

ОТЗЫВ

на автореферат диссертационной работы 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)"

Массивы текстовых ключей являются составной частью информационно-поисковых систем, которые широко используются в интернет. Естественным шагом при построении интерфейса с пользователем в таких системах является применение классификаторов. Автоматическая классификация и выделение категорий известны давно. Однако автоматическое построение оптимального классификатора по лексикографическому признаку является новым шагом в построении интерактивных методов поиска данных. Основой для построения такого классификатора является префиксное дерево, используемое в цифровом поиске. Аналогично "побуквенным меткам" в больших словарях, которые позволяют быстро найти страницы со словами, начинающимися с определённой буквы, можно автоматически построить классификатор в электронных изданиях баз данных. В отличие от книг в базах данных может быть значительно большие объёмы данных, что, в свою очередь, ставит задачу оптимизации доступа к таким массивам данных. В автореферате приведён краткий обзор префиксных деревьев и рассмотрена специфика применения этой структуры для построения оптимального классификатора по лексикографическому признаку. Решение данной задачи является актуальным для обзора наличия определённого класса ключей и обеспечения быстрого перехода на искомый ключ с использованием буквенных префиксов.

На основании результатов диссертационной работы, приведённых в автореферате, можно сделать вывод о том, что чётко сформулирован метод построения оптимального классификатора при минимизации общего числа операций в классификаторе и поставленная цель и задачи в диссертационной работе выполнены. Для достижения основной цели было проведено исследование на равномерность распределения ключей по буквенным префиксам посредством модельных распределений, построены функции плотности распределения для длины префикса и числа ключей с различными префиксами, построена регрессионная зависимость средней длины ключа префикса от максимального числа ключей на любой префикс. На основе префиксного дерева можно выделить целый класс классификаторов по лексикографическому признаку, характеристики которых определяются из данной регрессионной зависимости. Наконец, методом минимизации

