

ЗАКЛЮЧЕНИЕ ДИССЕРТАЦИОННОГО СОВЕТА Д 002.073.04
на базе федерального государственного учреждения
«Федеральный исследовательский центр «Информатика и управление»
Российской академии наук» (ФИЦ ИУ РАН) по диссертации
на соискание ученой степени кандидата технических наук

аттестационное дело № _____
решение диссертационного совета от 15 марта 2021 № 11

О присуждении Тищенко Владимиру Александровичу, гражданину Российской Федерации, ученой степени кандидата технических наук.

Диссертация «Методы построения многоуровневого классификатора по лексикографическому признаку применительно к ключевому уровню массива ООСУБД НИКА» по специальности 05.13.01 «Системный анализ, управление и обработка информации (информационно-вычислительное обеспечение)» принята к защите 28 декабря 2020 г., протокол №5, диссертационным советом Д 002.073.04 на базе федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН), 119333, Москва, ул. Вавилова, д. 44, корп. 2, приказ Министерства образования и науки Российской Федерации № 747/нк от 22 июня 2016 г.

Соискатель Тищенко Владимир Александрович, 1970 года рождения, занимает должность научного сотрудника ФИЦ ИУ РАН.

В 1993г. окончил МИФИ по специальности автоматизированные системы обработки информации и управления с квалификацией инженер-системотехник. С 07.09.2018 по 06.02.2019 был прикреплен к аспирантуре ФИЦ ИУ РАН в качестве экстерна и сдал кандидатские экзамены по специальности 05.13.01 - «Системный анализ, управление и обработка информации». С 1994г. по февраль 2015г. работал в ИСА РАН. С марта 2015г. по настоящее время работает в ФИЦ ИУ РАН, с января 2017 года состоит в должности научного сотрудника Отдела 94 «Организация

банков данных» Отделения 9 «Математическое обеспечение вычислительной техники» ФИЦ ИУ РАН.

Диссертация выполнена в Отделе 94 «Организация банков данных» Отделения 9 «Математическое обеспечение вычислительной техники» ФИЦ ИУ РАН.

Научный руководитель - доктор технических наук, Соловьёв Александр Владимирович, главный научный сотрудник ФИЦ ИУ РАН.

Официальные оппоненты:

1. Портнов Евгений Михайлович, доктор технических наук, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет «Московский институт электронной техники», профессор Института системной и программной инженерии и информационных технологий Национального исследовательского университета «МИЭТ».

Отзыв оппонента положительный, имеются замечания по диссертации:

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.
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.

2. Сафонова Ирина Евгеньевна, доктор технических наук, Федеральное государственное автономное образовательное учреждение высшего образования «Российский университет транспорта», профессор кафедры вычислительные системы, сети и информационная безопасность.

Отзыв оппонента положительный, имеются следующие замечания по диссертации:

1. Основной результат диссертации опирается во многом на публикации конца 50-60-х годов, таких авторов как Фредкин, Брианде, Моррисон, Сассенгат, а современные публикации используются только для сравнения полученных результатов; ссылки на них в основном сделаны в обзоре префиксного дерева сочетаний (в подразделе 2.1), например, при сравнении с деревом LC-trie.
2. В работе до конца не формализовано математически понятие «сжатое по поддеревьям» префиксное дерево и не определяется то, как оно соотносится с деревом LC-trie;
3. С математической точки зрения является некорректным делать обобщение на основе частных примеров без формулировки теорем и их доказательств - в частности, в работе рассмотрен только отдельный пример нахождения минимума функционала общего числа операций в классификаторе на заданных диапазонах и его существование обобщается на любой классификатор.

Ведущая организация - Федеральное государственное автономное образовательное бюджетное учреждение высшего образования «Южный федеральный

университет», г. Ростов-на-Дону, в своем положительном заключении, подписанном Геннадием Анатольевичем Угольницким, доктором физико-математических наук, профессором, заведующим кафедрой прикладной математики и программирования Института математики, механики и компьютерных наук им. И.И. Воровича ФГАОУ ВО «Южный федеральный университет», указал, что «диссертация Тищенко Владимира Александровича является завершённой научно-квалификационной работой, в которой содержатся решения и разработки в области автоматизированных гипертекстовых информационно-поисковых систем в глобальных сетях, внедрение которых вносит значительный вклад в развитие отрасли информационных технологий страны. Диссертация соответствует специальности 05.13.01 — «Системный анализ, управление и обработка информации» (информационно-вычислительное обеспечение), в частности, пунктам 2, 4 и 12 перечня областей исследования». Отзыв обсужден и утверждён на заседании кафедры прикладной математики и программирования Института математики, механики и компьютерных наук им. И.И. Воровича Южного федерального университета 28 января 2021 г., протокол №1.

Замечания:

1. Временная сложность полученного оптимального классификатора по лексикографическому признаку сравнивается с различными вариантами префиксных деревьев trie только на конкретном примере.
2. Метод построения ОКЛП не даёт решения задачи минимизации функционала S_{on} в общем виде, а реализован только в виде алгоритма перебора.
3. Неясно, насколько аналитический вид функций плотности отражает реальные максимумы случайной величины, полученные опытным путём.
4. В работе недостаточно чётко показано на каком регрессионном диапазоне применима регрессионная зависимость средней длины ключа классификатора от максимального числа ключей в классе при использовании её в точках вне диапазона, на котором построена зависимость.
5. Следовало бы привести более точный график зависимости $S_{on}(n)$, а не просто схематичную форму графика.

6. Метод построения ОКЛП никак не учитывает различные виды текстовых массивов, для которых строится классификатор.

Соискатель имеет 20 опубликованных научных работ по теме диссертации, в том числе: 10 статей в изданиях, рекомендованных ВАК РФ, 10 публикаций в изданиях, входящих в российские базы цитирования, патент на изобретение, а также свидетельство на программу для электронных вычислительных машин (ЭВМ). Наиболее значительные работы:

1. Тищенко В.А. Организация интерактивного доступа к ключевому массиву на основе классификатора по лексикографическому признаку // Материалы XVIII Международной научно-практической конференции “Advances in Science and Technology”, 31 января 2019, с.108-111.

2. Емельянов Н.Е., Тищенко В.А. Методология построения многоуровневого индекса ключевого массива по лексикографическому признаку на основе метода регрессионного анализа на примере СУБД НИКА // Труды Института системного анализа Российской академии наук (ИСА РАН) РАН. Обработка информационных и графических ресурсов. Под ред. чл.-корр. РАН Арлазарова В.Л. М.: URSS. 2010. -Т. 58 - С. 6-17.

3. Емельянов Н.Е., Тищенко В.А. Представление гипертекста в СУБД НИКА // Труды Института системного анализа Российской академии наук (ИСА РАН). Под ред. чл.-корр. РАН Арлазарова В.Л. и д.т.н. проф. Емельянова Н.Е. - М., 2009. -Т. 45 - С. 17-36.

4. Чернозуб С.П., Садовский В.Н., Емельянов Н.Е., Келле В.В., Тищенко В.И., Тищенко В.А., Чернышева И.Б., Богомолова Е.А., Никонова Т.В., Сергеев Д.И., Смирнова Н.С. О создании информационной системы «Философия и методология науки в журнале „Вопросы философии“ // Системные исследования. Методологические проблемы: Ежегодник 2011--2012. Вып.36/2011—2012. Под ред. чл.-корр. Попкова Ю.С., д.филос.н Садовского В.Н., к.филос.н. Тищенко В.И. М.: URSS, 2012. С.239-247.

5. Емельянов Н.Е., Садовский В.Н., Тищенко В.А., Чернышева И.Б. Построение web-сервера для периодических изданий на материале ежегодника “Систем-

ные исследования” // Системные исследования. Методологические проблемы. Ежегодник 1997, изд-во "Эдиториал УРСС", 1997, с. 313-323 (<http://sr.isa.ac.ru/sr-97/emelyan.html>).

6. В.Н. Воробьев, Н.В. Сомин, Л.С. Аристова, Н.Е. Емельянов, А.В. Мазырин, А.В. Соловьев, Н.С. Соловьева, В.А. Тищенко, И.В. Щелкачева. Информационная система по истории христианства в России в XX веке // Отчет о НИР № 97-07-90055 (Российский фонд фундаментальных исследований).

7. Emelyanov N.E., Muhanov I.V., Tishchenko V.A. "Web server on the basis of NIKA DBMS" // proceedings of the third international workshop on "Advances in databases and information systems", ACM SIGMOD, Moscow, sep. 10-13, 1996, Vol.2, p.58-59.

8. Тищенко, В.А. Выбор оптимального алфавитного классификатора при минимизации общего числа операций // Труды Института системного анализа Российской академии наук (ИСА РАН). Под ред. чл.-корр. РАН Арлазарова В.Л. - М.: Поли Принт Сервис, 2018. -Т. 68 - №1. - С. 54-57.

На автореферат поступило три положительных отзыва, которые подписали:

1. Парамонов Николай Борисович, доктор технических наук, профессор, главный научный сотрудник Публичного акционерного общества "Институт электронных управляющих машин им. И.С. Брука". Замечания:

- Несмотря на то, что результаты работы были внедрены в реляционную объектно-ориентированную СУБД, не была исследована применимость разработанного метода для других типов СУБД, в частности нереляционных и NoSQL.

2. Шашков Олег Владимирович, кандидат физико-математических наук, доцент Департамента математики Федерального государственного образовательного бюджетного учреждения высшего образования «Финансовый университет при Правительстве Российской Федерации». Замечания:

- Не приведён пример применения функционала общего числа операций, записанного в общем виде, учитывающего переход по всем уровням много-

уровневого классификатора. В формуле функционала это учтено в первой сумме тройной суммы.

- Желательно было бы привести большее количество примеров применений существующих методов интерактивного доступа, используемых в современных поисковых системах в интернете, в словарях, википедиях, и дать более глубокую оценку того, как они соотносятся с предлагаемым методом построения оптимального классификатора.
- Не проанализированы причины того, что префиксное дерево не было востребовано до сих пор в качестве основы для построения оптимального классификатора ключевого массива.
- В работе никак не отражено как соотносится реализация префиксного дерева на основе ООСУБД НИКА с уже существующими реализациями префиксного дерева.
- Насколько применима реализация префиксного дерева на основе ООСУБД НИКА для построения классификаторов для ключевых массивов, которые являются частью других СУБД.

3. Королев Александр Владимирович, кандидат физико-математических наук, заведующий сектором прикладного программирования в отделе программирования и математических исследований Закрытого акционерного общества «Научно-технический центр «Реагент». Замечания:

- С точки зрения дальнейшего развития результатов, полученных в диссертационной работе, необходимо более глубокое исследование, связанное с определением временной сложности оптимального классификатора, т.к. в данной диссертационной работе классификатор сравнивается по сложности с подобными структурами на конкретном примере.
- Также возможно интересно получить различные типы классификаторов посредством выделения различных видов ключевых массивов, например, фамилии, должности, адреса и т.д.
- Отдельного рассмотрения требует вопрос, связанный с тем как соотносится оптимальный классификатор на основе префиксов с классификатором по

ключевым словам и возможно ли их объединение, например, в виде реализации функции автозаполнения.

Выбор официальных оппонентов и ведущей организации обосновывается их компетентностью в области кластерного анализа, цифрового поиска и системного анализа, что подтверждается направлениями их исследований и публикациями в высокорейтинговых научных журналах.

Диссертационный совет отмечает, что на основании выполненных соискателем исследований: **разработан** метод модельных распределений для анализа неравномерности распределения ключей массива по n -граммным префиксам на основе префиксного дерева; для модельных распределений получена средняя длина ключа; **получены** функции плотности распределений длины префикса класса и числа ключей в классе для алфавитного классификатора путём кумулянтного разложения в ряд Эджворта; **разработан** метод построения классификатора по лексикографическому признаку на основе регрессионной зависимости средней длины префикса алфавитного классификатора от максимального числа ключей в классе методом ортогональных полиномов Чебышева; **разработан** метод построения оптимального классификатора по лексикографическому признаку с использованием префиксного дерева на основе оптимизации функционала общего числа операций в классификаторе, что позволяет формализовать интерактивный способ доступа к ключевому массиву; **разработан** алгоритм расчёта функционала общего числа операций и его минимизации с использованием минимаксных методов размещения префиксов по группам.

Теоретическая значимость исследования заключается в том, что впервые **формализована** и **решена** задача адаптации структуры лучевого поиска Сассенгата для организации интерактивного доступа к ключевому массиву в глобальных гипертекстовых системах, впервые **предложен** метод модельных распределений для анализа неравномерности распределения ключей массива по n -граммным префиксам, впервые **представлены** в аналитическом виде характерные случайные распределения длины префикса класса и числа ключей в классе, выбранные из семейства случайных распределений, для фиксированного максимального числа

ключей в классе с использованием разложения в ряд Эджворта, впервые **предложена** модель регрессии на ортогональных полиномах для зависимости средней длины ключа от максимального числа ключей в классе для определения средней длины ключа оптимального классификатора, впервые **предложен** алгоритм построения оптимального классификатора по лексикографическому признаку на основе префиксного дерева сочетаний при минимизации функционала общего числа операций в дереве.

Практическая значимость полученных соискателем результатов исследования состоит в построении гипертекстовой системы для визуализации оптимального классификатора по лексикографическому признаку на основе префиксного дерева, реализованного в виде БД НИКА. Разработанные алгоритмы использованы в ряде глобальных гипертекстовых систем:

- Информационно-справочная система на основе электронной публикации материалов ежегодника “Системные исследования”.
- Автоматизированная система управления электронными публикациями баз данных «Философия и методология науки в журнале ”Вопросы философии”»
- Информационно-поисковая система по репрессированным за годы советской власти, содержащая на данный момент более 36 тыс. биографических справок о пострадавших

Достоверность полученных результатов исследования подтверждается согласованностью разработанных методов и алгоритмов с экспериментальными результатами и строгими математическими рассуждениями. Основные результаты работы прошли апробацию на профильных российских и международных научных конференциях и семинарах и были успешно внедрены в глобальные гипертекстовые системы.

Личный вклад соискателя состоит в:

- разработке метода построения классификатора по лексикографическому признаку на основе регрессионной зависимости средней длины ключа префикса классификатора от максимального числа ключей в классе,

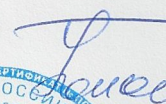
- разработке метода построения оптимального классификатора на основе минимизации функционала общего числа операций в классификаторе,
- разработке алгоритма расчёта функционала общего числа операций и его минимизации с использованием минимаксных методов размещения префиксов по группам,
- разработке гипертекстовой системы для визуализации оптимального классификатора по лексикографическому признаку на основе префиксного дерева, реализованного в виде БД НИКА.

На заседании 15 марта 2021 года диссертационный совет пришел к выводу о том, что диссертация представляет собой научно-квалификационную работу, которая соответствует критериям, установленным Положением о присуждении ученых степеней, и принял решение присудить Тищенко В.А. ученую степень кандидата технических наук.

При проведении голосования диссертационный совет в количестве 31 человека, из них 13 докторов наук по специальности защищаемой диссертации, участвовавших в заседании, из 42 человек, входящих в состав совета, проголосовали: за присуждение ученой степени 31, против присуждения ученой степени 0.

Председатель диссертационного совета Д 002.073.04

д.т.н., академик РАН

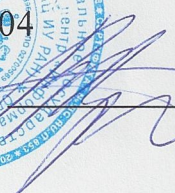


Ю.С. Попков

Ученый секретарь диссертационного совета Д 002.073.04

д.т.н., профессор





В.Н. Крутько

15 марта 2021 г.