

УДК 004.656

Анализ подходов систематизации и идентификации семантических отношений в онтологиях научно-технических предметных областей

Андропова В.В.

Студентка 4 курса факультета бизнес-информатики и управления комплексными системами Национального исследовательского ядерного университета «МИФИ»

115409, г. Москва, Каширское шоссе, 31, vero-nika_09@mail.ru

В статье рассматриваются онтологии как инструмент информационного моделирования, с помощью которого можно осуществить систематизацию и идентификацию отношений в текстах на естественном языке. Использование онтологий является перспективным подходом в области автоматизации информационно-поисковых систем. Онтологическое представление смыслового содержания документа или фрагмента предметной области, которое дает возможность их компьютерной обработки и учитывает динамическую взаимосвязь объектов предметного, базисного (концептуального) и знакового уровней, является потенциально успешным направлением в области формализации знаний. Приведен обзор существующих онтологических подходов, которые были проанализированы и для которых выявлены итоги на предмет использования в качестве инструмента информационного моделирования. В настоящее время, в лексической практике активно используются онтологические практики, и существует ряд разработанных онтологий, таких как язык Gellish, Фреймовый подход (проекты FrameNet, VerbNet), SUMO, DOLCE, YATO и т.д. В данной статье рассмотрены некоторые из перечисленных и сделаны выводы о корректности их использования. Выполненный анализ к различным подходам классификации отношений путем онтологического подхода подтвердил наличие существующих проблем и показал, что необходимо учитывать, что для деления отношений на классы нужно выделять значение классификационного признака, а также многозначность предикатной лексики.

Ключевые слова: информационное моделирование, онтологии, предметная область.

Analysis of approaches of systematization and identification of semantic relations in ontologies of scientific and technical subject areas

Andronova V.V.

4 th year student National Research Nuclear University.

115409, Moscow, Kashira Highway, 31, vero-nika_09@mail.ru

The article discusses the ontology as a tool for information modeling, which can be used to implement the systematization and identification of relations in natural language texts. The use of ontologies is a promising approach in the field of automation of information retrieval systems. Ontological representation of the semantic content of the document or a fragment of the subject area, which enables their computer processing and takes into account the dynamic relationship of objects of subject, basic (conceptual) and sign levels, is a potentially successful direction in the field of formalization of knowledge. The review of the existing ontological approaches, which were analyzed and for which the results were identified for the subject of use as an information modeling tool, is given. Currently, in lexical practice, ontological practices are actively used, and there are a number of developed ontologies, such as the language Gellish, Frame approach (projects FrameNet, VerbNet), SUMO, DOLCE, YATO, etc. in this article, we consider some of the above and draw conclusions about the correctness of their use. The analysis of various approaches to classification of relations of the autem ontological approach confirmed the existence of existing problems and showed that it is necessary to take into account that for the division of relations into classes it is necessary to distinguish the value of the classification characteristic, as well as the ambiguity of predicate vocabulary.

Key words: information modeling, ontology, subject area.

ВВЕДЕНИЕ

Неотделимой компонентой информационного моделирования является онтология, которая описывает в наиболее воспринимаемом виде информацию о предметной области. Онтология – это приблизительная спецификация некоторой предметной области. Формальное и декларативное представление, которое включает словарь (или имена) указателей на термины предметной области и логические выражения, которые описывают, что эти термины означают и как они соотносятся друг с другом. Таким образом, онтологии обеспечивают словарь для представления и обмена знаниями о некоторой предметной области и множество связей, установленных между терминами в этом словаре.

В настоящее время при автоматическом построении сложных онтологий при попытке учесть особенности естественного языка появляется сложность идентификации отношений. Чтобы идентифицировать отношения между сущностями, заданных в естественно-языковом тексте, необходимо определить основные группы отношений и типизировать предикатную лексику, с помощью которой они могут быть выражены, то есть провести классификацию отношений.

ОНТОЛОГИИ КАК ИНСТРУМЕНТ ИНФОРМАЦИОННОГО МОДЕЛИРОВАНИЯ

Использование онтологий является перспективным подходом в области автоматизации информационно-поисковых систем. Онтологическое представление смыслового содержания документа или фрагмента предметной области, которое дает возможность их компьютерной обработки и учитывает динамическую взаимосвязь объектов предметного, базисного (концептуального) и знакового уровней, является потенциально успешным направлением в области формализации знаний.



Рисунок 1. Представление предметной области

Онтология – это набор определений фрагмента декларативных знаний, ориентированный на совместное многократное использование различными пользователями. В онтологии вводятся термины, типы и соотношения (аксиомы), описывающие фрагмент знания. [4] В целом, основная цель создания онтологий – сделать возможным разделение и повторное использование знаний, и в частности, обеспечить интероперабельность между несовместимыми инструментальными средствами человека. [5] Практическое применение онтологии – это соглашение об использовании общего словаря в согласованной и логической манере. Таким образом, онтология выделяет из области знания отдельные объекты, сущности и отношения, что с точки зрения конкретной цели адекватно заменит саму предметную область.

В инженерной области подходы к онтологии сильно различаются. Так, к примеру, с точки зрения предмета концептуализации выделяют онтологии метауровня, прикладные онтологии, онтологии области знания и онтологии, ориентированные на задачу. Отдельно выделяют лингвистические онтологии (проекты WordNet, MikroKosmos, PyTез и др.), фиксирующие понятия (слова) вместе с их языковыми свойствами и отношениями [2].

Онтология – это моделирование предметной области на основе естественного языка. То есть, иными словами, онтология - это попытка подробной и полной формализации предметной области знаний с помощью построения концептуальной схемы, которая подразумевает семантическую сеть из взаимосвязанных понятий по некоторым принципам и концептам. В свою очередь, отношения в теории лингвистики в естественном языке

выражается связями и образуют в совокупности ту самую концептуальную схему, состоящий из объектов и связей между ними, то есть отношения можно формализовать, систематизировать и попытаться представить их в виде онтологии, поэтому рассмотрим в следующем пункте участников и способов выражения фрагмента предметной области в естественном языке.

Обзор и анализ классификаций отношений. Онтологические подходы к классификации отношений.

В настоящее время, в лексической практике активно используются онтологические практики, и существует ряд разработанных онтологий, таких как Фреймовый подход (проекты FrameNet, VerbNet), SUMO, DOLCE, YATO и т.д. Решения представляют разные таксономии отношений.

Рассмотрим «Фреймовый подход», относящийся к разделу компьютерной лексикографии, понимающийся в лингвистике как формализованное описание деятельности человека в определенной ситуации. Проект **FrameNet** – это английский электронный лексикографический ресурс, доступ к которому осуществляется онлайн, представляет собой базу данных, позволяющую определять семантические типы дефиниций, элементов «фрейма» и самих «фреймов» [4]. Глаголы восприятия разделяются на активные и пассивные (пример: видеть и смотреть). Под глаголами «слышать» и «видеть» понимается то, что случается с вами. А под глаголами «слушать» и «смотреть» подразумевается действие, которое совершаете именно вы. В базе данных FrameNet глаголы «слышать», «видеть» и использование других слов, таких как чувствовать ощущение вкуса и запаха, относятся к фрейму пассивного восприятия, а глаголы типа «смотреть» и «слушать» к фрейму активного восприятия.

В свою очередь, в английском языке есть много существительных, которые обозначают «дар» или «дарение» в каком-то большом событии, когда один человек дает (обычно) деньги другому человеку в определенной ситуации, но отдельные существительные вызывают очень разные эмпирические и институциональные надстройки: среди них чаевые, взятки, выкуп, гонорар, налог, каждый из которых может использоваться во вспомогательной конструкции для получения перевода денег, что вызывает различные схемы и результаты.

В российской компьютерной лексикографии нет специализированных компьютерных ресурсов аналогичных FrameNet. Поэтому особую актуальность приобретает вопрос, связанный с проектированием подобного лексикографического электронного ресурса, с учетом, к примеру, семантики англоязычной FrameNet с четким и однозначным обозначением классификационных признаков, а также русского языка.

Онтология **DOLCE** (Descriptive Ontology for Linguistic and Cognitive Engineering) – первая из онтологий в библиотеке базовых онтологий проекта WonderWeb [1].

Данная онтология создана для того, чтобы было применительно ее использовать для согласования между интеллектуальными агентами, использующими разную терминологию. При этом онтология не претендует на звание универсальной, стандартной или общей. Основная цель разработчиков - создать модель, помогающую при сравнении и объяснении связей с другими онтологиями библиотеки WFOL (базовой библиотеки онтологий WonderWeb), а также для выявления скрытых допущений, лежащих в основе существующих онтологий и лингвистических ресурсов, таких как WordNet. DOLCE имеет когнитивный уклон, поскольку фиксирует онтологические категории естественного языка и знания "здорового смысла".

В основу процесса проектирования легло фундаментальное философское разделение всех сущностей на универсалии (сущности, потенциально или реально имеющие экземпляры) и индивиды (или частности), которые не имеют и не могут иметь экземпляров. DOLCE- онтология индивидов, в том смысле, что область описания ограничена только ими. В качестве примера универсалии можно привести понятие Собака (оно имеет множество экземпляров, конкретных примеров в окружающем мире). В отличие от этого понятия, понятие Время скорее рассматривается как индивид (едва ли кому-то понадобится трактовать Время как множество различных сущностей, конечно, если речь не идет о параллельных мирах).

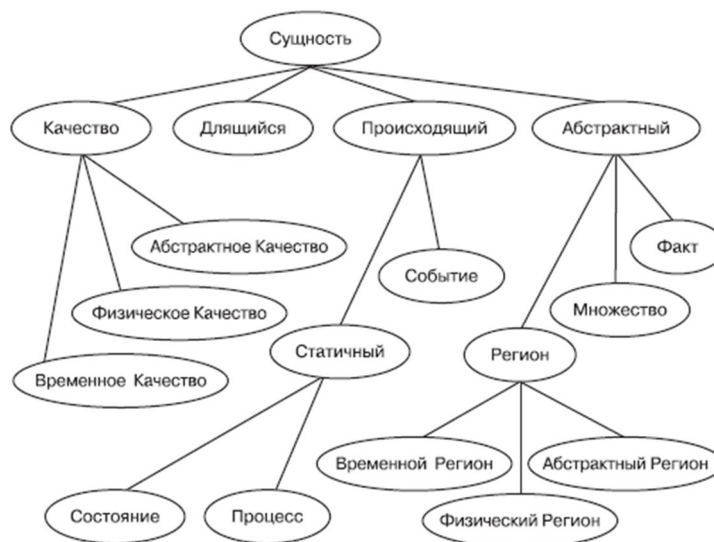


Рисунок 2 – Верхние уровни иерархии DOLCE

Немаловажной чертой DOLCE является разделение на «Постоянные» (есть в наличии целиком и неизменно в некотором фиксированном промежутке времени) и

«Происходящие» (разворачиваются во времени и в каждый момент они могут быть различными, по-разному себя проявляя и имея разный состав) сущности.

Такое разделение на «объект» и «процесс» условно. Это привело к тому, что в онтологии определены два типа отношения ЧАСТЬ-ЦЕЛОЕ. Первое никак не зависит от времени, второе имеет временной индекс, определяющий, в каких временных рамках отношение действует. Подобное "раздвоение" наблюдается и для отношения " КАЧЕСТВО - ОБЛАДАТЕЛЬ КАЧЕСТВА ". Другие базовые отношения онтологии: УЧАСТНИК-ПРОЦЕСС, КОМПОНЕНТ-ЦЕЛОЕ (компонент входит в состав целого) и отношение зависимости имеют временной индекс [3].

У рассмотренной онтологии несомненно есть свои плюсы, например, как разделение объектов на классы во времени, но деление классов не шаблонизировано, поэтому онтология может применяться только на верхнем поверхностном уровне. Для представления своей онтологии авторы DOLCE избрали гибкий подход: онтология фиксируется с использованием логики предикатов первого порядка. Затем описывается та часть утверждений, которая может быть представлена на языке OWL. Недосток такого подхода в том, что приложения, не имеющие информации о действительной структуре OWL-документа, не смогут получить доступ к "закомментированным" утверждениям.

Онтология **SUMO** (Standard Upper Merged Ontology) – онтология верхнего уровня, содержит наиболее общие и самые абстрактные концепты. Данная онтология предназначена для содействия улучшению интероперабельности данных, извлечения и поиска информации, автоматического вывода и обработки естественного языка [3]. Онтология охватывает такие области знания, как общие виды процессов и объектов, абстракции (теория множеств, атрибуты, отношения), числа и единицы измерения, временные понятия, части и целое, агенты и намерения; является канонической онтологией верхнего уровня: содержит обозримое число концептов и аксиом, имеет ясную иерархию классов, легко расширяется.



Рисунок 3 – Иерархия классов SUMO

Основными концептами являются Сущность и ее категории - Физический и Абстрактный. Первая категория включает всё, что имеет положение в пространстве-времени, а вторая - все остальное (а точнее только то, что существует в сознании). Физический делится на Объект и Процесс, что соответствует подходу, реализованному в DOLCE.

Таким образом, образом можно заметить, что иерархия классов в SUMO менее запутана, и, возможно, более удобна для практического применения, чем DOLCE.

Заключительным подходом в теории онтологий рассмотрим онтологию Джона Сова – **J. Sowa's ontology**. Новизна данного подхода в том, что онтология основана не на фиксированной иерархии концептов, а на каркасе, описывающем различия, по которым иерархия генерируется автоматически. В любом конкретном положении «концепты не определяются рисованием линий на диаграмме, а задаются выбором подходящего множества различий». Сова выявил следующие категории, представленные в каркасе ниже:

Таблица 1 – Онтологические категории верхнего уровня Джона Сова

	Физический		Абстрактный	
	Континуальный	Происходящий	Континуальный	Происходящий
Независимый	Объект	Процесс	Схема	Скрипт
Относительный	Слияние	Участие	Описание	Условие
Опосредованный	Структура	Ситуация	Причина	Цель

Кроме приведенных в таблице выше категорий онтологии есть еще два важных понятия. «Сущность» не определяется никаких отличительных признаков или различий и является надтипом всех концептов. Второе понятие – «Абсурдный тип», который наследует все возможные, в том числе противоречащие, различия. Его не может иметь ни один экземпляр.

Таким образом, можно сделать вывод, что онтология Джона Сова резко отличается подходом к построению классификации – от простой фиксированной иерархии он переходит к гибкому каркасу.

Итоги обзора существующих онтологических подходов к классификации отношений

Рассмотренные онтологические подходы говорят о том, что существуют неоднократные попытки унификации отношений.

В классификациях для автоматизированных систем отсутствует системность, они не перекрывают отношения предметной области, или иначе подходят только для верхнего

уровня онтологий. Касательно ресурса FrameNet, то он адаптирован полностью под англоязычную лексику, и то даже в нем не устранена задача однозначности понятий и событий. В русском языке лексикографический ряд намного шире, а данный подход не поддержит понимание разграничения синонимичных понятий, а, следовательно, исключается возможность определения точных признаков и построения классификация отношений для того, чтобы полностью покрыть предметную область.

Онтологии DOLCE и SUMO исключали попытку принятия универсальности и общности, отсутствует четкое деление на классы и обозначения классификационных признаков, что не дает нам уверенности в однозначности построения классификации и в попытке унифицировать отношения на различных уровнях детализации, но тем не менее, данные онтологии универсальны в использовании для семантики отношений верхнего уровня.

Онтология Джона Совы отличается подходом на разделения по классам: онтология основана не на фиксированной иерархии концептов, а на каркасе, описывающем различия, по которым иерархия генерируется автоматически. Такой подход является перспективным для развития построения шаблонов в идентификации отношений.

Список используемой литературы

1. Добров Б., Иванов В., Лукашевич Н., Соловьев В. Онтологии и тезаурусы: модели, инструменты, приложения – НОУ «Интуит» Лекция 4: Онтологии верхнего уровня: отличительные черты, 2017
2. Н.В. Максимов, О.Л. Голицына, Г.В. Тихомиров, П.Б. Храмцов: Информационные ресурсы и поисковые системы: учебное пособие / – М.: МИФИ, 2008. 400 с
3. A S Gavrilkina, O L Golitsyna and N V Maksimov Building a well-formalized conceptual semantic network based on scientific and technical texts – МЭФТИ, 2018
4. Charles J. Fillmore, Collin F. Baker, and Hiroaki Sato FrameNet as a "Net", International Computer Science Institute Berkeley, California, U.S.A, Senshu University Kawasaki, Japan, 2012
5. Uschold M., King M., Moralee S., Zorgios Y. The Enterprise Ontology // The Knowledge Engineering Review. – 1998. – Vol. 13, № 1. – P. 31 – 8