



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

*«Классификация запросов в предметной области на
примере анализа запросов к базам данных»*

Студент ИУ7-72Б
(Группа)

(Подпись, дата)

ИУ7-72Б
(И. О. Фамилия)

Руководитель НИР

(Подпись, дата)

А. В. Княжев
(И. О. Фамилия)

Консультант

(Подпись, дата)

Ю. В. Строганов
(И. О. Фамилия)

2023 г.

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ	4
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	5
ВВЕДЕНИЕ	6
1 Анализ предметной области	7
1.1 Виды запросов к базам данных	7
1.1.1 Виды отношений между сущностями	7
1.1.2 Виды использования связей при запросах к БД	8
1.2 Анализ СУБД PostgreSQL и Greenplum	8
1.3 Сравнение способов составления запросов к базам данных	8
2 Конструкторский раздел	9
3 Технологический раздел	10
4 Исследовательский раздел	11
ЗАКЛЮЧЕНИЕ	12
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	14
ПРИЛОЖЕНИЕ А	15

ОПРЕДЕЛЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

—

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие сокращения и обозначения.

БД — База Данных.

СУБД — Система Управления Базами Данных.

SQL — Structured Query Language.

ORM — Object-Relational Mapping.

ВВЕДЕНИЕ

Появление в середине XX века запоминающих устройств относительно большой емкости дало возможность создания долговременно хранимых структур данных в программах. Но реализация логики хранения данных на стороне приложения сильно усложнило код и привело к увеличению стоимости разработки. В связи с этим возникла идея централизовать организацию хранения информации с использованием систем управления базами данных [1].

Для управления СУБД и манипуляции данными, хранящимися в СУБД, используются специальные языки запросов. Изучение таких языков может быть довольно сложной задачей [2]. Недостаточное знание или понимание языка запросов может приводить к ошибкам при работе с данными [3].

Целью данной научно-исследовательской работы является классификация запросов к предметной области на примере анализа запросов к базам данных. Для достижения поставленной цели необходимо решить следующие задачи:

- провести обзор существующих видов запросов в произвольной предметной области;
- провести анализ СУБД PostgreSQL и Greenplum;
- сформулировать критерии сравнения способов составления запросов к базам данных;
- провести сравнение SQL, ORM, Prolog и Datalog для выявления критерия «понятности»;
- классифицировать запросы к предметной области на примере анализа запросов к базам данных.

1 Анализ предметной области

В данной части рассматриваются актуальность задачи, виды запросов к базам данных, производятся анализ СУБД PostgreSQL и Greenplum и сравнение способов составления запросов к базам данных.

В настоящее время веб-технологии переживают стремительный рост. Различные сервисы позволили значительно повлиять на многие сферы современной жизни [4]. В связи с этим хранение и обработка большого объема информации стало актуальной задачей [5]. СУБД используются с целью централизовать работу с данными [1].

СУБД предоставляют специальные языки запросов для получения, фильтрации и изменения данных [6]. Такие языки позволяют работать с базами данных, включающих в себя множественные отношения между сущностями, делать сложные запросы на получение информации [7]. Но, на практике, наиболее популярные языки запросов, такие как SQL, в случае сложных запросов оказываются очень сложными для понимания, что усложняет разработку и может приводить к возникновению ошибок [8].

Для упрощения работы с БД, в том числе для неквалифицированных пользователей, активно ведутся разработки таких методов, как, например, методы генерации запросов к СУБД на основе естественного языка, что говорит об актуальности задачи [9–11].

1.1 Виды запросов к базам данных

Для определения видов запросов к БД необходимо рассмотреть виды отношений между сущностями и то, каким образом будут использованы эти отношения при запросах к СУБД.

1.1.1 Виды отношений между сущностями

Отношения между сущностями представляют собой способ связывания данных, обеспечивающий целостность и согласованность хранимой информации [12]. Существует три вида отношений между сущностями. Ниже приведено описание каждого из них [1].

1. **Один-к-одному.** Данный тип связи подразумевает связь единственного экземпляра сущности с единственным экземпляром другой сущности.

2. **Один-ко-многим.** Данный тип связи подразумевает связь единственного экземпляра сущности с одним или более экземплярами другой сущности.
3. **Многие-ко-многим.** Данный тип связи подразумевает связь одного или более экземпляров сущности с одним или более экземплярами другой сущности.

1.1.2 Виды использования связей при запросах к БД

Использование связи при запросах к базам данных характеризуется количеством связанных сущностей, учитываемом при построении запроса. Запрос может осуществляться как относительно одной сущности без учета связей, так и относительно нескольких связанных сущностей [13].

запросы связи!!!!

1. **Запросы одной сущности.** Данный тип запросов не учитывает какие-либо связи сущности с другими. Условия выборки и набор запрашиваемых данных задаются относительно одной сущности.
2. **Запросы нескольких сущностей.** Данный тип запросов подразумевает использование в запросе объединения двух–трех сущностей.
3. **Запросы большого количества сущностей.** Данный тип запросов подразумевает использование в запросе объединения более, чем трех сущностей. В особо редких случаях, количество сущностей в запросе может достигать до 100 [14].

1.2 Анализ СУБД PostgreSQL и Greenplum

1.3 Сравнение способов составления запросов к базам данных

2 Конструкторский раздел

3 Технологический раздел

4 Исследовательский раздел

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Стасышин В.* Проектирование информационных систем и баз данных. — Litres, 2022.
2. *Невский А. А.* ПРИМЕНЕНИЕ МЕТОДА АГРЕГИРОВАНИЯ, ГРУППИРОВКИ И ОБЪЕДИНЕНИЯ ДАННЫХ С ПОМОЩЬЮ ЯЗЫКА ЗАПРОСОВ SQL В ЦЕЛЯХ АНАЛИЗА ПОВЕДЕНИЯ ПОЛЬЗОВАТЕЛЕЙ // *Univsum: технические науки.* — 2022. — 11—1 (104). — С. 50—52.
3. *Гарскова И.* Базы данных: создание и использование // Учебно-методическая разработка к практикуму по курсу "Информатика и математика. — 2005. — № 1.
4. *Murodilov K. T., Alisherov S. M.* WEB CARTOGRAPHY AT THE CURRENT STAGE OF DEVELOPMENT OF GEOINFORMATION RESOURCES // *Galaxy International Interdisciplinary Research Journal.* — 2023. — Т. 11, № 4. — С. 166—171.
5. *Картавец И. И.* Актуальность использования NOSQL баз данных // *Colloquium-journal.* — Голопристанський міськрайонний центр зайнятості= Голопристанский районный ... 2019. — С. 62—64.
6. Testing database systems via differential query execution / J. Song [и др.] // *Proceedings of IEEE/ACM International Conference on Software Engineering (ICSE).* — 2023.
7. *Hohenstein U., Engels G.* SQL/EER—Syntax and semantics of an entity-relationship-based query language // *Information Systems.* — 1992. — Т. 17, № 3. — С. 209—242.
8. *Date C. J.* SQL and relational theory: how to write accurate SQL code. — "O'Reilly Media, Inc.", 2011.
9. CatSQL: Towards Real World Natural Language to SQL Applications / H. Fu [и др.] // *Proceedings of the VLDB Endowment.* — 2023. — Т. 16, № 6. — С. 1534—1547.
10. An SQL query generator for cross-domain human language based questions based on NLP model / B. B. Naik [и др.] // *Multimedia Tools and Applications.* — 2023. — С. 1—24.

11. SQL-PaLM: Improved Large Language Model Adaptation for Text-to-SQL / R. Sun [и др.] // arXiv preprint arXiv:2306.00739. — 2023.
12. ЧАГЛЕЙ Г. Сравнение блокчейн и базы данных. — 2023.
13. *Li J., Lee L.* Teaching data joins: A conceptual approach using SQL, Alteryx, and Tableau // AIS Educator Journal. — 2021. — Т. 16, № 1. — С. 60—104.
14. Efficient massively parallel join optimization for large queries / R. Mancini [и др.] // Proceedings of the 2022 International Conference on Management of Data. — 2022. — С. 122—135.

ПРИЛОЖЕНИЕ А