

# Отчет Стариковской Т.А. за 2013 г.

## 1. Результаты, полученные в этом году:

- **Задача о вычислении  $k$ -ого суффикса подслова данного слова.**  
Лексикографический порядок на словах — это естественное обобщение порядка на буквах алфавита. Рассмотрим слово  $w$  длины  $n$ . Суффиксами  $w$  называются его подслова, заканчивающиеся в позиции  $n$  (например, суффиксами слова  $abc$  являются его подслова  $abc$ ,  $bc$  и  $c$ ). Для многих задач обработки слов очень важно уметь по слову быстро находить его  $k$ -й в лексикографическом порядке суффикс. Стандартным подходом к решению этой задачи является построение так называемого суффиксного дерева слова  $w$ , определяющего порядок на всех суффиксах слова, что занимает  $O(n)$  времени. Теперь предположим, что рассматриваемое слово  $w$  является подсловом другого слова  $u$ , которое дано заранее. В этом случае ту же задачу можно решить намного быстрее, если преобразовать слово  $u$  и хранить информацию о нем в некоторой структуре данных. Было рассмотрено два частных случая ( $k=1$  и  $k=n$ , поиск минимального и максимального суффиксов  $w$ ) и случай произвольного  $k$ . Для каждого из случаев предложена структура данных с линейной (от длины  $u$ ) памятью. Для  $k=1$  и  $k=n$  время запроса сублогарифмическое, для произвольного  $k$  время запроса чуть больше.
- **Задача о наибольшем общем подслове.**  
Пусть даны  $m$  слов  $w_1, w_2, \dots, w_m$  и число  $1 \leq d \leq m$ . Требуется найти максимальное по длине слово, входящее в хотя бы  $d$  из слов  $w_1, w_2, \dots, w_m$ . Как было показано Лукасом Чи Квонг Хьюи, эта задача может быть решена за линейное от суммарной длины слов  $w_1, w_2, \dots, w_m$  время с использованием линейной памяти. Были предложены новые решения для случая  $m=d=2$  и общего случая, использующие сублинейную память. Время работы алгоритмов больше линейного, но не существенно.
- **Задачи о максимальных и минимальных общих подсловах.** Предположим, что даны  $m$  слов  $w_1, w_2, \dots, w_m$ . Задача о максимальных общих подсловах заключается в том, чтобы по слову  $P$  и числу  $1 \leq d \leq m$  найти все максимальные слова, начинающиеся с  $P$  и встречающиеся в хотя бы  $d$  словах из  $w_1, w_2, \dots, w_m$ . Задача о минимальных общих подсловах формулируется аналогично: требуется найти все минимальные слова, начинающиеся с  $P$  и встречающиеся в не более чем  $d$  словах из  $w_1, w_2, \dots, w_m$ . Для обеих задач были предложены структуры данных с линейной памятью и оптимальным временем обработки запроса.

## 2. Опубликованные и поданные в печать работы:

- Minimal Discriminating Words Problem Revisited. Pawel Gawrychowski, Gregory Kucherov, Yakov Nekrich and Tatiana Starikovskaya. Lecture Notes in Computer Science, 2013, Volume 8214, 129-140. Proceedings of the 20th Symposium on String Processing and Information Retrieval.
- Cross-document Pattern Matching. Tsvi Kopelowitz, Gregory Kucherov, Yakov Nekrich and Tatiana Starikovskaya. Journal of Discrete Algorithms (2013), <http://dx.doi.org/10.1016/j.jda.2013.05.002>.
- On Minimal and Maximal Suffixes of a Substring. Maxim Babenko, Ignat Kolesnichenko, and Tatiana Starikovskaya. Lecture Notes in Computer Science, 2013, Volume 7922, 28-37. Proceedings of the 24th Symposium on Combinatorial Pattern Matching.
- Time-Space Trade-Offs for the Longest Common Substring Problem. Tatiana Starikovskaya and Hjalte Wedel Vildhøj. Lecture Notes in Computer Science, 2013, Volume 7922, 223-234. Proceedings of the 24th Symposium on Combinatorial Pattern Matching.

### **3. Участие в конференциях и школах:**

- 20<sup>th</sup> String Processing and Information Retrieval Symposium.
- Microsoft Summer School on Algorithms on Massive Data.
- 24<sup>th</sup> Annual Symposium on Combinatorial Pattern Matching.

### **4. Педагогическая деятельность:**

С сентября 2013 года я являюсь старшим преподавателем отделения прикладной математики и информатики НИУ ВШЭ. Руководжу курсовыми работами студентов второго, третьего и пятого курсов.