

縮退マップについて

2015年9月

独立行政法人 情報処理推進機構

経緯

IPA[®]

- 3月31日 MJ縮退マップVer.0.1(検証版)を公開
 - パブコメを実施
 - 専門家による検討を実施
- 7月31日 実験的コンテンツを「モジラボ」として公開
 - MJ縮退マップの開発中バージョン
 - 文字の関係性をグラフ化するツール

- 3月末に公開した「MJ縮退マップ」の正式版
 - 特徴：複数の縮退候補を根拠と共に示す
 - 縮退候補の追加
 - 法務省告示582号(外国人登録関連)別表第二を縮退根拠に追加
 - 表示する縮退根拠情報の追加
 - 「戸籍統一文字情報の親字・正字情報」を何段階辿ったかを示す「ホップ数」(本資料p.5-6参照)の表示を追加
 - 法務省告示582号別表第四を用いた縮退について、「別表第四の一」「別表第四の二」の区別の表示を追加
- 縮退先を一つに絞った縮退変換表（新規）
 - 特徴：MJ縮退マップを元に、縮退先を一つ以下に絞り込んだもの。

2

公開する形式

- JSON形式のテキスト（図形は含まない）
 - 意見募集の結果、表形式の提供を望む意見が寄せられた、システム開発者等現場からはコンピュータ処理で使いやすいという意見が多かった
 - 既に多くのベンダーが検証版を利用
- 簡易検索システムでの縮退情報の提示
 - 目視に適した情報提供への要望に対応



3

縮退マップのイメージ

実際のマップはJSON形式（図形無し）で公開予定

MJ文字図形	JIS包摂規準/ UCS統合規則	戸籍法関連通達	法務省告示582号	字辞典掲載 を確認	字辞典での関連字	音・字形等 からの推定
辻 MJ025760	辻 1-36-52 U+8FBB			○		
左 MJ000259		左 1-26-24 U+5DE6	尤 1-53-87 U+5C22	○	左, 中, 中 1-26-24, 1-54-05, 1-47-66 U+5DE6, U+5C6E, U+FA3C	
曉 MJ059692		曉, 曉 1-22-39, 1-58-92 U+6681, U+66C9		×		
擦 MJ012884				×		擦 1-58-22 U+64FD
皷 MJ068029				×		

4

戸籍統一文字情報のホップ数について

MJ030418の例

The screenshot shows two browser windows. The left window displays the '戸籍統一文字情報' (Unified Japanese Character Information) page for MJ030418. It shows the character '隠' (kaku) with its JIS level, Unicode, and other details. The right window shows a graph titled '戸籍統一文字情報 親字・正字' (Unified Japanese Character Information Parent Character/Standard Character). The graph illustrates the relationships between characters: '隠' (MJ030418) is the central node, connected to '隠' (MJ030418), '隠' (MJ030418), '隠' (MJ030418), '隠' (MJ030418), '隠' (MJ030418), '隠' (MJ030418), '隠' (MJ030418), and '雪' (MJ030418). The graph also shows '隠' (MJ030418) connected to '隠' (MJ030418) and '隠' (MJ030418). The graph is labeled with '1hop' and '2hop' and '3hop'.

<http://kosekimoji.moj.go.jp/kosekimojidd/mjko/PeopleTop>

<http://mojikiban.ipa.go.jp/lab/kosekiOyaji.html>

5

MJ030418 の例

検証版

```
{
  "MJ文字図形名": "MJ030418",
  "法務省戸籍法関連通達・通知": [
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "JIS X 0213": "2-01-15", "UCS": "U+4E5A"
    },
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "JIS X 0213": "1-17-03", "UCS": "U+96A0"
    },
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "JIS X 0213": "1-80-12", "UCS": "U+96B1"
    }
  ]
},
```

正式版

```
{
  "MJ文字図形名": "MJ030418",
  "法務省戸籍法関連通達・通知": [
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "ホップ数": 1,
      "JIS X 0213": "2-01-15", "UCS": "U+4E5A"
    },
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "ホップ数": 2,
      "JIS X 0213": "1-80-12", "UCS": "U+96B1"
    },
    {
      "種別": "戸籍統一文字情報 親字・正字",
      "ホップ数": 3,
      "JIS X 0213": "1-17-03", "UCS": "U+96A0"
    }
  ]
},
```

6

縮退先候補を1文字に絞った変換表について

• 目的

– 人手を介さない自動的変換

- 大きな文字集合を用いて作られたテキストを、小さな文字集合のみを用いたテキストへ自動的に変換する

– 変換前にあらかじめ縮退対応関係を吟味するよりも、変換後に対応する業務スタイルを想定

- 変換可能な文字の多さよりも、変換できないことを明示することが重要

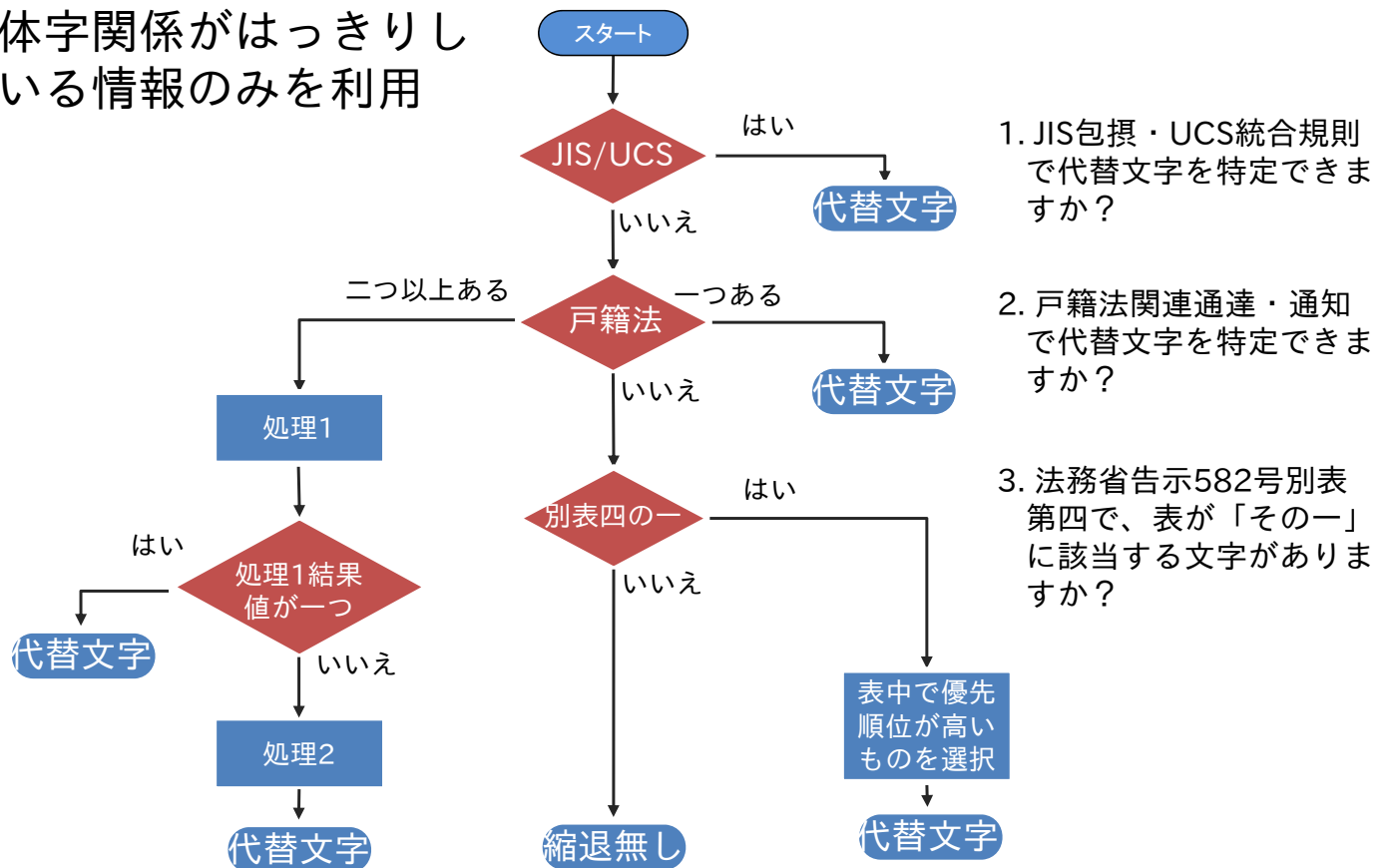
• 作成にあたっての基本的考え方

– 縮退はJISやISOにおける統合/包摂基準、辞書、告示文など、客観的裏づけのあるものに限る。

– MJ縮退マップからの派生物という位置づけとする。

7

異体字関係がはっきりしている情報のみを利用



8

「処理1」、「処理2」、「縮退無し」の説明

• 処理1

- ① 民二第5202号通知別表(正俗表)に掲載されているものがある。
- ② 戸籍統一文字情報 親字・正字ホップ数が最小のものが一つある。
- ③ 親字・正字ホップ数が最小のものが二つあり、一方が誤俗表にも掲載されているもの (ただし別字ではない)

• 処理2

- ① 候補が複数存在する場合、常用漢字、人名用漢字の順に選択
- ② ①で絞り込めない場合、JIS水準の低いものを選択

• 縮退無し

- 代替文字として「_」を仮置きし、縮退候補を示さなかったことを明示するとともに、後の処理を促す。

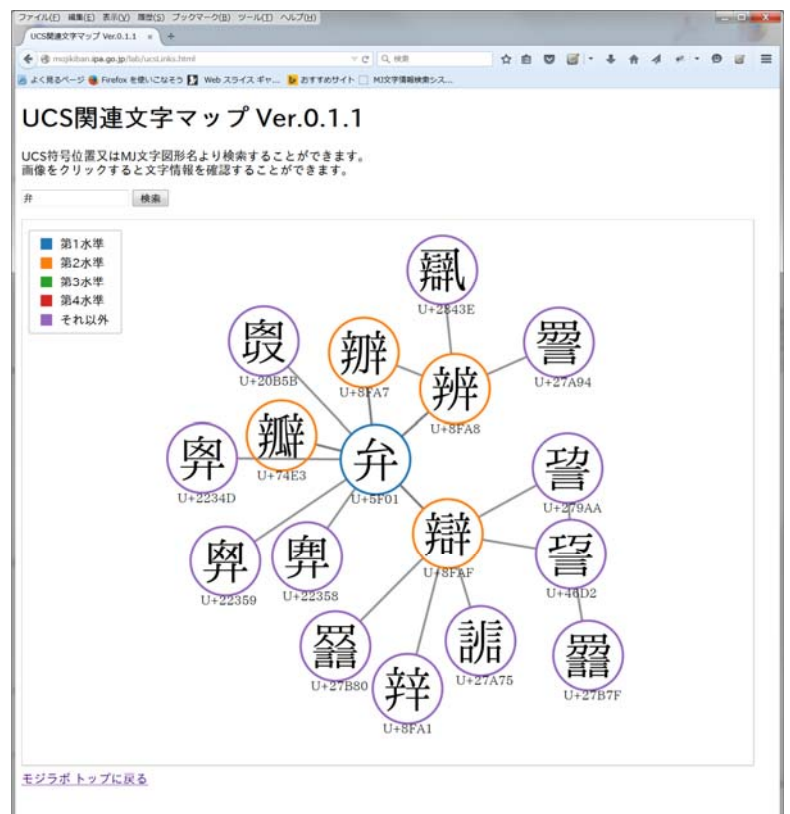
9

- 異体字等の関係を示す可視化ツール
 - 文字の関連性をグラフ形式で一覧できる
- 変換表作成ツール
 - MJ縮退マップ(JSON形式)から縮退候補を一つに絞った変換表を作成するためのプログラム
 - 複数の縮退候補から一つを選択するためのアルゴリズム(P.8)を組み込んだもの
 - IPAの提供する変換表の作成に使用
 - モジラボにてOSSライセンスにより提供
 - 複製、使用、改変が無償で可能

10

可視化ツール表示例

- 文字の関連性をグラフ構造で可視化
 - 字典から得られた文字の関係性をグラフ形式で表示
 - 縮退候補の関連性を一覧できる
 - JISの水準を色分けにより表示



<http://mojikiban.ipa.go.jp/lab/ucsLinks.html>

11

- 登記統一文字番号縮退マップ
 - 登記統一文字に固有な文字の一部を加え、MJ縮退マップと同様の処理により縮退候補情報を整備
 - JSON形式
- 上記縮退マップから縮退先候補を1文字に絞った変換表
 - 自動テキスト変換を想定し、本資料p.8のアルゴリズムにより縮退先を一つに絞ったもの。
 - 縮退候補がない場合は「_」を仮置き。
 - 併せて、上記縮退マップから上記変換表を作成するツールも提供。

- 登記統一文字は一般に公開されていない
 - IPAの公開する情報には登記統一文字番号と縮退候補情報のみを掲載
 - 国税庁が必要に応じ登記文字の図形を公開
- 登記統一文字は随時追加される
 - 追加された文字の縮退対応について、検討を続ける必要がある。

- 縮退マップの公開に併せて、MJ文字情報一覧表をマイナーバージョンアップ
 - JIS X 0213の対応情報の見直し（2文字追加）
 - MJ011199 強 (なし) → 1-22-15 (強)
 - MJ032950 壯 (なし) → 1-52-67 (壯)