Abstract

- This document specifies rules for deciding whether a codepoint, considered in isolation, is a candidate for inclusion in an Internationalized Domain Name.
What is this

- This document reviews and classifies the collections of codepoints in the Unicode character set by examining various properties of the codepoints. It then defines an algorithm for determining a derived property value. It specifies a procedure and not a table of codepoints so that the algorithm can be used to determine code point sets independent of the version of Unicode that is in use.
The list of codepoints that can be found in Appendix A is non-normative. Section 2 and Section 3 are normative.
Property values

- PROTOCOL VALID
- CONTEXTUAL RULE REQUIRED
- DISALLOWED
- UNASSIGNED
Category A

“Good codepoints”

- `generalCategory(cp)` is in
  `{Li, Lu, Lo, Nd, Lm, Mn, Mc}`
Category A

“Good codepoints”

- Ll - Lowercase_Letter
- Lu - Uppercase_Letter
- Lo - Other_Letter
- Nd - Decimal_Number
- Lm - Modifier_Letter
- Mn - Nonspacing_Mark
- Mc - Spacing_Mark
Category B
Normalization and Casefolding

- toNFKC(toCaseFolded(toNFKC(cp))) != cp
Category C
Properties to ignore

- property(cp) is in {
  Default_Ignorable_Code_Point,
  White_Space,
  Noncharacter_Code_Point
}
Category D
Blocks to ignore

- block(cp) in { Combining Diacritical Marks for Symbols, Musical Symbols, Ancient Greek Musical Notation, Private Use Area }

Category E
ASCII LDH

- cp is in {002D, 0030..0039, 0061..007A}
Category F

Exceptions

• \( \text{cp in \{ 002D, 00B7, 02B9, 0375, 0483, 05F3, 05F4, 3005, 3007, 303B, 30FB \}} \)
Category F
Exceptions

- 002D; CONTEXTO # HYPHEN-MINUS
- 00B7; CONTEXTO # MIDDLE DOT
- 02B9; CONTEXTO # MODIFIER LETTER PRIME
- 0375; CONTEXTO # GREEK LOWER NUMERAL SIGN (KERAIA)
- 0483; CONTEXTO # COMBINING CYRILLIC TILTO
- 05F3; CONTEXTO # HEBREW PUNCTUATION GERESH
- 05F4; CONTEXTO # HEBREW PUNCTUATION GERSHAYIM
- 3005; CONTEXTO # IDEOGRAPHIC ITERATION MARK
- 3007; PVALID # IDEOGRAPHIC NUMBER ZERO
- 303B; CONTEXTO # VERTICAL IDEOGRAPHIC ITERATION MARK
- 30FB; CONTEXTO # KATAKANA MIDDLE DOT
Category G
Backward compatibility

- cp in { }
Category H

- property(cp) is in {Join_Control, Resolution}

Require extended special treatment in Lookup and Resolution
Category I

Require special treatment in Lookup and extended special treatment in Resolution.

- generalCategory(cp) is in { Cf Resolution }
Category J
Unassigned codepoints

- cp is unassigned
Algorithm

- Category F, see table Exceptions
- Category G, see table Backward compatibility
- Category E, PVALID ASCII LDH
- Category H, CONTEXTJ Special for lookup, resolution
- Category I, CONTEXTO Lookup and extended resolution
- Category B, DISALLOWED Normalization, Casefolding
- Category C, DISALLOWED Properties to ignore
- Category D, DISALLOWED Blocks to ignore
- Category J, UNASSIGNED Unassigned codepoints
- Category A, PVALID Good codepoints
- not Category A, DISALLOWED “The rest”