

draft-ietf-iri-3987bis-13  
Overview and some Issues

IETF 85, Atlanta  
IRI WG Meeting  
2012-11-06

Martin J. Dürst, co-Editor

# Overview

- Background
- Some Open Issues
- Request for Review

# Background

- IRI: Internationalized Resource Identifier, currently [RFC 3987](#)
- Internationalized (i.e. not-ASCII-only) version of URI ([STD66, RFC 3986](#))
- Updating [draft-ietf-iri-3987bis-11.txt](#)
- List of open issues at:  
<http://trac.tools.ietf.org/wg/iri/trac/report/1>
- SVN revision log:  
<http://trac.tools.ietf.org/wg/iri/trac/log/draft-ietf-iri-3987bis/draft-ietf-iri-3987bis.xml>

# IRI Examples

- <http://台網中心.台灣>, <http://台网中心.台湾>
- <http://zh.wikipedia.org/wiki/臺北市>,  
<http://zh.wikipedia.org/zh/臺北市>
- <http://بوابة.تونس>
- <http://ja.wikipedia.org/wiki/青山学院大学>

# Please Don't Forget

- URIs/IRIs are a META-syntax
- Many pieces with different requirements get thrown together
- URIs/IRIs can be:
  - Absolute, complete from scheme to fragment id
  - Relative, just one or a few pieces
  - User-oriented (short, memorable)
  - Back-end (long, complicated)

# Query Part

- HTML uses
  - UTF-8 for URIs
  - Document encoding for forms
- Clash for query parts  $\Rightarrow$  accepted practice is document encoding
- Creates interoperability problems
- Possible solutions
  - Query part always ASCII
  - Query part always converted as UTF-8
  - Mixture:
    - Use ASCII for IRIs from non-Unicode documents
    - Use UTF-8 conversion for document-independent IRIs
    - Define in a way that allows HTML to reference
- Issues [#131](#), [#133](#)

# Character Coverage

- HTML URLs include/IRIs exclude U+FFF0-FFFD
- HTML URLs allow private-use characters in general, but IRIs only allow them in query part
- Potential for alignment
- Many other characters that are a bad idea but not excluded
- Issues [#135](#), [#136](#), related: [#93](#)

# Terminology

Issue #85: [Align terminology in 3987bis with RFC 6365](#)

- Partly implemented
- Partly [explained why not](#) (character encoding)
- Partly still to be done (UCS, octet,...)



# Please Help Reviewing

In particular:

- Section 5: Use of IRIs
- Section 8: URI/IRI Processing Guidelines (Informative)