УДК 004.04:004.9

МУЛЬТИАГЕНТНАЯ СИСТЕМА УПРАВЛЕНИЯ ЗНАНИЯМИ: МОДЕЛЬ ДОВЕРИЯ К ДЕЙСТВИЯМ ИНТЕЛЛЕКТУАЛЬНЫХ АГЕНТОВ

Бурый А.С., Российский институт стандартизации, г. Москва
Лопатин И.Н., Российский институт стандартизации, ПАО «Сбербанк», г. Москва
Митрофанов А.Д., Папруга А.А., ПАО «Сбербанк», г. Москва

Целью исследования является разработка архитектуры мультиагентной системы управления корпоративными знаниями (СУЗ) с механизмом динамического назначения уровня доверия интеллектуальным агентам (ИА) и источникам, обеспечивающей достоверность и управляемость корпоративной базы знаний в условиях распределенных источников данных.

Методы: аналитический обзор и синтез практик Data Governance, архитектурное проектирование мультиагентных систем, формализация метрик качества данных, разработка скоринговых функций и правил верификации знаний.

Результаты: предложена архитектура СУЗ с разделением на технический и бизнес-контуры. Определен состав и роли ИА (поставщик знаний, валидатор, суммаризатор, ассистент по знаниям). Сформулирована динамическая модель доверия и матрица прав для действий ИА в базе знаний. Описан двухступенчатый процесс валидации знаний и установлены требования к непрерывному мониторингу метрик качества и пересчету уровня доверия.

Ключевые слова: мультиагентная система, система управления знаниями, система доверия, интеллектуальный агент (ИА); мониторинг качества данных; уровень доверия к ИА.

Цитирование: Бурый А.С., Лопатин И.Н., Митрофанов А.Д., Папруга А.А. Мультиагентная система управления знаниями: модель доверия к действиям интеллектуальных агентов // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 5 (86). С. 83–91.

ПРИНЯТЫЕ СОКРАЩЕНИЯ

ДМД – Динамическая модель доверия
ИА – Интеллектуальный агент
ИС – Информационная система
ИИ – Искусственный интеллект
СУЗ – Системы управления знаниями
DKA – Domain knowledge agent

ВВЕДЕНИЕ

В современных организациях объем и разнообразие информации растут экспоненциально, что приводит к усложнению процессов поиска, интеграции и актуализации знаний. Эффективное управление корпоративными знаниями становится ключевым фактором конкурентоспособности, устойчивости и инновационного развития. Однако существующие системы управления знаниями (СУЗ) и метаданными[1] [1] часто не обеспечивают необходимый уровень достоверности, целостности и релевантности информации, ввиду отсутствия систем оценки и мониторинга контроля уровня качества знаний.

Одной из критических проблем является отсутствие комплексного механизма оценки доверия к источникам данных и агентам, участвующим в формировании и обновлении базы знаний любой организации. В качестве агентов обычно рассматриваются реактивные и когнитивные прикладные программные агенты [2], которые объединим понятием «интеллектуальные агенты» (ИА) [3]. При этом традиционные подходы к валидации данных [4] ориентированы либо на жесткие автоматические правила, либо на экспертную проверку, что приводит к низкой адаптивности системы в условиях динамично меняющейся информационной среды.

Задача представленной работы – разработка мультиагентной системы управления корпоративными знаниями, составляющими некоторый интеллектуальный капитал организации, который также требует измерения, накопления, придавая организации важные преимущества в конкурент-

[1] Метаданные – это данные, описывающее контекст (context – от лат. «связь») и контент (от англ. content – «содержание») объектов.

ной среде[2]. Система управления знаниями включает технический контур автоматизированного извлечения (модуль сбора знаний) и анализа метаданных, бизнес-контур экспертной валидации и семантического обогащения [2, 5], а также динамическую модель доверия, адаптирующую права ИА в зависимости от качества и достоверности предоставляемых ими знаний.

Поставленная задача декомпозируется в следующие подзадачи: проанализировать существующие подходы к управлению знаниями и метаданными; разработать мультиагентную архитектуру СУЗ с техническим контуром автоматизированного извлечения метаданных и бизнес контуром экспертного семантического обогащения; определить доменно-ориентированную модель представления знаний и принципы интеграции доменов; разработать модель доверия ИА и процедур валидации знаний; задать требования к разработке архитектуры мониторинга качества и доверия ИА.

Цель данного исследования – сформировать научно обоснованный алгоритмический аппарат, позволяющий рассчитывать уровень доверия к ответам ИА и их действиям в мультиагентной среде с применением СУЗ в качестве источника знаний.

Для этого предлагается интегрировать мультиагентную СУЗ с формализованной динамической моделью доверия (ДМД), которая адаптивно регулирует привилегии ИА на основе метрик качества и истории валидаций; предложить доменно ориентированную онтологическую модель представления знаний с обеспечением семантической совместимости между доменами; ввести правила пересчета доверия и их связь с матрицей прав на модификацию базы знаний ИА.

1. ЗНАНИЯ ОРГАНИЗАЦИИ

Под знаниями организации понимается целенаправленно актуализированная, обобщенная, систематизированная, структурированная и скоординированная информация. В системах организационного управления предназначение знаний – поддержка моделей принятия решений [6]. Составными частями знаний являются факты, концепции, объекты, субъекты, процессы и т. д. Информация представляется в виде формализации данных для их передачи, интерпретации и обработки.

Знания организации могут включать в себя, но не ограничиваться этим:

– знания информационных технологий – это знания об информационных системах (ИС) и процессах: код, архитектуры, конфигурации ИС, плейбуки. Такие знания хранятся в репозиториях, локальных wiki, поэтому быстро устаревают;

– бизнес-знания – знания, описывающие: клиентоориентированность, бизнес-цели, сведения о продуктах, ценах, каналах продаж, процессов сделки, об экономике и денежных потоках, тенденциях рынка и многое другое;

– знания кибербезопасности – это знания активов, рисков, реальных угроз, действующих политик и контролей [7]. Примеры знаний: архитектура средств защиты, плейбуки, отчеты работы команд по информационной безопасности, требования и стандарты кибербезопасности, распределение ролей, требования регуляторов.

На рис. 1 представлена общая схема взаимодействия элементов и подсистем в среде СУЗ.

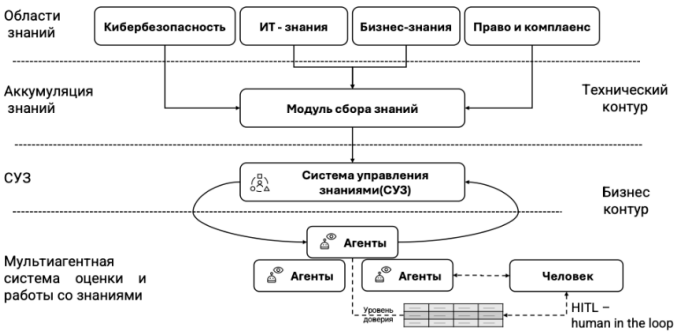

Рис. 1. Взаимодействие элементов и подсистем в среде СУЗ

2. СИСТЕМА УПРАВЛЕНИЯ ЗНАНИЯМИ

Система управления знаниями преобразует собранные метаданные в категоризированные области знаний [8]. Каждая область знаний разделена на внутренние и внешние домены, которые определяют ИА по знаниям (Domain Knowledge Agent, DKA) [9], который обеспечивает взаимодействие со знаниями своего домена на основе генеративного искусственного интеллекта (ИИ) [10, 11]. DKA берет факты из репозитория СУЗ, добавляет к ним контекст предметной области и синхронизирует с экспертами (людьми). Знания формализуются в онтологию домена, интегрированную в СУЗ, что обеспечивает семантическую совместимость и обмен между доменами. Представление в виде графов знаний [12] облегчает логический вывод и позволяет выявлять дополнительные связи между объектами знаний, дополнительный семантический слой фиксирует эквивалентные значения атрибутов разных источников.

Жизненный цикл знаний включает стадии представления знаний, моделирования знаний, приобретения знаний, хранения знаний, слияния знаний, вычислений знаний, применения и сопровождения знаний и т. д. [12].

[2] ГОСТ Р 54877–2016 Менеджмент знаний. Руководство для персонала при работе со знаниями. Измерение знаний. М.: Стандартинформ, 2020.

Качество и достоверность знаний – это обязательные метрики СУЗ, которые достигаются экспертной ручной валидацией и автоматическим контролем качества данных, интегрированным в сервисы СУЗ. Декомпозиция по доменам решает задачу масштабирования: уровень управляемости позволяет назначить одного ответственного за каждый домен и ускорить поиск (запросы могут ограничиваться доменом СУЗ), тем самым обеспечить необходимую управляемость.

Система управления знаниями [5, 12] представляет собой единую систему по работе с источниками знаний, сбору знаний, их систематизации, классификации и хранению.

Архитектура организации СУЗ включает в себя (см. рис. 2):

– инструменты взаимодействия с источниками знаний: коннекторы и API-интерфейсы к внутренним и внешним системам (репозитории кода, базы данных, wiki, тикеты);

– модули сбора информации: конвейеры извлечения – такие, как парсеры, нормализаторы, дедупликаторы, компоненты обогащения метаданных, работающие по расписанию или событиям;

– потоки наполнения СУЗ: автономные агенты с ручной модерацией, где ИА формируют знания, человек-эксперт подтверждает/исправляет автономные действия, повышая уровень доверия к действиям ИА [13];

– организацию хранения по доменам: логическая модель «домен / поддомен / подтип» с версиями, связями между артефактами (глоссарий, схемы, плейбуки), поиском по метаданным и разграничением доступа [14];

Рис. 2. Структура системы управления знаниями

– ИА валидации: формирование доверительных метрик (скоринг проверенных знаний) [15].

3. АГЕНТ ПО ЗНАНИЯМ

Для защиты корпоративной базы знаний данные от ИА проходят двойную проверку: автоматическими правилами и экспертами. Надежность источника оценивается моделью доверия [16], которая постоянно пересчитывается по метрикам системы многоуровневого мониторинга качества данных [15]: доля принятых знаний, количество ошибок, конфликтов и т. д.

При создании ИА получает доступ только в предметную область. Стартовый уровень доверия зависит от происхождения ИА: интеграция (использование инструментов «tools») с проверенной системой дает более высокий рейтинг.

Система мониторинга фиксирует результаты каждой валидации [7, 15], поднимая скоринг успешно пройденной проверки действия ИА человеком. Высокий скоринг действия ИА соответствует повышенному уровню доверия, что в свою очередь используется для передачи функции человека ИА [16, 17].

От уровня доверия зависят права и автономность ИА (рис. 3), включая:

– высокий уровень доверия – ИА может пополнять базу знаний, переопределять устаревшие записи, выступать арбитром при конфликте в случае мультиагентского исполнения;

– средний уровень доверия – допускается добавление и обновление некритичных данных, но требуется проверка человеком (HITL)[3];

– низкий уровень доверия – все сообщения маркируются «требуется проверка», удалять или менять критичные разделы запрещено.

[3] Аббревиатура от (англ.) Human In The Loop: модель, которая требует взаимодействия с человеком.

Рис. 3. Доступ к знаниям агентами с разным уровнем доверия

Права реализованы как «роль», автоматически меняющаяся при росте или падении рейтинга и мотивирующая ИА поддерживать репутацию.

Надежным ИА передают модерацию потоков: фильтрацию дубликатов, предварительное исправление метаданных, арбитраж конфликтов. Это масштабирует систему без постоянного участия людей.

Отображение доменов знаний в СУЗ представлено на рис. 4, где каждый из источников знаний получает автоматизированную скоринговую оценку.

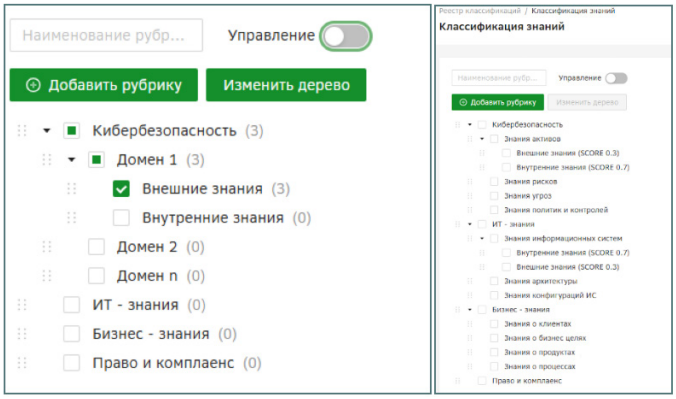

Рис. 4. СУЗ с отображением доменов

4. МУЛЬТИАГЕНТНОЕ ВЗАИМОДЕЙСТВИЕ

Каждый ИА наделен функциями в рамках своей роли. ИА по метаданным извлекает структурную информацию о таблицах и полях через API, а доменные ИА курируют содержательные аспекты своих областей [9, 14]:

– агент-валидатор автоматически проверяет новые знания на корректность и непротиворечивость (по правилам, онтологиям и истории изменений) и присваивает им статус доверия;

– агент-суммаризатор превращает длинные документы и потоковые логи в краткие, структурированные выжимки с ключевыми фактами, тегами и ссылками на источники;

– агент-ассистент по знаниям отвечает на вопросы пользователей [10], находит релевантные артефакты в СУЗ и объясняет контекст с учетом ролей, доменов и уровня доступа [14];

– агент-наполнения знаний подключается к источникам, извлекает и нормализует данные, создает черновики записей в СУЗ и обновляет их по расписанию или событиям.

Автономность сочетается с обменом сообщениями в общем информационном пространстве знаний, семантическую совместимость гарантирует единая онтология понятий и отношений.

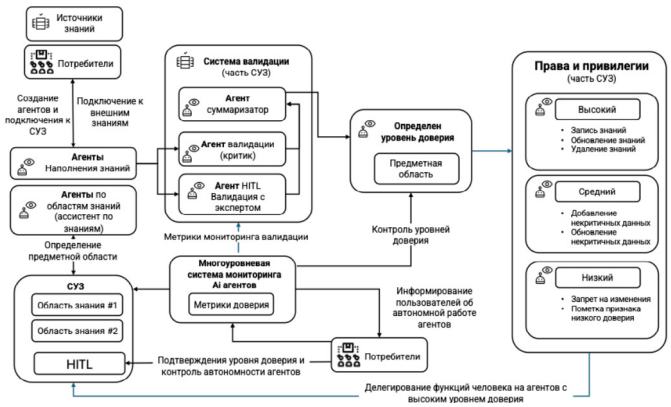

Рис. 5. Работа мультиагентной сети с областями знаний в СУЗ

Ниже представлена мультиагентная система записи и валидации знаний, которая включает механизмы оценки доверия и HITL (рис. 6):

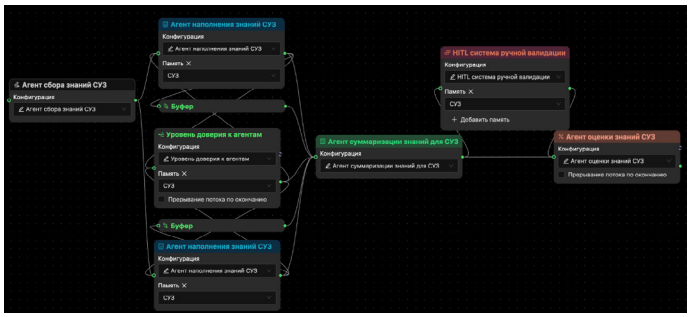

Рис. 6. Мультиагентная сеть агентов системы управления знаний

На рис. 6 представлены механизмы наполнения, валидации и оценки знаний ИИ агентами. Данный подход включает в себя модуль HITL, для ручной валидации действий ИА и поощрения ИА за успешно выполненные задачи.

5. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДОВЕРИЯ И СКОРИНГА

Эксплуатация мультиагентной СУЗ требует механизма динамического доверия к ИА, использующим разные домены знаний. В предлагаемой модели доверие обновляется на основе автоматической и экспертной валидации с учетом скоринга рейтинга ИА и текущего качества данных [17, 18].

Базовые параметры и исходное доверие к ИА

При инициализации каждому ИА назначаются предметный домен знаний D и исходная величина доверия T0 из диапазона [0; 1].

Исходное доверие рассчитывается как:

$$T0 = \alpha_{src} S + (1 - \alpha_{src}) C, \quad (1)$$

где

S – степень верификации источника ИА; например, $S = 0$ для внешнего домена знаний и $S = 1$ для внутреннего, корпоративного домена знания;

C – скоринговая метрика показателя ИА (отношение подтвержденных человеком действий ИА ко всем действиям за весь период работы: confirmed / total);

$\alpha_{src} \in [0; 1]$ – надежность источника знаний, определяемая человеком или в автономном режиме ИА относительно метрики C.

Пример 1. Внутренний домен знаний:

– источник: внутренний корпоративный домен знаний;

–степень верификации источника знания: $S = 1$ (полностью доверенный).

–портфель агента: $C = 0{,}6$ (то есть из 100 действий агентов – 60 были подтверждены человеком);

–вес источника: $\alpha_{src} = 0{,}7$ (эксперт считает происхождение источника более важным, чем данные в домене СУЗ).

Из выражения (1) получаем:

$$T0 = 0{,}7 \times 1 + (1 - 0{,}7) \times 0{,}6 = 0{,}88. \quad (2)$$

Агент получает высокий уровень доверия.

Пример 2. Внешний домен знаний:

– источник: внешний домен знаний;

– степень верификации источника: $S = 0$;

– портфель агента: $C = 0{,}75$ (75% действий подтверждены человеком);

– вес источника: $\alpha_{src} = 0{,}6$.

$$T0 = 0{,}6 \times 0 + (1 - 0{,}6) \times 0{,}75 = 0{,}3\,. \quad (3)$$

Агент получает низкий уровень доверия (L1).

На рис. 7 показаны графики зависимости величины доверия от уровня автономности ИА.

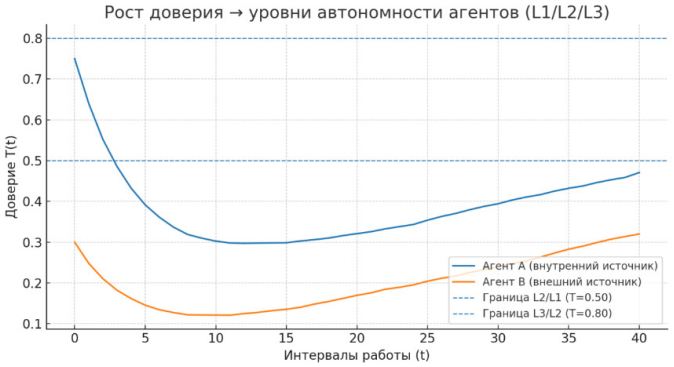

Рис. 7. Отношение доверия ИА к автономным действиям (работам)

Интервальные метрики и обновление доверия. Для адаптивного управления доверием система мониторинга на каждом интервале Δt собирает метрики активности ИА. Пересчет доверия выполняется по формуле экспоненциально-взвешенного усреднения:

$$T_{next} = g\,T + (1\text{-}g)\,[b\,(n_{val} / n_{sent}) - (1\text{-}b)\,(n_{ref} + n_{sent}) / n_{sent}]_{[1,0]}, \quad (4)$$

где T – доверие до пересчета; T_{next} – доверие после пересчета; $n_{sent}(D, t)$ – количество отправленных фактов; $n_{val}(D, t)$ – количество подтвержденных фактов; $n_{conf}(D, t)$ – количество коллизий; (D, t) – факторы, отправленные ИА в домене D за интервал времени t; $g \in [0; 1]$ – коэффициент памяти (инерционность модели); $b \in [0; 1]$ – коэффициент баланса между поощрением и штрафом; n_{val} / n_{sent} – доля подтвержденных фактов; $(n_{rej} + n_{conf}) / n_{sent}$ – доля отклоненных и конфликтных фактов.

Если $n_{sent} = 0$, допускается сохранение значения доверия без пересчета ($T_{next} = T$), либо применение полной памяти ($T_{next} = g\,T$ при $g = 1$).

Связь с базовой моделью доверия

Предлагаемая потоковая модель (4) опирается на классическую базовую схему пересчета доверия, используемую в моделях агентных систем [17, 18]. В общей форме она задается уравнением:

$$T_{now} = a\,T_{prev} + b\,Q + g_e\,E, \quad (4a)$$

где T_{prev} – доверие на предыдущем шаге; Q – автоматическая оценка качества фактов (например, доля подтверждений n_{val} / n_{sent} за интервал); E – экспертная оценка достоверности; a, b, g_e – весовые коэффициенты, нормированные по сумме, т.е. $(a + b + g_e = 1)$.

В данной схеме доверие ИА в текущий момент времени T_{now} рассчитывается по рекуррентному соотношению, представляющему собой линейную комбинацию нормирован-

ных факторов: предыдущего значения доверия T_{prev}, автоматической оценки качества Q и экспертной оценки E.

Потоковая модель (4) представляет собой конкретизацию общей схемы (4а) применительно к системам управления знаниями. В данной интерпретации автоматическая оценка качества Q определяется как отношение числа подтвержденных фактов к общему числу фактов за интервал наблюдения (n_{val} / n_{sent}). Экспертное влияние отражается косвенно: подтверждение увеличивает показатель n_{val}, а отклонение – показатель n_rej. Весовые коэффициенты общей модели (a, b, g_e) трансформируются в параметры g и b обеспечивая баланс между историческим значением доверия и результатами текущего наблюдения.

Уровни доверия и права агента. В базовом подходе предлагается использовать 3 уровня шкалы доверия агентов (см. табл. 1). При необходимости, в сложных ИС, градация может быть расширена.

Таблица 1

Градация уровней доверия и матрица прав

Уровень	Диапазон T	Действия в базе знаний	Основные функции
L3 (High)	T ⩾ 0.8	INSERT, UPDATE*, DELETE*	Арбитраж, модерация, редактура
L2 (Medium)	0.5 ⩽ T < 0.8	INSERT (tag=HITL), UPDATE (low-risk)	Ограниченное редактирование
L1 (Low)	T < 0.5	INSERT (HITL), SELECT	Без права публикации

Высокий уровень T_H. Агент с этим уровнем имеет привилегии на прямое внесение знаний, обновление существующих записей, а также арбитраж коллизий. Средний уровень T_M. Агент может добавлять знания только с дополнительной экспертизой, также может вносить редакцию в нечувствительные данные. Низкий уровень T_L. Сообщения агента проходят маркировку «требует проверки» (см. табл. 2).

Таблица 2

Права агентов на изменения знаний

Уровень доверия	Добавление знаний	Изменение знаний	Удаление знаний	Публикация в общую базу
Высокий	✔	✔	✔	✔
Средний	✔	✔	✖	✔
Низкий	✔	✖	✖	✖

При снижении уровня доверия $T \to T_L$ также происходит снижение уровня привилегий для исключения ошибок.

6. СИСТЕМА МОНИТОРИНГА ДЕЙСТВИЙ АГЕНТОВ И ДОВЕРИЯ

Функционирование мультиагентной системы управления знаниями невозможно без формализованного механизма мониторинга действий ИА и динамики доверия к ним. Мониторинг рассматривается как процесс непрерывного сбора, анализа и интерпретации данных о деятельности ИА в доменах знаний, обеспечивающий поддержание целостности и достоверности корпоративной базы знаний [7, 15].

Архитектура многоуровневой системы мониторинга ИА

Архитектура системы мониторинга базируется на принципах многоуровневого контроля и интеграции с СУЗ.

Для контроля агентов необходимы следующие уровни мониторинга:

– инфраструктурный уровень – регистрация инфраструктурных метрик ИА, контроль активности SQL запросов агентов к системе знаний;

– статистический уровень и уровень качества данных – контроль полноты и консистентности метаданных, контроль качества поставляемых знаний, контроль обязательных полей в системе знаний по каждому из доменов;

– уровень доверия – контроль и оценка доверия к ИА на основе формул (4) и (4а), отражающих соотношение успешных и ошибочных действий.

Данные уровни образуют целостную систему, обеспечивающую как технический, так и семантический контроль достоверности знаний [7, 15].

Метрики мониторинга агентов по знаниям.

Для каждого ИА $\mathcal{A}$ в домене D за интервал наблюдения t фиксируются следующие метрики из выражения (4):

$$\mathcal{A}(D, t) = \{ n_{sen}, n_{val}, n_{rej}, n_{conf} \}.$$

Совокупность указанных метрик используется при пересчете доверия по формуле (4). Кроме того, анализ временных траекторий $T(t)$ позволяет выявлять аномалии, например компрометацию ИА, либо систематические ошибки при обработке данных [7].

Механизм управления доверия к действиям ИА

Значение доверия определяет полномочия [14] ИА в системе:

– при **высоком доверии ($T \geqslant 0{,}8$)** агенту разрешается прямое внесение и модификация знаний;

– при **среднем доверии ($0{,}5 \leqslant T < 0{,}8$)** действия агента помечаются как требующие экспертной проверки;

– при **низком доверии ($T < 0{,}5$)** операции ограничиваются карантинным режимом и не влияют на общую базу без внешнего подтверждения.

Таким образом, доверие выполняет роль адаптивного регулятора прав доступа [14], а мониторинг обеспечивает своевременное их пересмотрение [7].

Мультиагентная организация мониторинга

Контроль достигается за счет мультиагентной организации системы мониторинга. В ее структуру входят ИА, предназначенные для контроля логов и метрик, а также выявления аномалий в работе ИА по знаниям:

– агенты-детекторы – фиксируют первичные метрики и выявляют аномалии;

– агенты-архитекторы – планируют реакцию системы на отклонения;

– агенты-мыслители – декомпозируют задачи по доменам знаний;

– агенты-исполнители – автоматизируют корректирующие действия и обновляют статусы доверия.

Коллективное функционирование данных типов ИА обеспечивает саморегуляцию мониторинга: факты классифицируются автоматически, отклонения направляются на повторную валидацию через систему HITL, а высоконадежные ИА принимают функции модерации потоков знаний [7].

ЗАКЛЮЧЕНИЕ

В ходе исследования выполнена разработка архитектуры мультиагентной системы управления корпоративными знаниями, обеспечивающей достоверность и управляемость корпоративной базы знаний в условиях распределенных источников знаний.

Решение позволило: выявить ограничения существующих подходов к управлению знаниями и метаданными; спроектировать систему с техническим контуром автоматизированного извлечения метаданных и бизнес-контуром экспертного семантического обогащения; определить онтологическую модель представления знаний и принципы интеграции доменов; предложить формализованную динамическую модель доверия ИА, основанную на скоринговых функциях и двухступенчатой процедуре валидации знаний; задать требования к мониторингу работы ИА и доверия к действиям; выявить влияния низкого качества данных при работе ИА по знаниям.

Полученные результаты формируют основу для создания корпоративных систем управления знаниями, способных к масштабированию, самоорганизации и повышению устойчивости к рискам использования недостоверных знаний. Перспективы дальнейших исследований связаны с расширением механизмов риск-ориентированного мониторинга, интеграцией с интеллектуальными ИС поддержки принятия решений и стандартизацией процессов оценки качества данных в мультиагентных средах.

***Рецензент: Сухов Андрей Владимирович,** доктор технических наук, профессор, старший научный сотрудник ФКУ «НПО «Специальная техника и связь», главный специалист ФГБУ «Институт стандартизации», г. Москва, Российская Федерация.*
E-mail: a.s.burji@gostinfo.ru

Список использованных источников и литературы / References

1. Тузовский А.Ф. Формирование семантических метаданных для объектов системы управления знаниями // Известия Томского политехнического университета. 2007. Т. 310, № 3. С. 108–112. / Tuzovsky A.F. Formirovanie semanticheskih metadannyh dlya obektov sistemy upravleniya znaniyami. Izvestiya Tomskogo politekhnicheskogo universiteta. 2007; 310(3): 108–112. (In Russ.).
2. Зайцев Е.И., Нурматова Е.В. О подходе к управлению знаниями и разработке мультиагентной системы представления и обработки знаний // Russian Technological Journal. 2023. Т. 11, № 4. С. 16–25. https://doi.org/10.32362/2500-316X-2023-11-4-16-25 / Zajtsev E.I., Nurmatova E.V. O podhode k upravleniyu znaniyami i razrabotke mul'tia-gentnoj sistemy predstavleniya i obrabotki znanij. Russian Technological Journal. 2023; 11(4):16–25. (In Russ.).
3. Бурый А.С., Фролов В.А., Куляница А.Л. Эволюция агентного моделирования. Часть 1. Архитектура интеллектуального агента // Информационно-экономические аспекты стандартизации и технического регулирования. 2023. № 5 (74). С. 38–47. / Buryi A.S., Frolov V.A., Kulyanitsa A.L. The evolution of agent-based modeling. Information and Economic Aspects of Standardization and Technical Regulation. 2023; 5 (74): 38–47. (In Russ.).

4. Бурый А.С., Погодин И.М. Оценка качества больших данных. Часть 1. Основные понятия и метрики // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 3 (78). С. 49–58. / Buryi A.S., Pogodin I.M. Ocenka kachestva bol'shih dannyh. Part 1. Osnovnye ponyatiya i metriki. Information and Economic Aspects of Standardization and Technical Regulation. 2024; 3(78): 49–58. (In Russ.).
5. Суслов Д.С. Управление знаниями в организации: основные модели // Креативная экономика. 2012. № 10 (70). С. 89–97. / Suslov D.S. Upravlenie znaniyami v organizacii: osnovnye modeli. Kreativnaya ekonomika. 2012; 10(70): 89–97. (In Russ.).
6. Бурый А.С. Информационно-поисковые социотехнические системы: термины и определения. – М.: «Горячая линия-Телеком», 2018. – 166 с. / Buryi A.S. Informacionno-poiskovye sociotekhnicheskie sistemy: terminy i opredeleniya. Moscow: "Goryachaya liniya-Telekom", Publ., 2018, 166 p. (In Russ.).
7. Лопатин И.Н. Многоуровневые системы качественных данных на основе моделей искусственного интеллекта: проблемы и решения // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 1 (82). С. 70–75. / Lopatin I. N. Multi-level data-quality systems based on artificial intelligence models: problems and solutions. Information and Economic Aspects of Standardization and Technical Regulation. 2025; 1 (82): 70–75. (In Russ.).
8. Aryal S., Do T., Heyojoo B., et al. Leveraging multi-AI agents for cross-domain knowledge discovery. arXiv preprint arXiv. 2024; 2404.08511.
9. Kostka A., Chudziak J.A. Synergizing logical reasoning, knowledge management and collaboration in multi-agent LLM system. arXiv preprint arXiv. 2025; 2507.02170.
10. Бурый А.С., Цаплина О.С. Генеративный искусственный интеллект цифрового университета // Информационно-экономические аспекты стандартизации и технического регулирования. 2024. № 5 (80). С. 85–91. / Buryi A.S., Tsaplina O.S. Generativnyj iskusstvennyj intellekt cifrovogo universiteta. Information and Economic Aspects of Standardization and Technical Regulation. 2024; 5(80): 85–91. (In Russ.).
11. Днепровская Н.В. Система управления знаниями как основа смарт-обучения // Открытое образование. 2018. Т. 22, № 4. С. 42–52. https://doi.org/10.21686/1818-4243-2018-4-42-52 / Dneprovskaya N.V. Sistema upravleniya znaniyami kak osnova smart-obucheniya. Otkrytoe obrazovanie. 2018; 22 (4): 42–52. (In Russ.).
12. Лахин О.И., Юрыгина Ю.С., Анисимов А.С. Принципы построения системы управления знаниями предприятий ракетно-космической промышленности // Онтология проектирования. 2017. Т. 7, № 3 (25). С. 270–283. https://doi.org/10.18287/2223-9537-2017-7-3-270-283 / Lakhin O.I., Yurygina Y.S., Anisimov A.S. Principy postroeniya sistemy upravleniya znaniyami predpriyatij raketno-kosmicheskoj promyshlennosti. Ontologiya proektirovaniya. 2017; 7 (3): 270–283. (In Russ.).
13. Гарбук С.В. Особенности применения понятия «доверие» в области искусственного интеллекта // Искусственный интеллект и принятие решений. 2020. № 3. С. 15–21. / Garbuk S.V. Osobennosti primeneniya ponyatiya "doverie" v oblasti iskusstvennogo intellekta. Iskusstvennyj intellekt i prinyatie reshenij. 2020; 3: 15–21. (In Russ.).
14. Shi T., He J., Wang Z., et al. Progent: Programmable privilege control for LLM agents. arXiv preprint arXiv. 2025; 2504.11703. https://doi.org/10.18287/2223-9537-2017-7-3-270-283
15. Лопатин И.Н. Методика скоринга источников и инцидентов в многоуровневых системах качества данных // Информационно-экономические аспекты стандартизации и технического регулирования. 2025. № 1 (84). С. 101–107. / Lopatin I.N. Scoring methodology for sources and incidents in multi-level data-quality systems. Information and Economic Aspects of Standardization and Technical Regulationю. 2025; 1 (84): 101–107. (In Russ.).
16. Al-Shamaileh M., Anthony P., Charters S. Agent-Based Trust and Reputation Model in Smart IoT Environments. Technologies. 2024; 12 (11): 208.
17. Бурый А.С., Фролов В.А., Куляница А.Л. Эволюция агентного моделирования. Часть 2. Имитационное моделирование // Информационно-экономические аспекты стандартизации и технического регулирования. 2023. № 6 (75). С. 46–52. / Buryi A.S., Frolov V.A., Kulyanitsa A.L. Evolyuciya agentnogo modelirovaniya. Part 2. Imitacionnoe modelirovanie. Information and Economic Aspects of Standardization and Technical Regulation. 2023; 6 (75): 46–52. (In Russ.).
18. DAMA-DMBOK: свод знаний по управлению данными / DAMA International; перевод с английского Г. Агафонов. – 2-е изд.. – М.: Олимп-Бизнес, 2023. – 828 с. / DAMA International. DAMA-DMBOK: Data Management Body of Knowledge. 2nd ed. Technics Publications. Moscow: Olimp-Biznes Publ., 2023, 828 p. (In Russ.).

MULTI-AGENT KNOWLEDGE MANAGEMENT SYSTEM: A MODEL OF TRUST IN THE ACTIONS OF INTELLIGENT AGENTS

Buryi A.S., Russian Standardization Institute, Moscow
Lopatin I.N., Russian Standardization Institute, Moscow, Russia, Sberbank of Russia, Moscow
Mitrofanov A.D., Sberbank of Russia, Moscow
Papruga A.A., Sberbank of Russia, Moscow

The purpose of the research is to develop an architecture for a multi-agent corporate knowledge management system (KMS) with a mechanism for dynamically assigning a level of trust to intelligent agents (IA) and sources, ensuring the reliability and manageability of the corporate knowledge base in a distributed data source environment. Methods: analytical review and synthesis of Data Governance practices, architectural design of multi-agent systems, formalization of data quality metrics, development of scoring functions and knowledge verification rules. Results: an architecture KMS is proposed with a division into technical and business contours. The composition and roles of the IA (knowledge provider, validator, summarizer, knowledge assistant) are defined. A dynamic model of trust and a matrix of rights for IA actions in the knowledge base are formulated. A two-stage knowledge validation process is described and the requirements for continuous monitoring of quality metrics and recalculation of the level of trust are established.

Keywords: multi-agent system, knowledge management system, trust system, intelligent agent (IA); data quality monitoring; level of trust in IA.

For citation: Buryi A.S., Lopatin I.N., Mitrofanov A.D., Papruga A.A. Multi-agent knowledge management system: a model of trust in the actions of intelligent agents. Information and Economic Aspects of Standardization and Technical Regulation. 2025; 5 (86): 83–91. (In Russ.).