

У цій статті розглядаються алгоритми роботи кешування даних та розглянута можливість застосування методу кластеризації K-means для оптимізації кешування в Collaborative Software

Ключові слова: k-means, алгоритм, кешування, кластеризація

В этой статье рассматриваются алгоритмы работы кэширования данных и рассмотрена возможность применения метода кластеризации K-means для оптимизации кэширования в Collaborative Software

Ключевые слова: k-means, алгоритм, кэширование, кластеризация

This article represents algorithms of caching data and examined the possibility of the k-means clustering method to optimize caching in Collaborative Software

Key words: k-means algorithm, caching, clustering

ПРИМЕНЕНИЕ МЕТОДА КЛАСТЕРИЗАЦИИ K-MEANS ДЛЯ ОПТИМИЗАЦИИ КЭШИРОВАНИЯ В COLLABORATIVE SOFTWARE

З. В. Дударь

Исполняющая обязанности заведующей кафедры,
профессор*

Контактный тел.: (057) 702-10-13

В. В. Глазунов*

*Кафедра программного обеспечения ЭВМ

Контактный тел.: 097-085-11-22

E-mail: int512@gmail.com

Харьковский национальный университет
радиоэлектроники

пр. Ленина, 14, г. Харьков, Украина, 61166

1. Введение

Collaborative Software - это область программного обеспечения, созданного для поддержки взаимодействия между людьми, которые совместно работают над общими задачами. Программные системы, такие как электронная почта, календарь, текстовый чат, вики, корпоративные закладки относятся к этой категории. В условиях всемирной глобализации информатизации вопросы разработки информационных систем входящих в область Collaborative Software, занимают одно из центральных мест во всем комплексе создания эффективных систем для ведения совместной работы.

2. Постановка проблемы, связь с научными и практическими задачами

При большом количестве пользователей этого программного обеспечения возникают проблемы с доступом к ресурсам, потому что производительности персональных компьютеров и серверов может не хватать для выполнения поставленной задачи. Для решения этой проблемы применяется технология кэширования данных.

Кэш – это промежуточный буфер, который обладает быстрым доступом и содержит информацию,

которая может быть запрошена с большой вероятностью. Получение данных их кэша может происходить быстрее чем выполнение запросов к СУБД(система управления базы данных) и чтение из файловой системы. За счет него уменьшается время доступа к информации, что приводит к увеличению общей производительности компьютерной системы. Но при всем этом размер памяти для кэша ограничен, поэтому разрабатываются алгоритмы для управления информацией в кэше. Кэширование наиболее эффективно, когда оно имеет высокий процент «попадания» кэша и данные находящиеся в нем еще актуальные. Актуальные данные на момент обращения в кэш располагают более рациональному использованию объема памяти отведенной для него.

3. Анализ последних достижений и публикаций по данной проблеме

Программы для управления кэшем используют различные алгоритмы для определения какие данные можно удалить для того чтобы освободить место для более актуальной информации. В сегодняшнее время активно используют LRU, MRU, LFU и ARC алгоритмы.

Least Recently Used (LRU) алгоритм работает следующим образом: сначала вытесняется неиспользо-

ванный дольше всех элемент данных [1]. Для работы этого алгоритма требуется отслеживать, что и когда использовалось, а это может оказаться несколько накладно. Реализация этого метода требует сохранения данных о возрасте данных в кэше и за счет этого происходит определение наименее использованной информации. В такой реализации, при каждом обращении к данным КЭШа, меняется возраст всего содержимого.

Most Recently Used (MRU): главным отличием от LRU, является то, что вытесняется последний использованный элемент [2].

Алгоритм MRU наиболее эффективен в случаях, когда чем старше данные, тем больше обращений к ним происходит.

Least-Frequently Used (LFU): в этом алгоритме считается частота запросов к данным в кэше. Те данные которые наименее часто запрашиваются – удаляются из памяти кэша.

Adaptive Replacement Cache (ARC): алгоритм разработан IBM и использует LRU и LFU, что улучшает итоговый результат [3].

Применение того или иного алгоритма зависит от стратегии кэширования данных. LRU наиболее эффективен, если данные с большой вероятностью будут повторно использованы в самое ближайшее время. MRU наиболее эффективен, если информация в кэше маловероятно будет повторно использована в ближайшее время.

4. Цель статьи

Целью статьи является расширение области применения метода кластеризации данных K-Means в алгоритмах управления кэша в Collaborative Software. Это позволит установить дополнительные условия попадания и не попадания данных в кэш с целью оптимизации процесса кэширования.

5. Основной раздел

Для определения актуальности информации помещаемой в кэш предлагается использовать алгоритм кластеризации K-means, который позволяет выявлять данные, которые часто запрашиваются и редко изменяются.

Под кластерным анализом принимается задача разбиения исходных данных на поддающиеся интерпретации группы, таким образом, чтобы элементы, входящие в одну группу были максимально схожи (по какому-то заранее определенному критерию), а элементы из разных групп были максимально различными друг от друга [4].

K-means основан на минимизации суммы квадратов расстояний между каждым элементом исходных данных и центром его кластера. В данном случае число кластеров задается 3. Так как необходимо разделить данные всего на 3 группы: данные, которые редко изменяются и часто запрашиваются (попадение в кэш желательно); данные, которые часто изменяются (попадение в кэш нежелательно); остальные данные.

Исходными данными для анализа служила пример протокола работы информационной системы форум.

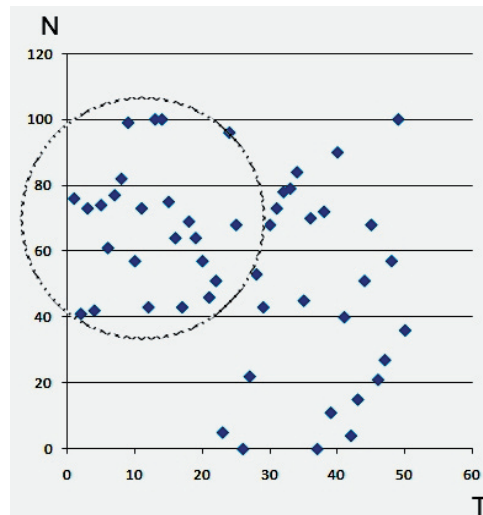


Рис. 1. Данные для анализа

Для работы алгоритма была взята следующая информация: количество запросов данных (N), количество изменений данных (T). Применение алгоритма K-means позволило выделить кластер данных, который редко изменяется и часто используется (внутри окружности). В ходе эксперимента было проведено сравнение результатов работы этого алгоритма с алгоритмами LRU, MRU, LFU.

Название алгоритма	% попадания в актуальные данные кэша
LRU	72
MRU	64
LFU	50
K-Means	67

Эксперимент показал, что K-Means алгоритм наиболее медленный из рассмотренных, но есть другой вариант реализации этого алгоритма, в целях повышения быстродействия можно ограничить число итераций [5]. Это необходимо для быстрой работы с большим количеством данных.

В результате эксперимента выявлено, что алгоритм K-means показал хорошие результаты работы по определению выборки данных, которая оптимальна для длительного хранения в кэше.

6. Вывод

Таким образом, исследована возможность применения метода кластеризации K-means для оптимизации кэширования в Collaborative Software. Этот алгоритм с ограниченным количеством итераций обеспечит быструю и эффективную работу кэширования в информационных системах входящих в область программного обеспечения Collaborative Software.

Література

1. LRU-WSR: integration of LRU and writes sequence reordering for flash memory [Електронний ресурс] / Hanyang University. – Режим доступу : \www/ URL: http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4637609 – 07.10.2008.
2. Hierarchical MRU policy for data cache [Електронний ресурс] / US Patent. – Режим доступу : \www/ URL: <http://goo.gl/EI8RB> – 19.05.2009.
3. Proceedings of FAST'08 [Електронний ресурс] / USENIX Association. – Режим доступу : \www/ URL: http://www.usenix.org/event/fast08/tech/full_papers/gill/gill.pdf – 30.05.2008.
4. Кластерний аналіз [Електронний ресурс] / Wikipedia. – Режим доступу : \www/ URL: <http://tinyurl.com/64gak9b> – 15.03.2011.
5. K-means clustering [Електронний ресурс] / Wikipedia. – Режим доступу : \www/ URL: <http://goo.gl/5r4Gs> – 15.03.2011.

УДК 004.045:621.396.96

ІНФОРМАЦІЙНЕ ЗАБЕЗПЕЧЕННЯ НА ОСНОВІ МЕРЕЖ СПОСТЕРЕЖЕННЯ ПОВІТРЯНОГО ПРОСТОРУ

О. П. Черних

Кандидат технічних наук

Центр навчання студентів на іноземній мові

Харківський національний університет радіоелектроніки

пр. Леніна, 14, м. Харків, Україна, 61166

Контактний тел.: (057) 702-14-27

E-mail: chernyh@ktute.kharkov.ua

І. І. Обод

Доктор технічних наук, професор*

Контактний тел.: (057) 702 14 29

E-mail: prof.obod@gmail.com

І. В. Свид

Старший викладач, аспірант*

Контактний тел.: (057) 702-14-29

*Кафедра «Мережі зв'язку»

Харківський національний університет радіоелектроніки

пр. Леніна, 14, м. Харків, Україна, 61166

E-mail: svid_iv@ktute.kharkov.ua

Наводяться структура та показники якості інформаційного забезпечення споживачів при розподіленій обробці інформації на рівні виявлення повітряних об'єктів у сполучених мережах систем спостереження повітряного простору

Ключові слова: інформаційне забезпечення, розподілена обробка

Приводятся структура и показатели качества информационного обеспечения потребителей при распределенной обработке информации на уровне обнаружения воздушных объектов в совместных сетях систем наблюдения воздушного пространства

Ключевые слова: информационное обеспечение, распределенная обработка

This article represents structure and quality indicators of consumers information providing in a distributed data processing at the level of air space objects detection in joint networks of airspace observing systems

Keywords: informative providing, distributed processing

1. Вступ

Підвищення надійності інформаційного забезпечення (ІЗ) користувачів системи контролю повітряного простору неможливо без використання інформаційних технологій (ІТ) у процесі отримання, збору, обробки, зберігання й розповсюдження аеронавігацій-

них даних. Дійсно, використання ІТ дозволить підвищити рівень інформаційного забезпечення, що забезпечить безпеку польотів, підвищення економічності й регулярності польотів цивільної й військової авіації в районі аеродрому, на повітряних трасах та у поза трасовому повітряному просторі. ІТ, у цій ситуації, припускають автоматизацію процесів отримання, збору,