

JEPA

製品紹介セミナー&ビジネスパートナー交流会
「未来志向の技術展」

EPUB3で外字異体字のバリアフリーを実現する

汎用的な文字図形番号をキーにして、
フォント実装環境に依存しない文字表示を実現する
「SVG画像フォント」の実際と課題を紹介

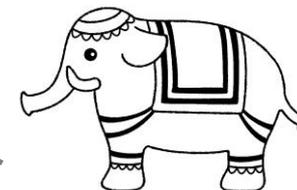
あらゆる歴史的な文字遺産を電子的に表現することを目指して活動する

特定非営利活動法人文字文化協會

<http://www.pcc.or.jp>

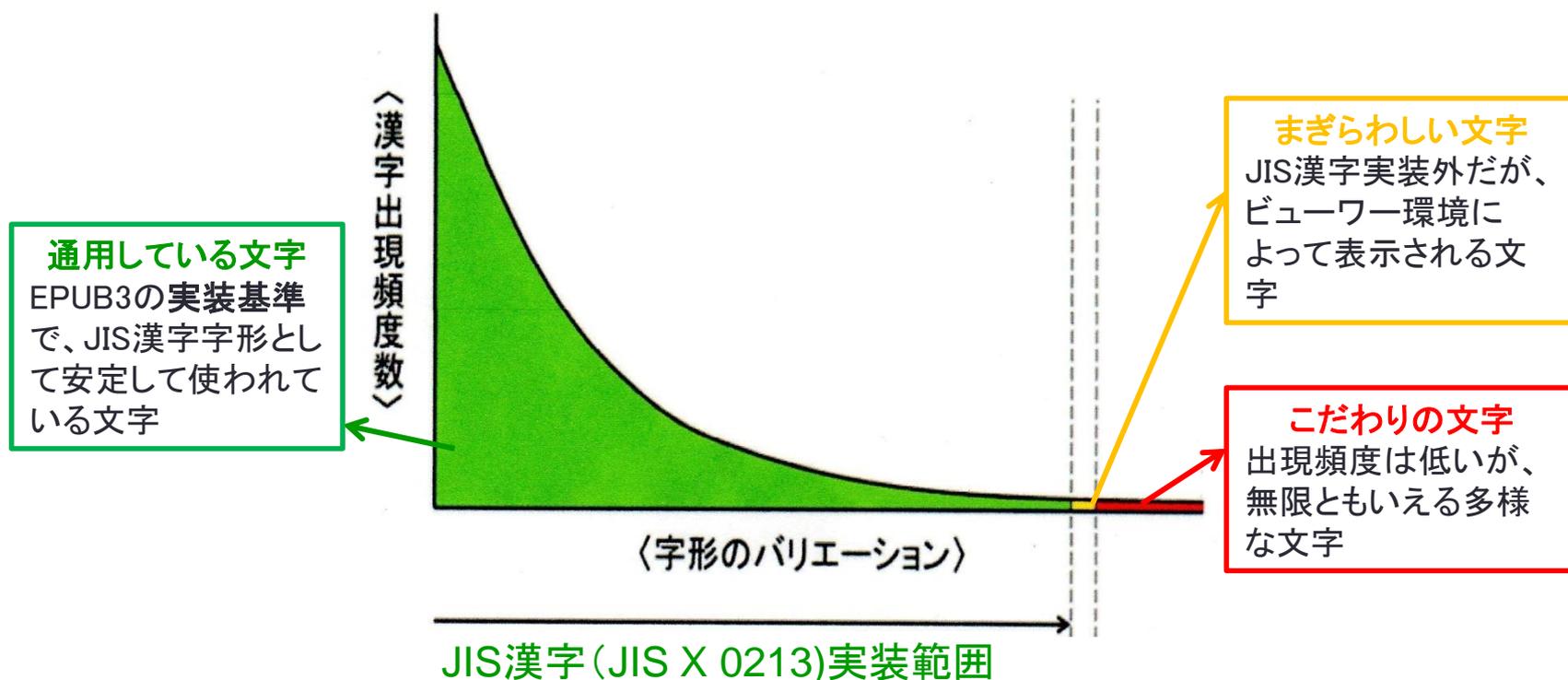
chair@pcc.or.jp

制作協力: ワイズネット



外字・異体字のロングテール

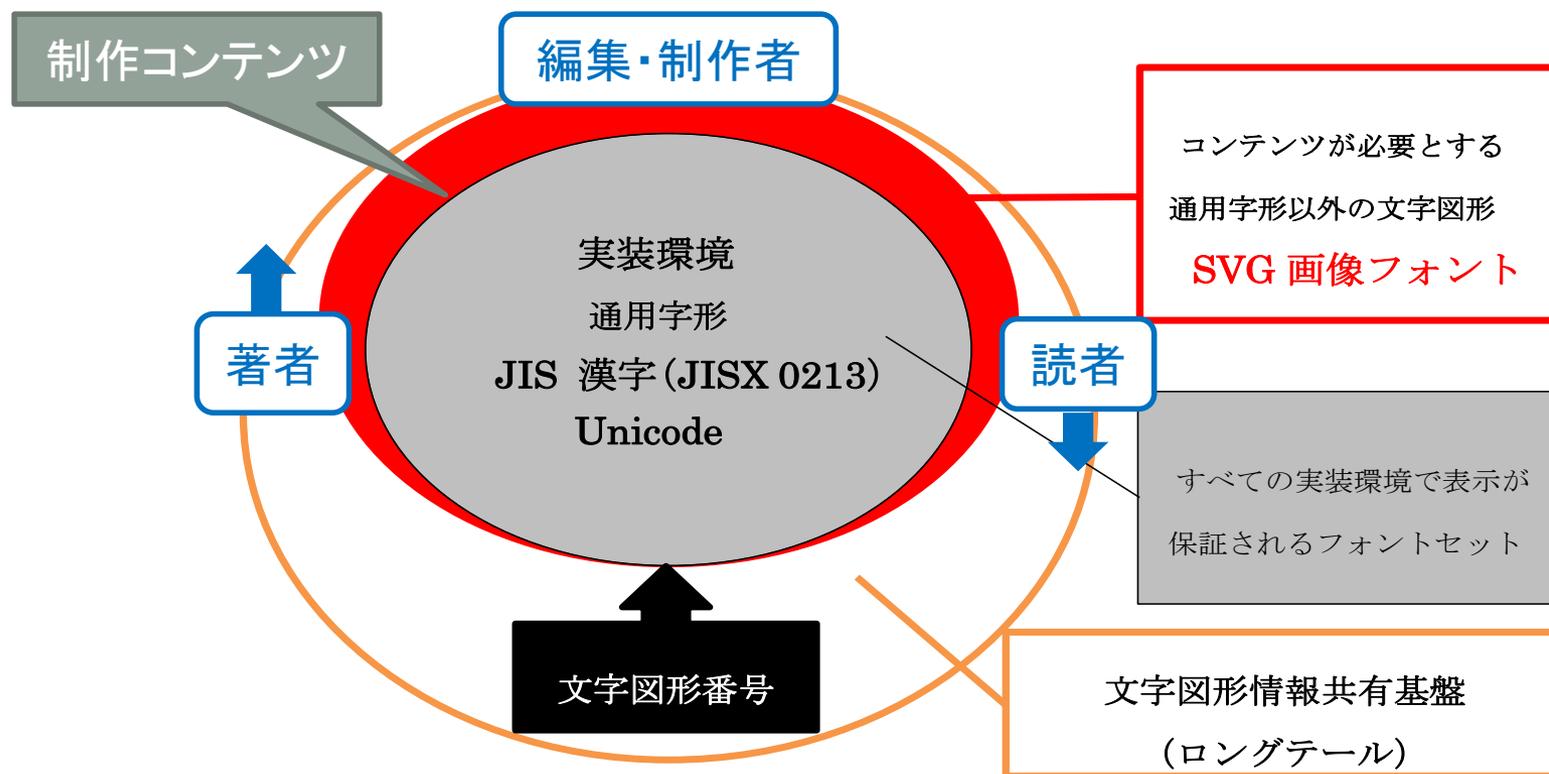
外字:規格(文字コード)／実装(ビューワー)
異体字⇔規格通用字／正字基本字



漢字出現頻度と字形のバリエーションのイメージ

文字図形番号で文字字形の共有化を実現

すべての環境で**通用字形**のフォント実装を保証し、**こだわりの文字**は各コンテンツに必要なものだけを**SVG画像フォント**を埋め込み、共有の文字図形番号で管理する仕組み
著者と読者、編集・制作者が字形情報を簡単に伝達できる



平成漢字基本字形集・和字正濫鈔での表示例(1)

例はすべてKinoppy for Windows(desktop) 游明朝体フォントがベース
SVG画像フォントはEPUB用文字鏡中太明朝体(126,000字)

侵 侵

SVG 画像フォント使用

正體 部首人 畫數09/07

【&M000646;】

つくり索引 侵 [208-01]

音 シン(シム)

訓 おかす(をかす)、みにくい

解説 [人(イ)部] 侵 [聲] 【會意】原字[侵]

侵 略字

【&M054175;】 JIS① <侵 :4FB5>

俣

正體 部首人 畫數09/07

【&M000718;】 JIS① <俣 :4FE3>

つくり索引 俣 [231-08]

訓 また

解説 [人(イ)部][矣]×[吳][呉][吞] 【國字】

註記 簡體字では[吳]は[呉]の略字。

わ	ワ	ら	ラ	や	ヤ	ま	マ	は	ハ	な	ナ	た	タ	さ	サ	か	カ	あ	ア
和		良		也		末		波		奈		太		左		加		安	
𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎	
va		ra		ya		ma		pa		na		ta		ca		ka		a	
ゐ	井	り	リ	い	イ	み	ミ	ひ	ヒ	に	ニ	ち	チ	し	シ	き	キ	い	イ
矣		矣		矣		矣		矣		矣		矣		矣		矣		矣	
𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎	
vi		ri		yi		mi		pi		ni		ti		ci		ki		i	
う	ウ	る	ル	ゆ	ユ	む	ム	ふ	フ	ぬ	ヌ	つ	ツ	す	ス	く	ク	う	ウ
掣		掣		掣		掣		掣		掣		掣		掣		掣		掣	
𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎	
vu		ru		yu		mu		pu		nu		tu		cu		ku		u	
ゑ	エ	れ	レ	い	イ	め	メ	へ	ヘ	ね	ネ	て	テ	せ	セ	け	ケ	い	イ
掣		掣		掣		掣		掣		掣		掣		掣		掣		掣	
𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎	
ve		re		ye		me		pe		ne		te		ce		ke		e	
お	オ	ろ	ロ	よ	ヨ	も	モ	ほ	ホ	の	ノ	と	ト	ろ	ロ	こ	コ	を	ヲ
製		製		製		製		製		製		製		製		製		製	
𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎		𠄎	
vo		ro		yo		mo		po		no		to		co		ko		o	

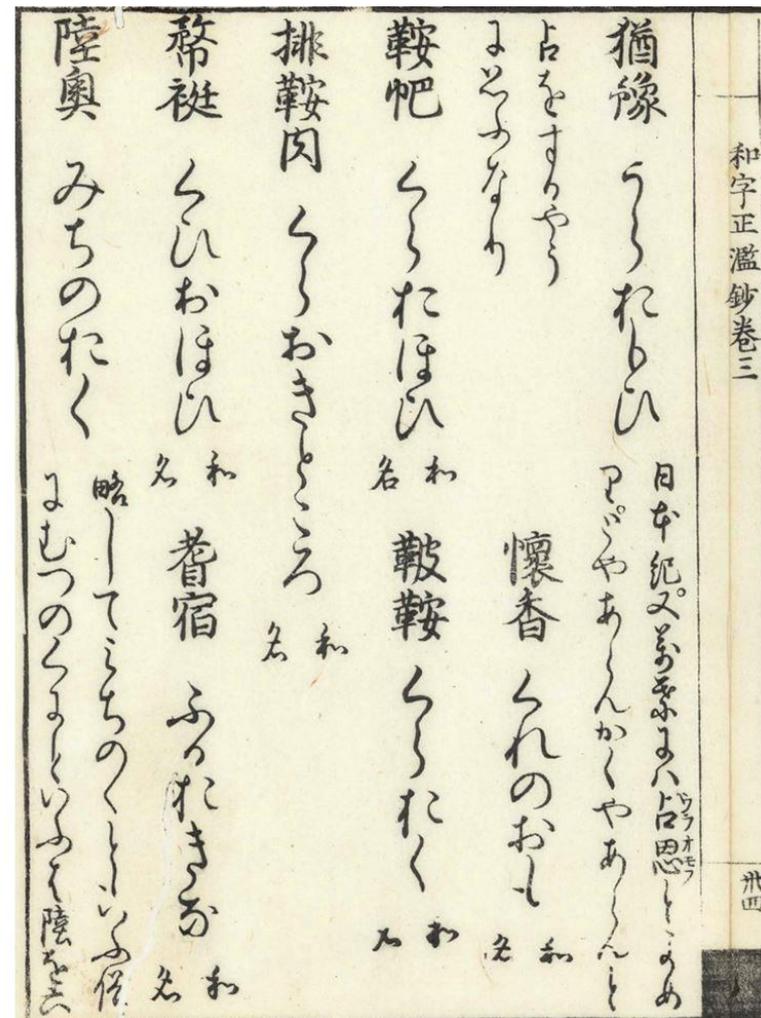
契沖五十音圖

平成漢字基本字形集・和字正濫鈔での表示例(2)

(例字)	(媒体)	(筆記具)
甲骨文	龜の甲羅、腹甲、牛や鹿の骨	青銅の三角鑿
金文	青銅の器 (鑄造用粘土、革)	鐵鑿 (たがね) か
篆書	石、木片、竹片、絹帛	鑿、毛筆
隸書	石、木片、竹片、絹帛、紙	毛筆、鑿
草書	石、紙	毛筆、鑿
行書	紙中心	毛筆
楷書	紙中心	毛筆

複雑な書寫のありやうを大まかに示したのですが、**甲骨文**の昔から、字源ともいふべき形で書かれ、意味が理解される文字作りになつてゐます。例字の〔夾〕は、甲骨文中でわかるやうに、〔人〕が〔手〕に挟まれてゐる様で、夾 (はさ) むといふ意味以外に「助けられる」といふ意味もあるのがこのやうな形から理解されます。

平成漢字基本字形集・和字正濫鈔での表示例(3)



猶豫 うらおもひ 日本紀。又萬葉には 占思ウラオモヒとよめり。とやあらんかく
 やあらんと占ウラをするやうに思ふなり

懷香 くれのおも 和名

鞍吧 *カハ くれのおもひ 和名

鞍鞍 くらおく 和名

排鞍肉 くらおきところ 和名

鞞衽 *カハ くれのおほひ 和名

耆宿 ふるおきな 和名

陸奥 みちのおく 略してミチのくともいふ。俗にむつを六のく
 は陸

画像参照
 MojikyoM_SVG_008808.svg

【入稿データ】

鞍 & M008808 ; *鞍吧 くらおもひ 和名

鞍鞍 くらおく 和名

排鞍肉 くらおきところ 和名

& M0088989 ; & M217859 ; *鞞□ くれのおほひ 和名

SVG画像フォントで出来ること

Scalable Vector Graphics(スケーラブル・ベクター・グラフィックス、SVG)は、XMLベースの2次元ベクターイメージ用の画像形式の1つで、W3Cによって開発されオープン標準として勧告されている。ベクター画像は画像を点の集合体ではなく、線や面などの図形の集合体として扱うため、見る環境に応じて最適な表示が可能となる。SVG画像フォントはTTFフォントをSVG形式に画像変換したもの。

【外字画像とSVG画像】

【太字】

祇 祇 祇 祇 祇

【文字色】

薩 薩 薩 薩 薩

【四角囲み】

鸞 鸞 鸞

【円囲み】

鸞 鸞 鸞

未来志向のSVG画像フォントの課題

SVG画像フォント集合

文字画像参照と未登録文字の継続的な登録体制

* 情報処理学会試行標準IP SJ-TS 002:2014「文字図形識別情報」

<https://www.itscj.ipsj.or.jp/ipsj-ts/ts02-2014/charsp-id.htm#6>. ⇒ISO/IEC10036

⇒6.文字図形集合の一覧(SVG画像フォント)

- a)文字図形番号順
- b)部首・部首内画数順

(文字検索ソフトは「インデックスフォント Version2.0」(フォント商用利用 ライセンス付):紀伊國屋書店発売)

各ビューワー上でのSVG画像加工

文字色、ボールド処理、書体切り替えなどの処理方法の共通化

既存ローカル外字活用のための名寄せテーブル作成

ローカルコード(外字登録エリア) ⇔ 文字図形番号 ⇔ Unicode

EPUB上での表示と検索等の課題

SVGデータ内への検索語埋め込み(←IME)

各種テーブルの整備(異体字 ⇔ 通用字Unicode)

テキスト読み上げでの処理

ご清聴ありがとうございました