

УТВЕРЖДАЮ:

Проректор по научной и
исследовательской деятельности
ФГАОУ ВО «Южный федеральный
университет» Доктор химических
наук

А.В. Метелица

« 29 »

2021 г.



Отзыв ведущей организации
федерального государственного автономного образовательного учреждения
высшего образования «Южный федеральный университет»
о диссертационной работе Тищенко Владимира Александровича на тему
"Методы построения многоуровневого классификатора по
лексикографическому признаку применительно к ключевому уровню массива
ООСУБД НИКА", представленной на соискание ученой степени кандидата
технических наук по специальности 05.13.01 — «Системный анализ,
управление и обработка информации» (информационно-вычислительное
обеспечение)

1. Актуальность темы диссертации

Распространение в интернет информационных систем, включающих в себя базы данных, в том числе, библиографические базы данных, различных википедий и энциклопедий, содержащих большие объёмы данных, а также существование информационно-поисковых систем таких, как google, yahoo, yandex и т.д., включающих в себя проиндексированные страницы, которые исчисляются миллиардами, приводит к проблеме, связанной с организацией интерактивного доступа к значительным объёмам данных, измеряемым петабайтами.

Коллекции текстовых ключей естественным образом разбиваются по буквам алфавита, что и является основой для создания интерфейса с пользователем в виде классификатора по лексикографическому признаку. Тривиальность построения такого классификатора становится неочевидной, если принять во внимание два фактора: неравномерность распределения

ключей по буквенным префиксам на всю длину ключа и классификатор должен предоставить пользователю возможность поиска ключа за минимальное число переходов по буквенным префиксам. В настоящее время сайты с энциклопедиями, словарями, библиографическими базами данных и используют тривиальные классификаторы — однобуквенные, двухбуквенные, диапазонные. При объёме коллекций в несколько десятков тысяч ключей встаёт проблема в организации более сложных классификаторов, т.к. тривиальные классификаторы становятся неоптимальными для доступа к таким коллекциям. Возникает задача, связанная с построением многоуровневого классификатора по лексикографическому признаку, позволяющему пользователю перейти на ключ за минимальное число переходов по префиксам в классификаторе.

Как элемент пользовательского интерфейса оптимальный классификатор по лексикографическому признаку (ОКЛП) существенно дополнит поисковое поле ввода, т.к. ОКЛП позволяет не только находить ключи, но и обнаруживает наличие или отсутствие целых классов ключей. При использовании мобильных устройств ОКЛП ещё более становится востребованным для упрощения ввода информации, предоставляя выбор алфавитных префиксов. ОКЛП актуален для построения интерфейса современных систем.

Добавление в информационную систему ОКЛП поможет сделать и интерфейс с пользователем более дружелюбным в отношении к навигации по коллекциям ключей.

Наличие огромных петабайтных массивов данных в поисковых системах в интернет, а также различных библиографических баз данных, энциклопедий и т.д. подтверждает актуальность темы диссертационного исследования.

Актуальность работы Тищенко В.А, подтверждается также соответствием работы пункту Приоритетных направлений развития науки, технологий и техники в РФ «Информационно-телекоммуникационные системы», и критической технологии в РФ «Технологии информационных, управляющих, навигационных систем» (Указ Президента РФ от 07.07.2011 г. № 899).

Работа проводилась в рамках государственного договора № 7/18 с Федеральной службой по интеллектуальной собственности «Федеральный институт промышленной собственности» (ФИПС) по созданию патента на изобретение «Способ построения оптимального многоуровневого алфавитного классификатора ключевого массива» заявка №2018109096/08(013945), в котором соискатель выступал в качестве исполнителя.

2. Структура и объем диссертации

Структура диссертации логична, дает возможность последовательно и полно исследовать предложенные в работе новые научные результаты. В работе обосновывается актуальность, новизна, теоретическая и практическая значимость диссертационного исследования, формируются цели и задачи, методологическая основа, излагаются основные положения, выносимые на защиту.

Диссертация состоит из введения, пяти глав, заключения, списка использованной литературы и четырех приложений. Ее содержание изложено на 4 страницах, проиллюстрировано 44 рисунками и 23 таблицами. Библиографический список литературы содержит 128 наименований.

3. Основные научные результаты и их значимость

I. В ходе решения научной задачи лично автором получены следующие новые научные результаты:

1. Впервые на основе регрессионного анализа и аппарата нечетких множеств получена зависимость, из которой оценивается средняя длина ключа классификатора $\bar{x}$ в зависимости от максимального числа ключей в классе на любой префикс n_{max} . Каждый ключ в классификаторе представляет собой префикс, который задает класс ключей, начинающийся на этот префикс.

Зависимость дает оценку средней длины ключа классификатора, которая определяет в среднем в префиксном дереве тот уровень, на который происходит переход из классификатора на список ключей массива. Классификатор использует комбинированную стратегию, сочетающую префиксное дерево с переходом на список ключей массива, Префиксное дерево представляет собой древовидную структуру trie без однопутевых ветвей. На основе сжатого дерева trie формируется многоуровневый классификатор с многозначным ветвлением. Входными параметрами для формирования классификатора являются максимальное число ключей в классе n_{max} и число ключей в группе n , на которые делится класс. Каждому классу сопоставлен уникальный буквенный префикс в классификаторе, средняя длина которого $\bar{x}$.

2. На основе разложения в ряд Эджворта получены функции плотности для основных характеристик классификатора (ОХК) — длины префикса класса и числа ключей в классе при заданном максимальном числе ключей в классе n_{max} .

Данные функции дают наиболее вероятные длины ключей префиксов и число ключей в классе и описываются функциями плотности, отличающимися от нормальной функции плотности. Для случайной длины ключа класса

распределение является мультимодальным, а для случайного числа ключей в классе — унимодально или не имеет характерного максимума.

3. На основе модельных распределений (МР), предложенных автором, показана неравномерность распределения ключей по префиксам на всю длину ключа.

МР считается неравномерным на заданную длину ключа, а свыше этой длины принимается равномерным. Основной вывод, который позволяет сделать применение метода МР, — это то, что в случае реальных приложений различные префиксы на любом уровне классификатора могут содержать различное случайное число ключей.

4. Впервые предложен метод построения оптимального классификатора по лексикографическому признаку (ОКЛП), который основан на формализации интерактивного доступа пользователя к классификатору в виде функционала общего числа операций в классификаторе $S_{оп}$, и, который реализуется посредством алгоритма выбора ОКЛП из класса возможных классификаторов.

Метод построения оптимального классификатора по лексикографическому признаку (ОКЛП) использует функционал $S_{оп}$ в качестве целевой функции. В результате формализации ОКЛП определяется как классификатор с минимальным значением функционала $S_{оп}$, который зависит от максимального числа ключей в классе n_{max} и числа ключей в группе n . Из регрессионной зависимости $\varpi(n_{max})$ определяется средняя длина ключа ОКЛП ϖ . ОКЛП позволяет находить искомым ключ за минимальное в среднем число переходов. Временные показатели классификатора получились лучше обычного дерева *trie*, но хуже сжатого по уровням дерева *LC-trie*, которое не применимо в виде интерфейса с пользователем в силу отсутствия многих префиксов в естественном языке.

Функционал $S_{оп}$ имеет глобальный минимум, хотя и может иметь незначительные колебания (т.е. локальные экстремумы) в зависимости от n_{max} и n . Это показывает также вид функционала для равномерного случая распределения ключей по префиксам.

5. Предложен алгоритм выбора ОКЛП из класса возможных классификаторов. Алгоритм содержит процедуру построения префиксного дерева с частотами встречаемости префиксов для исходного ключевого массива, а также содержит процедуру подсчёта значения функционала $S_{оп}$ для различных n_{max} и n посредством префиксного дерева, выборе минимального значения $S_{оп}$ на основе полного перебора значений для натуральных величин n_{max} и n на заданных диапазонах. Дополнительная процедура, отвечающая за визуализацию ОКЛП, позволяет отображать классификатор в виде

гипертекстового интерфейса с пользователем. Для оптимизации алгоритма расчёт функционала S_{on} для различных букв или двухбуквенных префиксов производится по нескольким параллельным каналам обслуживания и затем результат суммируется. Для распределения префиксов по каналам обслуживания с наиболее близкими общими средними частотами необходимо решить задачу, связанную с минимаксным размещением объектов.

Существует точное решение задачи и приближённое решение в виде аппроксимации $1+\varepsilon$, но оба класса алгоритмов являются NP-трудными. Также Гонзалесом было доказано, что все точные алгоритмы и алгоритмы с аппроксимирующим коэффициентом, меньше двух являются NP-трудными. Он предложил решение методом выделения отдалённых точек, который гарантирует удвоенное оптимальное решение за линейное время.

II. Практическая значимость полученных результатов для науки и практики подтверждена актами о принятии материалов к внедрению и использовании результатов диссертационного исследования, важнейшими из которых являются следующие:

1. Электронная публикация БД, содержащая данные о статьях материалов ежегодника “Системные исследования” в виде гипертекстовой системы в интернет. Модель БД в виде сети с выделенными иерархиями позволяет естественно применить к БД гипертекстовый интерфейс.

2. Автоматизированная система управления электронными публикациями баз данных применена для публикации БД, включающая электронные версии статей, для журнала “Вопросы философии”.

3. На основании предложенного алгоритма построения и визуализации ОКЛП разработана программа (свидетельство о государственной регистрации программ для ЭВМ №2019612352) и оформлен патент на изобретение (№ 2679967 С1 РФ). Программа, реализующая интерфейс в виде классификаторов, применяется в информационно-поисковой системе по репрессированным за годы советской власти на основе фактографической БД.

Каждая глава диссертации заканчивается подробными промежуточными выводами.

Публикации автора в количестве 20 статей, научных сообщений и тезисов докладов в материалах международных конференций, а также свидетельство о государственной регистрации программы для ЭВМ и патент на изобретение в достаточной мере отражают основное содержание диссертации, научную новизну и практическую значимость полученных результатов.

Заключение

Диссертация В.А. Тищенко является завершенной 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. По актуальности, научной новизне и значимости полученных результатов диссертация отвечает критериям Положения о присуждении ученых степеней (утв. Постановлением Правительства Российской Федерации от 24.09.2013 г. № 842), в том числе п. 9, а ее автор, Тищенко Владимир Александрович, заслуживает присуждения ученой степени кандидата технических наук по специальности 05.13.01 — «Системный анализ, управление и обработка информации» (информационно-вычислительное обеспечение).

Отзыв подготовлен профессором кафедры прикладной математики и программирования, доктором технических наук (05.13.18 - Математическое моделирование, численные методы и комплексы программ) Анатолием Борисовичем Усовым (344090, Ростов-на-Дону, ул. Мильчакова, 8А, тел. раб. +7(863)2975411, эл. почта: abusov@sfnedu.ru).

Отзыв обсужден и утвержден на заседании кафедры прикладной математики и программирования Института математики, механики и компьютерных наук им. И.И. Воровича Южного федерального университета от 28 января 2021 г., протокол №1. Присутствовало на заседании 7 чел.

Результаты голосования: «за» — 7 чел., «против» — нет, «воздержалось» — нет.

Заведующий кафедрой прикладной математики и программирования
Института математики, механики и компьютерных наук им. И.И. Воровича
ФГАОУ ВО «Южный федеральный университет»
доктор физико-математических наук,
профессор
344090, г. Ростов-на-Дону,
ул. Мильчакова, 8А,
тел. раб. +7(863)2975411
gaugolnickiy@sfnedu.ru

Геннадий Анатольевич Угольницкий

