

РАЗРАБОТКА И РЕАЛИЗАЦИЯ ПАРАЛЛЕЛЬНОГО ПОСТРОЕНИЯ SPLAYLIST И RADIXSPLINE

Бац К.А. (ИТМО)

Научный руководитель – доцент института прикладных компьютерных наук Аксенов В.Е. (ИТМО)

Введение. В условиях роста объемов данных и распространения многоядерных архитектур особое значение приобретают эффективные многопоточные индексы, то есть структуры данных, которые поддерживают вставку, удаление и поиск элементов. Существенную роль играют не только сами операции, но и скорость построения индексов на начальных данных. В этой работе мы рассматриваем две новые адаптивные структуры: SplayList [1] – конкурентная структура данных, адаптирующаяся к запросам, и RadixSpline [2] – компактный статический индекс с гарантией ошибки интерполяции. К сожалению, существующие последовательные алгоритмы их построения работают недостаточно эффективно на больших наборах данных. В этой работе мы рассказываем, как их сделать многопоточными и тем самым уменьшить время их построения на многопоточной машине.

Основная часть. В работе рассматривается задача многопоточного построения адаптивных и интерполяционных индексных структур по заранее известному набору данных. В качестве объектов исследования выбраны SplayList и RadixSpline. SplayList представляет собой поисковую структуру данных, похожую на SkipList, в которой уровень элементов определяется частотой обращений к ним, что позволяет учитывать реальное распределение запросов. RadixSpline является компактной статической индексной структурой, использующей битовые маски и аппроксимацию линейной интерполяцией для оценки позиции ключа с гарантированной погрешностью. Для SplayList разработан параллельный алгоритм построения по отсортированному массиву элементов с учетом частоты обращений, основанный на рекурсивной декомпозиции сегментов и использовании префиксных сумм. Для RadixSpline предложено два алгоритма построения: первый основан на бинарном поиске максимального отрезка допустимой интерполяции, второй – на построении выпуклой оболочки над множеством точек в отрезке интерполяции.

Заключение. В работе были предложены и реализованы параллельные алгоритмы построения структур SplayList и RadixSpline. Проведен анализ корректности и асимптотической сложности. Разработанные алгоритмы позволяют заметно сократить время построения индексов на многоядерных машинах и могут быть использованы в системах хранения и обработки больших данных.

Список использованных источников.

1. Vitaly Aksenov et al. The splay-list: A distribution-adaptive concurrent skip-list. // Distributed Computing 36.3 (2023), pp. 395–418.
2. Andreas Kipf et al. RadixSpline: a single-pass learned index. // Proceedings of the third international workshop on exploiting artificial intelligence techniques for data management. 2020, с. 1–5.