# Modulo

*Из рассылки softtalk #27093 Jun 5, 2000*

Из-за проблем с почтой только сейчас увидел это обсуждение:

> взял слово, перекодировал его из 1251 в
> КОИ, то, что получилось опять перекодировал
> из 1251 в КОИ ... В итоге получил исходное слово :-)

Никакой мистики здесь, конечно, нет - обычный результат при вычислениях по модулю.

И хотя тема относится скорее к Абстрактной Алгебре, нежели софту, ввиду распространенности и важности вычислений по модулю, решил разъяснить происходящее чуть подробнее.

В реальной жизни нас окружает множество устройств и явлений, связанных с операциями по некоторому модулю - практически всегда, если это связано с движением по окружности или измерением времени.

В школе мы привыкаем, что все вычисления проводятся над числами, представленными в виде отметок на бесконечной числовой оси.

Часто, однако, удобно использовать прямо противоположное представление, когда вместо бесконечной прямой, отметки располагаются на окружности, так что первое и последнее числа шкалы совпадают.

Простейший и привычный прибор для таких вычислений - циферблат часов.

Так, например, секунды и минуты отсчитываются по модулю 60, и отметка 60 одновременно является и отметкой 0.

Часы отсчитываются по модулю 12 или 24.

Другой распространенный прибор для расчетов по модулю - календарь.

Следуя восходящей к древнему Шумеру традиции, мы считаем дни недели по модулю 7 (соответственно фазам Луны).

Другие модули, используемые в современном Грегорианском календаре, это 3 для кварталов, 12 - для месяцев и 365/366 для отсчета лет.

В таблице ASCII используется 256 различных символов и, таким образом, различные конвертеры из одной кодировки в другую используют некоторые модули, не превышающие этого значения.

Важно отметить, что при любом обратимом перекодировании текста, каждый символ переходит либо сам в себя либо в другой символ того же самого 256-символьного

алфавита, но не во что-то иное, скажем число PI = 3.1415926...

В таких случаях говорят, что результаты перекодировок образуют группу, а сама перекодировка является групповой операцией.

Изучение общих свойств таких преобразований выделилось в раздел математики, называемый Теорией Групп.

Не вдаваясь далеко в теорию, посмотрим, что происходит при смене кодировки русского текста.

Russian Encoding Win-1251, в числе много другого, полностью заимствован с Macintosh.

Буквам русского алфавита соответствуют следующие позиции (символы) расширенной таблицы ASCII:

Для кодировки Win-1251:

```
000000C0: C0 C1 C2 C3 C4 C5 C6 C7 ¦ C8 C9 CA CB CC CD CE CF
АБВГДЕЖЗИЙКЛМНОП
000000D0: D0 D1 D2 D3 D4 D5 D6 D7 ¦ D8 D9 DA DB DC DD DE DF
РСТУФХЦЧШЩЪЫЬЭЮЯ
000000E0: E0 E1 E2 E3 E4 E5 E6 E7 ¦ E8 E9 EA EB EC ED EE EF
абвгдежзийклмноп
000000F0: F0 F1 F2 F3 F4 F5 F6 F7 ¦ F8 F9 FA FB FC FD FE FF
рстуфхцчшщъыьэюя
```

Для кодировки KOI-8:

```
000000A0: A0 A1 A2 A3 A4 A5 A6 A7 ¦ A8 A9 AA AB AC AD AE AF
-¦+ё++++++++++¦
000000B0: B0 B1 B2 B3 B4 B5 B6 B7 ¦ B8 B9 BA BB BC BD BE BF
¦¦¦Ё¦¦--------+++№
000000C0: C0 C1 C2 C3 C4 C5 C6 C7 ¦ C8 C9 CA CB CC CD CE CF
юабцдефгхийклмно
000000D0: D0 D1 D2 D3 D4 D5 D6 D7 ¦ D8 D9 DA DB DC DD DE DF
пярстужвьызшэщчъ
000000E0: E0 E1 E2 E3 E4 E5 E6 E7 ¦ E8 E9 EA EB EC ED EE EF
ЮАБЦДЕФГХИЙКЛМНО
000000F0: F0 F1 F2 F3 F4 F5 F6 F7 ¦ F8 F9 FA FB FC FD FE FF
ПЯРСТУЖВЬЫЗШЭЩЧЪ
```

Легко видеть, что буквы ё (A3) и Ё (B3) не содержатся в Win-1251 и слова с этими буквами не могут быть правильно конвертированы, поэтому в дальнейших рассуждениях мы будем использовать только усеченный русский алфавит из 32 букв.

Для кодировки Win-1251:

```
000000C0: C0 C1 C2 C3 C4 C5 C6 C7 ¦ C8 C9 CA CB CC CD CE CF
АБВГДЕЖЗИЙКЛМНОП
000000D0: D0 D1 D2 D3 D4 D5 D6 D7 ¦ D8 D9 DA DB DC DD DE DF
РСТУФХЦЧШЩЪЫЬЭЮЯ
```

```
000000E0: E0 E1 E2 E3 E4 E5 E6 E7 ¦ E8 E9 EA EB EC ED EE EF
абвгдежзийклмноп
000000F0: F0 F1 F2 F3 F4 F5 F6 F7 ¦ F8 F9 FA FB FC FD FE FF
рстуфхцчшщъыьэюя
```

Для кодировки KOI-8:

```
000000C0: C0 C1 C2 C3 C4 C5 C6 C7 ¦ C8 C9 CA CB CC CD CE CF
юабцдефгхийклмно
000000D0: D0 D1 D2 D3 D4 D5 D6 D7 ¦ D8 D9 DA DB DC DD DE DF
пярстужвьызшэщчъ
000000E0: E0 E1 E2 E3 E4 E5 E6 E7 ¦ E8 E9 EA EB EC ED EE EF
ЮАБЦДЕФГХИЙКЛМНО
000000F0: F0 F1 F2 F3 F4 F5 F6 F7 ¦ F8 F9 FA FB FC FD FE FF
ПЯРСТУЖВЬЫЗШЭЩЧЪ
```

В отличие от совершенно произвольной кодировки Win-1251, KOI-8 устроена очень продуманно, так, чтобы при потере старшего бита (обычная история при пересылке почты на Unix-машины), русские буквы перешли в соответствующие им английские (транслитерация) и текст все же остался читаемым. Этим объясняется неалфавитный порядок букв в таблице KOI-8.

Итак, мы видим, что преобразование текста Win-1251 -> KOI-8 затрагивает только 64 символа (или 32, если пренебречь регистром) и, таким образом, необходимо использовать арифметику по модулю 64.

Часто, окружность связанную с вычислениями по модулю представляют как окружность единичного радиуса на комплексной плоскости, а числовые отметки на ней (числа от 0 до N-1, где N - модуль), трактуют как комплексные корни степени N из единицы.

Мы не будем использовать это представление, но обычно оно весьма удобно и полезно.

Проще всего представить нашу шкалу как круглый циферблат с 64 отметками от 'А' (0) до 'я' (63).

Рассмотрим что происходит при перекодировке T (Win-1251 -> KOI-8) с первой буквой 'А'.

'А' (C0) из таблицы Win-1251 должно перейти в 'А' (E1) из таблицы KOI-8.

Если обозначить оригинальную литеру (Win-1251) как A0 и перекодированную - как A1, а оператор перекодировки как T, то операцию трансляции можно сокращенно записать как

A1 = T * A0

Но (E1) в таблице Win-1251 соответствует уже не 'А', так что если мы снова рассматриваем перекодированный текст как текст в кодировке Win-1251, то нам необходимо выяснить, какая литера сейчас находится в той же самой позиции, где была 'А'.

В таблице Win-1251, коду (E1) соответствует литера 'б', то есть запись

A1 = T * A0

эквивалентна

'б' = T * 'A'

Или, иными словами, преобразование T переводит 'A' в 'б'.

Если на нашем циферблате (0 - 63) нарисовать стрелку из центра (как на обычных часах), то преобразование A1 = T * A0 равносильно повороту этой стрелки на некоторый угол, от 'A' к 'б'.

Если мы будем измерять это расстояние (метрику) также, как мы это делаем на обычных часах - в отрезках окружности, то мы можем вычислить "сколько прошло" от 'A' до 'б' просто вычтя одно из другого.

'б' - 'A' = E1 - C0 = 21 // Все цифры - шестнадцатеричные

(десятичное значение 21 - 33)

Выясним, теперь, во что переходит 'б' при преобразовании T.

В таблице Win-1251 'б' имеет код (E1), а в таблице KOI-8
ему соответствует код (C2).

То есть, при следующем преобразовании T (Win-1251 -> KOI-8) на месте буквы 'A' будет стоять буква с кодом (C2) в таблице Win-1251, то есть 'B'

'B' - 'б' = C2 - E1 = - 1F

То есть, стрелка повернулась назад на 1F = 31 (десятичное).

Но так как все вычисления происходят на круглом циферблате (по mod 64), то поворот назад равносилен повороту вперед на дополнительное по модулю число, то есть на 33 (десятичное).

Таким образом, после двух преобразований T, исходная литера 'A' переместилась на 66 позиций, но так 64 из них составили полный оборот, то видимое перемещение составляет 66 по модулю 64, то есть 2.

33 + 33 = 66 (mod 64) = 2

Выясним, теперь, во что переходит 'B' при преобразовании T.

В таблице Win-1251 'B' имеет код (C2), а в таблице KOI-8 ему соответствует код (F7).

Коду (F7) в таблице Win-1251 соответствует литера 'ч'.

'ч' - 'B' = F7 - C2 = 35 // Все цифры - шестнадцатеричные

(десятичное значение 35 - 53)

Таким образом, после третьей перекодировки сдвиг составил

33 + 33 + 53 = 119 (mod 64) = 55

Легко понять, что через некоторое количество перекодировок сумма сдвигов окажется кратной 64, то есть буква 'А' перейдет сама в себя. (Если вы знакомы с фрезерным ремеслом, можете вспомнить устройство делительной головки).

То что такой момент ***обязательно*** наступит (существование решения, т.е. отсутствие замкнутых циклов) следует из взаимно-однозначного соответствия двух таблиц Win-1251 и KOI-8: каждой литере в Win-1251 соответствует ровно одна литера в KOI-8.

Выполнение последовательных перекодировок можно уподобить генерации последовательности псевдослучайных чисел, длиной 64 байта.

При хорошем выборе способа "перемешивания" (перекодировки), период последовательности мог бы оказаться достаточно большим, но в нашем случае это, кажется, не так.

Чтобы убедиться в этом окончательно, я написал маленькую программу - симулятор Modulo.com, которая выводит на экран (в Win-1251) результаты последовательных преобразований для каждой буквы и длину цикла.

Чтобы сохранить результат в файле, используйте перенаправление вывода:

modulo.com > result.txt

Ниже приведен полный вывод программы:

```
А: АбВчЮаБвЧюА
Б: БвЧюАбВчЮаБ
В: ВчЮаБвЧюАбВ
Г: ГзЪяСуХиЙкЛмНоПрТфЖцГ
Д: ДдД
Е: ЕеЕ
Ж: ЖцГзЪяСуХиЙкЛмНоПрТфЖ
З: ЗъЯсУхИйКлМнОпРтФжЦгЗ
И: ИйКлМнОпРтФжЦгЗъЯсУхИ
Й: ЙкЛмНоПрТфЖцГзЪяСуХиЙ
К: КлМнОпРтФжЦгЗъЯсУхИйК
Л: ЛмНоПрТфЖцГзЪяСуХиЙкЛ
М: МнОпРтФжЦгЗъЯсУхИйКлМ
Н: НоПрТфЖцГзЪяСуХиЙкЛмН
О: ОпРтФжЦгЗъЯсУхИйКлМнО
П: ПрТфЖцГзЪяСуХиЙкЛмНоП
Р: РтФжЦгЗъЯсУхИйКлМнОпР
С: СуХиЙкЛмНоПрТфЖцГзЪяС
Т: ТфЖцГзЪяСуХиЙкЛмНоПрТ
У: УхИйКлМнОпРтФжЦгЗъЯсУ
```

Ф: ФжЦгЗъЯсУхИйКлМнОпРтФ
Х: ХиЙкЛмНоПрТфЖцГзЪяСуХ
Ц: ЦгЗъЯсУхИйКлМнОпРтФжЦ
Ч: ЧюАбВчЮаБвЧ
Ш: ШыЩэЬшЫщЭьШ
Щ: ЩэЬшЫщЭьШыЩ
Ъ: ЪяСуХиЙкЛмНоПрТфЖцГзЪ
Ы: ЫщЭьШыЩэЬшЫ
Ь: ЬшЫщЭьШыЩэЬ
Э: ЭьШыЩэЬшЫщЭ
Ю: ЮаБвЧюАбВчЮ
Я: ЯсУхИйКлМнОпРтФжЦгЗъЯ
а: аБвЧюАбВчЮа
б: бВчЮаБвЧюАб
в: вЧюАбВчЮаБв
г: гЗъЯсУхИйКлМнОпРтФжЦг
д: дДд
е: еЕе
ж: жЦгЗъЯсУхИйКлМнОпРтФж
з: зЪяСуХиЙкЛмНоПрТфЖцГз
и: иЙкЛмНоПрТфЖцГзЪяСуХи
й: йКлМнОпРтФжЦгЗъЯсУхИй
к: кЛмНоПрТфЖцГзЪяСуХиЙк
л: лМнОпРтФжЦгЗъЯсУхИйКл
м: мНоПрТфЖцГзЪяСуХиЙкЛм
н: нОпРтФжЦгЗъЯсУхИйКлМн
о: оПрТфЖцГзЪяСуХиЙкЛмНо
п: пРтФжЦгЗъЯсУхИйКлМнОп
р: рТфЖцГзЪяСуХиЙкЛмНоПр
с: сУхИйКлМнОпРтФжЦгЗъЯс
т: тФжЦгЗъЯсУхИйКлМнОпРт
у: уХиЙкЛмНоПрТфЖцГзЪяСу
ф: фЖцГзЪяСуХиЙкЛмНоПрТф
х: хИйКлМнОпРтФжЦгЗъЯсУх
ц: цГзЪяСуХиЙкЛмНоПрТфЖц
ч: чЮаБвЧюАбВч
ш: шЫщЭьШыЩэЬш
щ: щЭьШыЩэЬшЫщ
ъ: ъЯсУхИйКлМнОпРтФжЦгЗъ
ы: ыЩэЬшЫщЭьШы
ь: ьШыЩэЬшЫщЭь
э: эЬшЫщЭьШыЩэ
ю: юАбВчЮаБвЧю
я: яСуХиЙкЛмНоПрТфЖцГзЪя

*** Relative Loop Size

А: 00010
Б: 00010
В: 00010
Г: 00020
Д: 00002
Е: 00002

Ж: 00020
З: 00020
И: 00020
Й: 00020
К: 00020
Л: 00020
М: 00020
Н: 00020
О: 00020
П: 00020
Р: 00020
С: 00020
Т: 00020
У: 00020
Ф: 00020
Х: 00020
Ц: 00020
Ч: 00010
Ш: 00010
Щ: 00010
Ъ: 00020
Ы: 00010
Ь: 00010
Э: 00010
Ю: 00010
Я: 00020
а: 00010
б: 00010
в: 00010
г: 00020
д: 00002
е: 00002
ж: 00020
з: 00020
и: 00020
й: 00020
к: 00020
л: 00020
м: 00020
н: 00020
о: 00020
п: 00020
р: 00020
с: 00020
т: 00020
у: 00020
ф: 00020
х: 00020
ц: 00020
ч: 00010
ш: 00010
щ: 00010
ъ: 00020

```
ы: 00010
ь: 00010
э: 00010
ю: 00010
я: 00020
```

Легко заметить, что длина цикла для различных букв принимает всего три различных значения: 2, 10 и 20.

Их наименьшее общее кратное - 20.

Таким образом, ровно за 20 последовательных перекодировок Win-1251 -> KOI-8, любой текст, не содержащий букв ё (A3) и Ё (B3), примет свой первоначальный вид.