

ОТЗЫВ

официального оппонента на кандидатскую диссертацию Меркурьева Олега Андреевича «Эффективные строковые алгоритмы в модели потока данных», представленную к защите на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 — теоретические основы информатики.

Актуальность темы: Тема диссертации относится к области алгоритмов на строках (в английском языке для области закрепился термин *stringology*). Это активная и бурно развивающаяся область с большим количеством теоретических результатов и практических приложений. Алгоритмы на строках используются, например, в информационном поиске, анализе данных, вычислительной биологии. Ежегодно проводится несколько специализированных международных конференций, посвященных алгоритмам на строках.

Предложенные в диссертации алгоритмы относятся к потоковой модели. Поточковые алгоритмы решают задачи, в которых данные поступают последовательно и не могут быть сохранены в связи с малым по сравнению с объемом входных данных объемом доступной памяти. Такие задачи накладывают на алгоритмы ограничения по доступной памяти и времени обработки каждого элемента последовательности. Строгие ограничения на время и память часто делают невозможным точное решение исследуемой задачи. Обычно потоковые алгоритмы являются вероятностными и дают приближенное решение.

Обзор диссертации по главам. В работе строятся новые алгоритмы для вычислительных задач, связанных с поиском регулярных структур в строках, таких как палиндромы, повторы и периодические подстроки. Все построенные алгоритмы являются рандомизированными алгоритмами типа Монте-Карло, т.е. решают задачи с высокой вероятностью, используя детерминированный объем памяти за детерминированное время.

Первая глава работы является вводной, в ней определяются все необходимые понятия, описывается модель вычисления, а также дается обзор исследований по тематике диссертации.

Во второй главе рассматривается задача о поиске палиндрома наибольшей длины (LPS) в модели потока данных. Автором найдены алгоритмы типа Монте-Карло, решающие задачу LPS максимально эффективно и по времени (алгоритмы работают в реальном времени), и по памяти. Используемая память совпадает с нижними оценками, полученными соавторами, с точностью до логарифмического множителя, а для широких диапазонов значений параметров совпадает точно. Рассмотрено несколько модификаций алгоритмов (с мультипликативной и аддитивной погрешностью.)

В третьей главе рассматриваются две схожие задачи: задачи поиска повтора (LRS) и обратного повтора (LRRS). Под повтором понимается два вхождения одной и той же подстроки, а под обратным повтором — вхождение строки и ее зеркального обращения. Задача поиска наибольшей повторяющейся подстроки в классической

модели хорошо изучена; в диссертации рассматривается модель потока данных. Приводится по два приближенных алгоритма типа Монте-Карло для решения каждой из задач. Кроме того, показано, что всякий алгоритм типа Лас-Вегас, решающий LRS с заданной аддитивной погрешностью, требует как минимум линейной памяти, и, следовательно, не является потоковым.

В четвертой главе рассматривается задача поиска максимальных периодических подстрок (runs). Экспонентой строки называется отношение ее длины к длине минимального периода. Максимальная периодическая подстрока (в английском языке используется термин run) — это подстрока экспоненты не менее двух, максимальная в том смысле, что при расширении ее влево или вправо на один символ период увеличивается. Максимальные периодические подстроки хорошо изучены с комбинаторной и алгоритмической точки зрения. Примечательно, что знаменитая гипотеза Колпакова–Кучерова о том, что количество не превосходит длину строки, была недавно доказана после почти двадцати лет усилий исследователей. В диссертации сформулирована приближенная задача поиска максимальных периодических подстрок, и найден эффективный потоковый алгоритм типа Монте-Карло для ее решения.

Методы и подходы. Для построения алгоритмов используются хэши Карпа–Рабина, а также различные структуры данных, в том числе словари и суффиксные деревья. При доказательстве корректности и вычислительной сложности алгоритмов используются основы комбинаторики слов. Доказательство нижних границ на сложность алгоритмов в работе основано на принципе Яо.

Научная новизна. Все результаты диссертации являются новыми.

Замечания к оформлению диссертации. В целом, диссертация оформлена хорошо, имеются лишь небольшие замечания по оформлению:

1. Большое количество ошибок в пунктуации, в частности, нехватка запятых, осложняет чтение текста.
2. Список литературы оформлен небрежно: иногда сначала идет название статьи, а потом авторы, иногда наоборот.
3. В работе есть некоторое количество опечаток, например, принцип Яо в одном месте назван алгоритмом Яо.
4. Некоторые термины следует пояснить, например, онлайн-алгоритм, сублинейная функция (термин имеет разные значения, например, линейные функции часто считаются сублинейными).

Указанные замечания не снижают значимости полученных результатов и не влияют на общую положительную оценку диссертационного исследования О.А. Меркурьева.

Общее заключение. Основные результаты диссертации опубликованы в 4 научных статьях в рецензируемых журналах и трудах международных конференций (Algorithmica, Combinatorial pattern matching, International Symposium on String Processing and Information Retrieval).

Автореферат правильно и полно отражает содержание диссертации, а ее основные результаты являются новыми и с надлежащей полнотой опубликованными в ведущих рецензируемых журналах из списка ВАК и приравненных к ним.

Содержание диссертации соответствует специальности, по которой она рекомендуется к защите. Диссертация Меркурьева Олега Андреевича соответствует требованиям п. 9 «Положения о присуждении ученых степеней в УрФУ» к диссертациям, представленным на соискание ученой степени кандидата физико-математических наук по специальности 05.13.17 — теоретические основы информатики.

Автор диссертации Меркурьев Олег Андреевич заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.17 — «Теоретические основы информатики».

Официальный оппонент, кандидат физико-математических наук по специальности 01.01.09 — «Дискретная математика и математическая кибернетика», доцент факультета математики и компьютерных наук ФГБОУ «Санкт-Петербургский государственный университет», 199034, Санкт-Петербург, Университетская наб., д. 7–9.

Контактные данные:

Электронная почта: s.puzynina@gmail.com.

Телефон: +79522292801.

Пузынина Светлана Александровна

3011 01.07.2020



С. Пузынина

С. Пузынина / О.А. Меркурьев