

## ОТЗЫВ

официального оппонента

на диссертацию Потапова Данила Романовича

«Разработка и исследование метода и алгоритмов адаптации ассоциативного контейнера данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «теоретические основы информатики».

В связи с информатизацией общества, ускорением технического прогресса в области информационных технологий и коммуникаций остро встают вопросы обработки и хранения информации, при этом явно или неявно требуются эффективно работающие контейнеры данных (структуры данных). Существует много различных типов таких контейнеров, однако каждый из них является оптимальным только для ограниченного числа задач. При смене задачи в процессе работы системы для максимально эффективной работы системы периодически необходима смена и типа контейнера. Универсальных контейнер к сожалению не существует. Таким образом, вполне естественно возникает необходимость создания самоадаптирующегося ассоциативного контейнера данных, который меняет логику своей работы в зависимости от нагрузки при обработке и хранении информации. При изменяющейся нагрузке выбирается контейнер, оптимальный либо с точки зрения быстродействия, либо – минимальных затрат оперативной памяти. При большом количестве данных и больших нагрузках помимо основного контейнера зачастую используется такой метод оптимизации как кэширование. Таким образом, помимо смены основного контейнера, самоадаптирующийся контейнер данных может изменять такие параметры кэша как алгоритм вытеснения или размер кэша. Этим и определяется актуальность рассматриваемого исследования.

Данная диссертационная работа Потапова Д.Р. изложена на 137 страницах текста, состоит из введения, четырех глав, каждая из которых разбита на параграфы, списка литературы, содержащего 151 наименование, и заключения. Содержит 33 рисунка, и 27 таблиц.

Во введении обоснованы актуальность темы, степень разработанности темы, цель и задачи исследования, выбор объекта и предмета исследования, методология и методы исследования, основные положения, составляющие научную новизну и практическую значимость диссертации, отражены данные об апробации и публикациях, вкладе автора в них, а также положения, выносимые на защиту, и основное содержание работы.

В первой главе представлен анализ и сравнение существующих методов построения контейнеров данных «ключ-значение» и методы их оптимизации (кэширование). Рассмотрены как одномерные, так и многомерные структуры данных. В результате проведенного в главе анализа выявлены основные ограничения, которые необходимо учитывать при реализации самоадаптирующихся ассоциативных контейнеров данных, основные сферы применения различных контейнеров при реализации самоадаптирующихся ассоциативных контейнеров данных, а также проанализированы существующие методы кэширования.

Изложению и обсуждению результатов диссертационного исследования посвящены последующие три главы.

Вторая глава посвящена разработке способа задания самоадаптирующегося контейнера данных и его компонентов. Разработана его структурная модель, способ задания, нагрузки на контейнер и процесса применения нагрузки на контейнер. Кроме того, описан критерий выбора оптимального контейнера и выдвинуты некоторые предположения о зависимости оптимального размера кэша от среднеквадратичного отклонения ключей в нагрузке и соотношения скоростей основного и кэширующего контейнеров. Также описан способ задания нагрузки из смеси нормальных распределений и адаптивного кэширующего контейнера с использованием интервального статистического ряда.

В третьей главе представлены алгоритмы работы модулей самоадаптирующегося контейнера. Описана общая схема работы кэширующего модуля, используемого для исследования зависимости

оптимального размера кэша от соотношения скоростей хранилищ и среднеквадратичного отклонения ключей в нагрузке. Алгоритм самоадаптирующегося контейнера данных, использующего программный кэш, был реализован с использованием архитектурного шаблона «кэш на стороне» (cache-aside). Помимо этого, в данной главе представлен алгоритм работы модуля определения параметров сложной нагрузки, состоящей из смеси нормальных распределений, а так же алгоритм кэширующего контейнера данных с использованием интервального статистического ряда.

Четвертая глава посвящена вычислительным экспериментам. В результате экспериментов были найдены множества  $K_d, K_b, K_0$  соотношения скоростей хранилищ ( $K_0$ - для таких скоростей использование кэша не целесообразно,  $K_b$  – при таких соотношений скоростей увеличение размера кэша не требуется,  $K_d = K \setminus (K_0 \cup K_b)$  – для таких соотношений скоростей необходима проверка дополнительных условий при решении вопроса о размере и использовании кэша), размер кэша  $M_0$  и значение среднеквадратичного отклонения  $\sigma_0$ , что позволило получить зависимость размера кэша от среднеквадратичного отклонения ключей в трассе и соотношения скоростей кэша и основного хранилища. Полученные закономерности могут быть использованы для динамического изменения размера кэша в зависимости от параметров нагрузки при реализации самоадаптирующегося контейнера данных.

В качестве основных результатов можно выделить следующие:

1. Разработан метод адаптации ассоциативного контейнера данных.
2. Определены области применения одномерных и многомерных ассоциативных контейнеров данных и алгоритмов кэширования с целью использования в самоадаптирующемся контейнере данных.
3. При исследовании зависимости размера кэша от среднеквадратичного отклонения нормального распределения и соотношения скоростей хранилищ была получена формула оптимального по времени размера кэша.

4. Разработан и реализован алгоритм определения параметров сложной нагрузки на самоадаптирующиеся контейнеры данных, который позволил достигнуть 100% точности для одного кластера и 96% для нескольких кластеров.

5. Разработан и реализован алгоритм адаптивного кэширующего контейнера данных с использованием интервального статистического ряда. Проведено сравнение с другими известными алгоритмами кэширования (LRU, MRU, ARC, LIRS). В результате экспериментов выявлено, что разработанный алгоритм во многих ситуациях превосходит данные алгоритмы по проценту попаданий в кэш.

Отметить, что разработанные алгоритмы доведены до реализации на языке программирования высокого уровня, что сделало возможным экспериментальную проверку сформулированных утверждений и эффективности работы самоадаптирующегося контейнера данных и алгоритма адаптивного кэширующего контейнера данных с использованием интервального статистического ряда, который во многих ситуациях превосходит данные известных алгоритмы по проценту попаданий в кэш верхней оценки.

Работа написана ясным языком, оформлена достаточно аккуратно. Но вместе с этим имеются некоторые недочеты.

- 1) Формулировки лемм 1 – 5 лучше дать как утверждения, т.к. они получены на основании анализа результатов вычислительного эксперимента, а не аналитическим выводом. А формулировки лемм 3 и 6 даны в неудачной форме.
- 2) Допущены неточности в записи математических форму, например, в формуле (2.4.4) (13 в автореферате)  $e_i$  может быть либо BL, либо STL, тогда как имеется в виду, что это элемент объединения этих множеств  $BL \cup STL$ , а в формуле (2.10.9) вместо неравенства  $\neq$  написано  $\neq$ . Запись  $g(c) = \{o_1, \dots, o_c\}$  (формула (2.4.7) на стр. 55 и формула (2.5.2) на стр.55) говорит о том, что это функция от одного аргумента, тогда как запись  $g: N \times R \times R$  (формула (2.5.2) на стр.55)

определяет эту функцию как функцию от трех аргументов. В формуле (2.10.6) на стр. 64 упорядоченная последовательность  $(min^i, max^i)$  рассматривается как начальный сегмент последовательности  $hc_i = (min^i, max^i, c_i)$ , что знаком принадлежности  $\in$  обозначать не корректно.

- 3) Допущены путаница в обозначениях, так на стр. 55, 58 одна и та же функция обозначается разными символами ( $g$  и  $g$ ), а один символ  $A$  используется для обозначения разных понятий «алгоритм кэширования» на стр. 46 и «архитектура информационной системы» на стр. 48.
- 4) Не достаточно четко описано как определяется, что время  $t_{SAOC}$  работы контейнера не достигло минимума (стр. 57).
- 5) И наконец, желательно было более четко указать тип задач, для которых адаптирующийся контейнер данных дает выигрыш по времени работы, или по размеру используемой памяти.

В целом на основании изучения диссертации, автореферата и работ, опубликованных Потаповым Д.Р. в печати по теме диссертации, можно сделать следующие выводы:

- тема диссертационного исследования является актуальной;
- научные положения и выводы, сформулированные в диссертации, являются обоснованными и достоверными;
- сделанные выше замечания не снижают ценности проведенного исследования;
- содержание автореферата и опубликованных работ соответствуют и в полной мере отражают содержание диссертации.

На основании выше изложенного считаю, что работа Потапова Данила Романовича «Разработка и исследование метода и алгоритмов адаптации ассоциативного контейнера данных», представленную на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 – «теоретические основы информатики», выполненная под руководством доктора физико-математических, профессора Артемова Михаила Анатольевича, содержит новое решение актуальной научно-практической

