

## Выбор кодировки

В. Лидовский

*Для текста на русском языке существует рекордное количество различных способов его компьютерного представления. Соответствующие программы и, в частности, Internet-браузеры предлагают большой выбор: cp866, cp1251, koi-8r, UTF-8, ISO 8859-5, ...*

Для компьютерного представления текстов необходимо каждому знаку текста сопоставить однозначно некоторое число, код знака. В настоящее время для кодирования знаков русского алфавита используются 7, 8 и 16-разрядные коды, а также транслитерация латинскими буквами. Кроме того, очевидно, что использовать на компьютерах кириллицу без латиницы невозможно.

Все современные таблицы кодировки происходят от возникшей еще в 60-е 7-разрядной таблицы ASCII (American Standard Code for Information Interchange), которая содержит 33 кода команд или управляющих символов, большая часть которых сегодня не используется (см. Таблицу 1), и 95 кодов для различных знаков (см. Таблицу 2), достаточных для работы по-минимуму с английскими текстами. При 7-разрядном кодировании каждому символу сопоставляется 7 бит кода, т. е. число в диапазоне от 0 до 127. В “чисто” текстовых (plain text) файлах допустимы только управляющие символы, отмечающие концы строк и переходы на позиции табуляции. Маркер конца строки в Linux — это код 10, в Macintosh OS — 13, в CP/M, DOS и Microsoft Windows — последовательность кодов 13, 10.

Для такого кодирования кириллицы используются два варианта таблицы КОИ-7 (Код для обмена информацией 7-битный): первый — это ASCII со строчными латинскими буквами и некоторые другими знаками, замененными на близкие фонетически заглавные буквы русского алфавита; второй — это ASCII, в котором вместо строчных латинских букв и 6 других символов размещены заглавные буквы кириллицы, а вместо заглавных — строчные. Для представления русских букв ныне используется естественно только второй вариант. У этого кодирования есть два важных преимущества:

- 1) оно наиболее компактно (для доступа к латинским и другим буквам и возврата обратно к кириллице используются специальные управляющие коды);
- 2) оно вводит естественную транслитерацию, что позволяет работать с текстами на русском единообразно на разных компьютерных системах.

ASCII коды все еще достаточно широко используются для текстов на английском, особенно в США, а также в системах e-mail. Стандартные текстовые утилиты Unix (grep, sed, awk, ...) вплоть до середины 90-х по умолчанию могли работать только с ASCII — подключение 8-го бита

требовало их перекомпиляции! 7-битная кириллица ныне используется, пожалуй, только в системе для подготовки научных публикаций T<sub>E</sub>X, да и то неявно. Кроме того, такое представление кириллицы используется в некоторых электронных системах, например, в телевизионном телетексте.

Сейчас для кодирования символов алфавитов большинства языков используется 8-разрядное кодирование, т. е. кодирование, при котором каждому символу сопоставляется 8 бит или 1 байт кода. Такой код — это число в диапазоне от 0 до 255. Первые 128 символов при таком кодировании — это ASCII.

Первая популярная международно признанная 8-разрядная кодировка используется на аппаратном уровне на компьютерах IBM PC и совместимых с ними с 1981 года (см. Таблицу 3). На ее основе была разработана кодовая страница 437, Microsoft Code Page (CP) 437.

Международный стандарт ISO/IEC 8859-1 (см. Таблицу 4) стал в наши дни де-факто заменой для ASCII. В этой кодировке первые 32 кода, числа 128–159, соответствуют практически неиспользуемым управляющим символам, общим для всех таблиц кодировки ISO. Все коды 8859-1, как и ASCII, совпадают с Unicode. Может использоваться для текстов на почти всех западноевропейских языках, хотя она и не полностью покрывает нужды французского и финского языков. Этот недостаток, а также отсутствие знака для новой общеевропейской валюты привели в 1999 г. к возникновению кодировки 8859-15, в которой изменены значения 8 кодов 8859-1: 164 (A4) — € (20ac), 166 (A6) — Š (0160), 168 (A8) — š (0161), 180 (B4) — Ž (017d), 184 (B8) — ž (017e), 188 (BC) — € (0152), 189 (BD) — œ (0153), 190 (BE) — Ÿ (0178).

Всего в ISO 8859 определяются 14 8-битовых таблиц кодировки, например, 8859-6 содержит набор арабских букв, 8859-7 — греческих, а 8859-8 — иврита.

Для представления русских букв используются, как правило, следующие пять таблиц:

- А. ISO 8859-5 — стандарт международной организации по стандартизации для кириллицы, в РФ используется редко. Разработана на основе основной кодировки ГОСТ. В ней отсутствует псевдографика (см. Таблицу 5);
- Б. Кодовая страница 866 разработана на основе альтернативной кодировки ГОСТ. Она создана специально для ОС MS-DOS, в которой часто используются символы псевдографики. В ней эти символы имеют те же коды, что и в CP437 (см. Таблицу 7). Практически вытеснена следующей таблицей;
- В. Кодовая страница 1251 для Microsoft Windows стала весьма популярной почти исключительно из-за огромного влияния фирмы Microsoft на рынок компьютерных технологий. Кроме того, в ней отсутствует ненужная в графических средах поддержка символов псевдографики

и наиболее полно по сравнению с другими кодировками представлены такие символы как ©, ®, №, различные виды кавычек и тире и т. п. (см. Таблицу 8). В ней пока нет знака для кода 152. Эта кодировка, несмотря на некоторые недостатки, например, наличие явно избыточных букв южнославянской кириллицы, была бы почти оптимальной, если бы она не была волонтеристки навязана пользователям Microsoft Windows в РФ точно также как и в 1995 была навязана отличная от стандартной машинописной раскладка клавиатуры. Используя CP1251, наверное невозможно не ощущать себя жителем “банановой” республики;

- Г. На базе ГОСТ КОИ-8, koï8-r (кодовая страница 878) — применяется в мире Linux/Unix. В ней естественным образом обеспечивается связь с 7-разрядным кодированием: отбрасыванием восьмого бита из кода буквы русского алфавита koï8-r получается код этой же буквы в КОИ-7. Эта кодировка поддерживает символы псевдографики, что требует около половины всех вводимых кодировкой кодов. В 1993 она стандартизирована в Internet документом RFC1489 (см. Таблицу 6). В начале XXI века появились несколько модификаций этой кодировки, koï8-u и koï8-ub, — с дополнениями до украинского и затем соответственно до белорусского. В них установлены следующие соответствия для специфических украинских букв: 164 (A4) — є (0454), 180 (B4) — Є (0404), 166 (A6) — і (0456), 182 (B6) — І (0406), 167 (A7) — ї (0457), 183 (B7) — Ї (0407), 173 (AD) — ґ (0491), 189 (BD) — Ґ (0490). В koï8-ub специфические белорусские буквы кодируются так: 174 (AE) — ў (045E), 190 (BE) — Ў (040E). Помимо них в koï8-ub включаются вместо символов псевдографики знаки разнообразных кавычек и тире, №, евро, иены и т. п., но эта кодировка находится на начальной стадии внедрения и её перспективы туманны;
- Д. Кодовая страница 10007 — используется на компьютерах под управлением Macintosh OS, по своему набору знаков почти совпадает с CP1251 (см. Таблицу 9).

В IBM OS/2 для русских букв использовалась кодовая страница 855, интересная тем, что в ней коды для каждой строчной и прописной буквы кириллицы идут подряд, что напоминает рисунок алфавита.

Хорошие текстовые редакторы, среди которых можно отметить, например, отечественный заменитель Microsoft Notepad, Aditor (<http://kulichki.rambler.ru/~kdavy/progs/aditor>), могут во многих случаях, анализируя частоты встречаемости символов в простом текстовом файле, автоматически определять используемую кодировку. Документы HTML, XML, RTF и других развитых форматов текста с разметкой включают в себя информацию об используемом кодировании.

Для кодирования русских букв могут использоваться еще пять таблиц, не имеющих международного статуса:

- Е. КОИ-8 — исторически ее варианты использовались на компьютерах под управлением ОС Unix, CP/M и некоторых других с середины 1970-х. На IBM PC не получила распространения из-за неалфавитного расположения букв кириллицы, чему придавалась возможно чрезмерное значение (действительно, такой порядок есть только для английского языка и ни для какого другого). Вместо нее на IBM PC в среде MS-DOS использовали в основном сначала болгарскую, затем основную и, наконец, варианты альтернативной (сначала базовую, затем модифицированную и затем CP866). В ней нет букв  $\ddot{E}\ddot{e}\ddot{Z}$ , а остальные буквы совпадают с koï8-r;
- Ж. Основная кодировка ГОСТ (государственный стандарт СССР) от 1987 года. Она совпадает с ISO 8859-5 только по всем знакам русских букв, кроме заглавной буквы  $\ddot{E}$ , расположенной как в CP866. Ее главный недостаток — символы псевдографики расположены не так, как на IBM PC. На практике эту кодировку можно встретить только на старых советских клонах IBM PC EC-1840 и в знакогенераторах принтеров, выпущенных до середины 90-х;
- З. Альтернативная кодировка ГОСТ отличается от CP866 только по позициям 242–251, содержащим символы  $\swarrow \searrow \nearrow \nwarrow \rightarrow \uparrow \downarrow \leftarrow \div \pm \text{—}$  первые 4 из них до сих пор не включены в Unicode;
- И. Модифицированная альтернативная кодировка ГОСТ отличается от CP866 только по позициям 242–247 и 252, 253, содержащим символы  $\geq \leq \int J \div \approx ^n 2$ ;
- Й. Болгарская кодировка, называемая также “старый вариант ВЦ АН СССР”, Interprog или MICS, получается механическим внесением блока из 64 букв русского алфавита (такого же как в CP1251) в позиции 128–191 CP437.

В кодировках А–В, Д, Ж–Й все буквы кроме  $\ddot{e}$  и  $\ddot{E}$  расположены в алфавитном порядке.

Буквы болгарского алфавита — подмножество русского. Кодировка Б позволяет работать также с текстами на украинском и белорусском, а кодировки А, В, Д еще и на македонском и сербском.

Ясно, что для кодирования символов некоторых языков, например, китайского, 8-разрядных чисел недостаточно. Кроме того, создание 8-разрядных таблиц стало практически неконтролируемым (каждый компьютерный шрифт, например, неявно вводил собой и таблицу кодировки). Поэтому был создан консорциум Unicode для введения единообразного кодирования всех возможных символов, что позволило, в частности, присваивать единообразным образом коды знакам компьютерных шрифтов. Кодировка Unicode опирается на каталог символов UCS (Universal Character Set) стандарта ISO 10646. UCS может содержать до  $2^{31} = 2\,147\,483\,648$  различных знаков и постоянно пополняется. Коды UCS-2 — 2-байтные, т. е. числа от 0 до 65535, а UCS-4 — 4-байтные, т. е. числа от 0 до

2 147 483 647. ASCII — это первые коды UCS-2, а UCS-2 — это начало UCS-4. Двух и четырехбайтные коды Unicode могут представляться двумя способами: байты располагаются слева направо от старшего к младшему (Big Endian, BE) или от младшего к старшему (Little Endian, LE). Второй способ используется в подавляющем числе случаев. Кроме того, для более компактного кодирования используются коды переменной длины UTF-8 (Unicode Transfer Format) — 1–6-байтные и UTF-16 — 2 или 4-байтные. Последние также существуют в двух видах (Little и Big Endian) и позволяют кодировать не более  $2^{20} + 2^{16} = 1\,114\,112$  знаков. Все коды UCS-2 являются также и кодами UTF-16. Есть еще формат UTF-7, который позволяет для большинства символов ASCII использовать 7-битные числа, — он обеспечивает доступ только к UCS-2. Коды Unicode 20 (пробел), A0 (неразрывный пробел) и AD (перенос), присутствующие во многих 8-разрядных таблицах, можно рассматривать как управляющие.

Часть кодов (диапазон E000–F8FF<sub>16</sub> или 57344–63743, т. е. всего 6400) в UCS зарезервированы для личного или ограниченного локальными рамками использования. Остальные коды также сгруппированы по назначению (см. Таблицу 11).

Все большее распространение, в частности, в Internet получает кодировка UTF-8, которая позволяет обходиться 8 битами для кодирования символов ASCII и 16 битами для кодирования символов большинства алфавитных письменностей, в частности, русской. Код UTF-8, последовательность байт, получается из кода каталога UCS по следующей однозначной схеме:

```

00000000–0000007F16 (0–127):
    0xxxxxxx
00000080–000007FF16 (128–2047):
    110xxxxx 10xxxxxx
00000800–0000FFFF16 (2048–65535):
    1110xxxx 10xxxxxx 10xxxxxx
00010000–001FFFFF16 (65536–2097151):
    11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
00200000–03FFFFFF16 (2 097 152–67 108 863):
    111110xx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
04000000–7FFFFFFF16 (67 108 864–2 147 483 647):
    1111110x 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx.

```

Позиции битов, обозначенные как буква *x*, заполняются соответствующими битами из кода символа в бинарном представлении. Используется кратчайшая из возможных многобайтовых последовательностей, которые могут представить код символа. Например, символ Unicode с кодом 169

=  $a_{9_{16}} = 1010\ 1001$  (знак ©) кодируется в UTF-8 как  $11000010\ 10101001$   
=  $c_{2_{16}}\ a_{9_{16}}$ , а символ с кодом  $8800 = 2260_{16} = 0010\ 0010\ 0110\ 0000$  (знак ≠)  
кодируется так:  $11100010\ 10001001\ 10100000 = e_{2_{16}}\ 89_{16}\ a_{0_{16}}$ .

Unicode достаточно полно поддерживается современными программами: Internet-браузерами, офисными пакетами и т. п. В Linux используется UTF-8, а в Microsoft Windows еще и UCS-2. Пока поддержка UCS в Linux и несколько слабее, чем в Microsoft Windows 2000/Me/XP. Главная проблема при использовании Unicode — отсутствие соответствующего полного набора шрифтов. Другая проблема — в сложности ввода. Как, например, российскому пользователю вводить специфические испанские буквы, математические знаки, китайские иероглифы и т. д.? Одна из программ, которая может во многих случаях удовлетворительно решать проблемы такого ввода — это текстовый редактор Yudit (<http://www.yudit.org>).

Как ни странно, но, казалось бы формальные таблицы 1–9, по своей сути принципиально неформальны — они образуют мост между формальной компьютерной и бескрайне-разнообразной реальной текстовой информацией. Действительно, одному и тому же коду Unicode 65 соответствует все возможные написания латинской буквы А — наклонной, готической, неразборчивой рукописной... Нужна какая-то неподдающаяся математическому моделированию “магическая” составляющая интеллекта, чтобы сопоставить рисунку знака его код в некоторой таблице. Программы распознавания символов могут в многих случаях удовлетворительно осуществлять такое сопоставление. Однако, их ограниченность широко используют в Internet для различения машинного и личного обращения, например, при попытке бесплатной регистрации сайта на поисковом сервере <http://www.altavista.com>.

Программа `iconv` (стандартная утилита Linux, легко переносимая в любую ОС) позволяет конвертировать тексты между практически всеми известными в мире кодировками для букв русского алфавита, кодировками А–Д, КОИ-7 (с ограничениями), `koj8`-и и всеми вариантами кодирования Unicode. Однако `iconv` не поддерживает кодировок Е–Й. Авторская программа `ruscv`, исходники которой на C++ доступны в Internet по адресу <http://litwr.boom.ru>, “знает” все кодировки А–Й, Unicode UTF-8 и UCS-2 Little Endian, а также может осуществлять несколько разновидностей прямой и обратной транслитерации кириллицы латинскими буквами. Опция *u* этой программы позволяет также работать с Unicode-текстами в обычном редакторе для 8-разрядных текстов.

Упомянутая транслитерация — это единственно возможный способ для работы с кириллицей в системе без соответствующих шрифтов. Она также бывает необходима при передаче собственных имен.

Для русского языка существует несколько систем транслитерации, разрабатываемых с начала XX века, — ни одна из них не стала при-

знанным стандартном. Главная проблема транслитерации заключается в том, что для представления 33 букв русского алфавита есть только 26 базовых букв латинского/английского алфавита. Решение этой проблемы — в привлечении диакритических знаков, например, обозначая э как *ê*, или в использовании буквенных последовательностей, например, используя *shch* для обозначения *щ*. Первый способ не вписывается в базовый латинский алфавит, а второй может приводить к неоднозначностям, например, учитывая, что *ч* обозначается как *ch*, а *ш* — *sh*, как транслитерировать в латиницу слово *веснушчатый*? Неоднозначности можно было бы избежать, заключая буквенные последовательности между знаками, например,  $\langle$  и  $\rangle$ .

Можно использовать “компьютерную” транслитерацию на основе ASCII (КОИ-7) — код получается вычитанием 128 от соответствующего кода в koï8-г, что, как правило, дает код латинской буквы, близкой фонетически к русской. Недостатком этой “естественной” транслитерации является представление 11 букв русского алфавита, *ЧчШшЩщЪЭэЮю*, в виде специальных знаков и отсутствие представления для букв *ЁёЪ*. Этот недостаток можно устранить, используя для этих букв заключенные между скобками (квадратными, круглыми, угловыми, ....) стандартные транслитерационные последовательности (см. Таблицу 10).

Стандарт ГОСТ 7.79 от 2000 года также вводит почти однозначную транслитерацию без использования диакритических знаков. В нем нет только возможности различать заглавные и строчные мягкий и твердый знаки.

#### Литература

1. Галенко В. Т. *Русский язык и латиница* //НТИ. Серия 1, 10/2000.
2. Лидовский В. В. *Первичная машинная обработка текстов: методика и проблематика* — Депон. в ИНИОН РАН №53656, 1998.
3. Соснин О. М. *О русско-латинской транслитерации личных имен, фамилий и географических названий в официальных документах* //НТИ. Серия 1, 2/2003.
4. Фигурнов В. Э. *IBM PC для пользователя* — М.: “Финансы и статистика”, 1990.
5. Gilyarevsky R. S., Grivnin V. S. *Language identification guide* — М.: “Nauka” Publishing House, 1970.
6. *ISO9: Information and documentation — Transliteration of Cyrillic characters by Latin characters — Slavic and non-Slavic Languages* — Geneve, 1995.
7. Материалы сайта <http://www.unicode.org>.

## ТАБЛИЦА 1

### Управляющие коды ASCII

Код			Полное имя кода в Unicode (краткое имя в ASCII)
10-й	16-й	Клавиатурный	
<i>Перевод имени кода</i> — описание использования кода.			

Выше представлен шаблон для таблицы управляющих символов. Под клавиатурным кодом подразумевается комбинация двух клавиш, Ctrl (Control, в таблице это знак ^) и приводимой, одновременное нажатие которых должно производить соответствующий код.

0	00	^@	NULL (NUL)
<i>Пусто</i> — этот код используется как завершающий в представлении строк многими системами программирования, например, Си, поэтому его использование в текстовых файлах крайне нежелательно.			
1	01	^A	START OF HEADING (SOH)
<i>Начало заголовка</i> — практически не используется.			
2	02	^B	START OF TEXT (STX)
<i>Начало текста</i> — практически не используется.			
3	03	^C	END OF TEXT (ETX)
<i>Конец текста</i> — в Unix и MS-DOS ввод этого символа с клавиатуры служит сигналом для прекращения выполнения программы.			
4	04	^D	END OF TRANSMISSION (EOT)
<i>Конец передачи</i> — в Unix и PostScript означает конец вводимых данных.			
5	05	^E	ENQUIRY (ENQ)
<i>Кто там?</i> — практически не используется.			
6	06	^F	ACKNOWLEDGE (ACK)
<i>Подтверждение, да</i> — практически не используется.			
7	07	^G	BELL (BEL)
<i>Звонок</i> — при его печати на консоли MS-DOS или Unix должен производиться звуковой сигнал.			
8	08	^H	BACKSPACE (BS)
<i>Возврат на шаг</i> — означает, что следующий символ следует печатать с предшествующей позиции.			
9	09	^I	HORIZONTAL TABULATION (TAB)
<i>Горизонтальная табуляция</i> — переход на следующую позицию табуляции.			
10	0A	^J	LINE FEED (LF)
<i>Подача новой строки</i> — переход на новую строку. В текстовых файлах MS-DOS и Microsoft Windows с сохранением текущей горизонтальной позиции. В текстовых файлах Unix с переходом на первую горизонтальную позицию.			



11	OB	^K	VERTICAL TABULATION (VT)
<i>Вертикальная табуляция</i> — используется очень редко, как правило, принтерами.			
12	OC	^L	FORM FEED (FF)
<i>Подача новой формы</i> — для консоли, как правило, означает очистку экрана, для принтера — завершение печати на текущем листе и запрос нового.			
13	OD	^M	CARRIAGE RETURN (CR)
<i>Возврат каретки</i> — переход на первую горизонтальную позицию строки. В текстовых файлах MS-DOS и Microsoft Windows с сохранением текущей строки, а в текстовых файлах Macintosh OS с переходом на новую строку. В текстовых файлах Unix не используется.			
14	OE	^N	SHIFT OUT (SO)
<i>Выход</i> — используется очень редко, как правило, принтерами.			
15	OF	^O	SHIFT IN (SI)
<i>Вход</i> — используется очень редко, как правило, принтерами.			
16	10	^P	DATA LINK ESCAPE (DLE)
<i>Авторегистр 1</i> — практически не используется.			
17	11	^Q	DEVICE CONTROL ONE (DC1)
Используется некоторыми телекоммуникационными протоколами как байт X-ON.			
18	12	^R	DEVICE CONTROL TWO (DC2)
Практически не используется.			
19	13	^S	DEVICE CONTROL THREE (DC3)
Используется некоторыми телекоммуникационными протоколами как байт X-OFF.			
20	14	^T	DEVICE CONTROL FOUR (DC4)
Практически не используется.			
21	15	^U	NEGATIVE ACKNOWLEDGE (NAK)
<i>Нет</i> — практически не используется.			
22	16	^V	SYNCHRONOUS IDLE (SYN)
<i>Синхронизация</i> — практически не используется.			
23	17	^W	END OF TRANSMISSION BLOCK (ETB)
<i>Конец блока</i> — практически не используется.			
24	18	^X	CANCEL (CAN)
<i>Аннулирование</i> — используется очень редко, как правило, принтерами.			
25	19	^Y	END OF MEDIUM (EM)
<i>Конец носителя</i> — практически не используется.			
26	1A	^Z	SUBSTITUTE (SUB)
<i>Замена</i> — в MS-DOS, Macintosh OS и CP/M — это маркер конца текстового файла.			
27	1B	^[	ESCAPE (ESC)
<i>Авторегистр 2</i> — указывает на то, что некоторое количество кодов после него и он сам образуют группу, рассматриваемую как один код.			

28	1C	^\ <i>Разделитель файлов</i>	FILE SEPARATOR (FS) — практически не используется.
29	1D	^] <i>Разделитель групп</i>	GROUP SEPARATOR (GS) — практически не используется.
30	1E	^^ <i>Разделитель записей</i>	RECORD SEPARATOR (RS) — практически не используется.
31	1F	^_ <i>Разделитель элементов</i>	UNIT SEPARATOR (US) — практически не используется.
127	7F		DELETE (DEL) <i>Забой</i> — удаление последнего видимого знака печатаемой строки.

## ТАБЛИЦА 2

### Видимые символы ASCII

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
20		!	"	#	\$	%	&	'	(	)	*	+	,	-	.	/
	32 0020	33 0021	34 0022	35 0023	36 0024	37 0025	38 0026	39 0027	40 0028	41 0029	42 002A	43 002B	44 002C	45 002D	46 002E	47 002F
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
	48 0030	49 0031	50 0032	51 0033	52 0034	53 0035	54 0036	55 0037	56 0038	57 0039	58 003A	59 003B	60 003C	61 003D	62 003E	63 003F
40	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	64 0040	65 0041	66 0042	67 0043	68 0044	69 0045	70 0046	71 0047	72 0048	73 0049	74 004A	75 004B	76 004C	77 004D	78 004E	79 004F
50	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
	80 0050	81 0051	82 0052	83 0053	84 0054	85 0055	86 0056	87 0057	88 0058	89 0059	90 005A	91 005B	92 005C	93 005D	94 005E	95 005F
60	'	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
	96 0060	97 0061	98 0062	99 0063	100 0064	101 0065	102 0066	103 0067	104 0068	105 0069	106 006A	107 006B	108 006C	109 006D	110 006E	111 006F
70	p	q	r	s	t	u	v	w	x	y	z	{		}	~	
	112 0070	113 0071	114 0072	115 0073	116 0074	117 0075	118 0076	119 0077	120 0078	121 0079	122 007A	123 007B	124 007C	125 007D	126 007E	

В этой и следующих таблицах 3–9 для каждого символа приводятся его 10-й и 16-й код в приводимой таблице кодировки и 16-й код в Unicode.

### ТАБЛИЦА 3

Кодировка CP437 и знакогенератор IBM PC

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
00		☺ 1 263A	☻ 2 263B	♥ 3 2665	♦ 4 2666	♣ 5 2663	♠ 6 2660	● 7 2022	◼ 8 25D8	○ 9 25CB	◉ 10 25D9	♂ 11 2642	♀ 12 2640	♪ 13 266A	🎵 14 266B	⚙ 15 263C
10	▶ 16 25BA	◀ 17 25C4	↕ 18 2195	!! 19 203C	☞ 20 00B6	§ 21 00A7	■ 22 25AC	↕ 23 21A8	↑ 24 2191	↓ 25 2193	→ 26 2192	← 27 2190	└ 28 221F	↔ 29 2194	▲ 30 25B2	▼ 31 25BC
70																🏠 127 2302
80	Ç 128 00C7	ü 129 00FC	é 130 00E9	â 131 00E2	ä 132 00E4	à 133 00E0	å 134 00E5	ç 135 00E7	ê 136 00EA	ë 137 00EB	è 138 00E8	ï 139 00EF	î 140 00EE	ì 141 00EC	Ä 142 00C4	Å 143 00C5
90	É 144 00C9	æ 145 00E6	Æ 146 00C6	ô 147 00F4	ö 148 00F6	ò 149 00F2	û 150 00FB	ù 151 00F9	ÿ 152 00FF	Ö 153 00D6	Ü 154 00DC	¢ 155 00A2	£ 156 00A3	¥ 157 00A5	Pt 158 20A7	f 159 0192
A0	á 160 00E1	í 161 00ED	ó 162 00F3	ú 163 00FA	ñ 164 00F1	Ñ 165 00D1	à 166 00AA	ó 167 00BA	¿ 168 00BF	┐ 169 2310	┌ 170 00AC	½ 171 00BD	¼ 172 00BC	¡ 173 00A1	« 174 00AB	» 175 00BB
B0	▒ 176 2591	▓ 177 2592	█ 178 2593	 179 2502	├ 180 2524	┌ 181 2561	┐ 182 2562	└ 183 2556	┘ 184 2555	┌ 185 2563	┐ 186 2551	└ 187 2557	┘ 188 255D	┌ 189 255C	┐ 190 255B	└ 191 2510
C0	┌ 192 2514	└ 193 2534	┐ 194 252C	┘ 195 251C	─ 196 2500	├ 197 253C	┌ 198 255E	┐ 199 255F	└ 200 255A	┘ 201 2554	┌ 202 2569	└ 203 2566	┐ 204 2560	─ 205 2550	├ 206 256C	┌ 207 2567
D0	┌ 208 2568	└ 209 2564	┐ 210 2565	┘ 211 2559	┌ 212 2558	┐ 213 2552	└ 214 2553	┘ 215 256B	┌ 216 256A	┐ 217 2518	└ 218 250C	▀ 219 2588	▁ 220 2584	▂ 221 258C	▃ 222 2590	▄ 223 2580
E0	α 224 03B1	β 225 00DF	Γ 226 0393	π 227 03C0	Σ 228 03A3	σ 229 03C3	μ 230 00B5	τ 231 03C4	Φ 232 03A6	Θ 233 0398	Ω 234 03A9	δ 235 03B4	∞ 236 221E	φ 237 03C6	ε 238 03B5	∩ 239 2229
F0	≡ 240 2261	± 241 00B1	≥ 242 2265	≤ 243 2264	∫ 244 2320	J 245 2321	÷ 246 00F7	≈ 247 2248	° 248 00B0	· 249 2219	· 250 00B7	√ 251 221A	n 252 207F	2 253 00B2	■ 254 25A0	 255 00A0

































































**ТАБЛИЦА 4**  
Кодировка ISO 8859-1

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A0	160 00A0	161 00A1	162 00A2	163 00A3	164 00A4	165 00A5	166 00A6	167 00A7	168 00A8	169 00A9	170 00AA	171 00AB	172 00AC	173 00AD	174 00AE	175 00AF
B0	176 00B0	177 00B1	178 00B2	179 00B3	180 00B4	181 00B5	182 00B6	183 00B7	184 00B8	185 00B9	186 00BA	187 00BB	188 00BC	189 00BD	190 00BE	191 00BF
C0	192 00C0	193 00C1	194 00C2	195 00C3	196 00C4	197 00C5	198 00C6	199 00C7	200 00C8	201 00C9	202 00CA	203 00CB	204 00CC	205 00CD	206 00CE	207 00CF
D0	208 00D0	209 00D1	210 00D2	211 00D3	212 00D4	213 00D5	214 00D6	215 00D7	216 00D8	217 00D9	218 00DA	219 00DB	220 00DC	221 00DD	222 00DE	223 00DF
E0	224 00E0	225 00E1	226 00E2	227 00E3	228 00E4	229 00E5	230 00E6	231 00E7	232 00E8	233 00E9	234 00EA	235 00EB	236 00EC	237 00ED	238 00EE	239 00EF
F0	240 00F0	241 00F1	242 00F2	243 00F3	244 00F4	245 00F5	246 00F6	247 00F7	248 00F8	249 00F9	250 00FA	251 00FB	252 00FC	253 00FD	254 00FE	255 00FF

**ТАБЛИЦА 5**  
Кодировка ISO 8859-5

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A0		Ë	Ђ	Ѓ	Є	Š	Ї	İ	Ј	Љ	Њ	Ћ	Ќ	-	Ў	Ц
	160 00A0	161 0401	162 0402	163 0403	164 0404	165 0405	166 0406	167 0407	168 0408	169 0409	170 040A	171 040B	172 040C	173 00AD	174 040E	175 040F
B0	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
	176 0410	177 0411	178 0412	179 0413	180 0414	181 0415	182 0416	183 0417	184 0418	185 0419	186 041A	187 041B	188 041C	189 041D	190 041E	191 041F
C0	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
	192 0420	193 0421	194 0422	195 0423	196 0424	197 0425	198 0426	199 0427	200 0428	201 0429	202 042A	203 042B	204 042C	205 042D	206 042E	207 042F
D0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
	208 0430	209 0431	210 0432	211 0433	212 0434	213 0435	214 0436	215 0437	216 0438	217 0439	218 043A	219 043B	220 043C	221 043D	222 043E	223 043F
E0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
	224 0440	225 0441	226 0442	227 0443	228 0444	229 0445	230 0446	231 0447	232 0448	233 0449	234 044A	235 044B	236 044C	237 044D	238 044E	239 044F
F0	№	ë	ђ	ѓ	є	š	ï	İ	ј	љ	њ	ћ	ќ	§	ў	ц
	240 2116	241 0451	242 0452	243 0453	244 0454	245 0455	246 0456	247 0457	248 0458	249 0459	250 045A	251 045B	252 045C	253 00A7	254 045E	255 045F

**ТАБЛИЦА 6**  
Кодировка koï8-r

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	 128 2500	 129 2502	 130 250C	 131 2510	 132 2514	 133 2518	 134 251C	 135 2524	 136 252C	 137 2534	 138 253C	 139 2580	 140 2584	 141 2588	 142 258C	 143 2590
90	 144 2591	 145 2592	 146 2593	 147 2320	 148 25A0	 149 2219	 150 221A	 151 2248	 152 2264	 153 2265	 154 00A0	 155 2321	 156 00B0	 157 00B2	 158 00B7	 159 00F7
A0	 160 2550	 161 2551	 162 2552	 163 0451	 164 2553	 165 2554	 166 2555	 167 2556	 168 2557	 169 2558	 170 2559	 171 255A	 172 255B	 173 255C	 174 255D	 175 255E
B0	 176 255F	 177 2560	 178 2561	 179 0401	 180 2562	 181 2563	 182 2564	 183 2565	 184 2566	 185 2567	 186 2568	 187 2569	 188 256A	 189 256B	 190 256C	 191 00A9
C0	Ю 192 044E	а 193 0430	б 194 0431	ц 195 0446	д 196 0434	е 197 0435	ф 198 0444	г 199 0433	х 200 0445	и 201 0438	й 202 0439	к 203 043A	л 204 043B	м 205 043C	н 206 043D	о 207 043E
D0	п 208 043F	я 209 044F	р 210 0440	с 211 0441	т 212 0442	у 213 0443	ж 214 0436	в 215 0432	ь 216 044C	ы 217 044B	з 218 0437	ш 219 0448	э 220 044D	щ 221 0449	ч 222 0447	ъ 223 044A
E0	Ю 224 042E	А 225 0410	Б 226 0411	Ц 227 0426	Д 228 0414	Е 229 0415	Ф 230 0424	Г 231 0413	Х 232 0425	И 233 0418	Й 234 0419	К 235 041A	Л 236 041B	М 237 041C	Н 238 041D	О 239 041E
F0	П 240 041F	Я 241 042F	Р 242 0420	С 243 0421	Т 244 0422	У 245 0423	Ж 246 0416	В 247 0412	Ь 248 042C	Ы 249 042B	З 250 0417	Ш 251 0428	Э 252 042D	Щ 253 0429	Ч 254 0427	Ъ 255 042A





**ТАБЛИЦА 8**  
Кодировка CP1251

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	Ђ 128 0402	Ѓ 129 0403	, 130 201A	ѓ 131 0453	„ 132 201E	… 133 2026	† 134 2020	‡ 135 2021	€ 136 20AC	% 137 2030	Љ 138 0409	‹ 139 2039	Њ 140 040A	Ќ 141 040C	ћ 142 040B	џ 143 040F
90	ђ 144 0452	‘ 145 2018	’ 146 2019	“ 147 201C	” 148 201D	● 149 2022	— 150 2013	— 151 2014	™ 152 2122	љ 154 0459	› 155 203A	њ 156 045A	ќ 157 045C	ћ 158 045B	џ 159 045F	
A0		Ў 161 040E	ў 162 045E	Ј 163 0408	Ѧ 164 00A4	Г 165 0490	І 166 00A6	§ 167 00A7	Ё 168 0401	© 169 00A9	€ 170 0404	« 171 00AB	¬ 172 00AC	- 173 00AD	® 174 00AE	İ 175 0407
B0	° 176 00B0	± 177 00B1	І 178 0406	і 179 0456	г 180 0491	μ 181 00B5	¶ 182 00B6	· 183 00B7	ё 184 0451	№ 185 2116	€ 186 0454	» 187 00BB	ј 188 0458	Ѕ 189 0405	ѕ 190 0455	ї 191 0457
C0	А 192 0410	Б 193 0411	В 194 0412	Г 195 0413	Д 196 0414	Е 197 0415	Ж 198 0416	З 199 0417	И 200 0418	Й 201 0419	К 202 041A	Л 203 041B	М 204 041C	Н 205 041D	О 206 041E	П 207 041F
D0	Р 208 0420	С 209 0421	Т 210 0422	У 211 0423	Ф 212 0424	Х 213 0425	Ц 214 0426	Ч 215 0427	Ш 216 0428	Щ 217 0429	Ъ 218 042A	Ы 219 042B	Ь 220 042C	Э 221 042D	Ю 222 042E	Я 223 042F
E0	а 224 0430	б 225 0431	в 226 0432	г 227 0433	д 228 0434	е 229 0435	ж 230 0436	з 231 0437	и 232 0438	й 233 0439	к 234 043A	л 235 043B	м 236 043C	н 237 043D	о 238 043E	п 239 043F
F0	р 240 0440	с 241 0441	т 242 0442	у 243 0443	ф 244 0444	х 245 0445	ц 246 0446	ч 247 0447	ш 248 0448	щ 249 0449	ъ 250 044A	ы 251 044B	ь 252 044C	э 253 044D	ю 254 044E	я 255 044F

## ТАБЛИЦА 9

Кодировка CP10007 (Macintosh Cyrillic)

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	А 128 0410	Б 129 0411	В 130 0412	Г 131 0413	Д 132 0414	Е 133 0415	Ж 134 0416	З 135 0417	И 136 0418	Й 137 0419	К 138 041A	Л 139 041B	М 140 041C	Н 141 041D	О 142 041E	П 143 041F
90	Р 144 0420	С 145 0421	Т 146 0422	У 147 0423	Ф 148 0424	Х 149 0425	Ц 150 0426	Ч 151 0427	Ш 152 0428	Щ 153 0429	Ъ 154 042A	Ы 155 042B	Ь 156 042C	Э 157 042D	Ю 158 042E	Я 159 042F
A0	† 160 2020	° 161 00B0	Г 162 0490	£ 163 00A3	§ 164 00A7	• 165 2022	¶ 166 00B6	І 167 0406	® 168 00AE	© 169 00A9	™ 170 2122	Ђ 171 0402	ђ 172 0452	≠ 173 2260	Ѓ 174 0403	ѓ 175 0453
B0	∞ 176 221E	± 177 00B1	≤ 178 2264	≥ 179 2265	ı 180 0456	μ 181 00B5	г 182 0491	Ј 183 0408	Є 184 0404	ё 185 0454	İ 186 0407	ï 187 0457	Љ 188 0409	љ 189 0459	Њ 190 040A	њ 191 045A
C0	ј 192 0458	Ѕ 193 0405	¬ 194 00AC	√ 195 221A	ƒ 196 0192	≈ 197 2248	Δ 198 2206	« 199 00AB	» 200 00BB	… 201 2026		Ѡ 203 040B	ѡ 204 045B	Ќ 205 040C	ќ 206 045C	ѕ 207 0455
D0	— 208 2013	— 209 2014	“ 210 201C	” 211 201D	‘ 212 2018	’ 213 2019	÷ 214 00F7	„ 215 201E	Ў 216 040E	ў 217 045E	Ѣ 218 040F	ѣ 219 045F	№ 220 2116	Ё 221 0401	ё 222 0451	я 223 044F
E0	а 224 0430	б 225 0431	в 226 0432	г 227 0433	д 228 0434	е 229 0435	ж 230 0436	з 231 0437	и 232 0438	й 233 0439	к 234 043A	л 235 043B	м 236 043C	н 237 043D	о 238 043E	п 239 043F
F0	р 240 0440	с 241 0441	т 242 0442	у 243 0443	ф 244 0444	х 245 0445	ц 246 0446	ч 247 0447	ш 248 0448	щ 249 0449	ъ 250 044A	ы 251 044B	ь 252 044C	э 253 044D	ю 254 044E	€ 255 20AC

**ТАБЛИЦА 10**

Латинская транслитерация

Буква	КОИ-7	ISO9	ГОСТ	Другие
А а	a A	A	A	
Б б	b B	B	B	
В в	w W	V	V	
Г г	g G	G	G	GH
Д д	d D	D	D	
Е е	e E	E	E	IE YE
Ё ё	з #	Ё	YO	IO E
Ж ж	v V	Ž	ZH	
З з	z Z	Z	Z	
И и	i I	I	I	
Й й	j J	J	J	Y
К к	k K	K	K	
Л л	l L	L	L	
М м	m M	M	M	
Н н	n N	N	N	
О о	o O	O	O	
П п	p P	P	P	
Р р	r R	R	R	
С с	s S	S	S	
Т т	t T	T	T	
У у	u U	U	U	
Ф ф	f F	F	F	PH
Х х	h H	H	X	KH
Ц ц	c C	C	C	CZ TS TZ
Ч ч	~ ^	Č	CH	
Ш ш	{ [	Š	SH	
Щ щ	} ]	Ŝ	SHH	SHCH SCH
Ъ ъ	-	”	“	,
Ы ы	y Y	Y	Y’	I Î
Ь ь	x X	,	‘	
Э э	\	È	E’	E Ê
Ю ю	‘ @	Û	YU	
Я я	q Q	Â	YA	

В КОИ-7 ЪЁё не входят — приведенные для них коды получены отбрасыванием 8-го бита у их кодов в koI8-r (код для Ъ получается 127 — это код невидимого управляющего символа ASCII).

## ТАБЛИЦА 11

### Структура расположения символов Unicode 3

В этой таблице приводится первый шестнадцатиричный код для каждой определяемой области. Последний код из этой области — это первый код следующей области минус 1, например, для кириллицы диапазон кодов — 0400–04FF.

L	0000 Basic Latin — базовая латиница
A	0080 Latin-1 Supplement — добавление 1 к латинице
A	0100 Latin Extended-A — латиница расширенная A
A	0180 Latin Extended-B — латиница расширенная B
a	0250 IPA <sup>1</sup> Extensions — расширения МФА
^	02B0 Spacing Modifier Letters — модификаторы промежутков
◌	0300 Combining Diacritical Marks — накладываемые диакритики
Ω	0370 Greek — греческие
Я	0400 Cyrillic — кириллица
U	0500 Cyrillic Supplement — добавление к кириллице
Թ	0530 Armenian — армянские
ה	0590 Hebrew — иврит
ع	0600 Arabic — арабские
ܣ	0700 Syriac — сирийские
𑄀	0780 Thaana — таана (Мальдивские острова)
क	0900 Devanagari — деванагари
ক	0980 Bengali — бенгали (бенгальские, ассамские)
ꣳ	0A00 Gurmukhi — гурмухи (синдхские, пенджабские)
ઉ	0A80 Gujarati — гуджаратские
ଞ	0B00 Oriya — ория (одри, уткали)
த	0B80 Tamil — тамильские
త	0C00 Telugu — телугу
ಕ	0C80 Kannada — каннара (каннада)
മ	0D00 Malayalam — малаялам
අ	0D80 Sinhala — сингальские
฿	0E00 Thai — тайские
ຮ	0E80 Lao — лаосские (лаотянские)
ཀ	0F00 Tibetan — тибетские
မ	1000 Myanmar — майанмар (бирманские)
Ⴐ	10A0 Georgian — грузинские
ᄀ	1100 Hangul Jamo — хангул (корейские) джамо (буквы)

---

<sup>1</sup> International Phonetic Alphabet — Международный фонетический алфавит

-  1200 Ethiopic — эфиопские
-  13A0 Cherokee — чероки (чироки)
-  1400 Unified Canadian Aboriginal Syllabic — унифицированные канадские аборигенные слоги
-  1680 Ogham — огамические (древне-ирландские)
-  16A0 Runic — руны
-  1700 Tagalog — тагальские (тагалогские)
-  1720 Hanunoo — хануну (Филиппины)
-  1740 Buhid — бухид, мангиан (Филиппины)
-  1760 Tagbanwa — тагбанва (Филиппины)
-  1780 Khmer — кхмерские
-  1800 Mongolian — монгольские
-  1E00 Latin Extended Additional — добавочные к расширенному латинскому
-  1F00 Greek Extended — расширенный греческий
-  2000 General Punctuation — пунктуация
-  2070 Superscripts and Subscripts — верхние и нижние индексы
-  20A0 Currency Symbols — валюты
-  20D0 Combining Marks for Symbols — накладываемые
-  2100 Letterlike Symbols — буквоподобные
-  2150 Number Forms — числовые формы
-  2190 Arrows — стрелки
-  2200 Mathematical Operators — математические операторы
-  2300 Miscellaneous Technical — смешанные технические
-  2400 Control Pictures — картинки управляющих кодов
-  2440 Optical Character Recognition — оптические символы распознавания
-  2460 Enclosed Alphanumerics — огороженные буквы и цифры
-  2500 Box Drawing — для рисования коробок
-  2580 Block Elements — элементы блоков
-  25A0 Geometric Shapes — геометрические формы
-  2600 Miscellaneous Symbols — разнообразные символы
-  2700 Dingbats — значки
-  27D0 Miscellaneous Mathematical Symbols-A — смешанные математические символы A
-  27F0 Supplemental Arrows-A — добавочные стрелки A
-  2800 Braille Patterns — шаблоны Брайля
-  2900 Supplemental Arrows-B — добавочные стрелки B
-  2980 Miscellaneous Mathematical Symbols-B — смешанные математические символы B
-  2A00 Supplemental Mathematical Operators — добавочные математические операторы

- 1 2E80 CJK<sup>2</sup> Radicals Supplement — дальневосточные корневые дополнения
- 2F00 Kangxi Radicals — кандзи корни
- 2FF0 Ideographic Description Characters — идеографического описания символы
- 3000 CJK Symbols and Punctuation — дальневосточные символы и пунктуация
- 3040 Hiragana — хирагана (японские)
- 30A0 Katakana — катакана (японские)
- 3100 Bopomofo — бопомофо (китайские)
- 3130 Hangeul Compatibility Jamo — ханул (корейские) джамо (буквы) совместимости
- 3190 Kanbun — канбун (японская идеографическая аннотация)
- 31A0 Bopomofo Extended — расширенные бопомофо (китайские)
- 3200 Enclosed CJK Letters and Months — огороженные дальневосточные буквы и месяцы
- 3300 CJK Compatibility — дальневосточные, совместимость
- 3400 CJK Unified Ideographs Extension A — дальневосточное унифицированное идеографическое расширение A
- 4E00 CJK Unified Ideographs — дальневосточные унифицированные идеографы
- A000 Yi Syllables — йи (ицзу) слоги
- A490 Yi Radicals — йи (ицзу) корни
- AC00 Hangeul Syllables — хангул (корейские) слоги
- D800 High Surrogates — верхние заменители (для UTF-16)
- DC00 Low Surrogates — нижние заменители (для UTF-16)
- E000 Private Use Area — для личного использования
- F900 CJK Compatibility Ideographs — дальневосточные идеографы совместимости
- fi FB00 Alphabetic Presentation Forms — буквенные формы
- FB50 Arabic Presentation Forms-A — арабские формы A
- FE00 Variation Selectors — селекторы вариантов
- FE20 Combining Half Marks — накладываемые полусимволы
- FE30 CJK Compatibility Forms — дальневосточные формы совместимости
- FE50 Small Form Variants — варианты маленьких форм
- FE70 Arabic Presentation Forms-B — арабские формы B
- FF00 Halfwidth and Fullwidth Forms — формы половинной и полной ширины

---

<sup>2</sup> China, Japan, Korea — Китай, Япония, Корея

- FFF0 Specials — специальные
- 10300 Old Italic — староитальянские
- 10330 Gothic — готические
- 10400 Deseret — дезерет (мормонские)
- 1D000 Byzantine Musical Symbols — византийские музыкальные
- 1D100 Musical Symbols — музыкальные
- 1D400 Mathematical Alphanumeric Symbols — математические буквенноцифровые
- 20000 CJK Unified Ideographs Extension B — дальневосточное унифицированное идеографическое расширение В
- 2F800 CJK Compatibility Ideographs Supplement — дополнительные дальневосточные идеографы совместимости
- E0000 Tags — ярлыки
- F0000 Supplementary Private Use Area-A — дополнительные для личного использования А
- 100000 Supplementary Private Use Area-B — дополнительные для личного использования В

---

















































Copyright © 2003 Лидовский Владимир Викторович.

Для подготовки материалов использовалась система Plain TeX

Опубликована в сильно переработаном и сокращенном варианте в журнале “КомпьюТерра” №42, 2003, с.38–40, в журнале “Магия ПК” №1, 2004, с.12–15 и в наиболее полном виде в журнале “Информационные технологии” №6, 2004, с.52–60.

# ПРИЛОЖЕНИЕ 1




## Основная кодировка ГОСТ

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	 128 2567	 129 2568	 130 2564	 131 2561	 132 2562	 133 2556	 134 2555	 135 2565	 136 2559	 137 2558	 138 2552	 139 255C	 140 255B	 141 255E	 142 255F	 143 2553
90	 144 2554	 145 2557	 146 255D	 147 255A	 148 2550	 149 2551	 150 2566	 151 2563	 152 2569	 153 2560	 154 256C	 155 2591	 156 2592	 157 2593	 158 256B	 159 256A
A0	 160 250C	 161 2510	 162 2518	 163 2514	 164 2500	 165 2502	 166 252C	 167 2524	 168 2534	 169 251C	 170 253C	 171 2588	 172 2584	 173 258C	 174 2590	 175 2580
B0	А 176 0410	Б 177 0411	В 178 0412	Г 179 0413	Д 180 0414	Е 181 0415	Ж 182 0416	З 183 0417	И 184 0418	Й 185 0419	К 186 041A	Л 187 041B	М 188 041C	Н 189 041D	О 190 041E	П 191 041F
C0	Р 192 0420	С 193 0421	Т 194 0422	У 195 0423	Ф 196 0424	Х 197 0425	Ц 198 0426	Ч 199 0427	Ш 200 0428	Щ 201 0429	Ъ 202 042A	Ы 203 042B	Ь 204 042C	Э 205 042D	Ю 206 042E	Я 207 042F
D0	а 208 0430	б 209 0431	в 210 0432	г 211 0433	д 212 0434	е 213 0435	ж 214 0436	з 215 0437	и 216 0438	й 217 0439	к 218 043A	л 219 043B	м 220 043C	н 221 043D	о 222 043E	п 223 043F
E0	р 224 0440	с 225 0441	т 226 0442	у 227 0443	ф 228 0444	х 229 0445	ц 230 0446	ч 231 0447	ш 232 0448	щ 233 0449	ъ 234 044A	ы 235 044B	ь 236 044C	э 237 044D	ю 238 044E	я 239 044F
F0	Ё 240 0401	ё 241 0451	/	\	/	\	→	←	↓	↑	÷	±	№	⊗	■	
	240 0401	241 0451	242	243	244	245	246 2192	247 2190	248 2193	249 2191	250 00F7	251 00B1	252 2116	253 00A4	254 25A0	255 00A0



















































## ПРИЛОЖЕНИЕ 2

“Болгарская” кодировка ВЦ АН СССР

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
	128 0410	129 0411	130 0412	131 0413	132 0414	133 0415	134 0416	135 0417	136 0418	137 0419	138 041A	139 041B	140 041C	141 041D	142 041E	143 041F
90	Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
	144 0420	145 0421	146 0422	147 0423	148 0424	149 0425	150 0426	151 0427	152 0428	153 0429	154 042A	155 042B	156 042C	157 042D	158 042E	159 042F
A0	а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
	160 0430	161 0431	162 0432	163 0433	164 0434	165 0435	166 0436	167 0437	168 0438	169 0439	170 043A	171 043B	172 043C	173 043D	174 043E	175 043F
B0	р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
	176 0440	177 0441	178 0442	179 0443	180 0444	181 0445	182 0446	183 0447	184 0448	185 0449	186 044A	187 044B	188 044C	189 044D	190 044E	191 044F
C0	└	┐	┘	┙	—	+	≡	∥	┌	└	≡	≡	≡	≡	≡	└
	192 2514	193 2534	194 252C	195 251C	196 2500	197 253C	198 2563	199 2551	200 255A	201 2554	202 2569	203 2566	204 2560	205 2550	206 256C	207 2510
D0					└	№	⊗	└	└	└	└	■	■	■	■	■
	208 2591	209 2592	210 2593	211 2502	212 2524	213 2116	214 00A4	215 2557	216 255D	217 2518	218 250C	219 2588	220 2584	221 258C	222 2590	223 2580
E0	α	β	Γ	π	Σ	σ	μ	τ	Φ	Θ	Ω	δ	∞	φ	ε	∩
	224 03B1	225 00DF	226 0393	227 03C0	228 03A3	229 03C3	230 00B5	231 03C4	232 03A6	233 0398	234 03A9	235 03B4	236 221E	237 03C6	238 03B5	239 2229
F0	≡	±	≥	≤	f	J	÷	≈	°	·	·	√	n	2	■	
	240 2261	241 00B1	242 2265	243 2264	244 2320	245 2321	246 00F7	247 2248	248 00B0	249 2219	250 00B7	251 221A	252 207F	253 00B2	254 25A0	255 00A0







### ПРИЛОЖЕНИЕ 3

#### Кодировка КОИ-8

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	 128 2514	 129 2534	 130 252C	 131 251C	 132 2500	 133 253C	 134 255E	 135 255F	 136 255A	 137 2554	 138 2569	 139 2566	 140 2560	 141 2550	 142 256C	 143 2567
90	 144 2568	 145 2564	 146 2565	 147 2559	 148 2558	 149 2552	 150 2553	 151 256B	 152 256A	 153 2518	 154 250C	 155 2588	 156 2584	 157 258C	 158 2590	 159 2580
A0	á 160 00E1	í 161 00ED	ó 162 00F3	ú 163 00FA	ñ 164 00F1	Ñ 165 00D1	а 166 00AA	о 167 00BA	ı 168 00BF	┐ 169 2310	┌ 170 00AC	½ 171 00BD	¼ 172 00BC	ı 173 00A1	« 174 00AB	» 175 00BB
B0	 176 2591	 177 2592	 178 2593	 179 2502	 180 2524	 181 2561	 182 2562	 183 2556	 184 2555	 185 2563	 186 2551	 187 2557	 188 255D	 189 255C	 190 255B	 191 2510
C0	Ю 192 044E	а 193 0430	б 194 0431	ц 195 0446	д 196 0434	е 197 0435	ф 198 0444	г 199 0433	х 200 0445	и 201 0438	й 202 0439	к 203 043A	л 204 043B	м 205 043C	н 206 043D	о 207 043E
D0	п 208 043F	я 209 044F	р 210 0440	с 211 0441	т 212 0442	у 213 0443	ж 214 0436	в 215 0432	ь 216 044C	ы 217 044B	з 218 0437	ш 219 0448	э 220 044D	щ 221 0449	ч 222 0447	ъ 223 044A
E0	Ю 224 042E	А 225 0410	Б 226 0411	Ц 227 0426	Д 228 0414	Е 229 0415	Ф 230 0424	Г 231 0413	Х 232 0425	И 233 0418	Й 234 0419	К 235 041A	Л 236 041B	М 237 041C	Н 238 041D	О 239 041E
F0	П 240 041F	Я 241 042F	Р 242 0420	С 243 0421	Т 244 0422	У 245 0423	Ж 246 0416	В 247 0412	Ь 248 042C	Ы 249 042B	З 250 0417	Ш 251 0428	Э 252 042D	Щ 253 0429	Ч 254 0427	

## ПРИЛОЖЕНИЕ 4

Кодировка CP850

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
80	Ç 128 00C7	ü 129 00FC	é 130 00E9	â 131 00E2	ä 132 00E4	à 133 00E0	å 134 00E5	ç 135 00E7	ê 136 00EA	ë 137 00EB	è 138 00E8	ï 139 00EF	î 140 00EE	ì 141 00EC	Ä 142 00C4	Å 143 00C5
90	É 144 00C9	æ 145 00E6	Æ 146 00C6	ô 147 00F4	ö 148 00F6	ò 149 00F2	û 150 00FB	ù 151 00F9	ÿ 152 00FF	Ö 153 00D6	Ü 154 00DC	ø 155 00F8	£ 156 00A3	Ø 157 00D8	× 158 00D7	ƒ 159 0192
A0	á 160 00E1	í 161 00ED	ó 162 00F3	ú 163 00FA	ñ 164 00F1	Ñ 165 00D1	à 166 00AA	ó 167 00BA	ı 168 00BF	® 169 00AE	¬ 170 00AC	½ 171 00BD	¼ 172 00BC	ı 173 00A1	« 174 00AB	» 175 00BB
B0				 179 2502	⊥ 180 2524	Á 181 00C1	Â 182 00C2	À 183 00C0	© 184 00A9	≡ 185 2563	∥ 186 2551	⊥ 187 2557	⊥ 188 255D	¢ 189 00A2	¥ 190 00A5	⊥ 191 2510
C0	⌒ 192 2514	⊥ 193 2534	⊥ 194 252C	⊥ 195 251C	— 196 2500	⊥ 197 253C	ã 198 00E3	Ã 199 00C3	⌒ 200 255A	⊥ 201 2554	⊥ 202 2569	⊥ 203 2566	⊥ 204 2560	⊥ 205 2550	⊥ 206 256C	⊕ 207 00A4
D0	ð 208 00F0	Ð 209 00D0	Ê 210 00CA	Ë 211 00CB	È 212 00C8	ı 213 0131	Í 214 00CD	Î 215 00CE	Ï 216 00CF	⊥ 217 2518	⊥ 218 250C			 221 00A6	Ì 222 00CC	
F0	Ó 224 00D3	ß 225 00DF	Ô 226 00D4	Ò 227 00D2	õ 228 00F5	Õ 229 00D5	μ 230 00B5	þ 231 00FE	ƒ 232 00DE	Ú 233 00DA	Û 234 00DB	Û 235 00D9	ý 236 00FD	Ý 237 00DD	- 238 00AF	' 239 00B4
90	- 240 00AD	± 241 00B1	= 242 2017	¾ 243 00BE	¶ 244 00B6	§ 245 00A7	÷ 246 00F7	˘ 247 00B8	° 248 00B0	¨ 249 00A8	· 250 00B7	1 251 00B9	3 252 00B3	2 253 00B2	■ 254 25A0	 255 00A0