**БРЭ и Викиданные с точки зрения онтологической структуры**

Панич Мария Борисовна, Садковский Фёдор Алексеевич

Студенты МГУ имени М.В. Ломоносова, Москва, Россия

Задача разработки искусственного интеллекта в настоящее время требует создания систем с четкой структуризацией знаний, которые были бы легкодоступны как пользователям, так и алгоритмам машинного обучения. На достижение этой цели направлено создание системы «Ковчег знаний», проектируемой специалистами МГУ. Она представляет собой хранилище знаний разных научных областей. Вместе с тем качество и достоверность собираемых знаний позволяют использовать систему как «инкубатор» текстов статей Большой Российской Энциклопедии (БРЭ). Структуризация знания в проекте достигается за счет разрабатываемой онтологической системы [Семенов и др. 2023: 233–234].

Создаваемая онтология должна учитывать основные понятия Стратегии научно-технического развития Российской Федерации, Программы развития МГУ имени М.В. Ломоносова, понятия, используемые для структуризации научной деятельности в системе наукометрических данных МГУ "Истина", в существующих рубрикаторах и онтологиях научной деятельности (ГРНТИ и др.), системах категорий научных энциклопедий и т.п.

Кроме того, онтология должна быть лингвистически ориентирована, поскольку энциклопедические знания, которые она призвана систематизировать, по большей части представляют собой текст на естественном языке. Принципы и методы построения такой онтологии можно найти в работах С. Ниренбурга и В. Раскина [Nirenburg, Raskin 2004: 192–228]. Онтология должна отвечать следующим требованиям:

1. однозначность понятий;
2. строгие определения и отсутствие омонимии;
3. задание области определения и области значений для отношений;
4. правильность логического вывода на основе системных онтологических отношений.

Поскольку база понятий создается не с нуля, а с опорой на содержание БРЭ, нужно прежде всего изучить, насколько устройство энциклопедии на данный момент далеко от приведенных выше онтологических принципов. Кроме того, с формальной и идеологической точки зрения система «Ковчег знаний» во многом опирается на принципы Википедии, одним из проектов которой является онтология Викиданные: при условии соблюдения важных для нас принципов она могла бы лечь в основу онтологии для БРЭ [Семенов и др. 2024].

В новой версии **БРЭ** (https://bigenc.ru) заметно стремление к онтологической организации содержания. Однако, на наш взгляд, используемый формат далек от идеальной понятийной структуры. БРЭ включает три подсистемы организации понятий: «категории», «термины» и «тэги». При этом категории малочисленны и лишены иерархической структуры, а термины чаще всего дублируют область знания, к которой независимо приписана каждая статья. Перемещение между категориям и терминами невозможно. Система тэгов устроена эффективнее с точки зрения навигации, но не лишена иных недостатков: тэги не универсальны, как и принципы их выделения и приписывания к каждой статье – наблюдаются непоследовательности даже в рамках одной категории. Например, статье «*сознание*» приписан следующий ряд тэгов:

#Психические процессы (1)

#Психические состояния (2)

#Психологические концепции (3)

#Психологические термины (4)

Последний тэг как маркер другого семантического отношения можно найти в статье «*когнитивная психология*», где список тэгов такой:

#Психологические дисциплины (5) – надкласс

#Психологические концепции (6) – часть

#Психологические термины (7) – часть, но ниже по уровню, чем (6)

Из этого примера видно, что один и тот же тэг может в разных случаях неэксплицитно отражать разные семантические отношения. Кроме того, ни один из тэгов (5-6) выше не соответствует какому-либо понятию из энциклопедии (т.е. статье). Тэги дополнительно разбивают категории, что, с одной стороны, является их достоинством, поскольку это создает иерархию классов внутри категории, но с другой, они могут и дублировать их или области знания, что избыточно. Таким образом, опираясь на структуру понятий в БРЭ, реализовать приведеннные выше требования к онтологическим системам будет крайне непросто.

**Викиданные** ближе к «идеальной» онтологии: элементы там связываются классическими онтологическими отношениями («класс-подкласс», «экземпляр» и др.) и набором новых, работающих по схожему принципу (они называются свойствами (properties)); для описания используются понятия транзитивности, симметричности и т.п. И все же можно выделить некоторые критические недостатки. В их числе:

1. **Проблемы с омонимией**. Например, в последовательности вложенных друг в друга подклассов *знание* – *память* – *информация* игнорируется очевидная многозначность всех трех терминов, хотя для многих комбинаций возможных значений такой ряд будет не будет соответствовать действительности.
2. **Недоопределенность обратных функций**. В правилах работы со свойствами в Викиданных (https://www.wikidata.org/wiki/Help:Properties) сказано, что при связывании сущностей одним из таких отношений (например, *родитель*) второе (например, *ребенок*) нужно добавить вручную, но на практике этого часто не происходит: *знание* – «подкласс от» *памяти,* но у *памяти* свойства «иметь подкласс» просто нет.
3. **Смешение слов и обозначаемых ими понятий**. Например, к *знанию* в Викиданных привязан аудиофайл произнесения слова «знание» на маратхи, а к *информации* – соответствующий жест французского жестового языка.

Тем не менее, если устройство БРЭ не соответствует целям создания онтологии «Ковчега знаний», то Викиданные при учете выявленных недостатков вполне могут послужить стартовой платформой для дальнейшей работы.

Исследование выполнено при финансовой поддержке Междисциплинарной научно-образовательной школы Московского университета “Математические методы анализа сложных систем” (грант № 23-Ш05-11)

**Литература**

Семенов А.Л., Гришин И.Ю., Тимиргалеева Р.Р., Бубнов А.С.. Портал «Ковчег знаний» на основе современных технологий онтологического моделирования // Дистанционные образовательные технологии : сборник трудов VIII Международной научно-практической конференции, Ялта, 19–21 сентября 2023 года. 2023. С.233-235.

Семенов А.Л., Гришин И.Ю., Кобозева И.М., Панич М.Б., Раевский Е.Н., Садковский Ф.А., Федорова О.В. Онтология как фундамент формализации энциклопедической деятельности и извлечения новых знаний // Журнал «Информационное общество». 2024. (в печати)

Nirenburg S., Raskin V. Ontological semantics. Cambridge MA, 2004.

Научно-образовательный портал «Большая российская энциклопедия»: <https://bigenc.ru/>

Правила работы со свойствами в Викиданных: <https://www.wikidata.org/wiki/Help:Properties>