

Онтологический подход в информационной системе «Ковчег знаний»: синтез знаний и формализация энциклопедической деятельности

А. Л. Семенов¹, Р. Р. Тимиргалеева^{1,2}, И. Ю. Гришин¹, Н. И. Галлини²

¹Московский государственный университет имени М.В. Ломоносова

²Крымский федеральный университет имени В.И. Вернадского

alsemno@ya.ru, renatimir@gmail.com, igrishin@sev.msu.ru, gallini.nadi@yandex.ru

Аннотация. Информационная система «Ковчег знаний» представляет собой онтологическую информационную систему, разрабатываемую Московским государственным университетом имени М.В. Ломоносова. В данной статье рассматривается использование онтологического подхода в рамках проекта «Ковчег знаний» для синтеза знаний из различных источников и формализации энциклопедической деятельности. Исследование описывает этапы создания онтологий, особенности и преимущества использования математических инструментов, таких как реляционная логика и теория графов, для построения онтологий. Также рассматривается значимость точности информации и гармонизации знаний при работе с обширными текстами. Разработка онтологий в рамках проекта «Ковчег знаний» объединяет усилия математиков-прикладников, кибернетиков, программистов и лингвистов-прикладников. Акцент делается на создание информационной системы «Ковчег знаний» на основе научно обоснованных онтологий и ее интеграцию с российским сообществом через портал «Большая российская энциклопедия».

Ключевые слова: онтологический подход; информационная система; «Ковчег знаний»; синтез знаний; формализация; энциклопедическая деятельность; МГУ; реляционная логика

I. ВВЕДЕНИЕ

В настоящее время ведется разработка Большой Российской Энциклопедии (БРЭ) в цифровом формате, что открывает новые перспективы для улучшения качества и доступности энциклопедических материалов. Московский государственный университет имени М.В. Ломоносова (МГУ) в сотрудничестве с широким научным и образовательным сообществом ведет работу над проектом «Ковчег знаний», который является ключевым элементом в развитии БРЭ. Целью проекта является создание семантической сети понятий и связей между ними, основанной на онтологии.

Онтологии стали неотъемлемой частью разработки знаний и обработки естественного языка, играя важную роль в синтезе энциклопедических словарей, тезаурусов и классификаторов. Для успешного построения онтологии необходимо разработать методологию и выбрать подходящие компьютерные инструменты.

Работа выполнена при финансовой поддержке Московского университета (грант № 23-Щ05-11) и государственного задания (регистрационный № 124020100068-4)

Применение технологий обработки естественного языка, включая обработку больших объемов текста, является перспективным направлением. Однако важно учитывать вопросы достоверности информации и доверия к искусственному интеллекту при интеграции различных знаний. Неоднозначность и противоречия в представленных знаниях могут возникать из-за терминологических несоответствий и концептуальных различий между экспертными видениями на изучаемую систему [1, 2].

В современном мире онтологии широко признаны как эффективный инструмент для формализации и систематизации знаний и данных в различных областях науки. Однако создание онтологии представляет собой сложный и трудоемкий процесс. Практика показывает, что при разработке онтологий в области программного обеспечения особенно полезно использование шаблонов онтологического проектирования. Это объясняется тем, что многие онтологии программного обеспечения содержат типичные элементы, которые могут быть успешно описаны с помощью таких шаблонов. Применение шаблонов значительно упрощает процесс разработки онтологий программного обеспечения и позволяет специалистам в данной области вносить свой вклад, даже без опыта моделирования онтологий [3–6].

Для создания онтологии программного обеспечения требуется анализировать большое количество публикаций, касающихся предметной области. Однако использование лексико-синтаксических шаблонов в онтологическом проектировании может ускорить и упростить этот процесс, путем обогащения онтологии информацией из различных источников. Таким образом, разработка онтологии является сложным, но важным этапом в систематизации знаний и обеспечении их семантической точности в научных областях.

Жизненный цикл информационной системы «Ковчег знаний» начался в рамках создания Большой российской энциклопедии в цифровом формате. Этапы включали определение направлений, разработку прототипа с поддержкой МГУ, построение семантической сети на основе онтологии и интеграцию цифровой платформы.

«Ковчег знаний» представляет собой инновационную платформу, построенную на принципах вики-системы, вдохновленную успешными аналогами, такими как Википедия и РуВики. Отличительной чертой «Ковчега

знаний» является фокус на создании и обновлении онтологий научных знаний на русском языке. В отличие от традиционных вики-платформ, где информация создается и редактируется пользователями без строгой формализации, «Ковчег знаний» предлагает новый уровень систематизации и точности благодаря использованию обоснованных онтологий.

II. ОНТОЛОГИЯ И СТРУКТУРИРОВАНИЕ ЗНАНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

В интеллектуальных системах знания не хранятся хаотически, а структурируются для облегчения поиска и поддержки баз данных. Для этого используются различные методы классификации, включая типы, такие как «часть-целое» или ситуационные классификации, где знания, связанные с конкретными сценариями, объединяются. В области исследований искусственного интеллекта и теории классификации исследования пересекаются, и процесс классификации часто включает абстрагирование элементов описаний или знаний. Этот процесс абстракции может включать несколько этапов, приводящих к абстрактным знаниям, не имеющим прямого отражения в реальном мире. Манипулирование абстрактными знаниями расширяет возможности систем, делая их применимыми к различным областям [7, 8].

Среди специалистов, занимающихся компьютерной лингвистикой, наиболее устоявшимся (классическим) определением онтологии, предложенным Грубером, является то, что она представляет собой «спецификацию концептуализации» [9]. Грубер предложил несколько определений онтологии:

- Онтология как явная концептуализация, описывающая объекты и их взаимосвязи в области.
- Онтология как формальное представление знаний с терминами, таксономией, определениями и правилами вывода.
- Онтология как спецификация общей концептуализации в контексте предметной области.
- Онтология как база знаний, описывающая всеобщие факты на основе широко принятых терминов.

Онтологии служат для формализации понятий, отношений и правил вывода. Однако помимо точного определения, важно учитывать сложности описания модели онтологии с использованием формального языка.

Рассмотрим классическое формальное описание онтологии: Онтология определяется как $O = \langle X, R, F \rangle$, где:

- X представляет собой конечный набор понятий в предметной области.
- R обозначает конечный набор отношений между понятиями.
- F включает в себя конечный набор функций интерпретации.

Онтологии – ключевой инструмент в развитии интеллектуальных информационных систем,

обеспечивая качественный скачок в их развитии. Создание и исследование онтологий осуществляется двумя подходами: формальным (основанным на логике) и лингвистическим (основанным на изучении естественного языка). Онтологии классифицируются на четыре уровня групп, определяемых объектом концептуализации.

Существуют два альтернативных подхода к созданию и исследованию онтологий:

- Формальный подход: этот подход основан на логике (такой как предикаты первого порядка, описательная логика, модальная логика и т. д.).
- Лингвистический подход: этот подход основан на изучении естественного языка (в частности, семантики) и включает создание онтологий из больших текстовых корпусов, известных как корпуса.

Следует отметить, что онтологии часто классифицируются на основе объекта концептуализации, что приводит к четырем уровням групп онтологий [10]. Характеристики и уровни онтологий приведены в табл. 1.

ТАБЛИЦА 1.

Уровень	Онтологическая группа	Описание
1	Онтологии представления	Формализмы представления знаний (Концептуализация)
2	Онтологии верхнего уровня	Используются в широком спектре предметных областей
3	Онтологии, специфичные для области	Используются в пределах одной и той же предметной области
4	Прикладные онтологии	Не могут быть повторно использованы

При создании онтологий на различных уровнях возникают особые трудности, особенно в случае формирования онтологий верхнего уровня, которые направлены на установление единой модели для охвата знаний, общих для нескольких областей. Разработка онтологий для конкретных групп фундаментальных наук (уровень 3) находится в стадии развития, начиная с математики как основной методологии. Методология построения онтологий в отдельных фундаментальных науках включает ряд этапов: определение целей, концептуализацию, формализацию, аксиоматизацию, реализацию, интеграцию с уже существующими онтологиями, оценку и документирование. В МГУ был разработан прототип системы Knowledge Ark System на основе указанных методов и подходов.

1. Создание онтологий на первом и втором уровнях представляет особые трудности, особенно в контексте создания онтологий верхнего уровня, цель которых заключается в установлении единой онтологии для захвата знаний, общих для нескольких предметных областей.
2. Разработка онтологий для конкретных групп фундаментальных наук (уровень 3) находится в процессе развития, начиная с математики как основной методологии.

