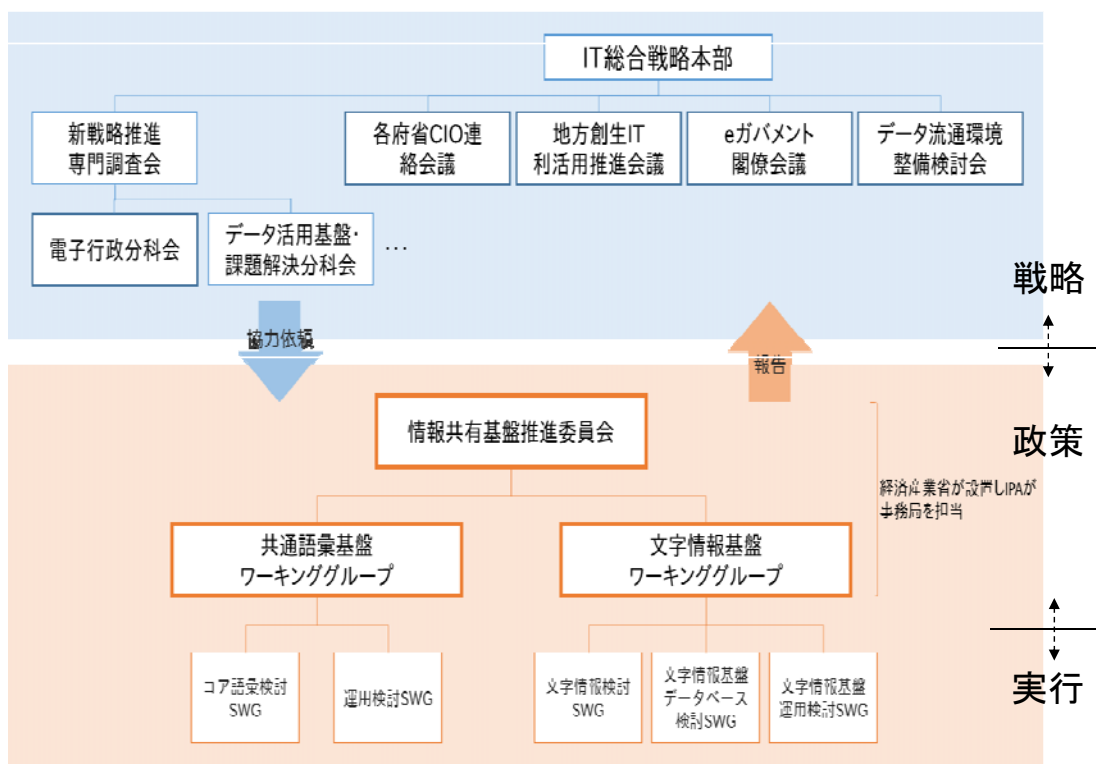


文字情報基盤 2016年度実施状況と2017年度実施方針

IPA技術本部国際標準推進センター



共通語彙基盤、文字情報基盤の事業実施体制



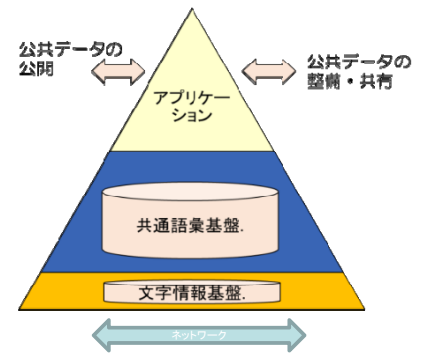
公共データの利活用等政府方針に基づく電子行政システムの構築支援

■ 文字情報基盤整備事業

- 人名の正確な表記を要する行政実務で必要とされる約6万の漢字を整備
 - ・ 国際標準化
 - ・ フォント、文字情報、縮退マップ等の提供
 - ・ データベース構築と運用

■ 共通語彙基盤整備事業

- オープンデータの公開、システム間の効率的な情報連携に必要な基本的情報の整備
 - ・ データ項目名等の標準化
 - ・ 情報連携用語彙データベースと、それを活用するためのツール群の構築
 - ・ 国際連携（欧州委員会、米国政府）



データに価値を生み出す基盤

- ・ 「コンピュータ、データ無ければただの箱」
- ・ 流通性の無いデータに対して「情報貸し金庫」は可能だが「情報銀行」は不可能。

閣議決定「世界最先端IT 国家創造宣言」

2016年5月改定

I.世界最先端IT 国家創造宣言に基づくこれまでの成果

I.-1. これまでの代表的な成果

I.-1-(1) 行政情報システム改革を通じた利用者志向の行政サービスの実現

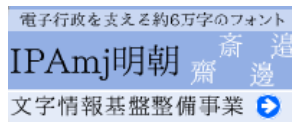
[主な成果]

- ・ **文字情報基盤**(注4)と**共通語彙基盤**(注5)を整備。国・地方公共団体がこれらの基盤を活用することで、組織や業務分野を超えた行政情報システム間の連携が可能となり、円滑な公共サービスの提供を期待。

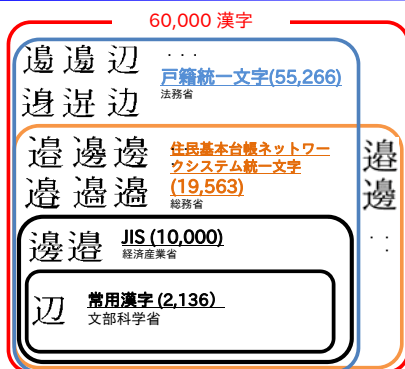
実施スケジュール (1. [重点項目1] 国・地方のIT化・業務改革(BPR)の推進)

	短期			中期			長期		
	2013	2014	2015	2016	2017	2018	2019	2020	2021
オープンな利用環境の整備	オープンな利用環境整備に向けたアクションプラン策定【内閣官房】	アクションプランを踏まえ、府省庁間連携を推進し、利便性の高い公共サービスを実現【内閣官房、関係府省庁】			法人名のフリガナ表記方針策定【内閣官房、関係府省庁】				
	文字情報基盤導入ガイドの作成【経済産業省】				文字情報基盤の導入【内閣官房、経済産業省、関係府省庁】				
	情報連携用語彙データベースの開発・実証【経済産業省】				用語の共通化基盤の推進【内閣官房、経済産業省、関係府省庁】				
	情報流通連携基盤共通API(データモデル、共通ボキャブラリ、標準API規格等)等を開発・実証【総務省】				共通化推進の機能整備の検討【内閣官房、経済産業省、関係府省庁】	共通化推進の機能整備に向けた必要な措置の実施【内閣官房、経済産業省、関係府省庁】			
	連携			データ構造の策定【内閣官房、総務省、関係府省庁】	ポータルサイトの実証【内閣官房、総務省、関係府省庁】	ポータルサイトの運用【内閣官房、総務省、関係府省庁】			

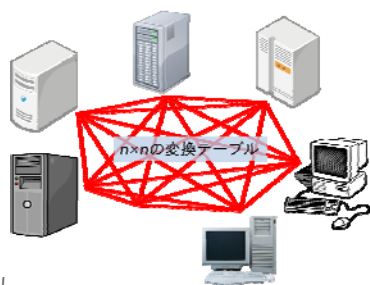
文字情報基盤 2016年度実施状況



文字情報基盤整備事業とは

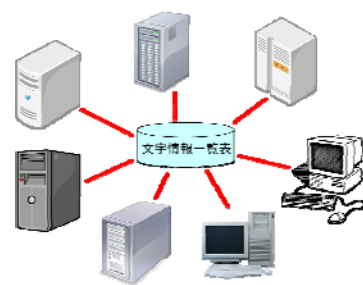


- 標準が無く、自治体ごとに独自の「外字」を作成し、運用せざるを得なかった状況を、国際標準化することで解決
- 業務に応じて適切な文字セットを使えるよう、「縮退マップ」を提供
- 戸籍 / 住基 / JIS / ISO 等の文字の関係を記述した「文字情報一覧表」を提供
- 自治体の現場やシステム開発者が参照する文字データベースを提供



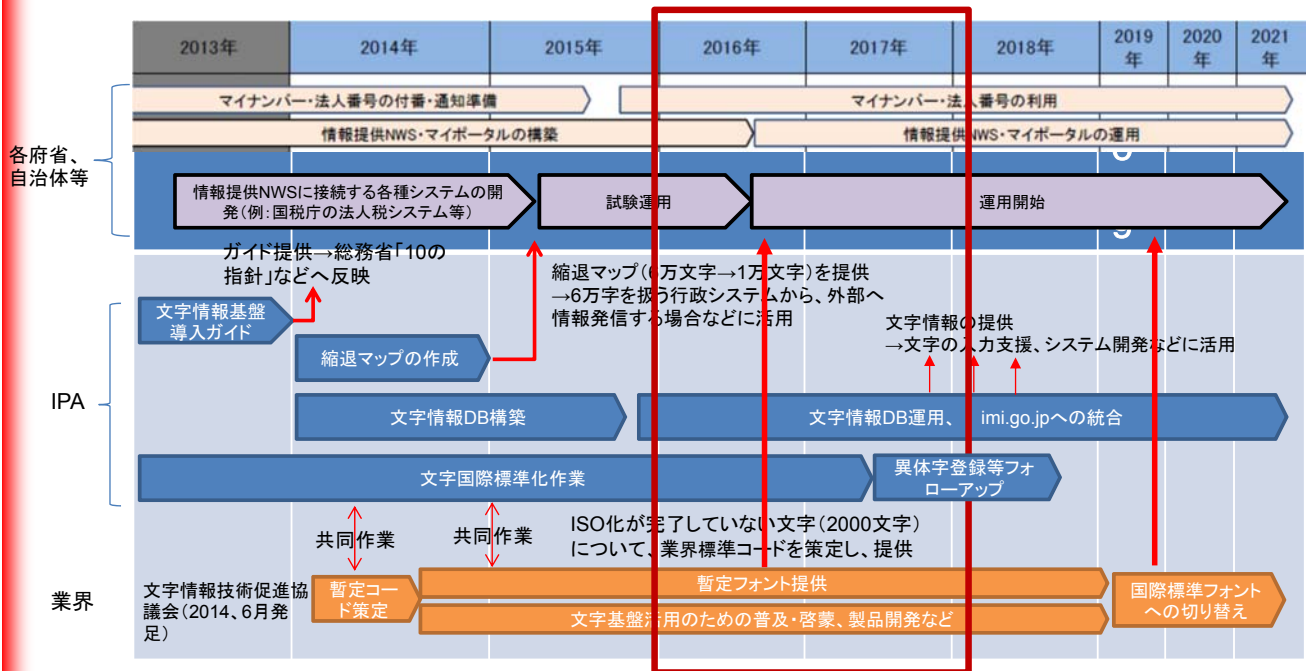
文字の標準なし
各自治体、システムが独自の文字体系を運用

- 情報連携が困難
- 外字の作成・運用に高いコスト



文字情報基盤の活用

- 情報連携が容易
- 外字の作成コストが無い



検討体制

文字情報基盤ワーキンググループ（2016年度）

委員長

林 史典 聖徳大学/聖徳大学短期大学部 教授
人文学部長/大学院言語文化研究科長/人文学部日本文化学科長

委員

大山 水帆 川口市役所 企画財政部情報政策課 課長
 杉谷 達哉 法務省 民事局 民事第一課 補佐官
 武田 康宏 文化庁 文化部国語課 国語調査官
 平本 健二 経済産業省CIO補佐官
 名越 一郎 総務省自治行政局 住民制度課 理事官
 三上 喜貴 長岡技術科学大学 副学長(国際交流担当)
 吉川 桂 地方公共団体情報システム機構住民基本台帳ネットワークシステム全国センターシステム部 担当課長

楠 正憲 文字情報基盤運用検討SWG主査
内閣官房 社会保障改革担当室 番号制度推進管理補佐官

高田 智和 文字情報基盤文字情報検討SWG主査
人間文化研究機構 国立国語研究所 理論・構造研究系 准教授

オブザーバ

総務省 行政管理局 行政情報システム企画課

- 5月 • ISO/IEC JTC1/SC2 WG2/IRG会議 (北京)
→CJK統合漢字拡張Fへのコメントへの対応
- 7月 • 文字情報基盤データベースを公開
• IPAmj明朝フォント、MJ文字情報一覧表およびMJ縮退マップをバージョンアップ
- 9月 • ISO/IEC JTC1/SC2会議 (サンノゼ)
→変体仮名議論
→常用漢字文字セットの提案
- 10月 • ISO/IEC JTC1/SC2 WG2/IRG会議 (松山)
• ISO/IEC JTC1/SC2 10646(ed5) DIS投票 →反対票0
→文字情報基盤 全漢字の符号化完了へ
- 11月 • ISO/IEC JTC1総会(リレハンメル)
- 3月 • ISO/IEC JTC1/SC2 DIS10646(ed5) PDAM 投票
→変体仮名、常用漢字文字セットが決着見込み

8

- 文字情報基盤成果物をバージョンアップし、MJ文字情報一覧表Ver.005.01、IPAmj明朝フォントVer.004.01およびMJ縮退マップVer.1.1.0を公開いたしました
 - MJ文字情報一覧表Ver.005.01 : 戸籍統一文字、住民基本台帳ネットワークシステム統一文字との対応関係を見直し、MJ文字図形名を追加
 - IPAmj明朝フォントVer.004.01 : MJ文字情報一覧表Ver.005.01で追加されたMJ文字図形と変体仮名文字図形のフォントを追加
 - MJ縮退マップVer.1.1.0 : MJ文字情報一覧表Ver.005.01で追加されたMJ文字図形名の対応とバグ修正を実施

9

- MJ文字図形と戸籍統一文字・住基ネット統一文字との対応関係を見直し、46文字のMJ文字図形を新規に追加した。

1. 視覚的な差異が大きかったものについて、新規のMJ文字図形名を追加し、戸籍統一文字、住基統一文字の対応関係を更新。

2. 戸籍統一文字と住基統一文字をそれぞれ比較した結果、対応関係を更新する必要が生じたもの。新規MJ文字図形名の追加が行われているものを含む。

Ver.005.01	Ver.004.03	戸籍統一文字	住基統一文字
MJ068056	MJ001937	184450	
MJ068057	MJ005507	489630	
MJ068058	MJ007341	018870	J+AEEE

	MJ文字図形	戸籍統一文字	住基統一文字
Ver.004.03			
	MJ000007	152700	
Ver.005.01			
	MJ000008	152770	J+3404
Ver.005.01			
	MJ000007	152700	J+3404
Ver.005.01			
	MJ000008		
Ver.005.01			
	MJ068055	152770	

MJ文字追加

- IPAmj明朝フォントVer.004.01
 - MJ文字情報一覧表Ver.005.01で追加されたMJ文字図形と変体仮名文字図形のフォントを追加
- MJ縮退マップVer.1.1.0
 - MJ文字情報一覧表Ver.005.01で追加されたMJ文字図形名の対応とバグ修正を実施

誤字俗字・正字一覧表

「正字等」漢字番号	「正字等」MJ別形名 (Subject)	アサの差	「誤字等」漢字番号	種別と表示記号 (Predicate)	「誤字等」MJ別形名 (Object)	アサの差
1554-1-1	MJ011752		1554-1-1-1	民事2842誤字		
1554-1-1	MJ011752	1				
1554-1-1	MJ011752	1				
1554-1-2	MJ010612		1554-1-2-1	民事2842誤字		
1554-1-2	MJ010612		1554-1-2-2	民事2842誤字		
1554-1-3	MJ010114		1554-1-3-1	民事2842誤字		
1554-1-4	MJ009693		1554-1-4-1	民事2842誤字		
1554-1-4	MJ009693	1				
1554-1-4	MJ009693		1554-1-4-2	民事2842誤字		
1554-1-5	MJ016350		1554-1-5-1	民事2842別字(▲)	MJ057990	

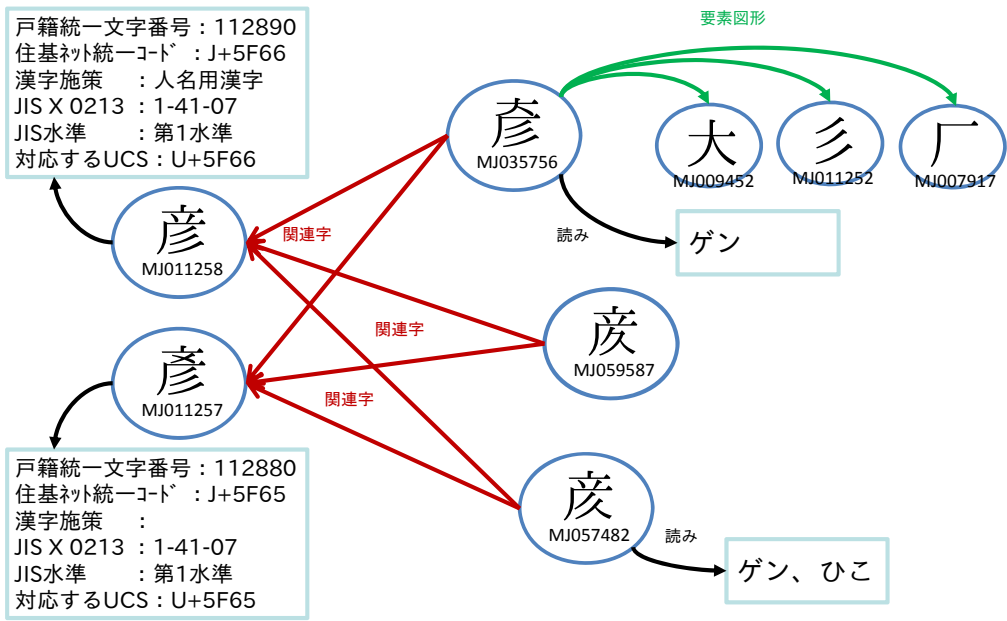
誤字俗字・正字一覧表における整理

辞書における関係性の整理

代表字 異体字

関係

Subject	Predicate	Object
8351	本字	8352
8351	別体	8353
8351	別体	8354
8351	別体	8355
8351	別体	8356
8351	別体	8357
8351	別体	8358
8351	別体	8359
8351	別体	8360
8351	別体	8361
8351	別体	8362



探したい文字

巛

読みが分からない
部首も分からない

従来の検索

総画数:18、部首:山で検索



みつからない



要素図形検索

要素図形:山、英で検索



検索結果:1件



U+908A(邊)と辞書上の関連がある漢字を検索

```

PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX class:<http://mojikiban.ipa.go.jp/terms/class/>
PREFIX mjp:<http://mojikiban.ipa.go.jp/terms/prop/>
PREFIX daikanwa:<http://mojikiban.ipa.go.jp/data/大漢和辞典/>
PREFIX nihongo:<http://mojikiban.ipa.go.jp/data/日本語漢字辞典/>
PREFIX shindaiji:<http://mojikiban.ipa.go.jp/data/新大字典/>
PREFIX daijigen:<http://mojikiban.ipa.go.jp/data/大辞源/>
PREFIX daikangorin:<http://mojikiban.ipa.go.jp/data/大漢語林/>
PREFIX ucs:<http://mojikiban.ipa.go.jp/data/crep/>
PREFIX mj:<http://mojikiban.ipa.go.jp/data/MJ/>
SELECT DISTINCT ?label WHERE {
    {
        VALUES ?対応UCS [ ucs:908A ]
        ?moji1 mjp:対応UCS ?対応UCS .
    }
    ?moji1 mjp:検字番号 ?dicmoto .
    ?node mjp:元検字番号 ?dicmoto .
    ?node mjp:先検字番号 ?dicsaki .
    ?moji2 mjp:検字番号 ?dicsaki .
    ?moji2 rdfs:label ?label .
}
ORDER BY ?label
LIMIT 2000
    
```

```

1 ?label rdfs:label "U+908A"
2 sparql select "http://www.w3.org/2005/sparql-result#"
3 <head>
4   <variable name="label" />
5 </head>
6 <result>
7   <binding name="label" />
8   <literal "U+908A" /literal>
9 </binding>
10 </result>
11 <binding name="label" />
12 <literal "U+9080" /literal>
13 </binding>
14 </result>
15 <binding name="label" />
16 <literal "U+9081" /literal>
17 </binding>
18 </result>
19 <binding name="label" />
20 <literal "U+9082" /literal>
21 </binding>
22 </result>
23 <binding name="label" />
24 <literal "U+9083" /literal>
25 </binding>
26 </result>
27 <binding name="label" />
28 <literal "U+9084" /literal>
29 </binding>
30 </result>
31 <binding name="label" />
32 <literal "U+9085" /literal>
33 </binding>
34 </result>
35 <binding name="label" />
36 <literal "U+9086" /literal>
37 </binding>
38 </result>
39 <binding name="label" />
40 <literal "U+9087" /literal>
41 </binding>
42 </result>
43 <binding name="label" />
44 <literal "U+9088" /literal>
    
```





縮退対応情報の表示

IPA MJ文字情報検索システム(簡易版)

詳細

ホームへ戻る 検索結果へ戻る

MJ053396

齎

MJ文字情報一覧表Ver.004.03

コピペフィールド	齎
戸籍統一文字番号	490050 (親字_正字を指す)
住基ネット統一コード	
入管正字コード	
入管外字コード	
漢字施項	
JIS X 0213	
包摂区分	
包摂標準	
JIS水準	---
JIS X 0212	
対応するUCS	U+29413 (関連字を指す)
対応する互換漢字	
UCS対応カテゴリ	A
実装したUCS	U+29413
実装したMojik_Joho	IVS
対応するHanyo	

MJ縮退マップVer.1.0.0

JIS包摂・UCS統合
 法務省戸籍法関連通達・通知
 法務省告示582号別表第四 その一
 法務省告示582号別表第四 その二
 辞書類等による関連字

齎 齎 齎

1-80-77 1-80-78 1-83-78
 読み・字形による類推
 参考情報



- 法人の商号及び所在地の文字を登記統一文字(7万)からJIS X 0213(1万)に変換するのに利用
- 縮退マップにより変換できないものは“ ”を仮置のうえ、画像で表示(約430万の法人のうち、変換できなかった法人は約1700(仮名、記号等や中国等の法人を含む))



国税庁法人番号公表サイト

国際標準化の状況（2017年2月現在）

ISO/IEC JTC1 SC2会議（2016年9月26日～30日 於サンノゼ）の成果

- 変体仮名
- 常用漢字に対応する文字集合（‘Japanese core Kanji’）をISO/IEC 10646 第5版のAmendment 1投票原案に追加する
→ 投票中(3月7日締め切り)

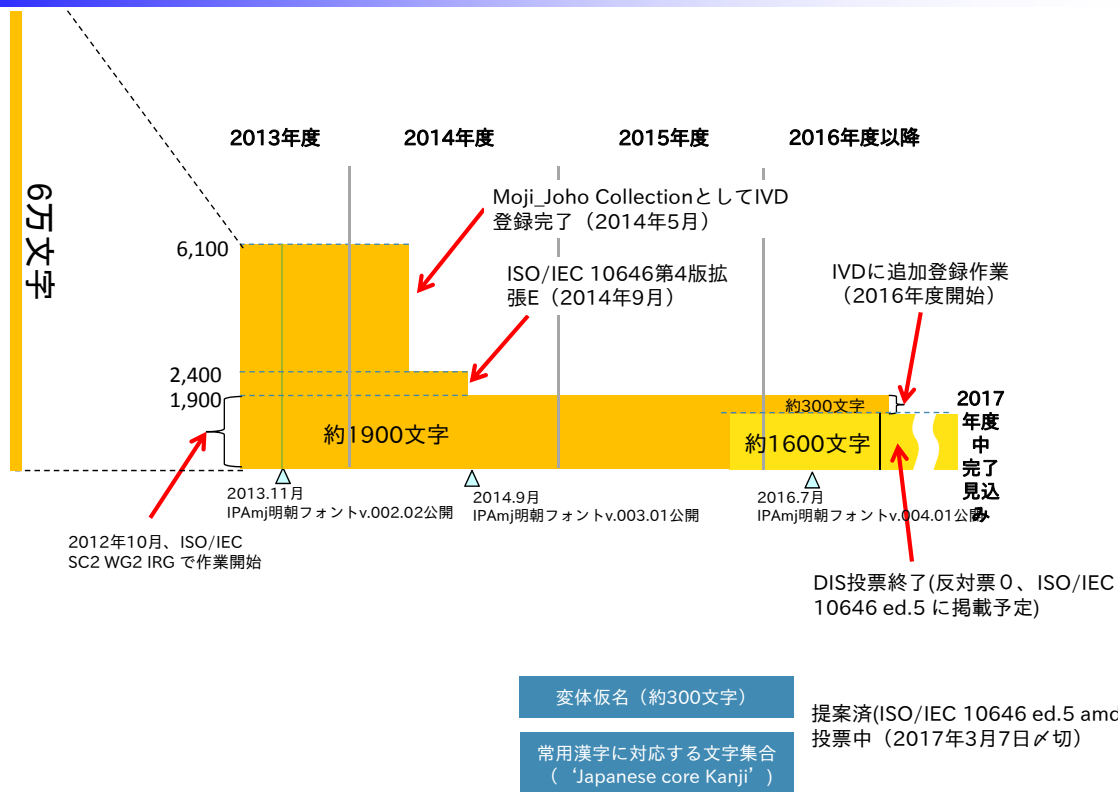
ISO/IEC 10646 第5版 の投票(DIS投票)結果 (10月締め切り)

- 文字情報基盤から提案した最後の文字を含む CJK Unified Ideograph extension F が反対票0で可決

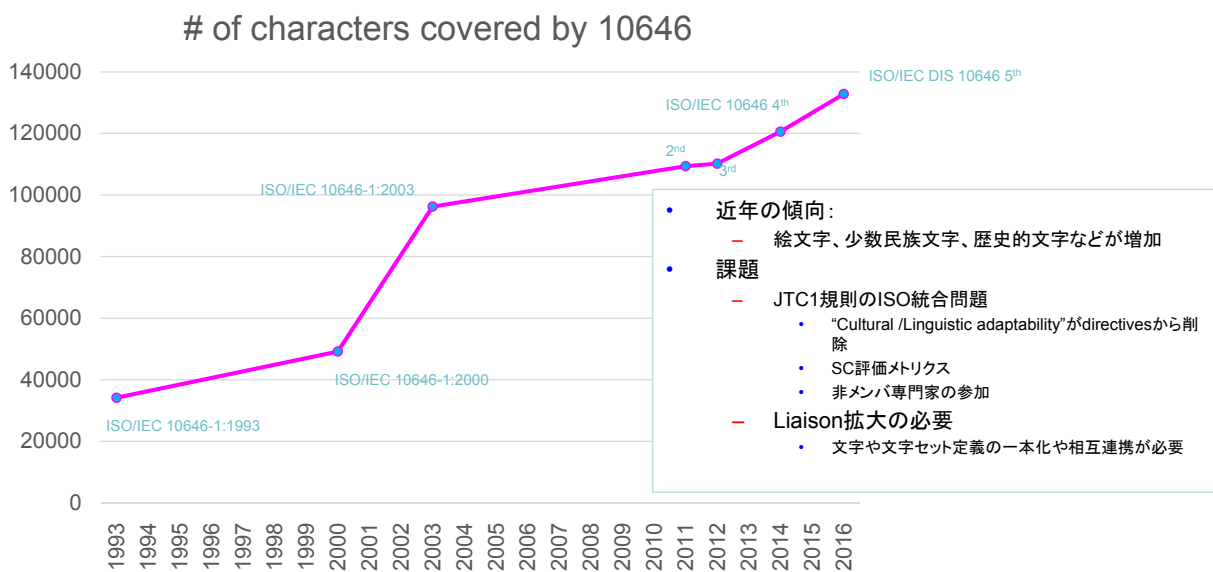


日本提案分の漢字標準化が全て完了の見込み

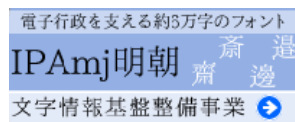
2017年中に発行の見込み



ISO/IEC 10646 文字数の変遷



文字情報基盤 2017年度実施方針



文字情報基盤 2017年度実施ポイント

開発・標準化から普及のフェーズへ

- 文字情報データベースの改善
- 国際標準化の完了を受けた
 - フォント、一覧表のバージョンアップ、
 - 異体字識別符号のIVD※登録
→情報処理学会情報規格調査会で議論
- 文字の修正
 - より戸籍文字に近い図形を一つ新たに追加
 - UCS符号位置への対応が適切でない3文字について、新たにUCS符号を提案（迅速プロセスで提案）
- 普及へ向けた協力
 - 省庁、自治体
 - 民間団体（文字情報技術促進協議会）との連携
- 国際標準完了報告イベントの実施
→年度末近くを目処

※IVD：Unicode consortiumが運用する異体文字符号のデータベース。ISO/IEC 10646から公式に参照されている。

文字の追加(1文字)

MJ 文字図形	戸籍統一文字
 MJ059043	 499580

問題点：戸籍統一文字番号：499580に対応するものとしてい
るMJ059043の図形が、不適切



対応策：戸籍統一文字番号：499580により類似した図形を作成し、新
たなMJ文字図形名を与える。国際符号としては、IVS（異体
字識別符号）の追加で対応

符号位置の追加提案(3文字)

MJ 文字図形名	現在の符号位置
 MJ013582	6730 <small>木 75.1</small>  <small>GK-682E</small> U+6730
 MJ057449	5ED7 <small>广 53.11</small>  <small>G3-3A62</small> U+5ED7
 MJ057586	23106 <small>方 70.8</small>  <small>UCS2003</small> U+23106

問題点

これら3文字について、現在対応付
けているUCS符号位置は適切でない。



対応策

これら3つに対応する新たな符号位置を、
ISO/IEC JTC1/SC2へ提案する。
提案に当たっては迅速法の適用を要請する。

IVD Moji_Johoコレクションの登録追加

- 以下を含む673文字について、2017年度に登録作業を開始。登録手続きの主体は、前回登録時と同様、情報規格調査会SC2専門委員会となる。
 - 前回の登録（2014年5月）後に出版されたISO/IEC 10646 ed4で追加された文字（CJK拡張E）に対応するもの・・・27文字
 - ISO/IEC 10646 ed5で追加される文字（CJK拡張F）に対応するもの・・・192文字
 - MJ文字情報一覧表に追加した文字に対応するもの・・・47文字
 - 同じ字形で字義が異なる文字(同形異字)・・・28文字(15組)
- UCSの一つの符号位置に、一つのグリフしか存在しない文字について、IVDへ登録するか、他の方法をとるかについて、情報規格調査会SC2専門委員会で議論を行う。

28

文字情報基盤データベースの改良

1. 文字情報基盤DBと文字情報検索システム(簡易版)の統合
 - 現在は、別々のサービスとして提供している2システムを統合
2. ユーザーインターフェースの充実化
 - 関連性ビジュアルライザーの追加
モジラボにおいて好評を得ている文字ふわふわマップを取り込んだ、関連性ビジュアルライザーを開発
 - SPARQLクエリービジュアルライザーの追加
SPARQL検索の良さを生かし取っ付きにくさを改善したSPARQLクエリービジュアルライザーを開発
 - 文字検索UIの改良
複雑な検索ができる一方で、操作手順数が煩雑で一般的な利用に向かないという意見があるため、文字検索UIの改良
3. コンテナ型仮想化によるオフラインシステム構築
 - オフライン環境化でも文字情報基盤DBを利用できる様にする。
行政機関では、オフラインで文字検索を行ないたいというニーズが強いため、セキュアな環境を意識しつつ、オフライン環境下でもある程度簡易に展開できる様な仕組みを開発し提供する

29

文字情報基盤データベース

MJ文字情報検索システム(簡易版)

簡易版が提供するAPIを利用

- RDF・文字の関係性情報を提供
- 要素図形(文字部品検索)を提供
- MJ縮退マップ 縮退情報を提供
- MJ文字情報APIを提供

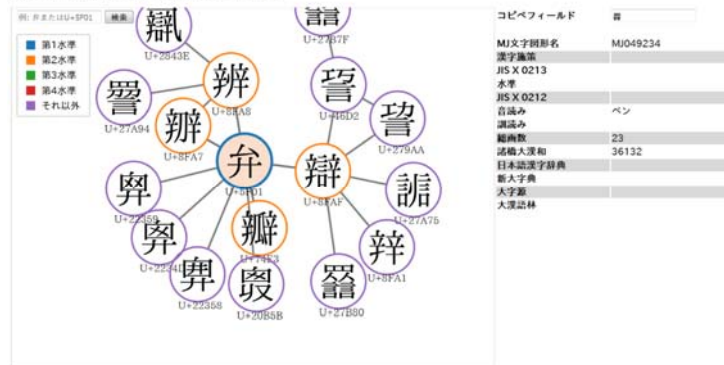
文字検索インタフェースの充実化

- 漢字検索機能の改善
 - 現在の基本検索は、複雑な検索ができる一方、検索条件の入力ステップ数が多く不便とのご意見
 - 検索条件の入力項目の見直し
 - 分かりやすいインタフェースに改良
- 関係性ビジュアルライザーの導入
 - モジラボで公開してる文字ふわふわマップを統合
 - MJ縮退マップの理解、利用促進
 - 文字の関係性検索するためのSPARQLクエリ検索にビジュアルライザーを付加
 - SPARQL検索の利用を促進

- 文字の関係性可視化ツール (試験公開中)
 - 文字の関係性把握や関連から文字を検索
 - 文字実務者からとても便利と高い評価
- 機能拡張し、MJ文字情報DBに統合

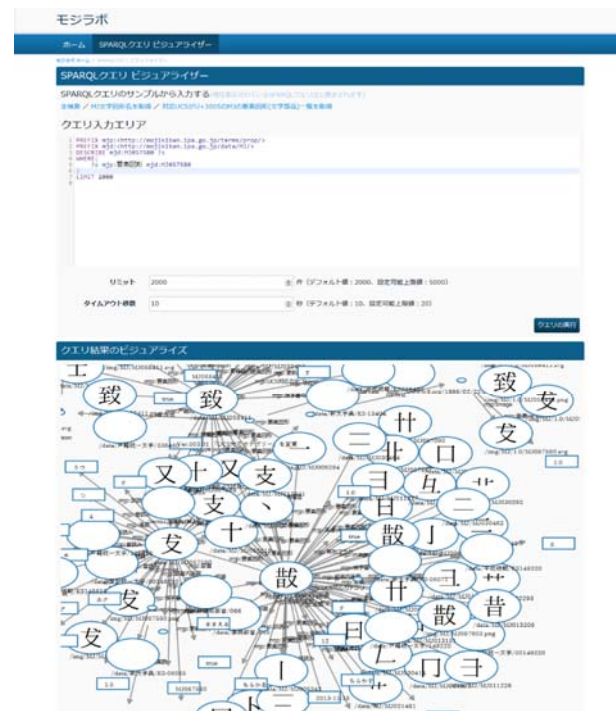
UCS関連文字マップ Ver.0.2.0

UCS符号位置又はMJ文字図形名より検索することができます。



画像をダブルクリックすると文字情報を確認することができます。
モジラボトップに戻る

- 試験的に公開中のツール
- SPARQLクエリ検索の結果をビジュアライズし、DBに格納された関係性情報を可視化
 - 現在は検索結果がファイルで返されるため、利用が少ない
 - ビジュアライザーと典型的なクエリで活用を促進



- 人名用漢字の検索を頻繁に利用する行政機関等の業務システムは、ネットワークに接続していないことが多い(オフラインシステム)
 - 文字情報基盤の検索システムをスタンドアロンで使いたいというニーズがある。(IPA内も同様の要望)
- 近年普及し始めているコンテナ型仮想化を利用し、スタンドアロンでの利用を希望するユーザーに配布可能な形態で開発

コンテナ型仮想化のイメージ

