Анализ Алгоритмов

Лабораторная работа №7

По теме “*Методы поиска подстроки в строке*”

Студент: Юмаев Артур

Группа: ИУ7-55

Оглавление

[Введение 3](#_Toc27488360)

[1. Аналитическая часть 4](#_Toc27488361)

[Обзор алгоритма Кнута-Мориса-Прата 4](#_Toc27488362)

[Обзор алгоритма Бойера-Мура 5](#_Toc27488363)

[Вывод 5](#_Toc27488364)

[2. Конструкторская часть 6](#_Toc27488365)

[Вывод 6](#_Toc27488366)

[3. Технологическая часть 7](#_Toc27488367)

[Вывод 7](#_Toc27488368)

[4. Исследовательская часть 8](#_Toc27488369)

[Вывод 9](#_Toc27488370)

[Заключение 10](#_Toc27488371)

[Литература 11](#_Toc27488372)

# Введение

Поиск подстроки в строке — одна из простейших задач поиска информации. Применяется в виде встроенной функции в текстовых редакторах, СУБД, поисковых машинах, языках программирования и т. п. В задачах поиска традиционно принято обозначать шаблон поиска как needle (с англ. — «иголка»), а строку, в которой ведётся поиск — как haystack (с англ. — «стог сена»). Обычно через Σ обозначается алфавит, на котором проводится поиск.

# 1. Аналитическая часть

В данном разделе будeт дано полное описание алгоритма Кнута-Морриса-Прата и алгоритма Бойера-Мура и их математические описания.

## Обзор алгоритма Кнута-Мориса-Прата

В задачах поиска информации одной из важнейших задач является поиск точно заданной подстроки в строке. Примитивный алгоритм поиска подстроки в строке основан на переборе всех подстрок, длина которых равна длине шаблона поиска, и посимвольном сравнении таких подстрок с шаблоном поиска.

Алгоритм Кнута-Морриса-Пратта является одним из первых алгоритмов с линейной оценкой в худшем случае.

Обозначим через:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

где означает длину слова *word*.

Префикс-функция строки – это длина наибольшего префикса строки , который не совпадает с этой строкой и одновременно является ее суффиксом. Проще говоря, это длина наиболее длинного начала строки, являющегося также и ее концом. Для строки *S* удобно представлять префикс функцию в виде вектора длиной . Можно рассматривать префикс-функцию длины , положив . Пример приведен на таблице 1.

Таблица 1

Пример префикс функции для строки «abcdabca»

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | a | b | c | d | a | b | c | a |
|  | 0 | 0 | 0 | 0 | 1 | 2 | 3 | 1 |

Наиболее полный алгоритм вычисления префикс-функции на псевдокоде показан на листинге 1.

Листинг 1. Псевдокод функции поиска префикса

|  |  |
| --- | --- |
|  | **алг** префикс\_функция (**арг** строка )  // Предположим, что  **Если** то    **Иначе**  **Если** , то  .  **Иначе**  Положить и перейти к шагу 3. |

Рассмотрим алгоритм Кнута-Морриса-Пратта. Пусть – образец, – строка, в которой ведется поиск. Рассмотрим сравнение строк на позиции *i*, то есть образец сопоставляется с частью строки . Предположим, первое несовпадение произошло между символами и , где . Обозначим . При сдвиге можно ожидать, что префикс *S* сойдется с каким-либо суффиксом строки *P*. Поскольку длина наиболее длинного префикса, являющегося одновременно суффиксом, есть префикс-функция от строки *S* для индекса *j*, приходим к следующему алгоритму.

1. Построить префикс-функцию образца *S*, обозначим ее *F*.
2. Положить .
3. Сравнить символы . Если символы равны, увеличить *k* на 1. Если при этом *k* стало равно длине образца, то вхождение образца *S* в строку *T* найдено, индекс вхождения равен  
   . Алгоритм завершается. Пока , присвоим и переходим в начало шага 3.
4. Пока , увеличиваем на 1 и переходим в шаг 3.

## Обзор алгоритма Бойера-Мура

## Вывод

В данном разделе было приведено общее и математическое описание муравьиного алгоритма, приведены примеры.

# 2. Конструкторская часть

В данном разделе будет приведена схема для муравьиного алгоритма (рисунок 1).

Рис. 1 – Схема муравьиного алгоритма

## Вывод

В данном разделе была приведена схема муравьиного алгоритма на псевдокоде.

# 3. Технологическая часть

В качестве языка программирования был выбран Python [3], так как имеется большой опыт работы с ним. На листингах 2-3 будут приведены алгоритмы Кнута-Мориса-Прата и Бойера-Мура поиска подстроки в строке.

Листинг 2. Алгоритм Кнута-Мориса-Прата

|  |  |
| --- | --- |
|  | # Алгоритм поиска префикса  **def** prefix(s):  v = [0] \* len(s)  **for** i in **range**(1, len(s)):  k = v[i - 1]  **while** k > 0 **and** s[k] != s[i]:  k = v[k - 1]  **if** s[k] == s[i]:  k = k + 1  v[i] = k  **return** v  # Алгоритм Кнута-Мориса-Прата  **def** kmp(s, t):  index = -1  f = prefix(s)  k = 0  **for** i in **range**(len(t)):  **while** k > 0 **and** s[k] != t[i]:  k = f[k - 1]  if s[k] == t[i]:  k = k + 1  if k == len(s):  index = i - len(s) + 1  break  **return** index |

## Вывод

В данном разделе была приведена реализация муравьиного алгоритма на языке программирования Python.

# 4. Исследовательская часть

Замеры времени проводились на 64-битной операционной системе Windows 10 и на x64 процессоре Inter Core i7 с 4 гб оперативной памяти для диапозона вершин графа от 1 до 10, так как на графе в 11 вершин мощности аппаратного обеспечения не хватало для решения задачи полным перебором. На момент замера времени работало в среднем 76 активных процессов.

Уже на графе с 9 вершинами видно, что решение полным перебором является неэффективным, несмотря на то, что в данном случае на количестве вершин < 9 полный перебор является более эффективным, так как меньше затраты алгоритма на память и операции.

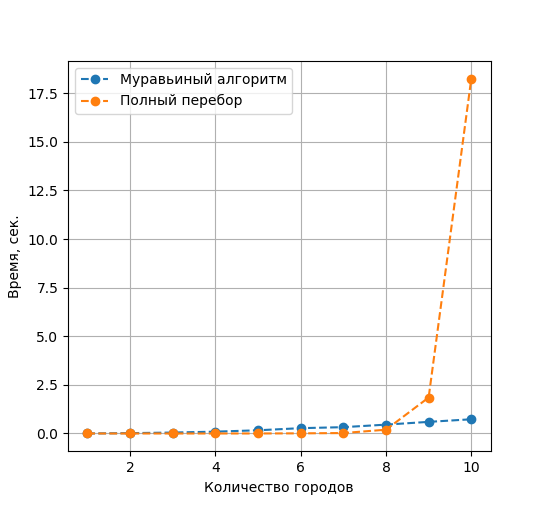


Рисунок 1. Сравнение времени работы муравьиного алгоритма с полным перебором в зависимости от количества вершин графа

Замер времени проводился с помощью стандартной библиотеки time в Python 3.7 и метода process\_time().

В качестве параметризации были такие параметры, как скорость испарения феромонов и параметр , отвечающий за то, насколько муравьи чувствительны к феромонам. Исследование проводилось на предмет сходимости алгоритма, выражающееся в количестве операций, необходимых для нахождения оптимального пути обхода всех вершин графа. Исследование проводилось на графе в 9 вершин с 9 муравьями.

Таблица 1

Параметризация по параметрам “скорость испарения” и

|  |  |  |
| --- | --- | --- |
| Скорость испарения |  | Итераций до сходимости |
| 0.0 | 0.0 | 0 |
| 0.0 | 0.5 | 5 |
| 0.0 | 1.0 | 26 |
| 0.25 | 0.0 | 0 |
| 0.25 | 0.5 | 5 |
| 0.25 | 1.0 | 5 |
| 0.75 | 0.0 | 0 |
| 0.75 | 0.5 | 4 |
| 0.75 | 1.0 | 1 |
| 1.0 | 0.0 | 0 |
| 1.0 | 0.5 | 1 |
| 1.0 | 1.0 | 1 |

Как видно из таблице, чем меньше скорость испарения, тем в среднем быстрее муравьи находят самый короткий путь. Если муравьи сильно восприимчивы к феромонам и скорость испарения не высока, они в среднем в 9 раз быстрее находят оптимальный путь.

## Вывод

В данном разделе был проведен анализ задачи коммивояжера при решении разными способами. Опыт выявил, что при небольшом количестве городов () быстрее работает метод полного перебора, но при количестве городов больше 9, муравьиный алгоритм начинает выигрывать по времени. Также опыт показал, что при большом количестве городов (> 100), алгоритм не гарантирует нахождение наикратчайшего пути.

# Заключение

В результате выполнения данной работы были рассмотрены способы решения задачи коммивояжера 2 способами: с помощью полного перебора и с помощью муравьиного алгоритма. Результаты опытов приведены в разделе 4. В целом для количества городов ( рекомендуется использовать перебор, иначе муравьиный алгоритм.

# Литература

[1] A. Colorni, M. Dorigo et V. Maniezzo, Distributed Optimization by Ant Colonies, actes de la première conférence européenne sur la vie artificielle, Paris, France, Elsevier Publishing, 134—142, 1991.

[2] M. Dorigo, Optimization, Learning and Natural Algorithms, PhD thesis, Politecnico di Milano, Italie, 1992.