         Internationalized Domain Names Registration and Administration
                 Guidelines for European Languages Using Cyrillic

Abstract

   This document is a guideline for registries and registrars on
   registering internationalized domain names (IDNs) based on (in
   alphabetical order) Bosnian, Bulgarian, Byelorussian, Kildin Sami,
   Macedonian, Montenegrin, Russian, Serbian, and Ukrainian languages in
   a DNS zone. It describes appropriate characters for registration and
   variant considerations for characters from Greek and Latin scripts
   with similar appearances and/or derivations.

Status of This Memo

   This document is not an Internet Standards Track specification; it is
   published for informational purposes.

   This is a contribution to the RFC Series, independently of any other
   RFC stream. The RFC Editor has chosen to publish this document at
   its discretion and makes no statement about its value for
   implementation or deployment. Documents approved for publication by
   the RFC Editor are not a candidate for any level of Internet
   Standard; see Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc5992.

Table of Contents

1. Introduction

   Cyrillic is one of a fairly small number of scripts that are used,
   with different subsets of characters, to write a large number of
   languages, some of which are not closely related to the others. When
   those languages might be used together in a zone (typical of generic
   TLDs (gTLDs) but likely in other zones both at and below the root),
   special considerations for intermixing characters may apply.
   Cyrillic also has the property that, while it is usually considered a
   separate script from the Latin (Roman) and Greek ones, it shares many
   characters with them, creating opportunities for visual confusion.
   Those difficulties are especially pronounced when "all of Cyrillic"
   is used rather than only the characters associated with a particular
   language.

   This specification provides guidelines for the use of Cyrillic, as
   encoded in Unicode [Unicode52] with internationalized domain name
   (IDN) labels derived from most "European" languages that use the
   script (use of the term "European" is a convenience, since there is
   disagreement about the relevant boundaries for different purposes
   and, of course, much of Russia lies within geological Asia).
   Specifically, it covers (in alphabetic order) Bosnian, Bulgarian,

Byelorussian, the Kildin member of the Sami (often written "Saami")
language family, Macedonian, Montenegrin, Russian, Serbian, and
Ukrainian. Supplemental tables, based on information in the Unicode
Standard and a recently completed Montenegrin government standard
[MontenegrinChars] are provided for use with Montenegrin. Moldovan
is no longer in official use with Cyrillic script: no registrations
are considered likely in Cyrillic, at least within the relevant
ccTLD, and it is not further discussed in this document. Languages
of Asia that use Cyrillic are not considered here and should be the
subject of separate specifications.

While Cyrillic script is the primary one used for many of the
relevant languages and countries, Latin script is often used instead
of, or in combination with, it. Standard keyboards used in most of
the countries have both Cyrillic and Latin characters. Therefore,
some registries could use Latin scripts for domain name registration
in their zones. From time to time, some registries and users have
claimed that there is a requirement for mixing Cyrillic and Latin
characters in the same label. We strongly recommend against such
mixing as user confusion is almost certain to result. In addition,
registries that support many scripts will probably encounter the need
to support labels in Greek or Latin scripts as well as Cyrillic, and
a large number of character forms are shared among those three
scripts.

Because the DNS has no way for the end user to distinguish among the
languages that might have been used to inspire a particular label, it
seems useful to treat the characters of a large number of languages
that use Cyrillic in their writing systems together, rather than
trying to differentiate them. The discussion and tables in this
specification should provide a foundation for developing more
restrictive rules for zones in which only a single language is likely
to be used, but it does not specify those language-specific rules.

Readers of this document should be aware that its recommendations are
about use in DNS labels. The orthography for some of the languages
involved, especially Kildin Sami, is not completely standardized and
local usage sometimes permits substitution of Latin-based characters
for their Cyrillic equivalents. Unless they are required by official
orthographies, those substitutions should generally be avoided in DNS
labels because of the risk of additional user confusion with the
Latin characters that are visually similar.

1.1. Similar Characters and Variants

For some human languages, there are characters and/or strings that
have equivalent or near-equivalent meanings. If someone is allowed
to register a name with such a character or string, the registry

   might want to automatically register all the names that have the same
   meaning in that language. Further, some registries might want to
   restrict the set of characters to be registered for language-based
   reasons.

   So-called "variant techniques", introduced in the JET specification
   for the CJK script [RFC3743] and its generalization [RFC4290],
   describe ways of registering IDNs to decrease the risk of
   misunderstandings, cybersquatting, and other forms of confusion.

   The tables below (Appendix A) identify confusable characters in Latin
   and Greek scripts that might be easily confused with Cyrillic ones.

   As with variant approaches for other scripts (e.g., see RFC 4713
   [RFC4713] for the Chinese language or RFC 5564 [RFC5564] for the
   Arabic language), this document identifies sets of characters that
   need special consideration and provides information about them. A
   registry that handles names using these characters can then make a
   policy decision about how to actually handle them. The options for
   those policy decisions would include automatically registering all
   look-alike strings to the same registrant, registering one such
   string and blocking the others, and so on.

1.2. Terminology

   The terminology that follows is derived from the JET specification
   for the CJK script [RFC3743] and its generalization [RFC4290], but
   this specification does not depend on them. All characters listed
   here have been verified to be "PVALID" under the IDNA2008
   specification [RFC5890] [RFC5892].

   A "string" is a sequence of one or more characters.

   This document discusses characters that have equivalent or near-
   equivalent characters or strings. The "base character" is the
   character that has one or more equivalents; the "variant(s)" are the
   character(s) and/or string(s) that are equivalent to the base
   character.

   A "registration bundle" is the set of all labels that comes from
   expanding all base characters for a single name into their variants.

   A registry is the administrative authority for a DNS zone. That is,
   the registry is the body that makes and enforces policies that are
   used in a particular zone in the DNS. The term "registry" applies to
   all zones in the DNS, not only those that exist at the top level.

2. Languages and Characters

   In the interest of clarity and balance, this document describes a
   "Base Cyrillic" set of 23 characters for use in comparing the
   character usage for Russian and Central European languages that use
   Cyrillic. The balance of this section compares the character usage
   of the individual languages in that group.

   "Base Cyrillic" consists of the following Unicode code points (names
   associated with these code points and those below appear in
   Appendix A): U+0430, U+0431, U+0432, U+0433, U+0434, U+0435, U+0436,
   U+0437, U+043A, U+043B, U+043C, U+043D, U+043E, U+043F, U+0440,
   U+0441, U+0442, U+0443, U+0444, U+0445, U+0446, U+0447, U+0448.

   In addition, modern writing systems that use Cyrillic do not have
   digits separate from the "European" ones used with Latin characters.
   For registries that permit digits to appear in domain name labels,
   the "Base Cyrillic" code point listed above should be considered to
   include U+0030, U+0031, U+0032, U+0033, U+0034, U+0035, U+0036,
   U+0037, U+0038, and U+0039 (Digit Zero, and Digit One through Digit
   Nine). The Hyphen-Minus character (U+002D) may also be used.

   It is worth noting that the EU top-level domain registry allows
   Cyrillic registrations using 32 code points [EU-registry]. That list
   is sufficient for some of the languages listed here but not for
   others.

   The individual languages that are the focus of this specification are
   discussed below (in English alphabetical order).

2.1. Bosnian and Serbian

   Bosnian and Serbian have 30 letters in the alphabet and the
   additional seven characters to the base of 23 shared Cyrillic
   characters: U+0438, U+0458, U+0452, U+0459, U+045A, U+045B, U+045F.

2.2. Bulgarian

   The Bulgarian alphabet has 30 characters, seven in addition to the
   basic 23: U+0438, U+0439, U+0449, U+044A, U+044C, U+044E, U+044F.

2.3. Byelorussian (Belarusian, Belarusan)

   The Byelorussian (now often spelled Belarusian or Belarusan) alphabet
   has 32 characters, i.e., nine characters in addition to the Base
   Cyrillic set of 23 characters: U+0451, U+0456, U+0439, U+044B,
   U+044C, U+045E, U+044D, U+044E, U+044F.

2.4. Kildin Sami

   The phonetics of the Kildin Sami are quite complex and not easily
   represented in Cyrillic (see, e.g., Kertom's work [Kert]). The
   orthography is not standardized and the writing system may best be
   thought of as an attempt to transcribe the language phonetically
   (primary in Latin script in the 1930s but in Cyrillic more recently).
   Different scholars have reported different numbers of phonemes,
   further complicating the transcription process. Kertom identifies 53
   consonants with long-short distinctions and, in many cases, hard-soft
   ones. He also identifies ascending and descending diphthongs and one
   triphthong as well as more common short and long vowels.

   The primary reference for Kildin Sami, widely circulated for some
   time but only in draft, is apparently used by Sami language(s)
   experts in Scandinavian countries [Riessl07]. It, and the references
   it cites, uses 56 characters, 33 of which do not appear in the basic
   set. Eight* of these characters have no precomposed forms in Unicode
   and hence must be written as a sequence of two code points with the
   second one being COMBINING MACRON (U+0304). Using parentheses to
   make the two-code-point sequences more obvious, the additional
   characters are: (U+0430 U+0304)*, (U+0435 U+0304)*, U+0438, U+0439,
   (U+043E U+0304), U+044A, U+044B, (U+044B U+0304), U+044C, U+044D,
   (U+044D U+0304), U+044E, (U+044E U+0304), U+044F, (U+044F U+0304),
   U+0451, (U+0451 U+0304), U+0458, U+048B, U+048D, U+048F, U+04BB,
   U+04C6, U+04C8, U+04CA, U+04CE, U+04D3, U+04E3, U+04E7, U+04ED,
   U+04EF, U+04F1, U+04F9.

   * These characters, CYRILLIC SMALL LETTER A (U+0430) with a
      COMBINING MACRON (U+0304) and CYRILLIC SMALL LETTER IE (U+0435)
      with a COMBINING MACRON (U+0304), respectively, have the same
      visual appearance as LATIN SMALL LETTER A WITH MACRON (U+0101) and
      LATIN SMALL LETTER E WITH MACRON (U+0113). There are no known
      keyboards designed specifically for Kildin Sami. If an extended
      Latin-based keyboard and associated software are used, these
      characters might appear with the code point based on Latin (e.g.,
      U+0113 for the second case). By contrast, keyboards and input
      software that are designed to be more Cyrillic-friendly are more
      likely to produce code points for the Cyrillic base characters.
      The use of a Latin character base for that second case occurs in
      some Western European sources including Riessler's work
      [Riessl07]. While we have not found explicit substitutions for A
      with Macron, we believe they might be found in practice. These
      alternatives are not mapped together by Unicode Normalization Form
      C (NFC) (or Normalization Form KC (NFKC)), so registries, and
      possibly applications software, should exercise some care about

      these coding variations. However, U+0101 and U+0113 are Latin
      Script characters so, if either is used, any tests on homogeneity
      of the script within a label need to be made with care.

      Similar issues may apply to other Kildin Sami characters
      constructed with combining sequences.

   The key references in Russian ([Anto90], [Kert86], [Kuru85]) all
   propose slightly different character tables relative to each other
   and to Riessler's list. Because the latter list appears to be more
   comprehensive and to represent more recent scholarship, we have based
   the tables in this document on it. We recommend, however, that
   registries review these recommendations and the relevant papers
   should registration requests for Kildin Sami actually appear.

   Additional perspectives on Kildin Sami can be found on the Omniglot
   Sami pages [OmniglotSaami].

2.5. Macedonian

   Macedonian has 31 characters in the alphabet. This is eight in
   addition to the basic set: U+0438, U+0458, U+0452, U+0459, U+045A,
   U+045C, U+045F, U+0491, U+0455.

2.6. Montenegrin

   According to the most recent, and now final, government specification
   [MontenegrinChars], Montenegrin has 32 characters in its alphabet,
   including two that have no precomposed forms in Unicode. This is
   nine in addition to the basic set and two in addition to Bosnian and
   Serbian: U+0437 U+0301, U+0438, U+0441 U+0301, U+0452, U+0458,
   U+0459, U+045A, U+045B, U+045F.

   See Bosnian, Section 2.1, above.

2.7. Russian

   The current Russian alphabet has 33 characters, consisting of the
   Base Cyrillic set plus an additional ten characters: U+0451, U+0438,
   U+0439, U+0449, U+044A, U+044B, U+044C, U+044D, U+044E, U+044F.

2.8. Serbian

   See Bosnian, Section 2.1, above.

2.9. Ukrainian

   The character list for modern Ukrainian has apparently not completely
   stabilized. Some references claim 31 characters and therefore an
   additional 8 characters to the Base Cyrillic set of 23. Others claim
   33, adding U+0438 and U+0439 and replacing U+044A (Hard Sign) with
   U+044C (Soft Sign), for a total of an additional 11 characters as
   compared to the Base Cyrillic set. Unless better information is
   available, the prudent registry should probably assume that all 34
   characters are in use, i.e., the Base Cyrillic set plus U+0438,
   U+0439, U+0454, U+0456, U+0457, U+0491, U+0449, U+044A, U+044C,
   U+044E, U+044F.

3. Language-Based Tables

   The registration strategy described in this document uses a table
   that lists all characters allowed for input and any variants of those
   characters. Note that the table lists all characters allowed, not
   only the ones that have variants.

4. Table Processing Rules

   The input to the process is called the "input label". The output of
   the process is either failure (the input label cannot be registered
   at all), or a registration bundle that contains one or more labels in
   A-label form.

5. Table Format

   The table in Appendix A consists of four columns. The first and
   second identify the Cyrillic character, and the third and fourth
   identify Latin or Greek characters that might be easily confused with
   them visually. If both a Latin and Greek character are present, the
   Greek one appears in the third and fourth columns on the subsequent
   line (with "..." in the first column to indicate more information
   about the character specified on the previous line). Variants needed
   only because of case folding are shown with "+++" in the first
   column, as noted in the table.

   Each character in the table is given in the "U+" notation for Unicode
   characters followed, in the next column, by its name as shown in the
   Unicode Standard. For easy reference, the characters are listed in
   the order in which they appear in the Unicode Standard.

   The table does not, and any future revision MUST NOT, have more than
   one entry for a particular base character.

6. Steps after Registering an Input Label

   A registry has at least three policy options for handling the cases
   where the registration bundle has more than one label. These
   options, and their key implications, are:

   o Allocate all labels to the same registrant, making the zone
      information identical to that of the input label.

      This option will cause end users to be able to find names with
      variants more easily, but will result in larger zone files. In
      principle, the zone file could become so large that it could
      negatively affect the ability of the registry to perform name
      resolution.

   o Block all labels so they cannot be registered in the future.

      This option does not increase the size of the zone file, but it
      may cause end users to not be able to find names with variants
      that they would expect.

   o Allocate some labels and block some other labels.

      This option is likely to cause the most confusion with users
      because including some variants will cause a name to be found, but
      using other variants will cause the name to be not found.

   With any of these three options, the registry MUST keep a database
   that links each label in the registration bundle to the input label.
   This link needs to be maintained so that changes in the non-DNS
   registration information (such as the label's owner name and address)
   are reflected in every member of the registration bundle as well.

7. Security Considerations

   The information provided in this document may assist DNS zone
   administrators and registrants in selecting names that are less
   likely to be confused with others and in adopting policies that help
   avoid confusion. It may also assist user-interface designers in
   identifying possible areas of confusion so that they can better
   protect users. The document otherwise has no consequences for the
   security of the Internet.

8. Acknowledgments

   Support from Afilias for a major portion of this work is appreciated.

   The material on Kildin Sami would not have been possible without the
   efforts of Cary Karp for his help directly and his pointer to
   Riessler's work [Riessl07] and from Vladimir Shadrunov and Sergey
   Nikolaevich Teryoshkin for their own analyses and references
   ([Anto90], [Kert86], and [Kuru85]) and partial translations from
   them. We are grateful for their efforts that facilitated treating it
   nearly the same way as other actively used European languages that
   use Cyrillic script.

   Careful reading of late drafts of this document by Bill McQuillan,
   Alexey Melnikov, and Peter Saint-Andre, identified a number of
   editorial problems, some of which might not have been caught
   otherwise.

9. References

9.1. Normative References

   [RFC5895]   Resnick, P. and P. Hoffman, "Mapping Characters
                         in Mapping Characters for Internationalized
                         Domain Names in Applications (IDNA) 2008",
                         RFC 5895, September 2010.

   [Unicode52]   The Unicode Consortium. The Unicode Standard,
                         Version 5.2.0, defined by: "The Unicode Standard,
                         Version 5.2.0", (Mountain View, CA: The Unicode
                         Consortium, 2009. ISBN 978-1-936213-00-9).
                         <http://www.unicode.org/versions/Unicode5.2.0/>.

9.2. Informative References

   [Anto90]   Antonova, A., "Primer for Sami schools first
                         grade: Sami language, 2nd edition", Leningrad:
                         Prosveshchenie, Leningrad department, 1990.
                         Published in Russian, no authoritative
                         translation is known.

   [EU-registry]   European Registry of Internet Domain Names
                         (EURid), ".eu Supported Characters",
                         January 2010, <http://www.eurid.eu/en/
                         eu-domain-names/technical-limitations/
                         supported-characters>.

   [Kert]   Kertom, G., "Kildin dialect of the Sami
                          language". Published in Russian, no
                          authoritative translation is known.

   [Kert86]   Kertom, G., "Sami-Russian and Russian-Sami
                          dictionary: textbook for primary school pupils",
                          Leningrad: Prosveshchenie Leningrad Department,
                          1986. Published in Russian, no authoritative
                          translation is known.

   [Kuru85]   Kuruch, R., "Sami-Russian dictionary: eight
                          thousand words", Moscow: Russkiy yazyk, 1985.
                          Published in Russian, no authoritative
                          translation is known.

   [MontenegrinChars]   Crna Gora Ministarstvo prosvjete i nauke
                          (Ministry of Science and Education, Montenegro),
                          "Pravopis Crnogorskoga Jezika I", 2009,
                          <http://www.gov.me/files/1248442673.pdf>. In
                          Montenegrin, no known English translation. See
                          especially the table on page 8.

   [OmniglotSaami]   Ager, S., "Sami (Saami)", 2009,
                          <http://www.omniglot.com/writing/saami.htm>.

   [RFC3743]   Konishi, K., Huang, K., Qian, H., and Y. Ko,
                          "Joint Engineering Team (JET) Guidelines for
                          Internationalized Domain Names (IDN) Registration
                          and Administration for Chinese, Japanese, and
                          Korean", RFC 3743, April 2004.

   [RFC4290]   Klensin, J., "Suggested Practices for
                          Registration of Internationalized Domain Names
                          (IDN)", RFC 4290, December 2005.

   [RFC4713]   Lee, X., Mao, W., Chen, E., Hsu, N., and J.
                          Klensin, "Registration and Administration
                          Recommendations for Chinese Domain Names",
                          RFC 4713, October 2006.

   [RFC5564]   El-Sherbiny, A., Farah, M., Oueichek, I., and A.
                          Al-Zoman, "Linguistic Guidelines for the Use of
                          the Arabic Language in Internet Domains",
                          RFC 5564, February 2010.

   [RFC5890]   Klensin, J., "Internationalized Domain Names for
                          Applications (IDNA): Definitions and Document
                          Framework", RFC 5890, August 2010.

   [RFC5892]   Faltstrom, P., "The Unicode Code Points and
                          Internationalized Domain Names for Applications
                          (IDNA)", RFC 5892, August 2010.

   [Riessl07]   Riessler, M., "Kola Saami character chart
                          (draft)", November 2007.

Appendix A. European Cyrillic Character Tables

   These tables are constructed on the basis of the characters that can
   actually occur in the DNS, i.e., those that are valid in U-labels as
   defined in RFC 5890. If the characters that can be mapped into those
   characters are to be considered instead, then the number of variants
   would increase considerably. For example, while CYRILLIC SMALL
   LETTER A (U+0430) and GREEK SMALL LETTER ALPHA (U+03B1) are readily
   distinguished visually, their capital letter equivalents are not, so,
   if case mappings such as those discussed in the IDNA2008 Mapping
   document [RFC5895] are considered, the two small letters must be
   considered variants of each other. Some of the variants have been
   selected on the assumption that unusual fonts may be used and that
   users will see what they expect to see; others, involving subtle
   decorations but considered more far-fetched out of context, have not
   been listed.

   These additional, possibly required, variants are shown below with
   "+++" in the first column of the table.

   "..." in the first column is used to indicate more information about
   the character specified on the previous line.

Characters needed for European languages, other than Montenegrin and
Sami, written in Cyrillic.

```
+----------+-------------------------+---------+------------------+
| Cyrillic | Unicode Name            | Variant | Unicode Name     |
| Char     |                         |         |                  |
+----------+-------------------------+---------+------------------+
| U+0430   | CYRILLIC SMALL LETTER A | U+0061  | LATIN SMALL      |
|          |                         |         | LETTER A         |
|          |                         |         |                  |
| +++      |                         | U+03B1  | GREEK SMALL      |
|          |                         |         | LETTER ALPHA     |
|          |                         |         |                  |
| U+0431   | CYRILLIC SMALL LETTER BE |        |                  |
|          |                         |         |                  |
| U+0432   | CYRILLIC SMALL LETTER VE | U+0062 | LATIN SMALL      |
|          |                         |         | LETTER B         |
|          |                         |         |                  |
| +++      |                         | U+03B2  | GREEK SMALL      |
|          |                         |         | LETTER BETA      |
|          |                         |         |                  |
| U+0433   | CYRILLIC SMALL LETTER   | U+0072  | LATIN SMALL      |
|          | GHE                     |         | LETTER R         |
|          |                         |         |                  |
| +++      |                         | U+03B3  | GREEK SMALL      |
|          |                         |         | LETTER GAMMA     |
|          |                         |         |                  |
| U+0434   | CYRILLIC SMALL LETTER DE |        |                  |
|          |                         |         |                  |
| +++      |                         | U+03B4  | GREEK SMALL      |
|          |                         |         | LETTER DELTA     |
|          |                         |         |                  |
| U+0435   | CYRILLIC SMALL LETTER IE | U+0065 | LATIN SMALL      |
|          |                         |         | LETTER E         |
|          |                         |         |                  |
| +++      |                         | U+03B5  | GREEK SMALL      |
|          |                         |         | LETTER EPSILON   |
|          |                         |         |                  |
| U+0436   | CYRILLIC SMALL LETTER   |         |                  |
|          | ZHE                     |         |                  |
|          |                         |         |                  |
| U+0437   | CYRILLIC SMALL LETTER ZE |        |                  |
|          |                         |         |                  |
| U+0438   | CYRILLIC SMALL LETTER I | U+0075  | LATIN SMALL      |
|          |                         |         | LETTER U         |
|          |                         |         |                  |
| U+0439   | CYRILLIC SMALL LETTER   |         |                  |
|          | SHORT I                 |         |                  |
```

U+043A	CYRILLIC SMALL LETTER KA	U+006B	LATIN SMALL LETTER K
...		U+03BA	GREEK SMALL LETTER KAPPA
U+043B	CYRILLIC SMALL LETTER EL		
+++		U+03BB	GREEK SMALL LETTER LAMBDA
U+043C	CYRILLIC SMALL LETTER EM	U+006D	LATIN SMALL LETTER M
+++		U+03BC	GREEK SMALL LETTER MU
U+043D	CYRILLIC SMALL LETTER EN	U+0048	LATIN CAPITAL LETTER H
+++		U+0068	LATIN SMALL LETTER H (in some fonts)
+++		U+03B7	GREEK SMALL LETTER ETA
U+043E	CYRILLIC SMALL LETTER O	U+006F	LATIN SMALL LETTER O
...		U+03BF	GREEK SMALL LETTER OMICRON
U+043F	CYRILLIC SMALL LETTER PE	U+006E	LATIN SMALL LETTER N
...		U+03C0	GREEK SMALL LETTER PI
U+0440	CYRILLIC SMALL LETTER ER	U+0070	LATIN SMALL LETTER P
...		U+03C1	GREEK SMALL LETTER RHO
U+0441	CYRILLIC SMALL LETTER ES	U+0063	LATIN SMALL LETTER C

U+0442	CYRILLIC SMALL LETTER TE	U+0074	LATIN SMALL LETTER T	
+++		U+03C4	GREEK SMALL LETTER TAU	
U+0443	CYRILLIC SMALL LETTER U	U+0079	LATIN SMALL LETTER Y	
+++		U+03C5	GREEK SMALL LETTER UPSILON	
U+0444	CYRILLIC SMALL LETTER EF	U+03D5	GREEK PHI SYMBOL	
+++		U+03C6	GREEK SMALL LETTER PHI	
U+0445	CYRILLIC SMALL LETTER HA	U+0078	LATIN SMALL LETTER X	
...		U+03C7	GREEK SMALL LETTER CHI	
U+0446	CYRILLIC SMALL LETTER TSE			
U+0447	CYRILLIC SMALL LETTER CHE			
U+0448	CYRILLIC SMALL LETTER SHA			
U+0449	CYRILLIC SMALL LETTER SHCHA			
U+044A	CYRILLIC SMALL LETTER HARD SIGN	U+0062	LATIN SMALL LETTER B	
U+044B	CYRILLIC SMALL LETTER YERU			
U+044C	CYRILLIC SMALL LETTER SOFT SIGN	U+0062	LATIN SMALL LETTER B	
U+044D	CYRILLIC SMALL LETTER E			
U+044E	CYRILLIC SMALL LETTER YU			

U+044F	CYRILLIC SMALL LETTER YA		
U+0451	CYRILLIC SMALL LETTER IO	U+00EB	LATIN SMALL LETTER E WITH DIAERESIS
U+0452	CYRILLIC SMALL LETTER DJE		
U+0453	CYRILLIC SMALL LETTER GJE		
U+0454	CYRILLIC SMALL LETTER UKRAINIAN IE	U+03B5	GREEK SMALL LETTER EPSILON
U+0455	CYRILLIC SMALL LETTER DZE	U+0073	LATIN SMALL LETTER S
U+0456	CYRILLIC SMALL LETTER BYELORUSSIAN-UKRAINIAN I	U+0069	LATIN SMALL LETTER I
+++		U+03B9	GREEK SMALL LETTER IOTA
U+0457	CYRILLIC SMALL LETTER UKRAINIAN YI	U+03CA	GREEK SMALL LETTER IOTA WITH DIALYTIKA
+++		U+00EF	LATIN SMALL LETTER I WITH DIAERESIS
U+0458	CYRILLIC SMALL LETTER JE	U+006A	LATIN SMALL LETTER J
...		U+03F3	GREEK LETTER YOT
U+0459	CYRILLIC SMALL LETTER LJE		
U+045A	CYRILLIC SMALL LETTER NJE		
U+045B	CYRILLIC SMALL LETTER TSHE		

Cyrillic Char	Unicode Name	Variant	Unicode Name
U+045C	CYRILLIC SMALL LETTER KJE		
U+045D	CYRILLIC SMALL LETTER I WITH GRAVE		
U+045E	CYRILLIC SMALL LETTER SHORT U		
U+045F	CYRILLIC SMALL LETTER DZHE		
U+0491	CYRILLIC SMALL LETTER GHE WITH UPTURN	U+0072	LATIN SMALL LETTER R
U+04C2	CYRILLIC SMALL LETTER ZHE WITH BREVE		

Additional characters needed for Montenegrin written in Cyrillic.

Cyrillic Char	Unicode Name	Variant	Unicode Name
U+0437 + U+0301	CYRILLIC SMALL LETTER ZE WITH ACUTE		
U+0441 + U+0301	CYRILLIC SMALL LETTER ES WITH ACUTE		

Additional characters needed for Kildin Sami written in Cyrillic.

Cyrillic Char	Unicode Name	Variant	Unicode Name
U+0430 + U+0304	CYRILLIC SMALL LETTER A WITH MACRON	U+0101	LATIN SMALL LETTER A WITH MACRON
...		U+03B1 + U+0304	GREEK SMALL LETTER ALPHA WITH MACRON
U+0435 + U+0304	CYRILLIC SMALL LETTER IE WITH MACRON	U+0113	LATIN SMALL LETTER E WITH MACRON
U+043E + U+0304	CYRILLIC SMALL LETTER O WITH MACRON	U+014D	LATIN SMALL LETTER O WITH MACRON
...		U+03BF + U+0304	GREEK SMALL LETTER OMICRON WITH MACRON
U+044B + U+0304	CYRILLIC SMALL LETTER YERU WITH MACRON		
U+044D + U+0304	CYRILLIC SMALL LETTER E WITH MACRON		
U+044E + U+0304	CYRILLIC SMALL LETTER YU WITH MACRON		
U+044F + U+0304	CYRILLIC SMALL LETTER YA WITH MACRON		
U+0451 + U+0304	CYRILLIC SMALL LETTER IO WITH MACRON	U+00EB + U0304	LATIN SMALL LETTER E WITH DIAERESIS AND MACRON
U+048B	CYRILLIC SMALL LETTER SHORT I WITH TAIL		

U+048D	CYRILLIC SMALL LETTER SEMISOFT SIGN		
U+048F	CYRILLIC SMALL LETTER ER WITH TICK		
U+04BB	CYRILLIC SMALL LETTER SHHA	U+0068	LATIN SMALL LETTER H
U+04C6	CYRILLIC SMALL LETTER EL WITH TAIL		
U+04C8	CYRILLIC SMALL LETTER EN WITH HOOK		
U+04CA	CYRILLIC SMALL LETTER EN WITH TAIL		
U+04CE	CYRILLIC SMALL LETTER EM WITH TAIL		
U+04D3	CYRILLIC SMALL LETTER A WITH DIAERESIS	U+00E4	LATIN SMALL LETTER A WITH DIAERESIS
U+04E3	CYRILLIC SMALL LETTER I WITH MACRON	U+016B	LATIN SMALL LETTER U WITH MACRON
U+04E7	CYRILLIC SMALL LETTER O WITH DIAERESIS	U+00F6	LATIN SMALL LETTER O WITH DIAERESIS
U+04ED	CYRILLIC SMALL LETTER E WITH DIAERESIS		
U+04EF	CYRILLIC SMALL LETTER U WITH MACRON		
U+04F1	CYRILLIC SMALL LETTER U WITH DIAERESIS		

```
| U+04F9   | CYRILLIC SMALL     |          |                      |
|          | LETTER YERU WITH   |          |                      |
|          | DIAERESIS          |          |                      |
+----------+--------------------+----------+----------------------+
```

Authors' Addresses

   Sergey Sharikov
   Regtime Ltd
   Kalinina str.,14
   Samara 443008
   Russia

   Phone: +7(846) 979-9039
   Fax: +7(846)979-9038
   EMail: s.shar@regtime.net

   Desiree Miloshevic
   Afilias
   Oxford Internet Institute, 1 St. Giles
   Oxford OX1 3JS
   United Kingdom

   Phone: +44 7973 987 147
   EMail: dmiloshevic@afilias.info

   John C Klensin
   1770 Massachusetts Ave, #322
   Cambridge, MA 02140
   USA

   Phone: +1 617 491 5735
   EMail: john-ietf@jck.com