

Network Working Group
Internet-Draft
Intended status: Informational
Expires: April 29, 2010

S. Sharikov
Regtime Ltd
D. Miloshevic
Afilias
J. Klensin
October 26, 2009

**Internationalized Domain Names Registration and Administration
Guideline for European languages using Cyrillic
draft-sharikov-idn-reg-01.txt**

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#). This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on April 29, 2010.

Copyright Notice

Copyright (c) 2009 IETF Trust and the persons identified as the

document authors. All rights reserved.

This document is subject to [BCP 78](http://trustee.ietf.org/license-info) and the IETF Trust's Legal Provisions Relating to IETF Documents in effect on the date of publication of this document (<http://trustee.ietf.org/license-info>). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

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. For completeness of the "European" languages, it also discusses the additional characters needed for Moldovan when it is written in Cyrillic script. It describes appropriate characters for registration and variant considerations for characters from Greek and Latin scripts with similar appearances and/or derivations.

Table of Contents

| | | |
|-----------------------------|--|--------------------|
| 1. | Introduction | 4 |
| 1.1. | Similar Characters and Variants | 5 |
| 1.2. | Terminology | 5 |
| 2. | Languages and Characters | 6 |
| 2.1. | Bosnian, Serbian, Montenegrin | 6 |
| 2.2. | Bulgarian | 6 |
| 2.3. | Byelorussian | 6 |
| 2.4. | Kildin Sami | 7 |
| 2.5. | Macedonian | 7 |
| 2.6. | Moldovan | 7 |
| 2.8. | Russian | 8 |
| 2.10. | Ukrainian | 8 |
| 3. | Language-based Tables | 8 |
| 4. | Table processing rules | 8 |
| 5. | Table Format | 8 |
| 6. | Steps after registering an input label | 9 |
| 7. | Acknowledgments | 9 |
| Appendix A. | European Cyrillic Character Tables | 10 |
| 8. | References | 15 |
| 8.1. | Normative References | 15 |
| 8.2. | Informative References | 16 |
| | Authors' Addresses | 16 |

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.

This specification provides guidelines for the use of Cyrillic, as encoded in Unicode [[Unicode51](#)] 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, the Omniglot discussion of Moldovan [[OmniglotMoldovan](#)], and a recently-issued Moldovan government standard [[MoldovanChars](#)] are provided for use with Moldovan. Moldovan is no longer in official use with Cyrillic script and no registrations are considered likely in Cyrillic, at least within the relevant ccTLD. 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 names registration in their zones. In some cases, there would even be a requirement for mixing Cyrillic and Latin characters in the same label although this is not generally considered desirable. 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.

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. In addition, IDNA [[RFC3490](#)] allows the use of thousands of non-alphanumeric characters, and some zone administrators will want to prohibit some or all of these characters.

So-called "variant techniques", introduced in [[RFC3743](#)] and generalized beyond East Asian language in [[RFC4290](#)], describe ways of registering IDN domain names 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.

1.2. Terminology

The terminology that follows is derived from [[RFC3743](#)] and [[RFC4290](#)], but this specification does not depend on them.

A "string" is an ordered set 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.

[[anchor4: Note in Draft: This specification is based on the original version of IDNA. Updates to any revision should be obvious, but the terminology should be adjusted in needed and special attention should be paid to the mapping-only variants listed in the Appendix. (RFC

Editor, if the I-D reaches you with this note in place, please just drop it.)]]

[2.](#) Languages and Characters

In the interest of clarity and balance, this document describes a "Base Cyrillic" set of twenty-three 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+0029) may also be used.

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

[2.1.](#) Bosnian, Serbian, Montenegrin

Bosnian, Serbian, and Montenegrin 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 thirty characters, seven in addition to the basic twenty-three: U+0456, U+0439, U+0449, U+044A, U+044C, U+044E, U+044F.

[2.3.](#) Byelorussian

Byelorussian alphabet has 32 characters, i.e., nine characters in addition to the base 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

According to the best information available (see [[Riessl07](#)] and the references he cites), the Kildin Sami orthography has __ characters, __ in addition to the basic set. Many of these characters have no precomposed forms in Unicode and hence must be written with combining characters. The 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 with a COMBINING MACRON and CYRILLIC SMALL LETTER IE with a COMBINING MACRON, respectively, have the same visual appearance as LATIN SMALL LETTER A WITH MACRON (U+0101) and LATIN SMALL LETTER E WITH MACRON (U+0113). The latter character is specified by some sources including Riessler [[Riessl07](#)] (the former doesn't appear to be, but is obvious). 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. If some input systems produce U+0113 (or U+0101) and others produce the two-character combining sequence, a variant approach may be appropriate.

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

[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.](#) Moldovan

According to the most recent government specification [[MoldovanChars](#)] Moldovan has 32 characters in its alphabet, including two that have no precomposed forms in Unicode. This is nine in addition to the basic set: U+0437 U+0302, U+0438, U+0441 U+0302, U+0452, U+0458, U+0459, U+045A, U+045B, U+045F.

[2.7.](#) Montenegrin

See Bosnian, [Section 2.1](#), above.

[2.8.](#) 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.

[2.9.](#) Serbian

See Bosnian, [Section 2.1](#), above.

[2.10.](#) Ukrainian

Ukrainian has 31 characters and therefore an additional 8 characters to the base of 23: U+0454, U+0456, U+0457, U+0491, U+0449, U+044A, 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 that have been processed with ToASCII.

[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) is reflected in every member of the registration bundle as well.

7. Acknowledgments

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

Cary Karp pointed us to authoritative information on Kildin Sami; his help for our being able to treat it the same way other actively-used European languages that use Cyrillic script is greatly appreciated.

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 can be obtained by applying the ToUnicode operation of [RFC 3490](#) to an ACE-encoded label as defined there. 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 and Greek Small Letter Alpha are readily distinguished visually, their capital letter equivalents are not, so, if the extended set of Nameprep [[RFC3491](#)] mappings are considered, the two small letters must be considered variants of each other.

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

Characters needed for European languages, other than Moldovan 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+03B0 | GREEK SMALL | |
| | | | LETTER ALPHA | |
| U+0431 | CYRILLIC SMALL LETTER BE | | | |
| U+0432 | CYRILLIC SMALL LETTER VE | U+0062 | LATIN SMALL | |
| | | | LETTER V | |
| +++ | | 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+039B | 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+0068 | LATIN SMALL LETTER H |
| +++ | | 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+044B | CYRILLIC SMALL LETTER | | | |
| | YERU | | | |
| U+044C | CYRILLIC SMALL LETTER | | | |
| | SOFT SIGN | | | |
| 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 | U+03B5 | GREEK SMALL | |
| | UKRAINIAN IE | | LETTER EPSILON | |
| U+0455 | CYRILLIC SMALL LETTER | U+0073 | LATIN SMALL | |
| | DZE | | LETTER S | |
| U+0456 | CYRILLIC SMALL LETTER | U+0069 | LATIN SMALL | |
| | BYELORUSSIAN-UKRAINIAN I | | LETTER I | |
| +++ | | U+03B9 | GREEK SMALL | |
| | | | LETTER IOTA | |
| U+0457 | CYRILLIC SMALL LETTER | U+03CA | GREEK SMALL | |
| | UKRAINIAN YI | | 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 | | | |
| 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+04C2 | CYRILLIC SMALL LETTER | | |
| | ZHE WITH BREVE | | |
| +-----+-----+-----+ | | | |

Additional characters needed for Moldovan written in Cyrillic.

| | | | |
|---------------------|---------------------------|---------|---------|
| Cyrillic | Unicode Name | Variant | Unicode |
| Char | | | Name |
| +-----+-----+-----+ | | | |
| U+04C2 | CYRILLIC SMALL LETTER ZHE | | |
| | WITH BREVE | | |
| U+0437 + | Cyrillic Small Letter ZE | | |
| U+0302 | with Acute | | |
| U+0441 + | Cyrillic Small Letter ES | | |
| U+0302 | 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+03B0 + 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+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 | U+00E4 | LATIN SMALL LETTER A | |
| | LETTER A WITH | | WITH DIAERESIS | |
| | DIAERESIS | | | |
| U+04E3 | CYRILLIC SMALL | U+016B | LATIN SMALL LETTER U | |
| | LETTER I WITH | | WITH MACRON | |
| | MACRON | | | |
| U+04E7 | CYRILLIC SMALL | U+00F6 | LATIN SMALL LETTER O | |
| | LETTER O WITH | | WITH DIAERESIS | |
| | 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 | | | |
| +-----+-----+-----+-----+ | | | | |

8. References

8.1. Normative References

- [RFC3490] Faltstrom, P., Hoffman, P., and A. Costello,
"Internationalizing Domain Names in Applications (IDNA)",
[RFC 3490](#), March 2003.
- [RFC3491] Hoffman, P. and M. Blanchet, "Nameprep: A Stringprep
Profile for Internationalized Domain Names (IDN)",
[RFC 3491](#), March 2003.
- [Unicode51]
The Unicode Consortium, "The Unicode Standard, Version
5.1.0", 2008.
- Defined by: The Unicode Standard, Version 5.0, Boston, MA,
Addison-Wesley, 2007, ISBN 0-321-48091-0, as amended by
Unicode 5.1.0
(<http://www.unicode.org/versions/Unicode5.1.0/>).

8.2. Informative References

[MoldovanChars]

Crna Gora Ministarstvo prosvjete i nauke, "Pravopis Crnogorskoga Jezika I", 2009, <<http://www.gov.me/files/1248442673.pdf>>.

In Moldovan, no known English translation.

[OmniglotMoldovan]

Ager, S., "Moldovan", 2009, <<http://www.omniglot.com/writing/moldovan.htm>>.

[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.

[Riessl07]

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

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

