

ПРИМЕНЕНИЕ ГРАФОВЫХ БАЗ ДАННЫХ В КОНКУРЕНТНОЙ РАЗВЕДКЕ

Т.Е. Белогривая, Р.А. Фортель

Развитие методов и средств мониторинга, адаптивного агрегирования и обобщения потоков информации из глобальных компьютерных сетей для поддержки информационно-аналитической деятельности в различных прикладных сферах является весьма актуальной проблемой, требует использования специальных технических средств и соблюдения определенных этических норм. Подобный вид деятельности, известный как конкурентная разведка по открытым источникам данных, часто ставит своей целью исследование рынка для развития бизнеса и разработки стратегии его дальнейшего продвижения.

В докладе представлен пример реализации методов конкурентной разведки по открытым источникам - веб-сайтам, публикующим объявления о вакансиях. Целью ставилось изучение потребностей белорусских работодателей в специалистах ИТ-сферы и динамики изменения таких потребностей за последнее время. Учитывая объем информации и ее сетевой характер, в качестве среды анализа и визуализации была выбрана графовая база данных Neo4j, решение с открытым кодом.

В качестве содержательных и надежных общедоступных источников данных были использованы сайты, публикующие в статической форме информацию из внутренних баз данных (подобно сайтам rabota.by, praca.by). Скрейпинг (парсинг) информации веб-страниц производился средствами библиотек `requests`, `bs4` и `selenium` языка Python. Сайты, использующие динамическое формирование контента средствами JavaScript и работу с объектной моделью документа, не рассматривались. Типовые поля данных о вакансии: ИТ-компания, адрес, период, должность, зарплата, опыт работы.

Собранная информация была представлена в формате кортежей вида `<object, relation, object>`, что позволило анализировать и визуализировать ее средствами графовой базы данных Neo4j, а также приложениями, реализующими технологию гиперболического браузера. Были получены эффективные представления о конкурирующих на рынке компаниях, приоритетах ИТ-отрасли, изменениях атрибутов вакансий и динамике спроса на специалистов. Использование средств графовой БД показало их высокую эффективность в сравнении со средствами реляционных СУБД.