

КОИ-8 — Lurkmore



A long time ago, in a galaxy far, far away...

События и явления, описанные в этой статье, были давно, и помнит о них разве что пара-другая олдфагов. Но Анонимус не забывает!

«Как то раз два гусара в Йоркшире

Переслать КОИ-8 решили Но проклятый модем Выдал им КОИ-7
КОИ-6, КОИ-5 и четыре

»

— Анонимус вспоминает

КОИ-8 (код обмена информацией, 8 бит, *KOI8*) — восьмибитная *ASCII*-совместимая кодовая страница, разработанная для кодирования букв [кириллических алфавитов](#). Имеет настолько широкое распространение на просторах [России](#), что даже [медведи](#) на улицах свободно на ней говорят^[1].

История и особенности

КОИ-8 (а сначала КОИ-7) родилось из практических соображений во времена глубокой древности, когда системы, не поддерживающие кириллицу даже теоретически (то есть владеющие только латинским алфавитом и использующие 5-битные таблицы символов, на 32 знака, а позднее 7-битные, на 128 знаков) были не исключением, а повседневной нормой (в частности, таковыми были почтовые серверы).

Но [146%](#) населения и тогда не хотели читать на [мунспике](#), а требовали русских букв. Увы, полный набор из латинских (26 больших и 26 маленьких), русских (по 33), 10 цифр и пачки знаков препинания никак не лез в 7 бит, в результате родилось гениальное решение — все *malen'kie latinskie* принудительно заменили на *БОЛЬШИЕ РУССКИЕ*, получилась кодировка КОИ7Н2 (набор 2), в которой **вообще** не было маленьких букв, **ОДИН СПЛОШНОЙ КАПСЛОК**. Для совместимости с буржуями скопировали их раскладку — только латинские, большие и маленькие — в КОИ7Н0 (набор 0), и по аналогии сделали КОИ7Н1 — только русские, но сначала маленькие, а затем БОЛЬШИЕ — костыль, чтобы сохранить совместимость наборов Н2 и Н1.

Порядок русских букв в КОИ-7 выбрали исходя из максимальной похожести на латинские, так русская буква «А» превратилась в латинскую «А», но буква «Я» — почему-то в «Q» (см. [Куковлев](#)). Таким образом, при случайном выводе русского 7-битного текста на буржуйской 7-бит аппаратуре, он без особых потерь превращался в [транслит](#) капслоком. Это породило такой мем, как [иНЖАЛИД ДЕЖИЦЕ](#), являющийся сообщением об ошибке (Invalid device) терминала Видеотон, а также некоторых [спектрумов](#), в которых маленькая латиница отсутствовала. Ныне этот мем благополучно забыт, но нестыковки между порядком букв в КОИ-8 и других кодировках кириллицы порождают новые лулзы и по сей день, см. [БНОПНЯ](#).

КОИ-8 прост как мычание — незатейливо склеили два набора КОИ7Н0 и КОИ7Н1 и получили полезную 8-битную кодировку, которая при потере 8-го бита сохраняла читаемость в транслите, хотя и с инвертированным регистром.

В результате, от всех прочих кодировок, применяемых для кириллицы, КОИ-8 отличается тем, что порядок букв в ней не соответствует алфавиту ни одного из использующих кириллицу человеческих языков. Слава (почтовым) роботам!

В сети

В своё время кодировка КОИ-8 люто, бешено форсилась [Андреем «ache» Черновым](#), [сатанистом](#) и



KOI8-R, MOTHERFUCKER, DO YOU SPEAK IT?!!

80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF
№	№	№	№	№	№	№	№	№	№	№	№	№	№	№	№
В0	В1	В2	В3	В4	В5	В6	В7	В8	В9	ВА	ВБ	ВС	ВД	ВЕ	ВФ
Ю	я	б	ц	д	е	ф	г	х	и	й	к	л	м	н	о
сч	с1	с2	с3	с4	с5	с6	с7	с8	с9	сА	сБ	сС	сD	сE	сF
П	я	р	с	т	у	ж	в	ь	ы	з	ш	э	щ	ч	ъ
Д0	Д1	Д2	Д3	Д4	Д5	Д6	Д7	Д8	Д9	ДА	ДБ	ДС	ДД	ДЕ	ДФ
Ю	А	Б	Ц	Д	Е	Ф	Г	Х	И	Й	К	Л	М	Н	О
В0	В1	В2	В3	В4	В5	В6	В7	В8	В9	ВА	ВБ	ВС	ВД	ВЕ	ВФ
П	Я	Р	С	Т	У	Ж	В	Ь	Ы	З	Ш	Э	Щ	Ч	Ъ
В0	В1	В2	В3	В4	В5	В6	В7	В8	В9	ВА	ВБ	ВС	ВД	ВЕ	ВФ

A quick brown fox jumps over the lazy dog.

юбдфхйлнпртжъзэчЮБДФХЙЛНПРТЖЪЗЭЧ

A quick brown fox jumps over the lazy dog.

юбдфхйлнпртжъзэчЮБДФХЙЛНПРТЖЪЗЭЧ

любителем Кроули, что какбе намекает. Аше же был автором древнего и в то время практически единственного пакета электронной почты для DOS — UUPC (не путать с протоколом *uucp*), так что ему можно. А во время инсталляции онога пакета **удивлённый** юзер мог лицезреть автора с черепом в руках, ага. Тот же Аше пропихнул KOI8-R в RFC 1489, что дало повод считать KOI8-R «**стандартной** кодировкой русского интернета» и увеличило **ЧСВ** создателя и первых юзеров на 9000.

Это вызвало глубокое **неприятие** со стороны пользователей DOS, **Windows** и даже **Mac** (у которых тоже была своя кодировка mac-cyrillic) и вызвало войну всех против всех. Бедные вебмастера делали по четыре копии сайта в разных кодировках, а беспощадные русские быдлокодеры ковали свою вундервафлю — Russian Apache, который эвристически пытался угадать кодировку на стороне юзера (иногда случайно перекодирюя невинные GIF-ки с **проном**). Только изготовители перекодировщиков ловили свой мутный PROFIT на войне кодировок и наплодили over 9000 быдло-перекодировщиков.

Чуть позже подтянулись **укры**, которым аше позабыл выдать расовых хохло-букв и запилили незалежный KOI8-U (RFC 2319), а позже ещё over 9000 разных **малых наций** с алфавитами на основе кириллицы.

Надо сказать, что правоверные **линуксоиды** встретили кодировку KOI-8 как родную — в те далекие времена многие программы для Линуха страдали 7-битностью и KOI-8 пришёлся очень кстати. А у **бэдунов** KOI8-R это вообще генеральная линия партии и ставится принудительно автоматически при выборе русского, хотя для эстетов Юникод всё же имеется, начиная с FreeBSD 7.

Нынче

Первый гвоздь в гроб KOI-8 забил стандарт MIME. Выяснилось, что нормальному почтовому клиенту не важно в какой кодировке написан текст, лишь бы в нём был правильно выставлен Charset. Почтовик знает их все, и ему похуй. Конечно, если содержимое Windows-1251 промаркировать KOI8-R, получится **бНОПНЯ**. Точно также во всех **браузерах** главное — верно промаркировать на сервере, а уж браузер покажет.

REN-TV. Фантастика под грифом Секретно: НЛО из принтера.

В смене кодировок можно увидеть не только ЭТО!

Второй и окончательный гвоздь забил Юникод в формате UTF-8, который содержит **все** символы всех мунспиков. Нынче только желающие быстро работать со строками через MMX/GPGPU **сопротивляются** UTF-8 и настаивают на тотальном использовании KOI-8 или другого восьмибита с фиксированной длиной символа, остальные хомячки просто юзают Юникод, даже не замечая его недостатков.

Интересные факты

Так как русские буквы в KOI-8 занимают те же коды, что и в **cp1251**, после насильного двадцатикратного прогона

```
iconv -f cp1251 -t koi8-r
```

получается **та же изначальная строка**.

См. также

- бНОПНЯ**
- иНЖАЛИД ДЕЖИЦЕ**

Примечания

- ↑ На самом деле пиндосы считали, что медведи должны говорить на ISO-8859-5, и которую в СССР даже объявили госстандартом в 1980-х из-за тупой копипасты совковым комитетом по стандартизации документов, в которых они не шарили. Но IRL всё оказалось иначе.



Языки программирования

++i ++i 1C AJAX BrainFuck C Sharp C++ Dummy mode Erlang Forth FUBAR
God is real, unless explicitly declared as integer GOTO Haskell Ifconfig Java JavaScript LISP
My other car Oracle Pascal Perl PHP Prolog Pure C Python RegExp Reverse Engineering
Ruby SAP SICP Tcl TeX Xyzy Анти-паттерн Ассемблер Быдлокодер
Выстрелить себе в ногу Грязный хак Дискета ЕГГОГ Индусский код Инжалид дежице
Капча KOI-8 Костыль Лог Метод научного тыка Очередь Помолясь Проблема 2000
Программист Процент эс Рекурсия Свистелки и перделки Спортивное программирование

w:КОИ-8