

JTC1/SC2/WG2 N2477

Title: W3C Liaison Report
To: ISO/IEC JTC1 SC2 WG2, Mike Ksar, Convener
From: W3C I18N WG, Martin Duerst
Date: 2002-05-16

The W3C Internationalization Working Group (with the help of Internationalization Interest Group) mainly worked on the "Character Model for the World Wide Web 1.0". This document went to Last Call on April 30, 2002. Please see below for more information.

Apart from the work on the Character Model, the WG has taken on responsibility for moving the work on Internationalized Resource Identifiers (IRI) (<http://www.ietf.org/internet-drafts/draft-duerst-iri-00.txt>). Comments on this draft are encouraged.

We are also working together with other W3C WGs on various issues related to internationalization. Among else, there is the XML 1.1 proposal (also in last call) that extends the repertoire of characters that can be used in XML names (element names, attribute names,...).

The Last Call version of the "Character Model for the World Wide Web 1.0", available at <http://www.w3.org/TR/2002/WD-charmod-20020430>. The newest version of this document is always available at <http://www.w3.org/TR/charmod>.

The Last Call is open until May 31. The Last Call is the stage in the W3C process where we are looking for comments from anybody interested, and of course in particular from SC2/WG2 and/or individual members of SC2/WG2.

Below is the Abstract, Status of the Document (which includes instructions for how to send in comments) and the Table of Contents.

Abstract

This Architectural Specification provides authors of specifications, software developers, and content developers with a common reference for interoperable text manipulation on the World Wide Web. Topics addressed include encoding identification, early uniform normalization, string identity matching, string indexing, and URI conventions, building on the Universal Character Set, defined jointly by Unicode and ISO/IEC 10646. Some introductory material on characters and character encodings is also provided. Status of this Document

This section describes the status of this document at the time of its publication. Other documents may supersede this document. The latest status of this series of documents is maintained at the W3C.

This is a second Last Call Working Draft for review by W3C Members and other interested parties. The Last Call period begins 30 April 2002 and ends 31 May 2002.

This working draft attempts to address review comments that were received during the initial Last Call period, which started 26 January 2001, and also incorporates other modifications resulting from continuing collaboration with other working groups and continuing work within the W3C Internationalization Working Group (I18N WG) (Members only). A list of comments (Members only) with their status is available.

The I18N WG invites comments on this specification. Due to the architectural nature of this document, it affects a large number of W3C Working Groups, but also software developers, content developers, and writers and users of specifications outside the W3C that have to interface with W3C specifications. Because review comments play an important role in ensuring a high quality specification, we encourage readers to review this Last Call Working Draft carefully. Comments should preferably be submitted via the Last Call Comment Form (<http://www.w3.org/2002/05/charmod/LastCall>). Comments may alternatively be submitted by email to www-i18n-comments@w3.org (public archive). In this case, please send one email per comment where possible, otherwise number comments clearly.

This document is published as part of the W3C Internationalization Activity by the Internationalization Working Group, with the help of the Internationalization Interest Group. The Internationalization Working Group will not allow early implementation to constrain its ability to make changes to this specification prior to final release. Publication as a Working Draft does not imply endorsement by the W3C Membership. It is inappropriate to use W3C Working Drafts as reference material or to cite them as other than "work in progress". A list of current W3C Recommendations and other technical documents can be found at <http://www.w3.org/TR/>.

Table of Contents

- 1 Introduction
 - 1.1 Goals and Scope
 - 1.2 Background
 - 1.3 Terminology and Notation
- 2 Conformance
- 3 Characters
 - 3.1 Perceptions of Characters
 - 3.2 Digital Encoding of Characters
 - 3.3 Transcoding

- 3.4 Strings
- 3.5 Reference Processing Model
- 3.6 Choice and Identification of Character Encodings
- 3.7 Character Escaping
- 4 Early Uniform Normalization
 - 4.1 Motivation
 - 4.2 Definitions for W3C Text Normalization
 - 4.3 Examples
 - 4.4 Responsibility for Normalization
- 5 Compatibility and Formatting Characters
- 6 String Identity Matching
- 7 String Indexing
- 8 Character Encoding in URI References
- 9 Referencing the Unicode Standard and ISO/IEC 10646
- Appendices
 - A References
 - B Examples of Characters, Keystrokes and Glyphs (Non-Normative)
 - C Composing Characters (Non-Normative)
 - D Resources for Normalization (Non-Normative)
 - E Acknowledgements (Non-Normative)
 - F Change Log (Non-Normative)