МИНОБРНАУКИ РОССИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВПО «ВГУ»)

Факультет прикладной математики, информатики и механики

Кафедра программного обеспечения и администрирования информационных систем

**Визуализация нагрузки на структуры хранения данных и адаптивного выбора оптимальных структур хранения данных на основе линейного классификатора**

Дипломная работа

по специальности 010503 – математическое обеспечение и администрирование информационных систем,

специализация 351504 – администрирование информационных систем

Допущено к защите в ГАК

|  |  |  |
| --- | --- | --- |
| Зав. Кафедрой  \_\_.\_\_.201\_ г. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись) | д. ф.-м. н., проф. Шашкин А.И. |
| Руководитель  \_\_.\_\_.201\_ г. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись) | к. т. н., доц. Селезнев К.Е. |
| Студент  \_\_.\_\_.201\_ г. | \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись) | Потапов Д.Р. |

Воронеж – 2015

Оглавление

[**1.Постановка задачи** 5](#_Toc420405031)

**Введение**

Происходит информатизация, т.е. развитие информационных технологий, обеспечивающих достижение и поддержание уровня информированности членов общества. Увеличивается количество знаний, получаемых человечеством. Несколько лет назад объем создаваемых во всем мире данных превысил 1 зеттабайт – это примерно миллиард целиком заполненных жестких дисков емкостью 1 Тбайт, и уже превышает все доступное на сегодняшний день пространство хранения. Следовательно, возникает необходимость эффективной организации их хранения и управления доступом к ним.

Существует множество различных вариантов хранения информации, каждый их которых обладает собственными характеристиками. Кроме того, существуют методы для комбинирования контейнеров (модулей хранения информации). Таким образом, возможно построение огромного числа различных контейнеров, отвечающих за хранение информации и выполняющих операции поиска, вставки и удаления данных.

В ходе анализа работы тех или иных хранилищ данных при различной нагрузке и параметрах самих контейнеров, можно сделать выводы об их эффективности. Такие выводы являются важной составляющей при выборе из всего многообразия контейнеров оптимального для каждого конкретного случая, в зависимости от нагрузки и других факторов, возникающих при работе с контейнерами.

К примеру, в базах данных существует такой объект как индекс. Индекс – объект, создаваемый с целью повышения производительности поиска данных. Ускорение работы с использованием индексов достигается в первую очередь за счёт того, что индекс имеет структуру, оптимизированную под поиск. В большинстве случаев при построении индекса используются такие структуры данных, как сбалансированные деревья (B-деревья) или хеш-таблицы [9]. Но сбалансированные деревья в данном случае имеют один большой недостаток – достаточная трудоемкость перестроения. Также существуют системы, у которых сильно ограничены ресурсы, в связи с этим в них должны использоваться наиболее подходящие структуры хранения данных. Разумеется, вывод об оптимальном контейнере хранения данных можно сделать на основе анализа работы системы. Но в таком случае встает вопрос, что будет, если возникнет ситуация, при которой данный контейнер окажется не оптимальным? Поэтому данное исследование является актуальным. Адаптивный выбор структуры хранения данных позволит системе выбирать оптимальный контейнер на статистики запросов к контейнеру данных.

# **1.Постановка задачи**

Произвести визуализацию нагрузки на структуры хранения данных и получить линейный классификатор для адаптивного выбора оптимальных структур хранения данных в зависимости от нагрузки. В связи с этим был поставлен ряд задач:

1. Разработать способ визуализации нагрузки на хранилище данных по параметру – количество операций.
2. Разработать программное обеспечение, в котором должны быть реализованы:
   1. Визуализация нагрузки на одном хранилище данных.
   2. Визуализация сравнения нескольких хранилищ данных.
   3. Алгоритм получения линейного классификатора для адаптивного выбора оптимального способа хранения данных в зависимости от нагрузки.

# **2.Анализ задачи**