Specification Information Note
WAP-183_005-ProvCont-20020411-a
Version 11-Apr-2002

for

Wireless Application Protocol
WAP-183-ProvCont-20010724-a
WAP Provisioning Content
Version 24-July-2001

This document is available online in PDF format at http://www.wapforum.org/.

Known problems associated with this document are published at http://www.wapforum.org/.

Comments regarding this document can be submitted to the WAP Forum™ in the manner published at http://www.wapforum.org/.

Contents

1. Scope

This document provides changes and corrections to the following document files:

- WAP-183-ProvCont-20010724-a

It includes changes from the following change requests:

- CR-PROVCONT-20020409-NOKIA-1

2. Notation

In the subsections describing the changes new text is <u>underlined</u>. Removed text has ~~strikethrough~~ marks. The presented text is copied from the specification. Text that is not presented is not affected at all. The change descriptions may also include editor's notes similar to the one below. The notes are not part of the actual changes and must not be included in the changed text.

> **Editor's note:** Framed notes like these only clarify where and how the changes shall be applied.

3. Character encoding used in connectivity documents

3.1 Change Classification

Class 2 – Bug Fixes

3.2 Change Summary

The current connectivity document specification states that the same rules as in WML 1.x apply to the choice of the used character encodings. Unfortunately the WML 1.x specification does not explicitly specify which character encoding is to be used. It is merely stated that some encoding that is capable of expressing characters in the Unicode character set must be used. UTF-8 and UTF-16 are called out as encodings that satisfy this requirement, but use of other suitable encodings is also permitted. This is not a problem in the case of browsing, because content negotiation between the client and server allows the server to select a character encoding actually supported by the client. However, in the case of WAP bootstrap provisioning the used delivery mechanisms do not permit any kind of negotiation. One example is delivery of the connectivity document on a smart card. As a result, the current specification must be considered faulty, since it permits interoperability problems involving the used character encoding.

The simplest means of preventing interoperability problems caused by different character encodings is mandating a single one. The Provisioning Content specification is part of the WAP 2.0 release. In the case of WAP 2.0 devices are already required to support UTF-8 and UTF-16, since these are mandated by the XHTML specification. As a result, there will be no additional implementation overhead, if the character encoding used in connectivity documents is chosen to be one of these. However, restricting the implementations to use a single character encoding will simplify things even further. Most of the content in the connectivity document can be expected to be composed of tokens and parameter values that can be encoded using a single octet when UTF-8 is used. Only the few parameters that contain strings intended to be shown to human users may include characters which use two or three octets when encoded using UTF-8. On the other hand, all the strings in the document will use two octets for each character when UTF-16 is used. Since UTF-8 and UTF-16 encode the same character set, the use of UTF-8 will in all practical cases result in a smaller document. As a result, the optimal solution seems to be to mandate the exclusive use of UTF-8.

Another problem in the existing specification is the implicit claim that named character entities are supported as they are supported in WML 1.x. This would require these named character entities to be defined as part of the `wap-provisioningdoc` DTD. This is currently not the case, so the named entities do not work. The simplest solution is forbidding use of named character entities, since the need for them is very limited in connectivity documents and numeric character entities can always be substituted for the named ones.

3.3 Change Description

> **Editor's note:** Section 3.1 on page 7, Normative References.

[RFC2279]   "UTF-8, a transformation format of ISO 10646", ed. F. Yergeau, 1998,
             URL:http://www.ietf.org/rfc/rfc2279.

[WML]   "Wireless Markup Language", WAP Forum, WAP-191-WML,
             URL:http://www.wapforum.org/

[XML]   "Extensible Markup Lanaguage (XML)", W3C Recommendation 10-February-1998,
             REC-xml-19980210, T. Bray, et al, February 10, 1998, URL:http://www.w3.org/TR/REC-xml

[XML]   "Extensible Markup Language (XML) 1.0 (Second edition)", W3C Recommendation
             6 October 2000, REC-xml-20001006, URL:http://www.w3.org/TR/2000/REC-xml-20001006

Editor's note: Section 5.7 on page 29, Connectivity Character Set.

The provisioning document uses an XML language. It inherits the XML document character set and the rules for handling from XML [XML]. Provisioning documents MUST be encoded using UTF-8 [RFC2279]. Numeric character entities are supported, as well as the predefined entities amp, lt, gt, apos, quot that XML processors MUST recognise. ~~The connectivity content type MUST use the same character set rules as specified in [WML], except the rules for meta-information placed within the content, since such information is not supported in connectivity.~~

Editor's note: Section A.1.1 on page 42, Character Set and Encoding.

Item	Function	Reference	Status	Requirement
ProvCont-CSE-C-001	UTF-8 Encoding	5.7 ~~, [WML]~~	~~O~~M	
~~ProvCont-CSE-C-002~~	~~UTF-16 Encoding~~	~~5.7, [WML]~~	~~O~~	
~~ProvCont-CSE-C-003~~	~~UCS-4 Encoding~~	~~5.7, [WML]~~	~~O~~	
~~ProvCont-CSE-C-004~~	~~Other character encoding~~	~~5.7, [WML]~~	~~O~~	
~~ProvCont-CSE-C-005~~	~~Reference processing (no meta-information)~~	~~5.7, [WML]~~	~~M~~	
ProvCont-CSE-C-006	Character entities	5.7 ~~, [WML]~~	M	