

IMAPEXT
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2009

A. Melnikov
Isode Limited
February 14, 2009

Additional collation algorithms for use in IMAP and Sieve
draft-ietf-morg-collations-00

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

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/lid-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 August 18, 2009.

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

Abstract

This document defines extra collation that were found useful when searching for text in email messages.

Internet-Draft

Collation Algorithms

February 2009

Note

A revised version of this draft document will be submitted to the RFC editor as a Proposed Standard for the Internet Community. Discussion and suggestions for improvement are requested, and should be sent to morg@ietf.org.

Table of Contents

1.	Conventions used in this document	3
2.	ASCII Signed Numeric Collation Description	3
2.1.	ASCII Signed Numeric Collation Registration	4
3.	ASCII Punctuation Ignore Numeric Collation Description . . .	4
3.1.	ASCII Punctuation Ignore Numeric Collation Registration . . .	5
4.	Other Collations	5
5.	Formal Syntax	5
6.	Security Considerations	5
7.	IANA Considerations	5
8.	Acknowledgements	6
9.	References	6
9.1.	Normative References	6
9.2.	Informative References	6
	Author's Address	6

1. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[Kwds](#)].

2. ASCII Signed Numeric Collation Description

The "i;ascii-signed-numeric" collation is a simple collation intended for use with arbitrarily-sized, signed decimal integer numbers stored as octet strings. US-ASCII digits (0x30 to 0x39) represent digits of the numbers. The numbers can have an arbitrary number of leading US-ASCII whitespace characters (0x20), optionally followed by the sign ("+", "-") and then by digits of the number. If the "+" sign is omitted, the number is considered to be positive.

Before converting from string to integer, all leading whitespace characters are removed. If no "+" or "-" character is found, "+" is then prepended to the string. And finally, the input string is truncated at the first non-digit character, not counting the "+"/"-" prefix. All input is valid for this collation; strings that do not start (after whitespace removal) with a "+", "-", or a digit represent positive infinity.

The collation supports equality and ordering, but does not support the substring operation.

The equality operation returns "match" if the two strings represent the same number (i.e., leading whitespaces, zeroes and trailing non-digits are disregarded), and "no-match" if the two strings represent different numbers.

The ordering operation returns "less" if the first string represents a smaller number than the second, "equal" if they represent the same number, and "greater" if the first string represents a larger number

than the second.

Some examples: "-500" is less than "-90", "0" is less than "1M", and "1" is less than "4294967298". "4294967298", "04294967298", and "4294967298b" are all equal. "04294967298" is less than ".", "+", "-", "", "x", and "y" are equal.

[2.1.](#) ASCII Signed Numeric Collation Registration

```
<?xml version='1.0'?>
<!DOCTYPE collation SYSTEM 'collationreg.dtd'>
<collation rfc="XXXX" scope="other" intendedUse="common">
  <identifier>i;ascii-signed-numeric</identifier>
  <title>ASCII Signed Numeric</title>
  <operations>equality order</operations>
  <specification>RFC XXXX</specification>
  <owner>IETF</owner>
  <submitter>alexey.melnikov@isode.com</submitter>
</collation>
```

[3.](#) ASCII Punctuation Ignore Numeric Collation Description

The "i;ascii-punc-ignore-numeric" collation is a collation intended for use with arbitrarily-sized, unsigned decimal integer numbers stored as octet strings. US-ASCII digits (0x30 to 0x39) represent digits of the numbers. Any digit of such numbers can be followed or preceded by any number of the following US-ASCII characters, which are ignored for the purpose of comparison: " " (0x20), "-", "+", ",", ";", ".". [[anchor5: The list of characters to ignore is to be discussed.]]

Before converting from string to integer, all characters to be ignored are removed from the string. After that, the input string is truncated at the first non-digit character. All input is valid for this collation; strings that do not start (after removal of all

characters to ignore) with a digit represent positive infinity.

The collation supports equality and ordering, but does not support the substring operation.

The equality operation returns "match" if the two strings represent the same number (i.e., leading whitespaces, zeroes and trailing non-digits are disregarded), and "no-match" if the two strings represent different numbers.

The ordering operation returns "less" if the first string represents a smaller number than the second, "equal" if they represent the same number, and "greater" if the first string represents a larger number than the second.

Some examples: "+1-500" is less than "2 50.0", " 0" is less than "1M", and "1" is less than "4294967298". "4294967298", "04294967298", and "4294967298b" are all equal. "04294967298" is less than "". "+", "-", "", "x", and "y" are equal.

[3.1.](#) ASCII Punctuation Ignore Numeric Collation Registration

```
<?xml version='1.0'?>
<!DOCTYPE collation SYSTEM 'collationreg.dtd'>
<collation rfc="XXXX" scope="other" intendedUse="common">
  <identifier>i;ascii-punc-ignore-numeric</identifier>
  <title>ASCII Punctuation Ignore Numeric</title>
  <operations>equality order</operations>
  <specification>RFC XXXX</specification>
  <owner>IETF</owner>
  <submitter>alexey.melnikov@isode.com</submitter>
</collation>
```

[4.](#) Other Collations

[[anchor8: Other collations to consider: case preserving version of i;unicode-casemap defined in [RFC 5051](#).]]

[5.](#) Formal Syntax

The following syntax specification uses the augmented Backus-Naur Form (BNF) as described in [\[ABNF\]](#). Terms not defined here are taken from [\[ABNF\]](#).

```
ascii-signed-numeric = *SP ["-" / "+"] 1*DIGIT
    ; This production defines valid string prefixes
```

```
punc-ignore-unsigned-numeric = *punctuation DIGIT
    *(punctuation / DIGIT)
    ; This production defines valid string prefixes
```

```
punctuation = SP / "+" / "-" / "." / "," / ";"
```

[6.](#) Security Considerations

[[anchor9: TBD.]]

[7.](#) IANA Considerations

TBD.

[8.](#) Acknowledgements

TBD.

[9.](#) References

[9.1.](#) Normative References

- [ABNF] Crocker, D., Ed. and P. Overell, "Augmented BNF for Syntax Specifications: ABNF", [RFC 5234](#), January 2008.
- [Kwds] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [RFC 2119](#), March 1997.

[RFC4790] Newman, C., Duerst, M., and A. Gulbrandsen, "Internet Application Protocol Collation Registry", [RFC 4790](#), March 2007.

[9.2](#). Informative References

[RFC3501] Crispin, M., "INTERNET MESSAGE ACCESS PROTOCOL - VERSION 4rev1", [RFC 3501](#), March 2003.

Author's Address

Alexey Melnikov
Isode Limited
5 Castle Business Village
36 Station Road
Hampton, Middlesex TW12 2BX
UK

Email: Alexey.Melnikov@isode.com

URI: <http://www.melnikov.ca/>