовую онтологию информационных ресурсов [8]. Все базовые онтологии имеют спецификации на языке OWL. Для наиболее важных понятий базовых онтологий разработаны паттерны содержания и включены в репозиторий ОП.

В качестве редактора онтологий в системе используется популярный редактор $Prot\acute{e}g\acute{e}$. Для удобного использования паттернов ОП в систему включён редактор данных, позволяющий пополнять онтологию НПрО путём конкретизации паттернов содержания, включённых в репозиторий паттернов ОП.

Словарь общенаучной лексики содержит семантически-размеченные термины, используемые в научных текстах для описания сущностей различных НПрО. Он используется для извлечения предметной лексики из текстов и автоматической генерации словаря ПрО, а также для последующего автоматического анализа текстов с помощью ЛСП.

Подсистема автоматического пополнения онтологии предназначена для занесения в онтологию НПрО информации, извлекаемой из текстов на естественном языке. Для этого используются ЛСП, построенные на основе паттернов содержания и словаря общенаучной лексики.

Информационно-аналитический Интернет-ресурс (ИАИР) предназначен для систематизации информации о паттернах ОП и обеспечения содержательного доступа к ней (см. рисунок 4). Работа ресурса организована на основе онтологии паттернов ОП, которая является его концептуальной основой.



Рисунок 4 – Информационно-аналитический Интернет-ресурс

В левой части рисунка 4 показана иерархия классов онтологии ОП, а в правой части представлено описание паттерна содержания *Научная деятельность*, которое включает название паттерна, описание его назначения, ссылку на *OWL*-представление, ссылку на графическое представление, набор вопросов оценки компетентности, а также связи с проектами, в рамках которых он разрабатывался и используется.

Кроме того, ИАИР «Паттерны онтологического проектирования» является пользовательским интерфейсом САПО, обеспечивающим пользователям доступ ко всем репозиториям и редакторам, поддерживающим разработку онтологии НПрО, а также к подсистеме автоматического пополнения онтологии на основе ЛСП.

3 Методика построения онтологии НПрО инженерами знаний

Онтология любой НПрО содержит не только описания присущих ей системы понятий, задач и методов обработки и анализа информации, но и описания релевантных ей информационных ресурсов. В связи с этим онтологию НПрО удобно представлять в виде системы взаимосвязанных онтологий: онтологии области знаний, онтологии задач и методов, онтологии научных Интернет-ресурсов.

Онтология области знаний задаёт систему понятий и отношений, предназначенных для детального описания моделируемой НПрО и выполняемой в её рамках научной и исследовательской деятельности. Онтология задач и методов описывает задачи, решаемые в данной НПрО, и методы их решения. Онтология научных Интернет-ресурсов служит для описания, представленных в сети Интернет информационных ресурсов, релевантных данной НПрО.

Построение онтологии конкретной НПрО с использованием базовых онтологий и системы паттернов ОП включает два этапа.

- 1) построение компонентов онтологии НПО на основе базовых онтологий путём их достройки и развития. На этом этапе выполняется специализация представленных в базовых онтологиях структурных логических паттернов и паттернов содержания на конкретную НПрО.
- 2) пополнение онтологии НПрО путём конкретизации структурных логических паттернов и паттернов содержания, представленных в базовых онтологиях или полученных из них путём их специализации на конкретную НПрО.

Специализация паттерна может состоять в переименовании, в уточнении имён и значений его свойств (атрибутов и отношений). Специализация паттернов на примере структурного логического паттерна "Бинарное атрибутированное отношение" приведена на рисунке 5.

Центральное место в этом паттерне занимает служебный класс *Отношение с атрибутами*, с которым связываются базовые классы, моделирующие аргументы бинарного отношения, посредством связей "является Аргументом" и "имеет Аргумент". При этом в паттерне (в метках связей) указывается, что таких аргументов должно быть по одному. Атрибуты бинарного атрибутированного отношения моделируются свойствами класса *Отношение с атрибутами* "имеет Атрибут" и "имеет Атрибут из Домена". В общем случае такое отношение может не иметь атрибутов, что отражено в метках связей, представляющих эти свойства.

Конкретизация (означивание) паттерна состоит в подстановке в него конкретных значений свойств.

На рисунке 6 представлен пример конкретизации паттерна содержания "Метод". (Этот паттерн был использован для представления информации о методе недоопределённых вычислений, предложенном А.С. Нариньяни в 1986 г. и реализованном в *UniCalc* [23]).

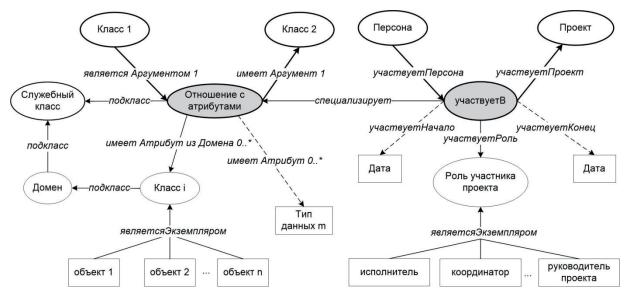


Рисунок 5 – Паттерн бинарного атрибутированного отношения и пример его специализации

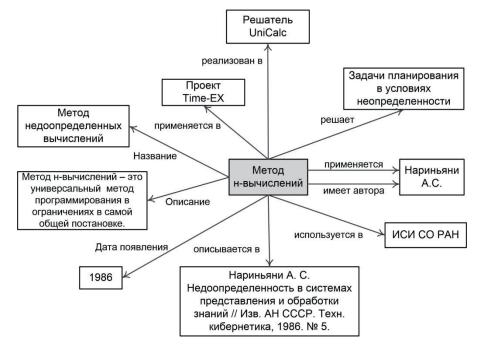


Рисунок 6 – Конкретизация паттерна содержания "Метод"

Использование паттернов содержания при пополнении онтологии НПрО поддерживается специальным редактором данных (см. рисунок 7), который позволяет специалистам в ПрО пополнять онтологию фактическими данными — объектами классов и их свойствами. При пополнении онтологии пользователь из представленной ему иерархии классов онтологии выбирает нужный класс, редактор по имени класса находит соответствующий паттерн и на его основе строит форму, содержащую поля для заполнения свойств объекта этого класса пользователем. Редактор может интерпретировать отношения с атрибутами, описанные паттерном на рисунке 5. Благодаря этому пользователь может работать с задаваемыми с помощью таких отношений свойствами объекта как с «обычными» объектными свойствами (Object Property). Отличие состоит в необходимости задания соответствующих значений атрибутов такого отношения.



Рисунок 7 – Редактор данных

4 Представление ЛСП

Для реализации автоматического пополнения онтологии для каждого паттерна содержания строится набор ЛСП, описывающих различные способы представления соответствующей ему информации в научных текстах, на основе которых осуществляется извлечение информации.

ЛСП организуются в многоуровневую систему, включающую терминологические паттерны (Т-ЛСП) и информационные паттерны (И-ЛСП). Т-ЛСП предназначены для описания элементарных языковых конструкций и извлечения новых терминов. С помощью И-ЛСП задаётся схема извлечения из текста фактов и генерации соответствующих им элементов онтологии. Под фактами здесь понимаются тройки вида <Object, Property, Value>, где Object — это сущность НПрО, найденная в тексте, Property — имя свойства данной сущности, а Value — значение этого свойства. При сопоставлении с онтологией Object соотносится с индивидом (экземпляром) какого-либо класса онтологии, Property — с названием свойства этого класса (это может быть отношение "type", связывающее индивид с классом, имя атрибута (Datatype Property) или отношения (Object Property)), Value — с соответствующим онтологическим значением.

Каждый ЛСП реализует модель вида *<Arguments, Constraints, Results>*, где *Arguments* – множество семантических аргументов факта, которым сопоставляются либо термины НПрО, либо объекты (при условии, что объекты уже извлечены с помощью других И-ЛСП), *Constraints* – семантические, синтаксические и/или позиционные условия на аргументы, а Results описывает результат применения ЛСП, которым может быть либо новый термин (для Т-ЛСП), либо генерируемый фрагмент онтологии (для И-ЛСП).

4.1 Предметный словарь

Для автоматического пополнения онтологии с помощью ЛСП требуется обеспечить извлечение из текста специфических терминов данной НПрО. Для этого используется словарь предметной лексики (предметный словарь) и терминологические паттерны, позволяющие извлекать новые термины (в частности, наименования объектов НПО или специфичных предикатных слов).

Предметный словарь создаётся как расширение словаря общенаучной лексики и включает слова и словосочетания (термины), организованные в соответствии с семантикой ПрО (см. рисунок 8). В словарной статье хранится вся информация, необходимая для извлечения термина из текста и поддержки последующих этапов анализа текста. Каждый термин предметного словаря, найденный в тексте, снабжается морфологической и семантической информацией, которая в дальнейшем используется при применении ЛСП.

Словарь описывается системой вида $D = \langle W, P, M, G, S \rangle$, где W – множество лексем, каждой из которых сопоставлена информация обо всей совокупности её форм; P – множество многословных терминов, описываемых парой $\langle N$ -грамма, тип структуры \rangle , где N-грамма задаёт последовательность лексем, а тип структуры определяет вершину и правила согласования элементов N-граммы; M – морфологическая модель языка, включающая описание морфологических классов и признаков; G – множество правил согласования для извлечения многословных терминов; S – проблемно-ориентированная система лексикосемантических признаков.

Семантический компонент предметного словаря включает две независимые иерархии лексико-семантических классов: универсальную иерархию, унаследованную из словаря общенаучной лексики, и предметно-ориентированную иерархию, создаваемую на основе онтологии НПрО. Для автоматической генерации словаря создана методика автоматического формирования системы лексико-семантических характеристик на основе имён элементов онтологии. На рисунке 8 представлена иерархия лексико-семантических признаков, созданная на основе фрагмента онтологии, относящегося к классу *Метод*. Все подклассы класса *Метод* сгенерированы автоматически из названий онтологических элементов в соответствии с шаблоном <имя класса.имя признака>.

Все термины словаря размечены признаками из предметно-ориентированной и/или универсальной иерархии. Термины конкретной НПрО в словаре общенаучной лексики получают синкретичный признак с одновременно выраженным значением универсального и предметного лексико-семантического класса. Так, для глагола *'использовать'* выделено четыре синкретичных признака (рисунок 8), каждый из которых включает универсальный класс *Применение* в сочетании с онтологически обусловленными признаками: *Метод.применяется_в, Метод.применяется_к, Метод.используется_в* и *Метод.реализует*.

Лексико-семантические признаки словаря используются при описании ЛСП (в аргументах и/или результате) как способ обращения к терминам НПрО с определённой семантикой, о которой в общем случае заранее ничего неизвестно.

4.2 Терминологические паттерны

Терминологические паттерны используются для извлечения из текстов новых терминов НПрО, не заданных в словаре. Т-ЛСП представляют собой ЛСП, формируемые на основе опорных терминов, маркеров, семантических и синтаксических ограничений.

Для автоматической генерации Т-ЛСП разработан язык и предложена методика создания типовых паттернов (или мета-паттернов), в состав которых включаются переменные.

Создание ЛСП для конкретной онтологии осуществляется путём означивания переменных именами онтологических классов, атрибутов и отношений.

Для извлечения терминов предложены два типа паттернов.

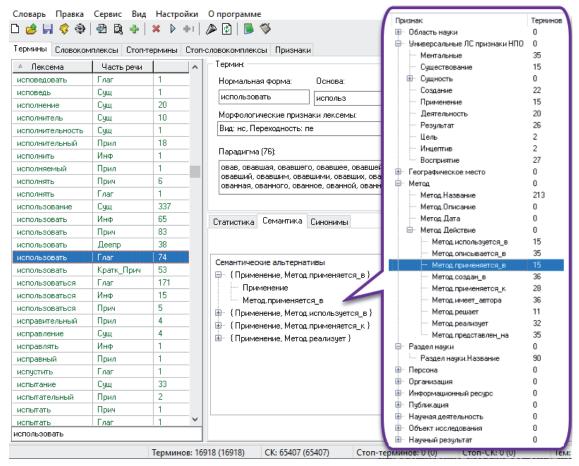


Рисунок 8 – Предметный словарь

Паттерны первого типа позволяют извлекать имена индивидов на основе центрального слова или термина с привлечением синтаксических правил сборки именных групп:

$$[\langle Adj \rangle^*, X, [\langle Adj \rangle^*, \langle N, GEN \rangle]^*] \Rightarrow X.$$
Название. (1)

Данный Т-ЛСП включает три аргумента: 1) группу прилагательных <Adj>*, 2) термин лексико-семантического класса (признака) X и 3) именную группу в родительном падеже, собранную по вложенному подшаблону вида [<Adj>*, <N, GEN >]*. Ограничениями здесь являются указание семантического класса для 2-го аргумента и указание падежа для вложенного Т-ЛСП, используемого в качестве 3-го аргумента. Результат Т-ЛСП определяет лексико-семантический признак X. Название для всех терминов, извлекаемых с помощью данного Т-ЛСП.

Данный типовой паттерн позволяет генерировать конкретные Т-ЛСП путём подстановки в качестве X имён классов онтологии. Например, для извлечения названия метода может быть автоматически сгенерирован следующий паттерн:

 $[< Adj > *, Meтoд, [< Adj > *, < N, GEN >] *] \Rightarrow Meтoд. Название.$

Этот паттерн позволяет извлечь такие термины: 'метод опорных векторов', 'метод анкетного опроса', 'метод медиан рангов', 'метод нейронных сетей', 'метод самооценки' и т.п.

Вторая группа паттернов позволяет извлекать новые термины на основе контекста, в котором присутствуют известные термины НПрО. Данный контекст сопоставляется с онтологическим отношением следующим образом:

$$[X, X.Rel, $t < NP >] \Rightarrow Y.Hasbahue$$

$$[X, $t < VP >, Y] \Rightarrow X.Rel.$$
(2)

Данные типовые паттерны позволяют генерировать Т-ЛСП, в которых классы X и Y связаны в онтологии отношением Rel₂ и приписывать извлечённым терминам семантику соответствующего онтологического отношения. С помощью второго паттерна можно анализировать вопросы оценки компетентности, представленные в паттернах содержания.

При означивании переменных (X = Memod, Y = 3adaчa, Rel = pemaem) могут получиться следующие Т-ЛСП:

[Метод, Метод.решает, t< NP>] \Rightarrow Задача.Название [Метод, x< VP>, Задача] \Rightarrow Метод.Решает.

С помощью этих паттернов можно извлечь такие термины как: 'задача классификации', 'нахождение оптимальной проекции', а также предикатные термины: 'ищут решения', 'позволяет решать', 'позволяют построить', 'позволил выделить' и др.

4.3 Информационные паттерны

Информационные паттерны предназначены для извлечения фактов и создания новых объектов для пополнения онтологии. Предложены три вида типовых И-ЛСП.

И-ЛСП первого вида (инициализирующие И-ЛСП) обеспечивают создание объектов на основе терминов, имеющих лексико-семантический признак, совпадающий либо с именем класса, либо с именем ключевого атрибута этого класса.

$$[X.\text{Hазваниe}] \Rightarrow \text{create } X(\text{Названиe: arg1}).$$
 (3)

Данные типовые паттерны позволяют генерировать конкретные И-ЛСП путём подстановки в качестве X имён классов онтологии, по которым впоследствии будут извлекаться объекты.

Примеры таких паттернов:

[Метод.Название] ⇒ create Метод(Название: arg1) [Задача.Название] ⇒ create Задача(Название: arg1).

Второй вид И-ЛСП предназначен для извлечения отношений между объектами (*Object Property*) на основе предикатных слов. Данный тип паттерна описывается тремя аргументами: двумя объектами соответствующих классов X и Y и связывающим их термином с лексико-семантическим признаком X.Rel.

$$[X(), X.Rel, Y()] \Rightarrow set X(Rel: arg3). \tag{4}$$

Данный паттерн позволяет генерировать И-ЛСП, в которых классы X и Y связаны в онтологии отношением Rel. Паттерн задаёт общую структуру И-ЛСП, которую требуется уточнить набором синтаксических ограничений, определяемых предикатным термином. Данные ограничения могут быть получены на основе корпуса текстов, используя уже установленные на уровне онтологии связи, или на основе статистических критериев.

Пример такого паттерна:

[Метод(), Метод.решает, 3адача()] \Rightarrow set arg1(решает: arg3).

Следующий вид И-ЛСП обеспечивает заполнение атрибутов объектов (*Datatype Property*) либо на основе предикатных терминов, сигнализирующих о наличии рядом с упоминанием объекта его атрибута, либо при контактном способе выражения значения атрибута:

$$[X(), X.A.type, X.A] \Rightarrow set X(A: arg3)$$

$$[X(), X.A] \Rightarrow set X(A: arg2).$$
(5)

Данные паттерны позволяют генерировать И-ЛСП, которые обеспечивают заполнение атрибута объекта X значением термина с семантикой X.А. Предикатные термины являются либо именами атрибутов, либо глагольными группами, вычисленными на основе вопросов оценки компетентности с семантикой Prop.A. Данные паттерны так же, как и в предыдущем случае, могут уточняться синтаксическими ограничениями.

Примеры таких паттернов:

[Метод(), Метод.Описание.type, Метод.Описание] ⇒ set Метод (Описание: arg3) [Персона(), Персона.Ученая_степень] ⇒ set Персона(Ученая_степень: arg2).

5 Автоматическая генерация ЛСП

ЛСП автоматически строятся на основе паттернов онтологического проектирования, словарей общена-учной и предметной лексики и текущей версии онтологии НПрО. На рисунке 9 представлена схема взаимосвязей компонентов системы, участвующих в генерации ЛСП.

Процесс генерации ЛСП начинается с создания и наполнения предметного словаря. Из онтологии и описания паттерна содержания извлекаются термины (лексемы и терминоподобные N-граммы) и формируются лексико-семантические классы. Все термины размечаются соответствующими семантическими признаками.

На основе анализа структуры паттерна содержания осуществляется означивание переменных в метапаттернах и формируются Т-ЛСП и И-ЛСП.

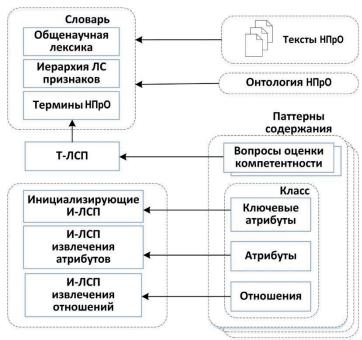


Рисунок 9 — Схема взаимосвязей компонентов системы, участвующих в генерации ЛСП

С помощью созданных Т-ЛСП анализируются вначале вопросы оценки компетентности, а затем тексты научного корпуса. Вопросы оценки компетентности, выраженные на естественном языке, позволяют не только извлечь предикатные термины, но и задать начальные синтаксические ограничения на извлекаемые факты, которые могут быть впоследствии уточнены на основе корпуса текстов.

При генерации И-ЛСП необходима информация о ключевых атрибутах классов онтологии (особенно для инициализирующих паттернов). В приведённых выше примерах использо-

ван атрибут *Название*, однако у классов онтологии могут быть и другие ключевые атрибуты (например, у класса *Персона*), для которых потребуется создание других типовых паттернов. Для И-ЛСП также требуется уточнение синтаксических и позиционных ограничений на основе примеров вхождений И-ЛСП в корпус текстов.

Таким образом, из онтологических компонентов знаний можно выделить знания о языке ПрО, необходимые для извлечения информации из текста и пополнения онтологии, и возможные способы языкового описания онтологических сущностей в текстах. Формализация этих знаний в виде системы ЛСП позволит применить существующие технологии автоматической обработки текстов для автоматического пополнения онтологии НПрО.

6 Пополнение онтологии НПрО на основе ЛСП

6.1 Архитектура подсистемы пополнения онтологии

При разработке подсистемы применены следующие инструменты и технологии: система извлечения из текстов предметной лексики и построения словарей KLAN [22], система анализа текста на основе шаблонов PatTerm [24] и система фактографического анализа текста FATON [25]. Система KLAN позволяет проводить морфологический и поверхностный синтаксический анализ текстов, создавать на основе результатов этого анализа предметноориентированные словари и извлекать словарные термины из текстов. На базе системы KLAN создаются словари общенаучной и предметной лексики. Система PatTerm используется для поиска в тексте терминов на основе $T-JIC\Pi$, а также для генерации $T-JIC\Pi$ по строковым константам. Система FATON позволяет собирать факты по заданным $II-JIC\Pi$ и множеству терминов, найденных в тексте системами $II-IIC\Pi$ и $II-IIC\Pi$ и множеству терминов, найденных в тексте системами $II-IIC\Pi$ и $II-IIC\Pi$ и множеству терминов, найденных в тексте системами $II-IIC\Pi$ и $II-IIC\Pi$ и II-

Таким образом, подсистема автоматического пополнения онтологии объединяет различные технологии, обеспечивая их взаимодействие. Она включает модуль взаимодействия с онтологиями и две подсистемы: подсистему генерации ЛСП и подсистему извлечения информации из корпуса текстов на основе ЛСП (см. рисунок 10).



Рисунок 10 – Архитектура подсистемы пополнения онтологии

Подсистема генерации ЛСП обеспечивает построение всех видов ЛСП – терминологических и информационных, а также построение предметного словаря, необходимого для их работы.

Подсистема пополнения онтологии НПрО на основе ЛСП организована в виде конвейера из нескольких последовательных обработчиков и реализует алгоритм пополнения онтологии в несколько этапов.

На этапе 1 из текста с помощью системы *KLAN* извлекаются словарные термины, которые снабжаются грамматическими и семантическими характеристиками. На этапах 2 и 3 система *PatTerm* извлекает построенные подсистемой генерации ЛСП N-граммы и новые термины, на основе Т-ЛСП. Найденные термины используются на этапе извлечения фактов на основе И-ЛСП (этап 4). Мультиагентный алгоритм, реализованный в системе FATON, подбирает аргументы для заданных И-ЛСП, проверяет ограничения и собирает объекты для пополнения онтологии. На последнем шаге (этап 5) найденные в тексте объекты идентифицируются, т.е. сопоставляются с индивидами онтологии, при необходимости объединяются и либо обновляют данные для уже имеющихся в онтологии индивидов, либо генерируют новые индивиды.

Взаимодействие подсистем с онтологией паттернов и онтологией НПрО обеспечивает специально разработанный модуль, использующий средства поддержки онтологически ориентированного программирования из библиотеки *owlready2* [26]. Для обработки онтологий НПрО, спроектированных с использованием структурных логических паттернов, таких как "Бинарное атрибутированное отношение" или "Область допустимых значений", модуль был расширен дополнительными функциями, позволяющими работать со свойствами, описанными сложными структурными паттернами, как и со стандартными свойствами из языка *OWL*.

6.2 Экспериментальная проверка подсистемы пополнения онтологии

Экспериментальная проверка подсистемы автоматического пополнения онтологии была проведена на корпусе русских текстов, являющихся краткими описаниями различных методов (длиной от 1 до 5 предложений) из НПрО «Поддержка принятия решений в слабоформализованных областях». На основе онтологии данной НПрО [26] и паттерна содержания *Метод* были автоматически созданы словарь ПрО, включающий 214 терминов, размеченных с помощью 21 лексико-семантического признака, 34 Т-ЛСП для извлечения новых терминов (названий экземпляров классов и предикатных слов) и 82 И-ЛСП для извлечения экземпляров класса *Метод* и их свойств. Для генерации иерархии лексико-семантических признаков использовались метки (*rdfs:label*) атрибутов и отношений класса *Метод*, а также значения атрибутов для экземпляров этого и связанных с ним классов.

С помощью Т-ЛСП были сгенерированы и внесены в словарь термины: 'аддитивная свертка' (Метод.Название), 'искусственный интеллект' (Раздел_науки.Название), 'задача о ранце' (Задача.Название), 'решает' (Метод.решает) и т.п.

Подсистемой генерации ЛСП были построены следующие И-ЛСП:

[Метод.Название]
$$\Rightarrow$$
 create Метод (Название: arg1) (6.1)

$$[3адача. Название] \Rightarrow$$
 create $3адача$ (Название: arg1) (6.2)

[Метод (), Метод.решает, Задача ()]
$$\Rightarrow$$
 arg1::Метод (Решает: arg3) (6.3)

На рисунке 11 представлены результаты обработки следующего текста из корпуса:

"Memod опорных векторов (Support Vector Machines) решает задачи классификации и регрессионного анализа путём построения нелинейной плоскости, разделяющей решения".

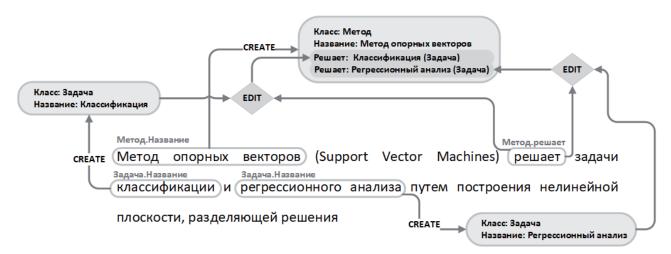


Рисунок 11 – Пример анализа текста на основе ЛСП

В процессе анализа этого текста были найдены четыре термина с лексикосемантическими признаками *Метод.Название* ('метод опорных векторов'), Задача.Название ('классификация' и 'регрессионный анализ') и *Метод.решает* ('решает'). Затем, в соответствии с инициализирующими И-ЛСП (6.1) и (6.2) были созданы объекты классов *Метод* и Задача; согласно шаблону (6.3) были порождены связи между объектом класса *Метод* и двумя объектами класса Задача.

Заключение

В статье описан предложенный подход к автоматизации построения и пополнения онтологий НПрО, базирующийся на разнородных паттернах ОП, и реализующая его САПО.

Особенностью данного подхода является то, что разработка и начальное наполнение онтологии НПрО выполняется инженерами знаний и экспертами в НПрО с использованием паттернов содержания и структурных паттернов. Дальнейшее пополнение онтологии НПрО выполняется в автоматическом режиме с использованием ЛСП, построенных на основе включённых в репозиторий САПО паттернов содержания и текущей версии онтологии НПрО. Главное отличие предложенного подхода от подходов, также использующих ЛСП [16, 17, 28], заключается в том, что в нём ЛСП генерируются автоматически.

Благодарности

Работа выполнена при финансовой поддержке РФФИ (проект № 19-07-00762).

Список источников

- [1] Загорулько, Ю.А. Методы и методологии разработки, сопровождения и реинжиниринга онтологий / Ю.А. Загорулько // Симпозиум «Онтологическое моделирование: состояние и направления исследований и применения» (Звенигород, 20-21 мая). М.: ИПИ РАН, 2008. С.135–162.
- [2] *Sattar, A.* Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review / A. Sattar, E. Salwana, M. Surin, M.N. Ahmad, M. Ahmad, A.K. Mahmood // International Journal of Advanced Computer Science and Applications. 2020. Vol.11(5). P.99–108.
- [3] *Noy, N.* Ontology Development 101: A Guide to Creating Your First Ontology / N. Noy, D. McGuinness // Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001.
- [4] *Brusa*, *G*. Towards ontological engineering: a process for building a domain ontology from scratch in public administration / G. Brusa, M. L. Caliusco, and O. Chiotti // Expert Systems. 2008. Vol.25. P.484-503.

- [5] *De Nicola, A.* A Lightweight Methodology for Rapid Ontology Engineering / A. De Nicola, M. Missikoff // Com. ACM. 2016. Vol.59. P.79–86.
- [6] *Загорулько, Ю.А.* Технология построения онтологий для порталов научных знаний / Ю.А. Загорулько, О.И. Боровикова // Вестник НГУ. Серия: Информационные технологии (ISSN 1818-7900). 2007. Т.5, №2. С.42-52.
- [7] *Gangemi, A.* Ontology Design Patterns / A. Gangemi, V. Presutti // In: Staab S., Studer R. (eds) Handbook on Ontologies. IHIS. Springer, Berlin, Heidelberg, 2009. P.221–243.
- [8] Загорулько, Ю.А. Технология создания тематических интеллектуальных научных интернет-ресурсов, базирующаяся на онтологии / Ю.А. Загорулько, Г.Б. Загорулько, О.И. Боровикова // Программная инженерия. – 2016. Т.7. №2. С.51-60.
- [9] *Asim, M.-N.* A survey of ontology learning techniques and applications / M.-N. Asim, M. Wasim, M.U.G. Khan et al. // Database: The Journal of Biological Databases and Curation. 2018. Vol.2018. doi:10.1093/database/bay101.
- [10] *Blomqvist, E.* Engineering Ontologies with Patterns: The eXtreme Design Methodology / E. Blomqvist, K. Hammar, V. Presutti // In: Hitzler, P., Gangemi, A., Janowicz, K., Krisnadhi, A., Presutti, V. (eds.) Ontology Engineering with Ontology Design Patterns. Studies on the Semantic Web. Vol.25. Amsterdam, IOS Press, 2016. P.23–50.
- [11] *Ломов, П.А.* Применение паттернов онтологического проектирования для создания и использования онтологий в рамках интегрированного пространства знаний / П.А. Ломов // Онтология проектирования. 2015. Т.5. №2(16). С.233-245.
- [12] Загорулько, Ю.А. Применение паттернов онтологического проектирования при разработке онтологий научных предметных областей / Ю.А. Загорулько, О.И. Боровикова, Г.Б. Загорулько // Selected Papers of the 19th International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2017. Vol.2022. P.258-265. CEUR Workshop Proceedings (CEUR-WS.org), 2017.
- [13] *Zagorulko, Yu.* Pattern-Based Methodology for Building the Ontologies of Scientific Subject Domains / Yu. Zagorulko, O. Borovikova, G. Zagorulko // In: New Trends in Intelligent Software Methodologies, Tools and Techniques. Proceedings of the 17th International Conference SoMeT_18. H. Fujita and E. Herrera-Viedma (Eds.). Series: Frontiers in Artificial Intelligence and Applications. Vol. 303. Amsterdam: IOS Press, 2018. P.529-542.
- [14] Petasis, G. Ontology Population and Enrichment: State of the Art / Petasis, G., Karkaletsis, V., Paliouras, G., Krithara, A., Zavitsanos, E. // In: Paliouras, G., Spyropoulos, C.D., Tsatsaronis, G. (eds) Knowledge-Driven Multimedia Information Extraction and Ontology Evolution). Lecture Notes in Computer Science, vol. 6050. Springer, Berlin, Heidelberg, 2011. P.134–166.
- [15] *Ganino, G.* Ontology population for open-source intelligence: a GATE-based solution / G. Ganino, D. Lembo, M. Mecella, F. Scafoglieri // Software: Practice and Experience. 2018. Vol.48(12).
- [16] *Maynard, D.* Using Lexico-Syntactic Ontology Design Patterns for Ontology Creation and Population / D. Maynard, A. Funk, W. Peters // In: Proc. Workshop on Ontology Patterns (WOP 2009), collocated with the 8th Int. Semantic Web Conf. (ISWC-2009). Vol.516. P.39–52. CEUR Workshop Proceedings (CEUR-WS.org), 2009.
- [17] *Ijntema, W.* A lexico-semantic pattern language for learning ontology instances from text / W. Ijntema, J. Sangers, F. Hogenboom, F. Frasincar // Journal of Web Semantics. 2012. Vol.15. P.37–50.
- [18] NeOn project home. http://www.neon-project.org
- [19] Association for Ontology Design & Patterns (ODPA). http://ontologydesignpatterns.org/wiki/ODPA
- [20] *Karima, N.* How to Document Ontology Design Patterns / Karima, N., Hammar, K., Hitzler, P. // In: Advances in Ontology Design and Patterns. Studies on the Semantic Web. vol.32. P.5–27. IOS Press, Kobe, Japan (2017).
- [21] *Antoniou, G.* Web Ontology Language: OWL / G. Antoniou, F. Harmelen // In: Handbook on Ontologies. Staab, S. and R. Studer. (eds.) Berlin: Springer Verlag, 2009. P.91-110.
- [22] *Сидорова, Е.А.* Подход к моделированию процесса извлечения информации из текста на основе онтологии / Е.А. Сидорова // Онтология проектирования. 2018. Т.8, №1(27). С.134-151. DOI: 10.18287/2223-9537-2018-8-1-134-151.
- [23] *Нариньяни, А.С.* Неопределённость в системах представления и обработки знаний / А.С. Нариньяни // Изв. АН СССР. Техн. кибернетика, 1988. №5.
- [24] *Sidorova, E.* The Software Environment for Multi-Aspect Study of Lexical Characteristics of Text / E. Sidorova, I. Akhmadeeva // In: Selected Papers of the XX International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2019). CEUR Workshop Proceedings, 2019. Vol.2523. P.306–315.
- [25] *Garanina, N.* A Multi-agent Text Analysis Based on Ontology of Subject Domain / N. Garanina, E. Sidorova, E. Bodin // In: Voronkov, A., Virbitskaite, I. (eds.) Perspectives of System Informatics (PSI 2014). LNCS, vol. 8974. Springer, Heidelberg 2015. P.102–110.
- [26] *Lamy, J.-B.* Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies / J.-B. Lamy // Artificial Intelligence In Medicine. 2017. Vol.80. P.11–28.
- [27] *Загорулько, Г.Б.* Разработка онтологии для интернет-ресурса поддержки принятия решений в слабоформализованных областях / *Г.Б. Загорулько* // Онтология проектирования. 2016. Т.б. №4(22). С.485-500.

[28] *de Cea, G.A.* Using Linguistic Patterns to Enhance Ontology Development / G.A. de Cea, A. Gomez-Perez, E. Montiel-Ponsoda, M. C. Suarez-Figueroa // In: Proc. Int. Conf. on Knowledge Engineering and Ontology Development (KEOD 2009) (Funchal - Madeira, Portugal, October 6-8, 2009). P.206–213. INSTICC Press, 2009.

Сведения об авторах



Загорулько Юрий Алексеевич, 1957 г. рождения. Окончил Новочеркасский политехнический институт им. С. Орджоникидзе (1979), к.т.н. (1989). Заведующий лабораторией Института систем информатики им. А.П. Ершова СО РАН, доцент кафедры программирования и кафедры систем информатики Новосибирского государственного университета. Член Российской и Европейской ассоциаций искусственного интеллекта. В списке научных трудов более 280 публикаций в области искусственного интеллекта, разработки интеллектуальных систем, инженерии знаний, онтологического моделирования и компьютерной лингвистики. ORCID: 0000-0002-7111-6524; Author ID (Scopus): 23394231500;

Researcher ID (WoS): R-1826-2016. zagor@iis.nsk.su.

Сидорова Елена Анатольевна, 1977 г. рождения. Окончила Новосибирский государственный университет в 2000 г., к.ф.-м.н. (2006). Старший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова (Новосибирск), старший преподаватель кафедры программирования Новосибирского государственного университета, член Российской и Европейской ассоциаций искусственного интеллекта. В списке научных трудов более 140 работ в области компьютерной лингвистики, мультиагентных систем, представления знаний и онтологического инжиниринга. ORCID: 0000-0001-8731-3058; Author ID (Scopus): 41961707000; Researcher ID (WoS): K-2432-2018. *lsi-dorova@iis.nsk.su*.





Загорулько Галина Борисовна, 1963 г. рождения. Окончила Новосибирский государственный университет (1985), к.т.н. (2020). Научный сотрудник Института систем информатики им. А.П. Ершова СО РАН, старший преподаватель кафедры программирования Новосибирского государственного университета. Член Российской ассоциации искусственного интеллекта. В списке научных трудов более 120 публикаций в области искусственного интеллекта, разработки интеллектуальных систем, инженерии знаний, онтологического моделирования, поддержки принятия решений. ORCID: 0000-0003-2155-1357; Author ID (Scopus): 35249924100; Researcher ID (WoS): K-2472-2018.

gal@iis.nsk.su.

Ахмадеева Ирина Равильевна, 1991 г. рождения. Окончила Новосибирский государственный университет в 2015г. Младший научный сотрудник Института систем информатики им. А.П. Ершова СО РАН, ассистент кафедры программирования Новосибирского государственного университета. В списке научных трудов более 20 работ в области искусственного интеллекта, разработки интеллектуальных систем и компьютерной лингвистики. ORCID:



0000-0002-7371-1087; Author ID (RSCI): 874172; Author ID (Scopus): 57188681471; Researcher ID (WoS): K-3145-2018. *i.r.akhmadeeva@iis.nsk.su*



Серый Алексей Сергеевич, 1987 г. рождения. Окончил Новосибирский государственный университет в 2010 г. Младший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова (Новосибирск). В списке научных трудов более 20 работ в области представления знаний и компьютерной лингвистики. ORCID: 0000-0001-8275-4700; Author ID (Scopus): 56403204900; Researcher ID (WoS): K-1557-2018. alexev.seryj@iis.nsk.su

Поступила в редакцию 12.11.2021, принята к публикации 08.12.2021.

Automation of the development of ontologies of scientific subject domains based on ontology design patterns

Yu.A. Zagorulko, E.A. Sidorova, G.B. Zagorulko, I.R. Akhmadeeva, A.S. Sery

A.P. Ershov Institute of Informatics Systems of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

Abstract

At present, ontologies are recognized as the most effective means of formalizing and systematizing knowledge and data in scientific subject domains (SSDs). However, the development of an ontology is a rather complicated and time-consuming process. All indications are that when developing SSDs ontologies, it is especially effective to use ontology design patterns (ODPs). This is due to the fact that the SSD ontology, as a rule, contains a large number of typical fragments, which are well described by the ODPs. In addition, due to the fact that the use of ODPs greatly facilitates the development of an SSD ontology, it is possible to involve experts in a modeled SSD not possessing the skills of ontological modeling. To obtain an ontology that adequately describes the SSD, it is necessary to process a huge number of publications relevant to the modeled SSD. It is possible to facilitate and accelerate the process of populating the ontology with information from such sources by using the lexical and syntactic patterns of ontological design. The paper presents an approach to the automated development of SSDs ontologies based on a system of heterogeneous ODPs. This system includes both ODPs intended for ontology developers and lexical and syntactic patterns built on the basis of the above-mentioned types of the ODPs and the current version of the SSD ontology.

Key words: ontology design patterns, scientific subject domains, content patterns, automatic generation of lexical and syntactic patterns, ontology population.

Citation: Zagorulko YuA, Sidorova EA, Zagorulko GB, Akhmadeeva IR, Sery AS. Automation of the development of ontologies of scientific subject domains based on ontology design patterns [In Russian]. Ontology of designing. 2021; 11(4): 500-520. DOI: 10.18287/2223-9537-2021-11-4-500-520.

Acknowledgment: This work was supported by the Russian Foundation for Basic Research (project No. 19-07-00762).

List of figures

- Figure 1 Fragment of the ODP ontology
- Figure 2 ODP for representing the "Method" concept
- Figure 3 Architecture of the system for the automated construction of ontologies
- Figure 4 Information and analytical Internet resource
- Figure 5 ODP for representing the attributed binary relation and an example of its specialization
- Figure 6 the "Method" content pattern specification
- Figure 7 Data editor
- Figure 8 Subject-specific vocabulary
- Figure 9 Diagram of the interconnections of the system components involved in the generation of the LSP
- Figure 10 Architecture of ontology population system
- Figure 11 An example of text analysis based on LSP

References

- [1] **Zagorulko YuA.** Methods and methodologies for the development, maintenance and reengineering of ontologies [In Russian]. Symposium "Ontological Modeling: State and Directions of Research and Application" (Zvenigorod, May 20-21). Moscow: IPI RAN, 2008: 135–162.
- [2] *Sattar A, Salwana E, Surin M, Ahmad M, Ahmad M, Mahmood A.* Comparative Analysis of Methodologies for Domain Ontology Development: A Systematic Review. International Journal of Advanced Computer Science and Applications. 2020; 11(5): 99–108.

- [3] *Noy N, McGuinness D.* Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March. 2001.
- [4] **Brusa G., Caliusco ML, Chiotti O.** Towards ontological engineering: a process for building a domain ontology from scratch in public administration. Expert Systems. 2008; 25: 484-503.
- [5] De Nicola A, Missikoff M. A Lightweight Methodology for Rapid Ontology Engineering. Com. ACM. 2016; 59: 79–86.
- [6] **Zagorulko YuA, Borovikova OI.** Technology of development of ontology for scientific knowledge portals [In Russian]. Vestnik NSU. Series: Information Technologies. 2007; 5(2): 42-52.
- [7] *Gangemi A, Presutti V.* Ontology Design Patterns. In: Staab S., Studer R. (eds): Handbook on Ontologies. IHIS. Springer, Berlin Heidelberg, 2009: 221–243.
- [8] Zagorulko YuA, Zagorulko GB, Borovikova OI. Technology for building subject-based intelligent scientific internet resources based on ontology [In Russian]. Software Engineering. 2016; 7(2): 51-60.
- [9] Asim M-N, Wasim M, Khan MUG et al. A survey of ontology learning techniques and applications. Database: The Journal of Biological Databases and Curation. 2018, vol. 2018. doi:10.1093/database/bay101
- [10] Blomqvist E, Hammar K, Presutti V. Engineering Ontologies with Patterns: The eXtreme Design Methodology. In: Hitzler P., Gangemi A., Janowicz K., Krisnadhi A., Presutti V. (eds.): Ontology Engineering with Ontology Design Patterns. Studies on the Semantic Web, vol. 25. Amsterdam, IOS Press, 2016: 23–50.
- [11] *Lomov PA*. Application of ontology design patterns to development and use of ontologies in an integrated knowledge space [In Russian]. Ontology of designing. 2015; 2(16): 233-245.
- [12] Zagorulko Yu, Borovikova OI, Zagorulko GB. Application of ontology design patterns in the development of the ontologies of scientific subject domains [In Russian]. In: Selected Papers of the 19th International Conference on Data Analytics and Management in Data Intensive Domains, DAMDID/RCDL 2017. CEUR Workshop Proceedings (CEUR-WS.org). 2017; 2022: 258-265.
- [13] Zagorulko YuA, Zagorulko GB, Borovikova OI. Pattern-Based Methodology for Building the Ontologies of Scientific Subject Domains. In: H. Fujita and E. Herrera-Viedma (Eds.): New Trends in Intelligent Software Methodologies, Tools and Techniques. Proceedings of the 17th International Conference SoMeT_18. Series: Frontiers in Artificial Intelligence and Applications, vol. 303. Amsterdam: IOS Press, 2018: 529-542.
- [14] *Petasis G, Karkaletsis V, Paliouras G, Krithara A, Zavitsanos E.* Ontology Population and Enrichment: State of the Art. In: Paliouras, G., Spyropoulos, C.D., Tsatsaronis, G. (eds): Knowledge-Driven Multimedia Information Extraction and Ontology Evolution). Lecture Notes in Computer Science, vol. 6050. Springer, Berlin, Heidelberg, 2011: 134–166.
- [15] *Ganino G, Lembo D, Mecella M, Scafoglieri F.* Ontology population for open-source intelligence: a GATE-based solution. Software: Practice and Experience. 2018; 48(12).
- [16] *Maynard D, Funk A, Peters W*. Using Lexico-Syntactic Ontology Design Patterns for Ontology Creation and Population. In: Proc. Workshop on Ontology Patterns (WOP 2009), collocated with the 8th Int. Semantic Web Conf. (ISWC-2009). CEUR Workshop Proceedings (CEUR-WS.org). 2009; 516: 39–52.
- [17] *Ijntema W, Sangers J, Hogenboom F, Frasincar F.* A lexico-semantic pattern language for learning ontology instances from text. Journal of Web Semantics. 2012; 15: 37–50.
- [18] NeOn project home. http://www.neon-project.org
- [19] Association for Ontology Design & Patterns (ODPA). http://ontologydesignpatterns.org/wiki/ODPA
- [20] *Karima N, Hammar K, Hitzler P.* How to Document Ontology Design Patterns. In: Advances in Ontology Design and Patterns. Studies on the Semantic Web, vol. 32. IOS Press, Kobe, Japan, 2017: 15–27.
- [21] *Antoniou G, Harmelen F.* Web Ontology Language: OWL. In: Staab, S. and R. Studer. (eds.): Handbook on Ontologies. Berlin: Springer Verlag, 2009: 91-110.
- [22] *Sidorova EA*. Approach to modeling the process of information extracting based on the ontology [In Russian]. Ontology of designing. 2018; 1(27): 134-151. DOI: 10.18287/2223-9537-2018-8-1-134-151.
- [23] *Narinyani AS.* Uncertainty in knowledge representation and processing systems. Izv. Academy of Sciences of the USSR. *Tech. cybernetics*, 1988. No. 5.
- [24] Sidorova E, Akhmadeeva I. The Software Environment for Multi-Aspect Study of Lexical Characteristics of Text. In: Selected Papers of the XX International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2019). CEUR Workshop Proceedings, 2019; 2523: 306–315.
- [25] Garanina N, Sidorova E, Bodin E. A Multi-agent Text Analysis Based on Ontology of Subject Domain. In: Voronkov A., Virbitskaite I. (eds.): Perspectives of System Informatics (PSI 2014). LNCS, vol. 8974. Springer, Heidelberg, 2015: 102–110.
- [26] *Lamy J-B*. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies. Artificial Intelligence In Medicine. 2017; 80: 11–28.

- [27] **Zagorulko GB.** Development of an ontology for an Internet resource for decision support in weakly formalized areas [In Russian]. Ontology of designing. 2016; 4(22): 485-500.
- [28] de Cea GA, Gomez-Perez A, Montiel-Ponsoda E, Suarez-Figueroa MC. Using Linguistic Patterns to Enhance Ontology Development. In: Proc. Int. Conf. on Knowledge Engineering and Ontology Development (KEOD 2009) (Funchal Madeira, Portugal, October 6-8, 2009). INSTICC Press, 2009: 206–213.

About the authors

Yury Aleksevich Zagorulko (b.1957) graduated from the Novocherkassk Polytechnic Institute in 1979, PhD (1989). He is a Head of Laboratory at A.P. Ershov Institute of Siberian Branch of Informatics Systems of Russian Academy of Science, Associate Professor at Novosibirsk State University. He is a member of Russian and European Associations for Artificial Intelligence. He is the author of more than peer-reviewed 270 publications in the fields of AI, Knowledge and Ontology Engineering, Intelligent System Development and Computational Linguistics. ORCID: 0000-0002-7111-6524; Author ID (Scopus): 23394231500; Researcher ID (WoS): R-1826-2016. *zagor@iis.nsk.su*

Elena Anatolievna Sidorova (b. 1977) graduated from the Novosibirsk State University in 2000, PhD (2006). She is a Senior Researcher of the Laboratory of Artificial Intelligence at the A.P. Ershov Institute of Informatics Systems (Novosibirsk, Russia), Senior Lecturer at Novosibirsk State University. She is a member of Russian and European Associations for Artificial Intelligence. Dr. Sidorova has about 140 peer-reviewed publications in the field of NLP Systems, Multi-agent Systems, Knowledge Representation and Ontology Engineering. ORCID: 0000-0001-8731-3058; Author ID (Scopus): 41961707000; Researcher ID (WoS): K-2432-2018. Isidorova@iis.nsk.su

Zagorulko Galina Borisovna (b.1963) graduated from the Novocherkassk State University in 1985, PhD (2021). She is a researcher at A.P. Ershov Institute of Siberian Branch of Informatics Systems of Russian Academy of Science (Novosibirsk city), and a senior lecturer at Novosibirsk State University (Department of Mechanics and Mathematics). She is a member of the Russian Association of Artificial Intelligence. She is the author of more than 120 peer-reviewed publications in the field of AI, intelligent system development, knowledge engineering, ontological modeling and decision-making support. ORCID: 0000-0003-2155-1357; Author ID (Scopus): 35249924100; Researcher ID (WoS): K-2472-2018. gal@iis.nsk.su.

Irina Ravilevna Akhmadeeva (b.1991) graduated from the Novosibirsk State University in 2015. She is a Junior Researcher at A.P. Ershov Institute of Informatics Systems of Siberian Branch of Russian Academy of Science, Assistant Lecturer at Novosibirsk State University. She is the author of more than 20 peer-reviewed publications in the fields of AI, Intelligent System Development and NLP. ORCID: 0000-0002-7371-1087; Author ID (RSCI): 874172; Author ID (Scopus): 57188681471; Researcher ID (WoS): K-3145-2018. *i.r.akhmadeeva@iis.nsk.su*

Alexey Sergeevich Sery (b.1987) holds a master's degree in mathematics from Novosibirsk State University (2010) and a position of junior research at the A.P. Ershov Institute of Informatics Systems. He is the author of more than 20 peer-reviewed papers in the fields of NLP systems and Knowledge Representation. ORCID: 0000-0001-8275-4700; Author ID (Scopus): 56403204900; Researcher ID (WoS): K-1557-2018. alexey.seryj@iis.nsk.su

Received November 12, 2021. Accepted December 08, 2021.