

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА НА ДИССЕРТАЦИОННУЮ РАБОТУ

Тищенко Владимира Александровича

"Методы построения многоуровневого классификатора по лексикографическому признаку применительно к ключевому уровню массива ООСУБД НИКА",
представленной на соискание учёной степени кандидата технических наук по
специальности

05.13.01 – «Системный анализ, управление и обработка информации»
(информационно-вычислительное обеспечение)

Актуальность темы диссертационной работы

Появление за последние 30 лет сложных графических пользовательских интерфейсов, а также распространение широкого спектра мобильных устройств являются естественными предпосылками для применения классификаторов в различных предметных областях для организации интерфейса с пользователем. Известным видом классификатора являются алфавитные индексы и проблемой здесь является построение таких классификаторов для больших коллекций упорядоченных по алфавиту записей, которые встречаются часто в базах данных. В этом случае встаёт вопрос об оптимизации числа переходов пользователя при поиске ключа в алфавитном классификаторе. Алфавитные классификаторы, не требующие использование поля ввода, являются наиболее актуальными при использовании мобильными устройствами.

Структура работы и ее содержание

От соискателя была получена диссертационная работа, состоящая из введения, пяти глав, заключения, списка принятых сокращений, списка литературы и четырех приложений. Общий объем диссертации составляет 207 страниц, список литературы состоит из 128 источников.

Во *введении* даётся обоснование актуальности проблемы связанной с организацией интерактивного доступа к ключевому массиву сложноструктурированной БД на основе оптимального классификатора, формулируются задачи, которые необходимо решить для достижения цели

построения многоуровневого классификатора, аргументируется научная новизна исследования, показывается теоретическая и практическая значимость полученных результатов.

В *первой главе* автор приводит постановку задачи построения оптимального классификатора по лексикографическому признаку на основе структуры цифрового поиска Сассенгата, описывается научный вклад в область интерактивных методов доступа к базам данных, а также рассматривается современное состояние дел в указанной научной области. В конце раздела соискатель формулирует значимость предлагаемого метода в сравнении с существующими методами интерактивного доступа, расширяет область применения префиксного дерева и предлагает использовать его для организации интерфейса для интерактивного доступа к ключевым массивам в виде оптимального классификатора, после чего актуализирует цель и задачи научного исследования.

Во *второй главе* Тищенко В.А. описывает метод построения классификатора на основе регрессионной зависимости средней длины префикса классификатора от максимального числа ключей в классе. Модель регрессионной зависимости сводится к регрессии на ортогональных полиномах Чебышева. Ограничением метода является область определения максимального числа ключей в классе, на которой была построена зависимость. Предварительно вводится понятие префиксного дерева и даётся обзор существующих методов, связанных с этой структурой, на основе модельных распределений показывается, что ключи распределены неравномерно по префиксам на всю длину ключа, а также аппроксимируются функции плотности длины ключа класса и числа ключей в классе на основе разложения в ряд Эджворта. Построенная регрессионная зависимость задаёт целый класс классификаторов с различными параметрами и актуальна для получения характеристик оптимального классификатора, рассматриваемого в третьей главе.

В *третьей главе* автор решает задачу оптимизации целевой функции в виде функционала общего числа операций в классификаторе с использованием префиксного дерева. Оптимизация функционала даёт параметры оптимального классификатора — число ключей в классе и число ключей в группе. Из регрессионной зависимости из главы 2 по числу ключей в классе определяется

средняя длина ключа оптимального классификатора. Такой оптимальный классификатор позволяет находить ключ в массиве за минимальное в среднем число переходов в классификаторе и даёт глубину дерева, меньшую, чем обычное префиксное дерево, но большую, чем префиксное дерево LC-trie, сжатое по уровням, поскольку классификатор содержит только часть многозначных узлов. Во второй части главы автор приводит вид функционала общего числа операций в общем случае, а также алгоритм расчёта функционала с использованием параллельных вычислений для увеличения скорости вычислений. При размещении букв по каналам обслуживания могут применяться минимаксные алгоритмы размещения объектов для уменьшения разницы средних частот букв по каналам обслуживания.

В *четвёртой главе* описана программная реализация классификатора по лексикографическому признаку в рамках гипертекстовой системы для ООСУБД НИКА, модель данных которой очень близка к модели гипертекста. Ядро системы составляют методы отображения объектов, а надстройка над ядром в виде стилевых таблиц позволяет произвести тонкую настройку методов отображения. Префиксное дерево для данного ключевого массива строится в виде отдельной БД НИКА с частотами префиксов. Оно позволяет задать классификатор с различными длинами префиксов на одном уровне в зависимости от частоты данного префикса.

В *пятой главе* представлено практическое применения многоуровневого классификатора по лексикографическому признаку на основе префиксного дерева.

В *заключении* приведены основные результаты диссертации.

В приложения соискатель вынес пример описания данных для заданного массива, перечень методов отображения и их атрибутов и фрагмент оптимального классификатора для заданного ключевого массива.

Научная новизна работы

Основная новизна работы состоит в следующем:

1. Сформулирована задача оптимизации функционала общего числа операций в классификаторе по лексикографическому признаку для организации интерактивного доступа к ключевому массиву.

2. Предложен алгоритм нахождения оптимального значения функционала общего числа операций для построения оптимального классификатора по лексикографическому признаку.

3. Предложен способ нахождения оптимальной длины ключа классификатора с использованием регрессионной зависимости средней длины ключа класса от максимального числа ключей в классе.

Степень обоснованности и достоверности результатов исследования

Все положения и выводы диссертации достоверны и научно обоснованы. Достоверность указанных соискателем результатов подтверждается экспериментальными данными, полученными в процессе использования ряда глобальных гипертекстовых систем в интернет, задача поиска текстовых ключей в которых решается методами и алгоритмами, представленными в диссертации.

Практическая значимость

Разработанные соискателем методы реализованы в виде гипертекстовой системы для ООСУБД НИКА NKWSystem, предоставляющей возможность построения классификаторов по лексикографическому признаку на основе структуры цифрового поиска Сассенгата. В настоящее время система NKWSystem уже используется в глобальных гипертекстовых системах. Благодаря гибкой надстройке в виде стилевых таблиц над ядром, состоящим из методов отображения, круг применимости гипертекстовой системы NKWSystem может быть существенно расширен.

Рекомендации по использованию результатов работы

Результаты диссертационной работы представляют большой теоретический и практический интерес. Они могут быть использованы для дальнейшего изучения и развития науки в данной области. Рекомендуется также использовать разработанную гипертекстовую систему NKWSystem в научных работах, связанных с исследованием задач организации интерфейса к ключевым массивам для поиска ключей в массивах.

Замечания

1. Вторая глава содержит подраздел, связанный с модельными распределениями, использование и назначение которых недостаточно чётко

сформулировано в работе. Применение модельных распределений является искусственным приёмом и может не отражать действительную картину. Кроме того, связь между тремя подразделами второй главы — различные виды многоуровневого классификатора, модельные распределения ключей и регрессионная зависимость длины префикса класса от числа ключей в классе — является достаточно условной.

2. В работе не приведены результаты экспериментов с алгоритмом, который бы использовал точное решение и решение в виде аппроксимации $1+\varepsilon$ задачи минимаксного размещения букв по каналам обслуживания. В третьей главе, где описывается предложенный алгоритм построения оптимального классификатора, используется только метод выделения отдалённых точек или просто последовательное разбиение алфавита на части с наиболее близкими средними частотами.

3. Средние частоты групп букв по каналам обслуживания помимо сравнения с результатом применения метода выделения отдалённых точек корректнее было бы сравнить и с результатами применения методов, дающих точное решение за субэкспоненциальное время и решение в виде аппроксимации $1+\varepsilon$ за время $O(P^n)$.

4. Во второй главе были получены характеристики классификаторов по лексикографическому признаку в виде регрессионной зависимости, а в третьей главе — характеристики оптимального классификатора, но возникает вопрос, насколько эти характеристики обеспечивают возможность использования алгоритмов на практике для решения указанных задач. В работе не решён вопрос, является ли глобальным полученный минимум функционала общего числа операций.

Перечисленные замечания не влияют на результаты работы в целом и на корректность полученных результатов.

Заключение

На основании материалов диссертации считаю, что диссертация Тищенко Владимира Александровича «Методы построения многоуровневого классификатора по лексикографическому признаку применительно к ключевому уровню массива ООСУБД НИКА» является законченным научно-квалификационным трудом и удовлетворяет требованиям Положения о порядке

присуждения учёных степеней, а её автор, В.А. Тищенко, заслуживает присуждения искомой учёной степени кандидата технических наук по специальности 05.13.01 "Системный анализ, управление и обработка информации (информационно-вычислительное обеспечение)"

Доктор технических наук,
профессор кафедры "Информатика и
программное обеспечение
вычислительных систем"
Национального исследовательского
университета "МИЭТ"



Портнов Евгений Михайлович

Подпись профессора Портнова Е.М. удостоверяю

НАЧ. ОТД. КАДРОВ
С.В. ЗАБОЛСТРИН



Адрес: 124498, г. Москва, г. Зеленоград, площадь Шокина, дом 1
Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Московский институт электронной техники»

Телефон: (499) 731-44-41

Факс: (499) 710-22-33

Телеграф: 124498, Москва, АТ 205264

Электронная почта: netadm@miee.ru

<https://miet.ru/>