

Representing the Indian Region at the W3C

Richard Ishida
W3C Internationalization Lead

W3C Internationalization Activity Groups 


Core Working Group
Reviews, advice, and internationalization specifications

ITS (Internationalization Tag Set) Working Group
Elements and attributes for schema developers

GEO (Guidelines, Education & Outreach) Working Group
Making internationalization aspects of W3C technology better understood and more widely and consistently used

Interest Group
www-international@w3.org

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 2

I18n Overview: Document formats 
Current work items


Character model: Resource identifiers

Architectural Specification providing authors of specifications, software developers, and content developers with a common reference for the use of resource identifiers building on the Universal Character Set, defined jointly by the Unicode Standard and ISO/IEC 10646. .

Character model: Normalization

Architectural Specification providing authors of specifications, software developers, and content developers with a common reference for early uniform normalization and string identity matching to improve interoperable text manipulation on the World Wide Web.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 3

I18n Overview: Document formats 
Current work items


WS-I18n

Enhancements to SOAP messaging to provide internationalized and localized operation via locale and international preference negotiation, and a general-purpose mechanism for associating a "locale policy" with messages.

LTLI

How document formats, specifications, and implementations should implement language and locale identifiers, as well as data structures for describing international preferences.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 4

I18n Overview: Document formats 
Current work items


ITS

Organized by data categories, the ITS set of elements and attributes supports the internationalization and localization of schemas and documents. Implementations are provided for DTDs, XML Schema and Relax NG, and for existing vocabularies like XHTML, DocBook and OpenDocument.

Best practices and articles, tutorials, etc

Practical advice on how to use W3C technologies for international content creation.

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 5

W3C Internationalization Activity 
Objectives

- Help Working Groups understand international requirements as early as possible
- Check specifications in Working Drafts, especially at Last Call, for internationalization issues
- Define, or work with other Working Groups to define, behavior needed for support of international requirements
- Evangelize the need to consider multiple languages and scripts when developing Web technologies of any kind
- Helping users of Web technology understand what's available to them and how to use it

Copyright © 2005 W3C (MIT, ERCIM, Keio) slide 6

Ways to get involved



- Provide resources to collaborate in developing the world wide aspects of the Web
 - not just the i18n WG

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 7

W3C technologies



- Web services
- Semantic Web
- CSS, XSL
- Privacy
- HTML, XHTML, XForms
- XML, XLink
- SVG
- MathML
- Voice
- Accessibility
- ...

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 8

Ways to get involved



- Provide resources to collaborate in developing the world wide aspects of the Web
 - not just the i18n WG
- Raise issues / requirements specific to Indian languages
 - esp. style related

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 9

I18n Overview: Characters

Character sets & encodings



European alphabetic scripts	Middle East scripts	South & South East Asian scripts	
Latin	Hebrew	Devanagari	
Greek	Arabic	Bengali	
Cyrillic	Syriac	Gurmukhi	
Armenian	Thaana	Gujurati	
Georgian		Panjabi	
Runic		Oriya	
Ogham		Tamil	
Modifier letters	Symbols	Telugu	
Combining characters	Currency symbols	Kannada	
	Letter like symbols	Malayalam	
	Mathematic operators	Sinhala	
	Numeric forms	Thai	
	Technical symbols	Lao	
East Asian scripts	Geometrical symbols	Tibetan	
Han	Miscellaneous	Myanmar	
Hiragana	symbols & dingbats	Khmer	
Katakana	Enclosed & square		
Hangul	Braille		
Bopomofo			
Yi			
		Additional scripts	
		Ethiopic	
		Cherokee	
		Canadian Aboriginal	
		Syllabics	
		Mongolian	
		Tifinagh	
		Etc....	

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 10

I18n Overview: Document formats

Speech synthesis



這一晚會如常舉行

這一|晚會|如常|舉行

This banquet is held as usual.

這一|晚會|如|常|舉行

If this banquet is held frequently.

這一晚|會|如常|舉行

(An event) will be held tonight as usual.

Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 11

I18n Overview: Presentation matters

Script-specific typography



Copyright © 2005 W3C (MIT, ERCIM, Keio)

slide 12

One ought to know whether first letter styling has special implications for languages in non-Latin scripts.

स्थिति



- Working Groups at W3C still don't know the Indian specific issues that are outstanding, and we need your help to spot these
- Read the specifications and try out the technology
- Then report back to us on issues
- Then help us gather knowledge about how to address those issues
- This means participating in the work yourselves
- But be clear about what is relevant to the W3C and what is not

- Provide resources to collaborate in developing the World Wide aspects of the Web
 - not just the i18n WG
- Raise issues / requirements specific to Indian languages
 - esp. style related
- Encourage users to do the right thing
 - help create articles and advice
 - spread the word
 - translate

- Provide resources to collaborate in developing the World Wide aspects of the Web
 - not just the i18n WG
- Raise issues / requirements specific to Indian languages
 - esp. style related
- Encourage users to do the right thing
 - help create articles and advice
 - spread the word
 - translate
- Follow and influence groups external to the W3C
 - IDN and IRIs
 - language tags
 - Unicode

- Unicode characters and behaviors
- Fonts and rendering support
- Keyboard definitions
- Sort orders
- Local format registries
- Language values
- IDN and IRIs