

Секция «Математика и механика»

О сложности поиска подслова в слове

Перпер Евгений Михайлович

Студент

Московский государственный университет имени М.В. Ломоносова,

Механико-математический факультет, Москва, Россия

E-mail: e\_m\_perper@mail.ru

При пользовании словарями может возникнуть ситуация, когда слово содержится в словаре не в той форме, в которой его ищут. Например, если это глагол, то в словаре он будет содержаться, скорее всего, лишь в неопределённой форме. Если поиск в словаре осуществляет компьютер, это может оказаться для него достаточно большой проблемой. В связи с этим возникает задача нахождения всех слов в словаре, содержащих в себе заданное подслово, не осуществляя перебора всего словаря.

В данной работе позиция в слове, с которой должно начинаться нужное нам подслово, задаётся в запросе вместе с подсловом (в [1] она считалась произвольной). Это может позволить сократить время поиска, если, например, точно известно, что заданное подслово должно находиться в начале слов из словаря. В качестве базы данных будем рассматривать структуру, содержащую все слова из словаря, но устроенную иначе (для ускорения поиска). Запросом к базе данных будем считать следующую пару: подслово и позицию в слове. Нам нужно, чтобы ответом базы данных на этот запрос были те и только те слова, которые содержат данное подслово, начиная с данной позиции. Например, слово "разработал" удовлетворяет запросу ("работа" ,4). При обработке запроса алгоритму разрешено вычислять только достаточно простые функции: сравнивать два целых числа, проверять, совпадают ли две буквы, находить букву по её позиции в слове. Объём памяти, занимаемой базой данных, вычисляется как количество ссылок между элементами базы данных [1]. Наименьшее среди всех алгоритмов, требующих не более  $q$  памяти, время обработки запроса в худшем случае обозначим через  $\hat{T}(q)$ .

**Теорема** Пусть в словаре  $p$  слов, все они имеют длину  $n$ , в языке  $k$  различных букв и  $d$  — наибольшее по всем возможным запросам количество слов из словаря, удовлетворяющих запросу. Тогда

$$n + d - 1 \leq \hat{T}(q) \leq \begin{cases} pn, & \text{если } q \geq p \\ 3 + 4n + 2d, & \text{если } q \gtrsim pnk \\ 1 + 2n + 2d, & \text{если } q \gtrsim pn^2k/2 \\ 2 + n + d, & \text{если } q \gtrsim pn^3k/3 \end{cases}$$

при  $p \rightarrow \infty$ .

Литература

1. Перпер Е.М. О функциональной сложности поиска подстроки// Интеллектуальные системы. 2011. Т. 15, вып. 1-4. С. 553-570.

Слова благодарности

Автор выражает благодарность профессору Э.Э.Гасанову за научное руководство.