

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

Нестеров И.А. (Университет ИТМО)

Научный руководитель — Цопа Е.А. (Университет ИТМО)

Введение. В современном мире поиск недвижимости стал более чем актуальной задачей: ежегодно растёт как спрос на жильё, так и объёмы объявлений от владельцев. Традиционные реляционные базы данных не всегда справляются с динамичным расширением структуры данных и усложнённым анализом текстовых описаний. Чтобы повысить эффективность работы с большим массивом объявлений, актуально использовать семантические сети, позволяющие гибко и точно структурировать сведения об объектах недвижимости.

Основная часть. В данной работе представлена разработка модуля импорта объявлений из внешних источников (на примере «Циан») в систему хранения на базе семантической сети. Решение ориентировано на несколько ключевых этапов. Во-первых, проводится анализ структуры данных в релевантных внешних ресурсах для выявления основных и наиболее часто повторяющихся атрибутов объекта недвижимости, таких как площадь, количество комнат, высота потолков, вид из окон и другие. Во-вторых, разрабатывается и внедряется механизм конвейерной обработки объявлений, включающий в себя парсинг HTML-страниц, лемматизацию текста описания и преобразование полученной информации к семантическому формату. Особое внимание уделяется связям внутри семантической сети: каждая лексема (например, «площадь» или «балкон») дополняется информацией о типах отношений (property, гиперогипонимия, антонимия и др.). В итоге, при занесении объявлений в базу данных проект «SemNet» получает полную картину характеристик объекта с учётом внутренних связей и синонимии. Это даёт возможность быстро масштабировать решение и осуществлять более точную фильтрацию и поиск. Кроме того, благодаря гибкой архитектуре (Spring Batch, IoC, CDI) модуль легко расширять для интеграции с другими площадками объявлений, а транзакционный механизм обеспечивает надёжность при больших объёмах данных.

Выводы. Предложенный модуль импорта объявлений улучшает качество поиска объектов недвижимости за счёт полноценной семантической модели и даёт основу для дальнейших исследований — от интеллектуальной обработки текстовых описаний до внедрения нейросетевых алгоритмов. Решение способствует экономии времени пользователей, повышает точность рекомендаций и открывает перспективы для масштабирования в смежные сферы анализа больших текстовых массивов.

Список использованных источников:

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 журнал] – 2020. - № 2(62). – С. 27
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.