

Ukrainian Character Set KOI8-U

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Copyright (C) The Internet Society (1998). All Rights Reserved.

Abstract

This document provides information about character encoding KOI8-U (KOI8 Ukrainian) which is a de-facto standard in Ukrainian Internet community. KOI8-U is compatible with KOI8-R (RFC 1489) in all Russian letters and extends it with four Ukrainian letters which locations are compliant with ISO-IR-111. The official site of KOI8-U Working Group is <http://www.net.ua>.

Introduction

This document provides information about character encoding KOI8-U (KOI8 Ukrainian) widely used in Ukrainian Internet community for mail and news exchange as well as for presentation of WWW information resources in Ukrainian language.

Originally, specification of proposed standard koi8-u was officially adopted by the conference of Postmasters of Ukrainian Internet Service Providers in Slavsk in Autumn of 1992 presented by Igor Sviridov from Kiev and Stas Vorony from Kharkiv. Later in June 1995 this specification was completed with UKRAINIAN GHE WITH UPTURN.

KOI8-U (KOI8 Ukrainian) is a de-facto standard supported in many operation systems and Internet user applications including encoding tables, fonts, locale support for many operating systems and environments.

MIME character set name: koi8-u

Relation to other RFCs

This standard is based on several published standards: RFC1489 (it is fully compatible in all Russian letters), RFC-1345, ISO-IR-111, ISO 10646.

Compatibility with other character sets

The lower part of the KOI8-U Ukrainian Character Set is a complete copy of ASCII, just as it's used in KOI8-R and other non-ASCII codepages.

The upper part of the KOI8-U Character Set contains all Russian letters defined in KOI8-R and four Ukrainian letters (#164, #180 - ukr. ie, #166, #182 - ukr. i, #167, #183 - ukr. yi, #173, #189 - ukr. ghe with upturn) which locations are compliant with ISO-IR-111.

BOX DRAWINGS elements in the other positions (that are not used by Ukrainian letters) are the same as in KOI8-R character set. List of all differences between KOI8-R and proposed KOI8-U is given in APPENDIX A.

Specification of the upper part of KOI8-U codepage

The description of all characters of upper half part of KOI8-U codepage is given according to ISO 10646 Unicode Character Set (UCS).

KOI8-U charset table in RFC1345 format is given in Appendix B.

<decimal>	<hex-code>	<UCS>	<description>
128	80	U2500	BOX DRAWINGS LIGHT HORIZONTAL
129	81	U2502	BOX DRAWINGS LIGHT VERTICAL
130	82	U250C	BOX DRAWINGS LIGHT DOWN AND RIGHT
131	83	U2510	BOX DRAWINGS LIGHT DOWN AND LEFT
132	84	U2514	BOX DRAWINGS LIGHT UP AND RIGHT
133	85	U2518	BOX DRAWINGS LIGHT UP AND LEFT
134	86	U251C	BOX DRAWINGS LIGHT VERTICAL AND RIGHT
135	87	U2524	BOX DRAWINGS LIGHT VERTICAL AND LEFT
136	88	U252C	BOX DRAWINGS LIGHT DOWN AND HORIZONTAL
137	89	U2534	BOX DRAWINGS LIGHT UP AND HORIZONTAL
138	8A	U253C	BOX DRAWINGS LIGHT VERTICAL AND HORIZONTAL
139	8B	U2580	UPPER HALF BLOCK
140	8C	U2584	LOWER HALF BLOCK
141	8D	U2588	FULL BLOCK
142	8E	U258C	LEFT HALF BLOCK
143	8F	U2590	RIGHT HALF BLOCK

144	90	U2591	LIGHT SHADE
145	91	U2592	MEDIUM SHADE
146	92	U2593	DARK SHADE
147	93	U2320	TOP HALF INTEGRAL
148	94	U25A0	BLACK SQUARE
149	95	U2219	BULLET OPERATOR
150	96	U221A	SQUARE ROOT
151	97	U2248	ALMOST EQUAL TO
152	98	U2264	LESS THAN OR EQUAL TO
153	99	U2265	GREATER THAN OR EQUAL TO
154	9A	U00A0	NO-BREAK SPACE
155	9B	U2321	BOTTOM HALF INTEGRAL
156	9C	U00B0	DEGREE SIGN
157	9D	U00B2	SUPERSCRIFT TWO
158	9E	U00B7	MIDDLE DOT
159	9F	U00F7	DIVISION SIGN
160	A0	U2550	BOX DRAWINGS DOUBLE HORIZONTAL
161	A1	U2551	BOX DRAWINGS DOUBLE VERTICAL
162	A2	U2552	BOX DRAWINGS DOWN SINGLE AND RIGHT DOUBLE
163	A3	U0451	CYRILLIC SMALL LETTER IO
164	A4	U0454	CYRILLIC SMALL LETTER UKRAINIAN IE
165	A5	U2554	BOX DRAWINGS DOUBLE DOWN AND RIGHT
166	A6	U0456	CYRILLIC SMALL LETTER BYELORUSSIAN- UKRAINIAN I
167	A7	U0457	CYRILLIC SMALL LETTER YI (UKRAINIAN)
168	A8	U2557	BOX DRAWINGS DOUBLE DOWN AND LEFT
169	A9	U2558	BOX DRAWINGS UP SINGLE AND RIGHT DOUBLE
170	AA	U2559	BOX DRAWINGS UP DOUBLE AND RIGHT SINGLE
171	AB	U255A	BOX DRAWINGS DOUBLE UP AND RIGHT
172	AC	U255B	BOX DRAWINGS UP SINGLE AND LEFT DOUBLE
173	AD	U0491	CYRILLIC SMALL LETTER GHE WITH UPTURN
174	AE	U255D	BOX DRAWINGS DOUBLE UP AND LEFT
175	AF	U255E	BOX DRAWINGS VERTICAL SINGLE AND RIGHT DOUBLE
176	B0	U255F	BOX DRAWINGS VERTICAL DOUBLE AND RIGHT SINGLE
177	B1	U2560	BOX DRAWINGS DOUBLE VERTICAL AND RIGHT
178	B2	U2561	BOX DRAWINGS VERTICAL SINGLE AND LEFT DOUBLE
179	B3	U0401	CYRILLIC CAPITAL LETTER IO
180	B4	U0404	CYRILLIC CAPITAL LETTER UKRAINIAN IE
181	B5	U2563	BOX DRAWINGS DOUBLE VERTICAL AND LEFT
182	B6	U0406	CYRILLIC CAPITAL LETTER BYELORUSSIAN-UKRAINIAN I
183	B7	U0407	CYRILLIC CAPITAL LETTER YI (UKRAINIAN)
184	B8	U2566	BOX DRAWINGS DOUBLE DOWN AND HORIZONTAL
185	B9	U2567	BOX DRAWINGS UP SINGLE AND

			HORIZONTAL DOUBLE
186	BA	U2568	BOX DRAWINGS UP DOUBLE AND HORIZONTAL SINGLE
187	BB	U2569	BOX DRAWINGS DOUBLE UP AND HORIZONTAL
188	BC	U256A	BOX DRAWINGS VERTICAL SINGLE AND HORIZONTAL DOUBLE
189	BD	U0490	CYRILLIC CAPITAL LETTER GHE WITH UPTURN
190	BE	U256C	BOX DRAWINGS DOUBLE VERTICAL AND HORIZONTAL
191	BF	U00A9	COPYRIGHT SIGN
192	C0	U044E	CYRILLIC SMALL LETTER YU
193	C1	U0430	CYRILLIC SMALL LETTER A
194	C2	U0431	CYRILLIC SMALL LETTER BE
195	C3	U0446	CYRILLIC SMALL LETTER TSE
196	C4	U0434	CYRILLIC SMALL LETTER DE
197	C5	U0435	CYRILLIC SMALL LETTER IE
198	C6	U0444	CYRILLIC SMALL LETTER EF
199	C7	U0433	CYRILLIC SMALL LETTER GHE
200	C8	U0445	CYRILLIC SMALL LETTER KHA
201	C9	U0438	CYRILLIC SMALL LETTER I
202	CA	U0439	CYRILLIC SMALL LETTER SHORT I
203	CB	U043A	CYRILLIC SMALL LETTER KA
204	CC	U043B	CYRILLIC SMALL LETTER EL
205	CD	U043C	CYRILLIC SMALL LETTER EM
206	CE	U043D	CYRILLIC SMALL LETTER EN
207	CF	U043E	CYRILLIC SMALL LETTER O
208	D0	U043F	CYRILLIC SMALL LETTER PE
209	D1	U044F	CYRILLIC SMALL LETTER YA
210	D2	U0440	CYRILLIC SMALL LETTER ER
211	D3	U0441	CYRILLIC SMALL LETTER ES
212	D4	U0442	CYRILLIC SMALL LETTER TE
213	D5	U0443	CYRILLIC SMALL LETTER U
214	D6	U0436	CYRILLIC SMALL LETTER ZHE
215	D7	U0432	CYRILLIC SMALL LETTER VE
216	D8	U044C	CYRILLIC SMALL LETTER SOFT SIGN
217	D9	U044B	CYRILLIC SMALL LETTER YERU
218	DA	U0437	CYRILLIC SMALL LETTER ZE
219	DB	U0448	CYRILLIC SMALL LETTER SHA
220	DC	U044D	CYRILLIC SMALL LETTER E
221	DD	U0449	CYRILLIC SMALL LETTER SHCHA
222	DE	U0447	CYRILLIC SMALL LETTER CHE
223	DF	U044A	CYRILLIC SMALL LETTER HARD SIGN
224	E0	U042E	CYRILLIC CAPITAL LETTER YU
225	E1	U0410	CYRILLIC CAPITAL LETTER A
226	E2	U0411	CYRILLIC CAPITAL LETTER BE
227	E3	U0426	CYRILLIC CAPITAL LETTER TSE
228	E4	U0414	CYRILLIC CAPITAL LETTER DE
229	E5	U0415	CYRILLIC CAPITAL LETTER IE

230	E6	U0424	CYRILLIC CAPITAL LETTER EF
231	E7	U0413	CYRILLIC CAPITAL LETTER GHE
232	E8	U0425	CYRILLIC CAPITAL LETTER KHA
233	E9	U0418	CYRILLIC CAPITAL LETTER I
234	EA	U0419	CYRILLIC CAPITAL LETTER SHORT I
235	EB	U041A	CYRILLIC CAPITAL LETTER KA
236	EC	U041B	CYRILLIC CAPITAL LETTER EL
237	ED	U041C	CYRILLIC CAPITAL LETTER EM
238	EE	U041D	CYRILLIC CAPITAL LETTER EN
239	EF	U041E	CYRILLIC CAPITAL LETTER O
240	F0	U041F	CYRILLIC CAPITAL LETTER PE
241	F1	U042F	CYRILLIC CAPITAL LETTER YA
242	F2	U0420	CYRILLIC CAPITAL LETTER ER
243	F3	U0421	CYRILLIC CAPITAL LETTER ES
244	F4	U0422	CYRILLIC CAPITAL LETTER TE
245	F5	U0423	CYRILLIC CAPITAL LETTER U
246	F6	U0416	CYRILLIC CAPITAL LETTER ZHE
247	F7	U0412	CYRILLIC CAPITAL LETTER VE
248	F8	U042C	CYRILLIC CAPITAL LETTER SOFT SIGN
249	F9	U042B	CYRILLIC CAPITAL LETTER YERU
250	FA	U0417	CYRILLIC CAPITAL LETTER ZE
251	FB	U0428	CYRILLIC CAPITAL LETTER SHA
252	FC	U042D	CYRILLIC CAPITAL LETTER E
253	FD	U0429	CYRILLIC CAPITAL LETTER SHCHA
254	FE	U0427	CYRILLIC CAPITAL LETTER CHE
255	FF	U042A	CYRILLIC CAPITAL LETTER HARD SIGN

Security Considerations

This memo raises no known security issues.

Acknowledgments

The present edition of this document was prepared by joint activity of KOI8-U Working group and is a result of wide discussion in Ukrainian USENET newsgroup ukr.nodes and consensus reached among majority of Ukrainian ISPs.

Special acknowledges to:

Andrew Chernov <ache@astral.msk.su>, author of the first on Internet RFC 1489 describing KOI8-R Russian character set;

Igor Sviridov <sia@nest.org> for the initial work on establishing and support of KOI8-U character set and it's implementation in first e-mail products.

Many people have contributed to the early work on koi8-u encoding:

Stanislav V. Voronyi <stas@uanet.kharkov.ua>
Serge Vakulenko <vak@zebub.msk.su>
Lena Savchenko <epsav@eps.computerland.kiev.ua>
Igor Romanenko <igor@carrier.kiev.ua>
Ruslan Belkin <rus@UA.net>
Andrey Blohintsev <bag@UA.net>

References

- [1] Chernov, A., "Registration of a Cyrillic Character Set", RFC 1489, July 1993.
- [2] UNICODE 2.0 CHARACTER DATABASE. - <ftp://unicode.org/pub/2.0-Update/UnicodeData-2.0.14.txt>
- [3] Ukrainian letters in koi8-u and other character sets
ftp://ftp.ua.net/pub/info/encodings/koi8-u/ukr_chars_in_koi8-u_and_others.txt, June 1995.
- [4] ECMA-CYRILLIC. - <ftp://dkuug.dk/il8n/charmaps.all/ECMA-CYRILLIC>
- [5] Simonsen, K., "Character Mnemonics & Character Sets" RFC 1345, June 1992.

KOI8-U Working Group List

Coordinator:

Alexander Yeremenko <koi8-u@sita.kiev.ua>

Yuri Demchenko <demch@cad.ntu-kpi.kiev.ua>

Victor Forsyuk <victor@gu.net>

Taras Heychenko <tasic@lucky.net>

Pavel Gulchuk <gul@lucky.net>

Dmitry Kohmanyuk <dk@farm.org>

Boris Mostovoy <vms@breaker.gu.net>

Helen Panchenko <elena@alex-ua.com>

Igor Romanenko <igor@lucky.net>

Eugene Sherstobitov <gene@lucky.net>

Andrew Stesin <stesin@gu.net>

Igor Sviridov <sia@nest.org>

Roman A. Tkachuk <roman@bit.ternopil.ua>

APPENDIX A

DIFFERENCE OF KOI8-U from KOI8-R (RFC 1489)

KOI8-U is compatible with KOI8-R in all Cyrillic Letters and completes it with four Ukrainian letters UKRAINIAN IE #164, #180, CYRILLIC SMALL LETTER BELORUSSIAN-UKRAINIAN I #166, #182, UKRAINIAN YI #167, #183, UKRAINIAN GHE WITH UPTURN #173, #189.

<decimal>	<hex-code>	<UCS>	<description>
164	A4	U0454	CYRILLIC SMALL LETTER UKRAINIAN IE
166	A6	U0456	CYRILLIC SMALL LETTER BELORUSSIAN-UKRAINIAN I
167	A7	U0457	CYRILLIC SMALL LETTER YI (UKRAINIAN)
173	AD	U0491	CYRILLIC SMALL LETTER UKRAINIAN GHE WITH UPTURN
180	B4	U0403	CYRILLIC CAPITAL LETTER UKRAINIAN IE
182	B6	U0406	CYRILLIC CAPITAL LETTER BELORUSSIAN-UKRAINIAN I
183	B7	U0407	CYRILLIC CAPITAL LETTER YI (UKRAINIAN)
189	BD	U0490	CYRILLIC CAPITAL LETTER UKRAINIAN GHE WITH UPTURN

APPENDIX B

KOI8-U charset table in RFC1345 format

```
&charset KOI8-U
&rem source: RFC 2319
&rem Mibenum: 2088
&rem source: http://www.net.ua/KOI8-U/
&bits 8
&code 0
NU SH SX EX ET EQ AK BL BS HT LF VT FF CR SO SI
DL D1 D2 D3 D4 NK SY EB CN EM SB EC FS GS RS US
SP ! " Nb DO % & ' ( ) * + , - . / 0 1 2 3 4 5 6 7 8 9 : ; < = > ?
At A B C D E F G H I J K L M N O P Q R S T U V W X Y Z <( // )> '> _
'! a b c d e f g h i j k l m n o p q r s t u v w x y z (! !! !) '? DT
hh vv dr dl ur ul vr vl dh uh vh TB LB FB lB RB
.S :S ?S Iu fS Sb RT ?2 =< >= NS Il DG 2S .M -:
HH VV dR io ie DR ii yi LD uR Ur UR uL g3 UL vR
Vr VR vL IO IE VL II YI DH uH Uh UH vH G3 VH Co
ju a= b= c= d= e= f= g= h= i= j= k= l= m= n= o=
p= ja r= s= t= u= z% v= %' y= z= s% je sc c% ='
JU A= B= C= D= E= F= G= H= I= J= K= L= M= N= O=
P= JA R= S= T= U= Z% V= %" Y= Z= S% JE Sc C% ="
```


Full Copyright Statement

Copyright (C) The Internet Society (1998). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

