

Семантика и архитектура данных для агентных систем

А. Р. Мударова

Санкт-Петербургский горный университет императрицы Екатерины II

amimudarova@gmail.com

Аннотация. Статья посвящена теоретическим и практическим аспектам организации семантического взаимодействия и архитектуры данных в агентных системах. На основе анализа современных подходов к формализации знаний и моделей представления данных предложена классификация типов агентов и соответствующих им моделей знаний. Ключевой проблемой исследования является семантическая интероперабельность гетерогенных агентов. В связи с этим анализировались архитектурные решения для хранения и обработки данных, включая озера данных, базы знаний на основе онтологий и микросервисную архитектуру. Приведена сравнительная характеристика различных подходов к организации данных и показана их взаимосвязь с семантическими возможностями агентов. Предложены критерии выбора архитектурных решений в зависимости от типа агента и требований к семантической согласованности. Результаты применимы при проектировании интеллектуальных систем различного назначения.

Ключевые слова: агентные системы, семантика данных, архитектура данных, онтология, семантическая интероперабельность, база знаний, озеро данных

I. ВВЕДЕНИЕ

Развитие информационных систем на сегодняшний день отличается возрастающей сложностью взаимодействия между автономными компонентами. Это требует новых подходов к организации семантического уровня коммуникации и архитектуры данных. Агентные системы, которые представляют собой совокупность взаимодействующих интеллектуальных агентов, сталкиваются с фундаментальной проблемой согласованной интерпретации информации, которой они обмениваются. Цель статьи — анализ существующих подходов к организации семантического взаимодействия и архитектуры данных в агентных системах, а также разработка классификационной модели, связывающей типы агентов с соответствующими моделями знаний и способами хранения данных.

II. ОСНОВНАЯ ЧАСТЬ

Когда агенты различаются по архитектуре и моделям знаний, обеспечить их взаимопонимание становится сложнее. На способность к корректной интерпретации сообщений влияют два разнонаправленных фактора: включенность в глобальную сеть эту способность усиливает, а высокая специализация, наоборот, ослабляет. Это наблюдение имеет принципиальное значение для понимания архитектурных требований: глобализация требует формирования универсальных семантических моделей, в то время как специализация порождает замкнутые терминологические системы. Это

напрямую сказывается на том, как выстраивается архитектура данных.

К.И. Костенко и В.Ю. Белкин предложили фундаментальную модель интеллектуальной системы, которая базируется на формализмах представления знаний. Авторы определяют формализм как «четверку $F = (M, DM, \circ, \pi)$ », где M — множество абстрактных знаний, DM — фрагменты знаний, \circ — операция композиции, а π — отношение проекции. Это позволяет строить строгие математические модели памяти агентов. Интересен подход к использованию семантических иерархий: в этой модели «знания представляются конечными бинарными деревьями», а внутренние вершины размечены отношениями между поддеревьями [3]. Такое представление согласуется с современными подходами к организации баз знаний.

Архитектура системы требует четкого разделения декларативных и процедурных знаний. Декларативные знания — это «знания в форме описания фактов, событий или явлений». Процедурные знания определяются как «последовательность действий, которая используется при решении задач» [4]. Это различие влияет на выбор инструментов хранения. Декларативные знания хранятся в онтологиях и базах знаний, тогда как процедурные знания часто реализуются в виде программного кода или правил вывода.

Переход от традиционных хранилищ данных к озерам данных расширяет возможности для агентных систем. Р. Абу Хасан с соавторами отмечают, что озеро данных объединяет структурированные данные из реляционных баз данных (строки и столбцы), полуструктурированные данные (CSV, журналы, XML, JSON), неструктурированные данные (электронные письма, документы, PDF-файлы) и двоичные данные (изображения, аудио, видео). Для озер данных характерен подход *schema-on-read*: данные загружаются как есть, в своем «родном формате», а схема накладывается только на этапе чтения. Это дает гибкость, необходимую для динамичных агентных систем. При этом современные базы знаний функционируют не изолированно, а в тесной интеграции с системами поиска и извлечения информации. Структурой, позволяющей классифицировать понятия и унифицированно представлять знания, служит онтология. Это иерархическая модель, которая фиксирует концептуальное представление предметной области и задаёт логику взаимосвязей между её элементами. В гибридной архитектуре атрибутивная информация о документах хранится в онтологии, а полные тексты документов в отдельной поисковой системе. Такое разделение позволяет объединять

семантическую обработку запросов с классическим полнотекстовым поиском.

Для систематизации подходов к семантике и архитектуре данных в агентных системах была составлена табл. 1. В ней показаны модели знаний и формы их представления для разных типов агентов.

ТАБЛИЦА I. МОДЕЛИ ЗНАНИЙ И ФОРМЫ ПРЕДСТАВЛЕНИЯ

Тип агента	Модель знаний	Форма представления знаний
Человек-агент (ЧА)	Псевдофизическая, фреймовая	Естественный язык, документы, опыт
Когнитивный агент (КА)	Логическая, семантическая сеть	RDF-тройки OWL-онтологий
Реактивный агент	Производственная	Правила «ЕСЛИ-ТО»
Микросервис как агент	Комбинированная (процедурная + декларативная)	JSON, XML, API-спецификации
Гибридный агент (ЧА+КА)	Интегрированная	Различные форматы с единой семантической разметкой

[2; 3]

Для того чтобы проанализировать архитектурные решения, была составлена табл. 2. В ней указаны архитектурные решения по хранению данных. Также оценена степень семантической совместимости каждого подхода.

ТАБЛИЦА II. АРХИТЕКТУРА ХРАНЕНИЯ И СЕМАНТИЧЕСКАЯ СОВМЕСТИМОСТЬ

Тип агента	Архитектура хранения	Семантическая интероперабельность
Человек-агент (ЧА)	Файловые хранилища, СЭД, корпоративные порталы	Зависит от лингвистической совместимости, культурного контекста
Когнитивный агент (КА)	RDF-хранилища, графовые базы данных	Высокая при использовании общих онтологий
Реактивный агент	Реляционные СУБД, оперативная память	Низкая, ограничена фиксированным набором понятий
Микросервис как агент	Озера данных, NoSQL, распределенные системы	Средняя, обеспечивается через API-контракты
Гибридный агент (ЧА+КА)	Многоуровневые архитектуры	Высокая при наличии семантического слоя

[2; 3]

Классификация строится на основе двух ключевых критериев: уровня автономности агента и сложности обрабатываемой семантики. Для человека-агента достаточно минимальной формализации, но нужна максимальная гибкость интерпретации. Когнитивные и гибридные агенты, наоборот, нуждаются в строгих семантических моделях и соответствующих архитектурах хранения. Реактивные агенты эффективны в замкнутых предметных областях, а микросервисы в распределенных сценариях с четко определенными интерфейсами.

Приведенные таблицы демонстрируют, что выбор архитектуры данных напрямую зависит от двух факторов: требуемого уровня семантической интероперабельности и динамичности предметной области. Для систем с высокой степенью неопределенности и частыми изменениями требований

предпочтительны озера данных и NoSQL-решения, позволяющие быстро адаптировать схему. Для систем, где важна строгая согласованность и сложные логические выводы, оптимальны графовые базы и онтологии. При проектировании гибридных систем разумно сочетать разные подходы. Семантическое ядро, построенное на основе онтологий, дополняют периферийными подсистемами с более гибкой архитектурой. Важным аспектом архитектуры данных является управление потоками знаний. Они представляют собой именованные последовательности операций для переноса знаний между соседними компонентами, с учетом их внутренней трансформации. Перенос знаний — это перевод знания в формализме компоненты источника в формат знания компоненты, в которую переносится знание. Без такой трансформации семантическая совместимость невозможна. Поэтому архитектура должна содержать средства для конвертации форматов.

В агентных системах, работающих с большими данными, лямбда-архитектура применяется как перспективное решение. Она решает проблему выполнения произвольных функций над распределенными данными с поддержкой обработки в реальном времени. Эта архитектура включает три слоя: пакетной обработки, обработки в реальном времени и обслуживания. Пакетная обработка отвечает за анализ больших объемов накопленных данных, а потоковая обработка обеспечивает работу с данными в режиме реального времени. Слой обслуживания объединяет результаты, это позволяет агентам реагировать на изменяющуюся обстановку. Однако хранение данных невозможно без их описания. Метаданные формируют основу для работы с информационными массивами. Через них реализуется поиск информации, регулируются права доступа и выполняется подготовка данных для последующего анализа. Отсутствие продуманной системы метаданных создаёт серьёзную угрозу, так как озеро данных может превратиться в неструктурированное информационное болото. Чтобы избежать подобных рисков, каждому объекту сопоставляется набор метаданных, фиксирующих его ключевые характеристики: сведения о происхождении, типе и формате, пространственной привязке, а также о правах доступа [1].

Большие языковые модели применяют в агентных системах как инструмент семантической обработки. Через такой интерфейс агенты могут обрабатывать запросы на естественном языке, формулировать объяснения своих решений и осваивать новые предметные области без перестройки онтологий. Языковые модели не замещают формальные представления знаний, они снижают требования к подготовке пользователей. В результате упрощается взаимодействие разнородных агентов через единый языковой интерфейс. Пользователям становится проще работать с системой. Одновременно происходит смена парадигмы: статичные онтологии уступают место динамическим семантическим интерфейсам. Обновление графов происходит в реальном времени — на основе потоковых данных и сигналов от агентов. Кроме фактов и отношений, они фиксируют эволюцию знаний. Например: временные изменения, причинно-следственные связи, уровень достоверности данных. Подобное устройство позволяет системам работать в

условиях неопределенности, адаптируя семантическую модель по мере накопления информации и изменения целей. На этом фоне растет востребованность федеративной семантики. Универсальные онтологии не покрывают все сценарии взаимодействия: каждый агент или группа агентов оперирует собственной онтологией. Связность обеспечивается за счет механизмов сопоставления и трансформации данных на лету. Такой подход снижает требования к предварительному согласованию и позволяет системам масштабироваться более органично.

Важно не просто хранить данные, а управлять связями между ними. Графовые базы данных решают эту задачу. Они фиксируют сущности и отношения между ними: временные зависимости, транзитивные связи, вероятностные отношения. Для агентных систем такая структура необходима, так как решения редко принимаются на основе изолированных фактов. Обычно требуется проследить цепочку взаимосвязанных событий.

Опыт построения баз знаний на основе онтологий представлен в работе А.Н. Подобриня и соавторов, где выделяются ключевые сущности для описания бизнес-процессов: Class (класс) — «коллекция из элементов основных единиц хранения информации», Instance (экземпляр) — «коллекция из элементов хранения составных частей сущности Class», Attribute (атрибут) — «коллекция элементов, описывающих свойства объекта», а также сущности Group, User и Right для управления доступом. Формально модель данных хранилища онтологии представляется выражением $O = \{C, I, R\}$, где C — множество классов, I — множество объектов, R — множество отношений между классами. Также на практике часто применяются гибридные подходы к хранению. Они сочетают графовые базы для семантического слоя с векторными базами данных для работы с неструктурированной информацией. Векторные представления, получаемые через embedding-модели, позволяют агентам находить семантически близкие фрагменты текста, изображений или других данных без жесткой привязки к заранее заданной онтологии. Это достаточно востребовано в сценариях, где предметная область меняется быстрее, чем успевает развиваться формальная модель знаний. Необходимо различать понятия совместимость и интероперабельность. Совместимость относится к отдельным параметрам взаимодействия — целям, предметной области, контексту или моделям знаний. Семантическая интероперабельность достигается лишь тогда, когда совместимы все эти параметры одновременно. С архитектурной точки зрения это означает, что система должна обеспечивать не только техническую унификацию форматов, но и согласованность на уровне смыслов, целей и контекста обмена данными [4].

Условия, в которых происходит взаимодействие агентов, закладывают в основу архитектуры. К ним относят время, местоположение, текущую ситуацию, предметную область и состав участников. Наличие временных меток и пространственных привязок позволяет интерпретировать сообщения в динамике развивающейся ситуации и согласовывать действия агентов. Именно поэтому архитектура данных должна в первую очередь предусматривать хранение и обработку всех типов контекстов. В этой связи интерес

представляет концепция активных метаданных, когда метainформация сопровождает данные и участвует в процессе принятия решений. Например, метаданные могут содержать сведения о том, какие агенты уже обрабатывали данный объект, с какой степенью достоверности были получены выводы. Данные устаревают и требуют обновления. Учет актуальности данных позволяет выстраивать более гибкие политики управления данными и повышает общую надежность системы.

Мультиагентную парадигму и микросервисную архитектуру нередко применяют совместно. В этом случае организационно-техническую структура рассматривают как мультиагентную систему, где лица, принимающие решения, выступают в роли интеллектуальных агентов. Практическая реализация такого подхода представлена в прототипе веб-сервиса имитационного моделирования. Прототип поддерживает построение гибридных моделей, объединяющих агентное, имитационное и дискретно-событийное представления процессов. В нем также реализовано проведение имитационных экспериментов для оценки поведения систем в различных сценариях [2].

При использовании онтологий поиск в базах знаний учитывает смысловые связи между терминами, а не только прямые совпадения. Эффективность подхода подтверждается количественными оценками: полнота достигает 0,67, точность — 0,5, а интегральная F-мера составляет 0,57, что демонстрирует ощутимый прирост релевантности по сравнению с традиционным полнотекстовым поиском [5].

Современные подходы семантического поиска объединяют классические онтологические методы с возможностями больших языковых моделей. Альтернативой детерминированному расширению запроса через онтологию служат гибридные стратегии: языковая модель интерпретирует запрос, выделяя ключевые сущности и намерения, после чего уточненные данные обрабатываются с использованием онтологии и векторного индекса. Итоговые данные система ранжирует и агрегирует, опираясь на формальные семантические связи и статистическую близость в векторном пространстве. Реализация предложенной методологии предполагает три этапа проектирования. Первым шагом определяется тип агента или группы агентов, для которых создается архитектура. Затем оцениваются требования к семантической совместимости: необходимость сложных логических выводов, допустимая степень неопределенности, динамичность предметной области. Завершает процесс выбор соответствующего архитектурного решения из предложенной таблицы. Для гетерогенных систем, включающих агентов разных типов, рекомендуется выделять единый семантический слой на основе онтологий. Архитектуру при этом дополняют специализированными хранилищами для отдельных подсистем.

Архитектуру Data Lakehouse в агентных систем выбирают чаще других. Спрос объясняется сочетанием двух свойств. Data Lakehouse объединяет гибкость и масштабируемость озер данных с дисциплиной управления, которая свойственна классическим хранилищам: единым стандартам качества, версионности и контроля доступа. Для агентных систем это имеет

огромное значение. Архитектура должна обеспечивать работу с любыми типами данных — от структурированных таблиц до неструктурированных документов и потоковых событий. При этом сохраняется семантическая целостность в рамках единой системы [1]. Развитие этой концепции идет в сторону создания так называемых семантических озер данных (Semantic Data Lakes), где поверх гибридного хранилища надстраивается унифицированный семантический слой. Такой слой предоставляет агентам единое понятийное пространство независимо от физического формата и расположения данных. Решения дополняют механизмы аудита и управления доступом. Система удовлетворяет требованиям безопасности и прослеживаемости данных.

III. ЗАКЛЮЧЕНИЕ

Семантика и архитектура данных в агентных системах тесно связаны. Изменение в одном аспекте влияют на требования к другому. Поэтому требуются комплексный подход при проектировании. Чтобы агенты эффективно взаимодействовали, недостаточно семантической согласованности их моделей знаний. Для этого нужны архитектурные решения, обеспечивающие хранение, обработку и передачу данных с учетом контекста и целей взаимодействия. Таблицы «Модели знаний и формы представления» и «Архитектура хранения и семантическая совместимость» систематизируют подходы к организации семантики и архитектуры данных в агентных системах. Их можно использовать как ориентир для выбора архитектурных решений при разработке конкретных систем. Сформулированные критерии выбора могут помочь разработчикам обоснованно принимать решения в

зависимости от типа агентов и требований к семантической согласованности.

В дальнейшем стоит сосредоточиться на двух задачах. Первая — автоматизировать согласование разнородных онтологий. Вторая задача — отработать гибридные архитектуры, которые объединяют сильные стороны разных подходов в рамках единой системы.

БЛАГОДАРНОСТЬ

Выражаю благодарность своему научному руководителю Первухину Дмитрию Анатольевичу за ценные советы при планировании исследования и рекомендации по оформлению статьи.

СПИСОК ЛИТЕРАТУРЫ

- [1] Абу Хасан Р., Кириенко А.Б., Хомоненко А.Д. Метод перехода от хранилищ данных к озерам данных геоинформационных систем на основе Лямбда-архитектуры // Интеллектуальные технологии на транспорте. 2024. № 1(37). С. 45-55. DOI: 10.20295/2413-2527-2024-137-45-55
- [2] Аксенов К.А., Спицина И.А. Применение метода поддержки принятия решений при разработке информационных систем на основе мультиагентного подхода // Современные наукоемкие технологии. 2022. № 12. С. 14-19. DOI: 10.17513/snt.39430
- [3] Костенко К.И., Белкин В.Ю. Онтология пользовательских интерфейсов в кибернетической модели интеллектуальных систем // Онтология проектирования. 2021. Т. 11. № 1(39). С. 89-103. DOI: 10.18287/2223-9537-2021-11-1-89-103
- [4] Макаренко С.И., Соловьева О.С. Семантическая интероперабельность взаимодействия элементов в сетевых системах // Журнал радиоэлектроники. 2021. № 6. DOI: 10.30898/1684-1719.2021.6.3
- [5] Подобрый А.Н., Тимирязнов В.В., Перцев А.А. Архитектура построения базы знаний проектной организации // Автоматизация процессов управления. 2021. № 4(66). С. 28-38. DOI: 10.35752/1991-2927-2021-4-66-28-38