

Жадные алгоритмы: кодирование Хаффмана

Александр Куликов

Онлайн-курс «Алгоритмы: теория и практика. Методы»
<http://stepic.org/217>

Computer
Science
Center

всем привет продолжаем голову жадный алгоритм а сегодня у нас называется кодирование ахаха на мы будем с вами сегодня пытаться строить оптимальной бинарный код для сжатия данных строки более формально на на вход до на строка с этим каждому ее символов

Сжатие данных

Вход: строка s .

Выход: бинарный код символов строки s ,
обеспечивающий кратчайшее представление s .

Пример

$s = \text{abacabad}$

коды символов: a: 00, b: 01, c: 10, d: 11

закодированная строка: 0001001000010011 (16 битов)

a b a

Computer
Science
Center

присвоить бинарный код так чтобы после того как каждый символ мы заменили на его бинарный код получилось как можно более короткое это не единство и каша способ проделать строка но 1 из самых простых и уже с ним полезно разобраться как же мы ну например рассмотрим строчку а б ца бд тогда 1 из способов это просто каждому символу присвоить свой код состоящий из 2 битов например а закодировать как 0 0 b как 0 1 1 1 1 наша строка содержала 8 символов и от каждой строке это 2 бита поэтому закодирована строка будет содержать 16 битов более того что совсем приятно это то что распланировать я совсем очевидно такую страну мы просто идем читаем очередной 2 бита вот прочитали 0 0 что это а прочитали 0 1 следующие 2 символа это вы прочитали 0 1 прочитали 0 0 до 5 прошу прощения

а ну и так далее когда вот так называемой постоянно длиной почти у всех символов для на совсем просто мечтаем блоки 71
 другой страны примеру видно что буква а входит гораздо чаще чем например буква д 3 буквы а входит 4 раза буквы d входит всего 1 раз поэтому возникают это естественно и деле попытаться таким частым символом нашем примере буквы буквы а присвоить более

Коды переменной длины

- Естественная идея: присвоить более короткие коды более частым символам.



Computer
Science
Center

короткий отдых за счет того что более редким символом присвоить более длинной года например можно было бы сделать следующим образом букву а закодировать кодом 0 букву б закодировать кодом 1 0 ц 1 одинаково для 1 1 1

Коды переменной длины

- Естественная идея: присвоить более короткие коды более частым символам.
- $s = \text{abacabad}$
 коды символов: a: 0, b: 10, c: 110, d: 111
 закодированная строка: ~~011011001100111~~ (14 битов)

ab

Computer
Science
Center

чтобы тогда получены суд а мы записали на а мы заменили на 0 на общем получилась такая строка понятно как она получилась она содержит 14 битов 16 битов учатся более оптимальный вопрос только действительно ли так вот действительно ли можем раскодировать в данном конкретном случае да давайте разбираться как же это произойдет вот слева направо и пытаемся понять что же нам дали за симку сначала прочитали 0 в этот момент не понимаю что так вот символ а просто потому что никакой другой никакой другой код на 0 не начинайте так напишем о

и понимает что вы уже прочитали следующее следующей
 это единица единица мы пока не понимаем мы видим на примере что так вот
 либо букву б либо цели батыров ночей конкретно еще не понимаем дальше
 прочитали 0 и в этот то момент мы понимаем что букву б потому что все
 остальные
 надя 0 начинается этот б и мы прочитали но и так далее а теперь давайте
 сформулируем что же нам помогло в данной ситуации под однозначно это
 самогоном вот такого свойства код который здесь приведен является без
 префиксом это означает что ни 1 код символа не является префиксом кода другого
 символа это означает что когда мы будем читать хода мы всегда однозначно будет

Коды переменной длины

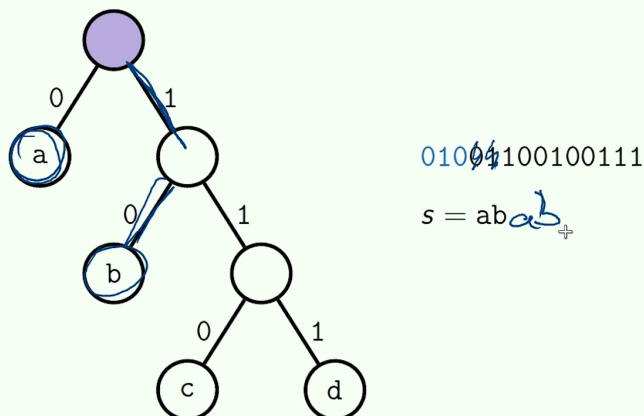
- Естественная идея: присвоить более короткие коды более частым символам.
- $s = abacabad$
 коды символов: a: 0, b: 10, c: 110, d: 111
 закодированная строка: 01001100100111 (14 битов)
- Код называется **беспрефиксным**, если никакой код символа не является префиксом кода другого символа.

Computer
Science
Center

понимать что же у нас за символ надеюсь по примеру уже было немножко
 понятно но сейчас будет еще 1 пример и так давайте все ту же строчку осмотром
 и рассмотрим ваш преданный кот оформленного беспечный кот оформлена
 видела такого дерева

Декодирование на примере

Пример



Computer
Science
Center

дерево задает под следующим образом дереву настроили двоичная это означает
 что каждая вершина либо 0 детей их либо ровно 2 ребенка сотряс на каждый раз
 на дом и земля в ребенком представляя что это соответствует нелегко он

оправдывает единицу от например
буквы а здесь соответствует код 0 а сколько тогда мы идем от корня буква
соответствует код 1 1 0 это то что мы читаем на пути от корня предположим теперь
что именно так заданном кот
он конечно же будет без префиксом поскольку совершенно у нас расположена в
личных и мы не может случиться так что на пути к 1 листу мы пройдем через
другой лист правильно это как раз и означает что
отметим кот не является префиксом другого тогда декодирование может быть
устроена следующим образом мы считаем нашу битву строчку изначально мы
находимся в корень и начинаем считать строчку прочитали 0 видим что при член
есть
после этого мы просто вы 8 symbol of и возвращаемся в корень и продолжаем
считать прочитали единицу серединами соответствующе вершину это не лист в
этом продолжаем читать читаем дальше прочитали б вы прочитали 0
означает что можно напечатать б и вернуться в корень на продолжая таким же
образом мы видим следующей субботы дальше мы спустимся в 0 80 а прочитав
этот 0 напиши мне о
прочитаем единица окажемся в этой вершине больше прочитай 0 а также в
совершении баре можно будет написать букву б но и так далее а что такое
представление куда удобно и для кодирования и для декорирования