Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 30, 2007

Y. YONEYA
JPRS

Feb 26, 2007

IRI recognition in Applications
draft-yoneya-iri-recognition-00.txt

Status of this Memo

Copyright Notice

Abstract

Nowadays access to the Internet is a part of daily life. Users see URIs written in various ways on various media, recognize them as "the Internet Addresses", and use them to access to the Internet. Many application programs recognize URIs automatically and make links to them, so the users can access to the URIs very easily. But, at this moment, most of application programs can't recognize Internationalized Domain Name (IDN) and Internationalized URI (IRI) correctly, so users will feel stress when using IDNs and IRIs.

Utilization of the IDNs and the IRIs are getting higher. Therefore, improvement of the application programs are highly recommended. This document is intended to be an application developpers' guideline for recognizing and corresponding to IDNs/IRIs correctly.

Table of Contents

1. Introduction

Since most of major Web browsers are compliant to Internationalized Domain Name (IDN) [RFC3490] and Internationalized URI (IRI) [RFC3987], utilization of the IDNs and the IRIs are getting higher. However, IDN/IRI aware application programs (hereinafter, applications) other than Web browsers are very few.

From users' point of view, applications recognizing IRIs in context and taking appropriate action as well as performed for URIs are prefered. For example, copy-and-paste the IRI from an application to another application, click the IRI to invoke external application, and so on. In principle, there is no difference between recognizing URI and IRI. But, in reality, many applications can't recognize IRIs although can recognize URIs.

General method for delimiting a URI in context is already defined in the Appendix C. of [RFC3986]. To apply the method to IRIs in applications, this document tries to clarify ambiguous point and relations between caller and callee applications.

2. Delimiting an IRI in Context

In Appendix C. of [RFC3986], main points to delimit a URI are as follows:

1. Wrap URI with double-quotes, e.g. "http://www.example.com/"
2. Wrap URI with angle brackets, e.g. <http://www.example.com/>
3. Delimit by using whitespace, e.g. http://www.example.com/

The first two points are clear. When applications find a URI scheme name with preceding double-quote or left angle bracket, then applications search for corresponding double-quote or right angle bracket and recognize the string within double-quotes or angle brackets is a URI. Note that the double-quotes and the angle brackets are not part of a URI. This method is also valid for an IRI.

The last point is ambiguous to two things. The first one is the definition of whitespace itself. In applications, characters marked as White_Space in [PropList] defined by [Unicode] should be recognized as whitespace. The seconde one is whether a URI have to be wrapped by whitespaces. The beginning of a URI scheme name is unambiguous, therefore, prepending whitespace is not necessary. When

applications find a URI scheme name without preceding double-quote or left angle bracket, then applications search for whitespace and recognize the string terminated by whitespace is a URI. Note that the whitespaces are not part of a URI. This method is also valid for an IRI.

3. Passing an IRI between applications

IDNs and IRIs have its ASCII compatible encoding (hereinafter, ACE) form according to its specification. For example, IDNs have Punycode form, and IRIs have %-encoding form. Generally, when passing an IRI as a protocol parameter between applications, it should be in ACE form. But, there is a big difference between applications being able to recognize IRIs and being able to convert IRIs to ACE form. At the User Interface level, passing an IRIs in IRI form is more convenient. For example, when copying the IRI which is recognized automatically by an application and pasting to another application, it should be in IRI form.

When an application passes an IRI to another application by users' action such as link click, there might be some combinations. Following is a matrix which describes combinations conceptually.

			Callee	
			IDN/IRI	
Caller		Form	compliant	uncompliant
IDN / IRI	recognize	IRI	OK	NG
	compliant	IRI	OK	NG
		ACE	OK	OK

passing form and access availability

For example, if caller application recognize IDN/IRI correctly according to this document but does not compliant to IDN/IRI standards, then it passes IRI to another (callee) application. The callee application can access to the passed IRI if it is IDN/IRI compliant, otherwise not.

Caller application does not know whether callee applications are IDN/IRI compliant. But it can be assumed under certain circumstances.

In general, the most probable callee application is the Web browser. As described in Section 1, most of major Web browsers on PCs are IDN/IRI compliant, so passing an IRI in IRI form is assumed to work. Right now, there is no comprehensive list of circumstances, therefore the assumption is an open issue.

4. Security Considerations

Without verifying IRI string well, it is dangerous to access automatically recognized IRI. See the security considerations in [RFC3490] and [RFC3986] for further information.

5. IANA Considerations

No IANA actions are required by this document.

6. Acknowledgements

The authors of normative references.

7. Change History

This section is used for tracking the update of this document. Will be removed after finalize.

7.1. draft-yoneya-ima-downgrade: Version 00

- Initial version

8. Normative References

[PropList] The Unicode Consortium, "PropList, Unicode Character Database", 2006, <http://www.unicode.org/Public/UNIDATA/PropList.txt>.

[RFC3490] Faltstrom, P., Hoffman, P., and A. Costello, "Internationalizing Domain Names in Applications (IDNA)", RFC 3490, March 2003.

[RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.

[RFC3987] Duerst, M. and M. Suignard, "Internationalized Resource Identifiers (IRIs)", RFC 3987, January 2005.

[Unicode] The Unicode Consortium, "The Unicode Standard, Version 5.0", 2006, <http://www.unicode.org/>.

Author's Address

Yoshiro YONEYA
JPRS
Chiyoda First Bldg. East 13F, 3-8-1 Nishi-Kanda
Chiyoda-ku, Tokyo 101-0065
Japan

Phone: +81 3 5215 8451
Email: yone@jprs.co.jp

Full Copyright Statement

Intellectual Property

Acknowledgment

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).