

Формализация информации и Big Data

<http://vikchas.ru>

Практическая работа 1-2026 «Формализация информации, BigData и технологии ИИ»

Часовских Виктор Петрович

д-р техн. наук, профессор кафедры ШИиКМ

ФГБОУ ВО «Уральский государственный экономический
университет»

Екатеринбург 2026

ВВЕДЕНИЕ

В современном мире объемы генерируемой информации растут в геометрической прогрессии. Согласно исследованиям IDC, к 2025 году общий объем мировых данных превысит 175 зеттабайт. Эта лавина информации требует новых подходов к её обработке, хранению и анализу. Именно здесь на передний план выходят три взаимосвязанных концепции: формализация информации, BigData и технологии искусственного интеллекта.

Цель данной лабораторной работы — исследовать взаимосвязь между формализацией информации, технологиями Big Data и развитием искусственного интеллекта, а также проанализировать, как эти направления трансформируют современные бизнес-процессы и общество.

Теоретические основы формализации информации

Формализация информации представляет собой процесс представления информации в виде, пригодном для машинной обработки. Этот процесс включает в себя структурирование данных, определение их типов, форматов, связей между элементами и правил их преобразования.

Согласно теории информации Клода Шеннона, формализация является ключевым этапом преобразования сырых данных в информацию, а затем — в знания. Основными принципами формализации выступают:

- ❖ Единообразие представления данных
- ❖ Недвусмысленность толкования
- ❖ Возможность алгоритмической обработки
- ❖ Минимизация избыточности данных

Формализация информации играет фундаментальную роль в развитии информационных технологий. Именно благодаря формализации стало возможным создание сложных информационных систем, баз данных и впоследствии — технологий Big Data и искусственного интеллекта.

Эволюция методов формализации информации напрямую коррелирует с развитием вычислительных возможностей. От простых структурированных данных в первых базах данных мы пришли к сложным семантическим моделям, позволяющим работать с неструктурированной информацией и извлекать из неё знания.

Big Data: от концепции к технологиям

Термин “Big Data” относится к наборам данных, которые настолько велики или сложны, что традиционные методы обработки данных неадекватны для их анализа. Концепция Big Data характеризуется через модель “5V”:

Volume(Объем) — огромное количество данных

Velocity(Скорость) — быстрое генерирование и необходимость быстрой обработки данных

Variety(Разнообразие) — различные типы и форматы данных

Veracity(Достоверность) — неопределенность и недостоверность данных

Value(Ценность) — потенциальная полезность извлекаемой информации

Несмотря на значительные достижения в области Big Data, существуют серьезные проблемы и вызовы:

Технические проблемы:

- ❖ Масштабируемость систем хранения и обработки
- ❖ Производительность алгоритмов анализа
- ❖ Интеграция гетерогенных данных

Нетехнические проблемы:

- ❖ Конфиденциальность данных
- ❖ Этические аспекты использования
- ❖ Правовое регулирование
- ❖ Недостаток квалифицированных специалистов

Технологии искусственного интеллекта в контексте Big Data

Искусственный интеллект прошел несколько этапов развития:

Символический ИИ(1950-е — 1980-е):

- Экспертные системы
- Логическое программирование
- Формальные грамматики

Статистический ИИ(1990-е — 2000-е):

- Байесовские методы
- Скрытые марковские модели
- Метод опорных векторов

Нейросетевой ИИ(2010-е — настоящее время):

- Глубокие нейронные сети
- Глубокое обучение с учителем
- Трансформеры и большие языковые модели (LLM)

Машинное обучение и Big Data

Машинное обучение является центральным компонентом современного ИИ и тесно связано с Big Data:

Типы машинного обучения:

- Обучение с учителем (supervised learning)
- Обучение без учителя (unsupervised learning)
- Обучение с подкреплением (reinforcement learning)
- Трансферное обучение (transfer learning)

Алгоритмы машинного обучения для работы с большими данными:

- Стохастический градиентный спуск
- Распределенные ансамблевые методы
- Алгоритмы для параллельной и распределенной обработки

Нейронные сети и глубокое обучение в эпоху Big Data

Глубокое обучение произвело революцию в области ИИ, во многом благодаря возможности обработки больших объемов данных:

Архитектуры нейронных сетей:

- Сверточные нейронные сети (CNN)
- Рекуррентные нейронные сети (RNN)
- Архитектуры на основе внимания (Attention)
- Трансформеры (Transformers)

Технологические аспекты глубокого обучения:

- Графические процессоры (GPU) и специализированные чипы (TPU, NPU)
- Распределенное обучение на кластерах
- Оптимизация гиперпараметров

Взаимосвязь формализации информации, Big Data и ИИ

Формализация информации создает необходимую инфраструктуру для развития технологий Big Data и ИИ:

От неструктурированных данных к структурированной информации:

- Методы извлечения структурированной информации из текста, изображений, видео
- Системы аннотирования и метаданных
- Технологии семантического веба и связанных данных (Linked Data)

Формализация знаний для ИИ:

- Представление знаний для экспертных систем
- Онтологии как основа для семантического анализа
- Графы знаний для современных систем ИИ

Big Data как источник для обучения систем ИИ

Большие данные служат фундаментом для обучения современных систем искусственного интеллекта:

Корпусы данных для обучения:

- Текстовые корпуса (Common Crawl, Wikipedia, WebText и др.)
- Наборы изображений (ImageNet, COCO и др.)
- Мультимодальные наборы данных

Методологии создания и разметки данных:

- Краудсорсинг для разметки данных
- Активное обучение (active learning)
- Самообучение (self-supervised learning)

Практическое применение технологий Big Data и ИИ

Технологии Big Data и ИИ трансформируют бизнес-процессы и промышленность:

Маркетинг и продажи:

- Персонализированные рекомендации
- Прогнозирование потребительского поведения
- Оптимизация ценообразования

Производство и логистика:

- Предиктивное обслуживание оборудования
- Оптимизация цепочек поставок
- Контроль качества с применением компьютерного зрения

Финансовый сектор:

- Алгоритмическая торговля
- Выявление мошенничества
- Оценка кредитных рисков
- Автоматизация комплаенса

Применение в науке и медицине

Научные исследования и медицина получили мощный импульс благодаря развитию технологий Big Data и ИИ:

Биоинформатика и геномика:

- Анализ геномных последовательностей
- Предсказание структуры белков (AlphaFold)
- Разработка новых лекарств

Медицинская диагностика:

- Анализ медицинских изображений
- Прогнозирование заболеваний
- Персонализированная медицина

Фундаментальная наука:

- Обработка данных с научных установок (коллайдеров, телескопов)
- Моделирование сложных физических и химических процессов
- Открытие новых материалов и соединений

Формализация информации, технологии Big Data и искусственный интеллект представляют собой взаимосвязанный триумвират, определяющий развитие информационных технологий в XXI веке.

Формализация создает основу для структурирования информации, Big Data обеспечивает масштабное хранение и обработку данных, а ИИ предоставляет методы для извлечения ценных знаний и принятия решений.

Дальнейшее развитие этих направлений будет определяться несколькими ключевыми тенденциями:

- Интеграция мультимодальных данных из разнородных источников
- Развитие методов федеративного машинного обучения, сохраняющего приватность
- Создание самообучающихся систем с минимальным участием человека
- Разработка объяснимых моделей ИИ с прозрачными механизмами принятия решений
- Увеличение энергоэффективности алгоритмов и инфраструктуры

Искусственный интеллект для анализа Big Data

Искусственный интеллект предоставляет инструменты для более эффективного анализа больших данных:

Интеллектуальная аналитика данных:

- Обнаружение аномалий
- Предиктивная аналитика
- Кластерный анализ
- Обработка естественного языка для извлечения знаний

Автоматизация процессов Data Science:

- AutoML (автоматическое машинное обучение)
- Интеллектуальная подготовка данных
- Автоматический выбор моделей и гиперпараметров

Темы для изучения

1. Как формализация информации влияет на эффективность работы с большими данными?
2. Какую роль играют большие данные в разработке современных систем ИИ?
3. Какие методы формализации используются для обучения нейронных сетей?
4. Как технологии ИИ помогают в анализе и структурировании больших объемов данных?

Задание по курсовой работе: Теоретически - практическое изучение тем и оформление по каждой теме реферата.

Благодарю за внимание!

