Network Working Group
Internet-Draft
Intended status: Informational
Expires: March 19, 2011

B. Hoehrmann
September 15, 2010

The i;codepoint collation
draft-hoehrmann-cp-collation-01

Abstract

This memo describes the "i;codepoint" collation. Character strings are compared based on the Unicode scalar values of the characters. The collation supports equality, substring, and ordering operations.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on March 19, 2011.

1. Introduction

The i;codepoint collation operates on Unicode strings and treats any and all differences between two strings as significant. Ordering of different strings is determined by the Unicode scalar values of the characters. It produces usable results where further information is unavailable. In that it is suitable as default collation.

The equality operation determines if two strings are identical. This makes the collation suitable for use with strings known to be in some canonical form. Similarily, applications that require strings to be in a canonical but otherwise arbitrary order may find this collation the most efficient as it requires no transformations.

2. Definition

The i;codepoint collation is the same as the i;octet collation except that it operates on sequences of Unicode scalar values, not octets. Note that by definition the set of Unicode scalar values excludes the surrogate code points and as such they do not occur in valid input.

3. Security Considerations

None beyond those in RFC 4790 [RFC4790].

4. IANA Considerations

This document defines the i;codepoint collation as per [RFC4790].

5. Registration document

```
<?xml version='1.0'?>
<!DOCTYPE collation SYSTEM 'collationreg.dtd'>
<collation rfc="XXXX" scope="global" intendedUse="common">
  <identifier>i;codepoint</identifier>
  <title>Unicode identity</title>
  <operations>equality order substring</operations>
  <specification>RFC XXXX</specification>
  <owner>IETF</owner>
  <submitter>bjoern@hoehrmann.de</submitter>
</collation>
```

6. References

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

Author's Address

Bjoern Hoehrmann
Mittelstrasse 50
39114 Magdeburg
Germany

EMail: mailto:bjoern@hoehrmann.de
URI: http://bjoern.hoehrmann.de

Note: Please write "Bjoern Hoehrmann" with o-umlaut (U+00F6) wherever possible, e.g., as "Björn Höhrmann" in HTML and XML.