

- 1:
2:
3:
4:
5: JGPとしての検討課題への対応
6:
7: ■本資料の目的
8:
9: JGP会合(11)資料4で示し合意を得た今後の検討課題について、IPとの調整結果
10: を反映し、次回CJKおよびIPと調整する際のJGP方新案としてまとめたので、再
11: 度合意を得たい。
12:
13: なお、本資料の方針案については、CGPおよびKGPのLGR- αが確定した後に影響
14: があると判断した場合は再度相談する。
15:
16: 以下、前回資料からの修正箇所は、行頭に縦棒(|)をつけて示している。
17:
18: ■検討課題への対応方針案
19:
20: 凡例 ★：JGP内で相談・検討し、方針を定める
21: ☆：JGPのこれまでの方針に従いCJK調整担当が主に検討・調整を
22: 進め、必要に応じJGPに相談する
23: -：JGPのこれまでの方針に従いCJK調整担当が検討・調整を進め
24: る
25:
26: A. JGPが独自にIPと進めるもの
27: (A1) 明らかな禁止事項等のルールをWLEで定義するか？(★)
28:
29: JGPでの仮合意は、オブジェクションのプロセスで排除できるので日本語LGR
30: では特別なWLEの定義は必要ないとしている。
31: 日本語JPドメイン名での使用状況を確認し、不自然な使われ方がないもしくは
32: は極端に少なければWLEを定義しないという方針を維持するのが妥当であると
33: 考えられる。
34:
35: 2014年10月時点での日本語JPドメイン名登録(125628ドメイン名)における、
36: 不自然と考えられる長音および踊り字の登録状況(数字はドメイン名数)を示
37: す。
38:
39: 文字列の先頭に長音： 0
40: 漢字の後に長音： 4
41: 文字列の先頭に平仮名踊り字： 0
42: 片仮名・漢字の後に平仮名踊り字： 0
43: 文字列の先頭に片仮名踊り字： 0
44: 平仮名・漢字の後に片仮名踊り字： 0
45: 文字列の先頭に漢字踊り字： 0
46: 平仮名・片仮名の後に漢字踊り字： 0
47:
48: (参考)
49: 平仮名の後に長音： 642
50: 文字列の先頭に小平仮名： 6
51: 文字列の先頭に小片仮名： 0
52:
53: ★JGPとしての対応方針(案)
54: 不自然と思われる文字列は「漢字の後ろに長音」の4例しか存在せず、こ
55: れは全体の0.003%と無視できる数であるため、従来のWLEを定義しないと

56: いう方針を維持する。

57:

58: (A2) 多すぎるallocatableを減らす方法は？ (★)

59:

60: | JGPでの仮合意は、C-LGR- α の影響は非常に限定的であり統合アルゴリズム案

61: | によって異体字とされる文字が登録可能(allocatable)の異体字タイプにな

62: | るのであれば許容できるとしている。

63:

64: 統合アルゴリズム案はこの考え方に基づいて設計されているが、IPはこのア

65: | ルゴリズムではJ-LGR- β のWLE適用後に多数の登録可能(allocatable)文字列

66: | (ラベル)ができることを懸念し、改善を要求している。

67:

68: IPからの要求に対してJGPが取り得る対応は以下の3つが考えられる。

69:

70: (1) RootLGRのプロセス外に、申請者が登録可能ラベルの中から登録するラ

71: ベルを選定するプロセスを作るようICANN/IPに要求する

72: (2) CGPに異体字の数を減らすよう要求する

73: | (3) J-LGR- β に基づいて受身的異体字定義および異体字サブタイプ定義をし

74: | た第2世代のJ-LGR- α を作成する

75:

76: | ICANNブエノスアイレス会合におけるCJKおよびIPとの調整において、(1)に

77: | ついてはICANNスタッフを交えて検討していくこととなった。次回ICANNダブ

78: | リン会合にて、CJKからの提案を議論する予定である。一方、CGPはCDNCでの

79: | 15年におよぶ異体字運用の経験に基づき可能な限り最小限の異体字定義を

80: | しており、そこからさらに異体字を減らせと要求する(2)は合意に至る可能

81: | 性は低い。

82:

83: | 並行して下記(B3)の分析を行った後、(3)も検討するのが妥当であると思

84: | えられる。この場合、コミュニティへはJ-LGR- α の作成は他GPやIPからの

85: | フィードバックを得て複数世代に渡って行われたものであり、基本概念は

86: | 第1世代J-LGR- α を作成したときのままであることを十分に説明する必要がある

87: | る(それに矛盾のない第n世代J-LGR- α を作成する必要がある)。

88:

89: | ★JGPとしての対応方針(案)

90: | 6/21-25のCJK調整委員会およびIPとの議論に基づき、2014年10月時点での

91: | 日本語JPドメイン名登録(125628ドメイン名)を対象にどれほどの異体字ラ

92: | ベルが「登録可能」(allocatable)の処置になるかを調査した(前回の調査

93: | は、生成される全異体字ラベル数の調査であったため)。

94: | その結果、1つのドメイン名から生成され「登録可能」となる異体字ラベ

95: | ル数の最大は540通りであること、および全体の95%のドメイン名は登録可

96: | 能な異体字ラベル数が6通り以内であることがわかった。

97: また、実際に登録されている

98: 125628ドメイン名の中で異体字ラベル同士のもの210組(2ラベルが207組、

99: 3ラベルが1組、4ラベルが2組の計425ドメイン名)あり、全体の0.3%であっ

100: た。なお、これら210組に共通のルール(文字列中のすべての新字体が旧字

101: 体に置き換わっているなどの規則性)を見出すことはできなかった。

102: | したがって、第2世代のJ-LGR- α (J-LGR- α 2)では「第1世代のJ-LGR- α

103: | (J-LGR- α 1)に、第1世代のJ-LGR- β (J-LGR- β 1)で異体字とされた文字の

104: | うち、J-LGR- α 1のレパートリーに入っていた文字のみを異体字タイプが

105: | allocatableの異体字として定義する」という方針をIPに示すのが妥当で

106: | あると考えられる(実質的には、J-LGR- α 1との違いはない)。

107: | ただし、IPからは可能性であったとしても登録可能540通りは多いので

108: | 更に減らす努力をして欲しいと求められる可能性がある。その要求への対

109: | 案として、日本語としてみたときに明らかに異体字でない組は、一方が申

110: | 請されたとき(CPに相当)に、他方(VPに相当)の異体字タイプはblockedと

111: | するという案も準備しておくのが妥当と考える。この案をCJK交渉担当3名
 112: | の直感に基づいて事前評価したところ、特に異体字の組が3文字以上にな
 113: | るものを個別の文字とした場合に効果が大きく、ラベル数の最大は256通
 114: | り、全体の97%が6通り以内であることがわかった。なお、この案に基づき
 115: | J-LGR- α 2を実装する場合は、オーソリティのもとで手間を要する異体字
 116: | タイプの分類作業が必要となる。
 117:
 118: B. CJKが一体となって検討するもの
 119: (B1) 用語及びその定義の合意 (－)
 120:
 121: 2015/05のCJK調整委員会での議論を踏まえ、米谷が修正を反映し、CJK GP
 122: MLおよびIPで共有する。
 123: ICANNブエノスアイレスまでに用語及びその定義の合意を得る。
 124:
 125: (B2) アルゴリズムのスペック(入出力仕様とアルゴリズム)の完成 (☆)
 126:
 127: 2015/05のCJK調整委員会での議論を踏まえ、米谷がアルゴリズムのスペック
 128: をドラフトし、CJK GP MLで共有する。
 129: ICANNブエノスアイレスまでにスペックの合意を得る。
 130:
 131: |(B3) 各言語LGR- α の提案が互いに与える影響を分析し、問題ある場合は必要に
 132: | 応じLGR- α を修正する (★)
 133:
 134: | JGPでの仮合意は、C-LGR- α の影響は非常に限定的であり統合アルゴリズム案
 135: | によって異体字とされる文字が登録可能な異体字タイプになるのであれば許
 136: | 容できるとしている。
 137:
 138: | 2015/05のCJK調整委員会用に提出されたCJK各LGR- α (*1)を統合アルゴリズム
 139: | 案で統合した結果導き出されるJ-LGR- β の異体字について、再度日本語JPド
 140: | メイン名登録における異体字とみなされる登録の状況を調査し、影響がない
 141: | ことを確認することが妥当と考えられる。
 142:
 143: | その上で、上記(A2)の問題を解決するための検討と併せてJ-LGR- β からの
 144: | フィードバックとして新たにJ-LGR- α を再定義する検討(J-LGR- β の異体字を
 145: | 使い異体字サブタイプを定義するか、等)を実施するのが妥当であると考え
 146: | られる。
 147:
 148: | (*1) C-LGR- α
 149:
 <<http://mm.icann.org/pipermail/japanesegp/attachments/20150430/4fd4c763/mss-var-20150430-0001.xml>>
 150: | J-LGR- α
 151:
 <<http://mm.icann.org/pipermail/japanesegp/attachments/20150427/3bbfc797/J-LGR-1-0001.xml>>
 152: | K-LGR- α
 153:
 <<http://mm.icann.org/pipermail/japanesegp/attachments/20150515/643c2fac/K2-KLGR-1-0001.xlsx>>
 154:
 155: ★JGPとしての対応方針(案)
 156: | 調査結果はA2で述べたとおりであり、C-LGR- α の影響は限定的であるとい
 157: | えるため、従来他LGR- α の異体字定義を受け入れる方針を維持する。
 158:
 159: C. その他

160: (C1) 現レパートリー案から不適切な文字を削除 (★)
161:
162: JGPでの仮合意は、JIS X 0208第1水準・第2水準の平仮名・片仮名・漢字お
163: よびそれらに準ずる文字を文字範囲(レパートリー)とし、それに対する文字
164: の追加・削除の合理的な理由はないとしている。
165: MSRのように、ICANN/IPが定める外的要因がない限り、この方針は変更しな
166: いのが妥当であると考えられる。JGPチェアからは「各言語GPはその言語お
167: よびScriptの範囲を超えた判断はすべきではなく、その判断はGPの外部で行
168: うべき」というアピールをICANNに対して実施しており、CGPチェアからも
169: ICANNスタッフがいる場で賛同を得ている。
170:
171: 日本語LGRに対しICANNが実施するパブリックコメントで指摘が行われる可能
172: 性があるので、回答は用意しておく必要がある。
173:
174: なお、2014年10月時点での日本語JPドメイン名登録(125628ドメイン名)中、
175: 以下の文字の登録状況(数字はドメイン名数)を示す。
176:
177: ㇿ : 2
178: あ : 19
179: い : 67
180: う : 6
181: え : 51
182: お : 61
183: ゐ : 1
184: ゑ : 3
185: 𛄟 : 3
186: 𛄠 : 8
187:
188: ★JGPとしての対応方針(案)
189: 上記文字を排除する合理的な理由は存在しないため従来の範囲を維持する。
190:
191: 以上