

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА МОДУЛЯ ИМПОРТА ОБЪЯВЛЕНИЙ ОБ
ОБЪЕКТАХ НЕДВИЖИМОСТИ ИЗ
ВНЕШНИХ ИСТОЧНИКОВ В СИСТЕМУ ХРАНЕНИЯ НА БАЗЕ
СЕМАНТИЧЕСКОЙ СЕТИ**

Нестеров И.А. (Университет ИТМО)

**Научный руководитель — старший преподаватель факультета программной
инженерии и компьютерной техники Цопа Е.А.** (Университет ИТМО)

Введение

В условиях стремительного роста объемов и разнообразия данных, сопровождающих рынок недвижимости, особое значение приобретает разработка эффективных средств структурирования, хранения и анализа информации. Традиционные реляционные модели, используемые в большинстве современных систем, оказываются недостаточно гибкими при работе с слабо формализованными и семантически насыщенными данными [1], особенно при необходимости последующего лингвистического анализа и интеграции данных из различных источников.

Основная часть

В представленной работе рассматривается разработка модуля импорта объявлений о недвижимости из внешних веб-источников (в частности, платформы ЦИАН) в систему хранения, построенную на основе семантической сети. Одной из ключевых задач стало проектирование онтологии предметной области, отражающей взаимосвязи между объектами недвижимости, типами сделок, характеристиками помещений и иными сущностями, присущими данному рынку. Онтология реализована в формате OWL и согласована с архитектурой графовой СУБД, применяемой в проекте. На основе анализа функциональных требований и особенностей предметной области был разработан модуль обработки данных, включающий этапы парсинга объявлений, извлечения структурированных атрибутов, лемматизации и токенизации текстовых описаний. Для промежуточной обработки данных выбран фреймворк Spring Batch, обеспечивающий транзакционную обработку потоков данных и интеграцию с семантической моделью. Разработана структура графowego представления объявления, адаптированная под специфику семантического хранилища, включая поддержку пользовательских связей между лексемами.

В рамках работы был подготовлен тестовый набор из 84 объявлений, очищенных и нормализованных для загрузки в базу. Кроме того, проведена сравнительная оценка графовых СУБД (Neo4j, ArangoDB, OrientDB) по параметрам масштабируемости, гибкости, скорости и удобства внедрения, а также протестирована применимость NLP-инструментов при работе с ними. Результаты подтвердили эффективность выбранных решений и позволили обосновать выбор технологий.

Выводы

Разработанная архитектура продемонстрировала стабильную работу в экспериментальном режиме и готова к расширению. Работа прошла апробацию в

рамках студенческой научной конференции и представляет собой основу для последующей интеграции в промышленную среду [2].

Литература

1. Клименков С. В., Николаев В. В., Харитонов А. Е., Гаврилов А. В., Письмак А. Е., Покид А. В. Применение семантической сети для хранения слабоструктурированных данных // Инженерный вестник Дона [Электронный ресурс]. 2020. № 2(62). С. 27. – Режим доступа: <https://cyberleninka.ru/article/n/primenenie-semanticheskoy-seti-dlya-hraneniya-slabostrukturirovannyh-dannyh> (Дата обращения 01.03.2026).
2. Письмак А. Е., Харитонов А. Е., Цопа Е. А., Клименков С. В. Метод автоматического формирования семантической сети из слабоструктурированных источников // Программные продукты и системы. 2016. № 3. С. 74–78.