

МОДЕЛИ ГРАФОВЫХ БАЗ ДАННЫХ ДЛЯ АНАЛИЗА ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ

Павлюченко К. А., Киселев А. И.

Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь

E-mail: mendorill@mail.ru, and.kis135@gmail.com

Модели графовых баз данных предоставляют мощный инструмент для анализа предпочтений пользователей. Их использование позволяет эффективно хранить и анализировать связи между объектами информационных систем, что может быть весьма полезным бизнесу для повышения заинтересованности пользователей и увеличения доходности. Данный исследовательский проект изучает, как графовые базы данных могут быть применены для анализа предпочтений пользователей и создания персонализированных рекомендаций.

ВВЕДЕНИЕ

Наличие возможности детального анализа предпочтений пользователей имеет существенное влияние на доходность предприятий в современном мире. Знание о том, что предпочитают пользователи, позволяет создавать персонализированные рекомендации, улучшать поисковую выдачу и повышать качество обслуживания. В данной работе рассматривается применение моделей графовых баз данных для анализа предпочтений пользователей, исследуется их реализация и полученные результаты.

Модели графовых баз данных предоставляют удобный и эффективный способ хранения и анализа данных, организованных в виде графа [1]. Модели графовых баз данных имеют следующие преимущества:

- Гибкость и масштабируемость создаваемых моделей;
- Интуитивно понятное моделирование сложных связей;
- Широкие возможности для глубокого анализа данных.

Граф представляет собой совокупность узлов и связей между ними. В контексте анализа предпочтений пользователей информационной системы, узлами могут быть аккаунты пользователей и другие объекты системы, а связи между ними будут отражать предпочтения конкретных пользователей [2].

I. МОДЕЛИРОВАНИЕ ГРАФОВОЙ БАЗЫ ДАННЫХ

Моделирование графовой базы данных начинается с анализа и получения доступа к источнику данных. Это включает в себя изучение документации и получение ключей публичного API сервиса. Если сервис не предоставляет пользовательских API, то для получения данных можно применить техники веб-скрейпинга или же использовать сторонние API для доступа к контенту веб-ресурса. Исходя из анализа структуры доступных данных происходит создание графовой модели, определяются цели и параметры сбо-

ра данных, (например, информация о подписках, лайках, комментариях, просмотрах и других активностях пользователей). В полученной модели каждый пользователь и объект (например, фотография или видео) представлены узлами, а связи между ними отражают действия пользователя по отношению к объектам (например, "пользователь-лайк-фото"). С использованием полученного доступа начинается сбор данных о пользователях, их профилях, подписках, а также данных о контенте, который представляет наибольший интерес (например, фотографии, видео, текстовые публикации). Используя полученную графовую базу данных, становится возможно определить, какие объекты наиболее популярны среди пользователей, какие пользователи имеют схожие интересы и какие темы наиболее обсуждаются. Этот анализ позволяет лучше понимать предпочтения пользователей.

II. РАЗРАБОТКА РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ

На основе моделей анализа данных разрабатываются рекомендательные системы [3], которые предоставляют пользователям персонализированный контент и рекомендации на основе их предпочтений. Это может включать в себя рекомендации для просмотра видео, фотографий, текстовых публикаций и других объектов системы.

Для удобства описания данных, используется язык запросов, который поддерживает графовую базу данных Neo4j [4]. Пример запроса для создания узлов пользователей представлен на рисунке 1.

```
graph.run("CREATE (u1:User {name: 'User1'})")  
graph.run("CREATE (u2:User {name: 'User2'})")
```

Рис. 1 – создание узлов пользователей

Пример создания узлов категорий представлен на рисунке 2.

```
graph.run("CREATE (c1:Category {name: 'Category1'})")
graph.run("CREATE (c2:Category {name: 'Category2'})")
```

Рис. 2 – создание узлов категорий

Создание связей между пользователями и их предпочитаемыми категориями продемонстрировано на рисунке 3.

```
graph.run("MATCH (u:User), (c:Category)
WHERE u.name = 'User1' AND c.name = 'Category1'
CREATE (u)-[:LIKES]->(c)")

graph.run("MATCH (u:User), (c:Category)
WHERE u.name = 'User1' AND c.name = 'Category2'
CREATE (u)-[:LIKES]->(c)")

graph.run("MATCH (u:User), (c:Category)
WHERE u.name = 'User2' AND c.name = 'Category1'
CREATE (u)-[:LIKES]->(c)")
```

Рис. 3 – создание связей между пользователями и их предпочитаемыми категориями

Полученная модель может быть представлена в виде графа, изображённого на рисунке 4.

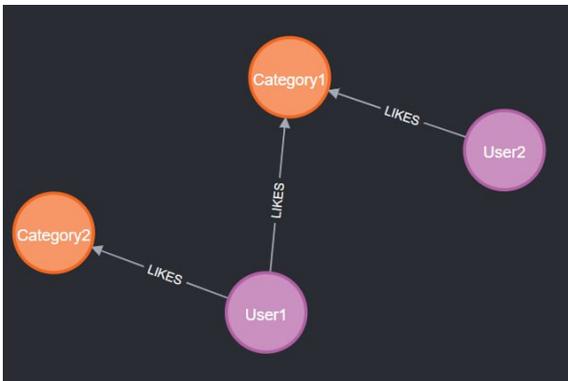


Рис. 4 – граф базы данных

Продемонстрировать работу созданной графовой модели можно на примере выполнения запросов для анализа предпочтений пользователей. Один из часто используемых запросов - поиск пользователей с общими предпочтениями. Можно найти всех пользователей, которые имеют общие предпочтения с «User1», запрос представлен на рисунке 5.

```
result = graph.run(
  "MATCH (u1:User {name: 'User1'})-[:LIKES]->(c:Category) "
  "MATCH (u2:User)-[:LIKES]->(c) "
  "WHERE u2 <> u1 " "RETURN u2.name, "
  "COLLECT(DISTINCT c.name) AS common_likes")

for record in result:
  print(f"{record['u2.name']} shares common likes with "
  "User1: {'', '.join(record['common_likes'])}")
```

Рис. 5 – нахождение всех пользователей, имеющих общие предпочтения с «User1»

Результатом выполнения данного запроса будет «User2 shares common likes with User1: Category1». Другие типы исследуемых запросов включают в себя анализ сетей, выявление лидеров мнений или создание рекомендаций на основе предпочтений.

III. Выводы

При реализации анализа предпочтений пользователей важно учесть оптимизацию запросов. Эффективные запросы позволяют быстро анализировать большие объемы данных. Для этого можно использовать кэширование и оптимизацию запросов к базе данных.

Помимо базовых запросов, можно рассмотреть применение более сложных аналитических методов, таких как выявление общих паттернов и предсказательное моделирование. Эти методы могут дополнительно улучшить результаты и эффективность анализа предпочтений пользователей.

Реализация анализа предпочтений пользователей с использованием графовых баз данных позволяет наглядно и эффективно моделировать связи между пользователями и объектами анализируемой системы. После успешной реализации и оптимизации анализа предпочтений пользователей, этот подход может быть интегрирован в реальные проекты и приложения в качестве рекомендательной системы для социальных сетей, интернет-магазинов и других сервисов. Полученные в ходе работы над исследовательским проектом результаты могут быть использованы для значительного повышения заинтересованности пользователей и увеличения доходности бизнеса, путём создания персонализированных рекомендаций.

СПИСОК ЛИТЕРАТУРЫ

1. Применение графовых баз данных [Электронный ресурс] / – Режим доступа: <https://nitrosdata.ru/2019/02/20/primenenie-grafovyh-baz-dannyh>. – Дата доступа: 11.10.2023.
2. Графовый анализ — обзор и области применения [Электронный ресурс] / – Режим доступа: <https://habr.com/ru/companies/glowbyte/articles/594221>. – Дата доступа: 13.10.2023.
3. Hybrid Recommender Systems: Survey and Experiments / Robin Burke. – User Modeling and User-Adapted Interaction 12(4), 2002.
4. Neo4j DBMS Documentation [Электронный ресурс] / – Режим доступа: <https://neo4j.com/docs/>. – Дата доступа: 12.10.2023.