

古い公共データあるある

朱成敏







■ 農作物名称:農作物語彙体系(2017)

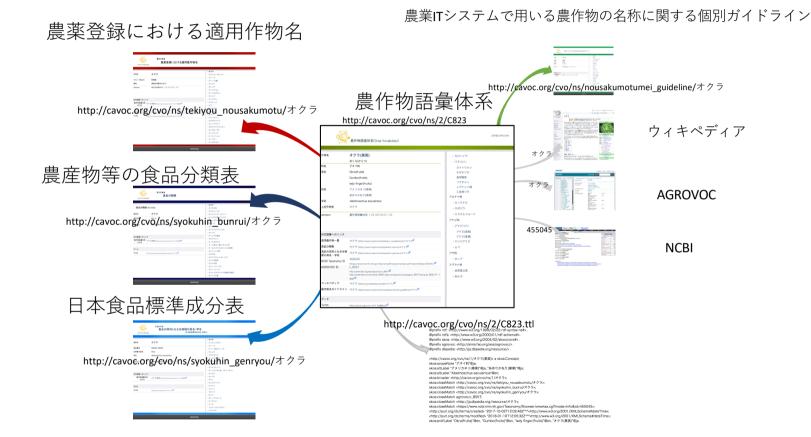
- 1. 植物学的分類に基づいた分類
- 2. 利用部位、栽培方法による定義
- 3. 属性を定義(用途、成熟度など)
- 4. 品種の区分
- 5. 既存語彙、外部情報とのリンク

■ データの提供

URI RDF/Turtle, CSV http://cavoc.org/







- 既存語彙についてもデータ化し、URIを構築。
- 農作物語彙体系を通じて語彙の変換、情報の問い合わせが可能。
- Wikipedia(DBpedia),AGROVOC,NCBIとの連携。



オクラの例

1. 農薬登録における適用作物名

大グループ名	中グループ名	作物名	作物名に含まれる別名, 地 方名,品種名等の例	備考
野菜類	-	オクラ		果実を収穫するもの。
野菜類	-	花オクラ		花を収穫するもの。

2. 農産物等の食品分類表

食品名	食品分類
オクラ	おくら
オクラの花	その他の野菜

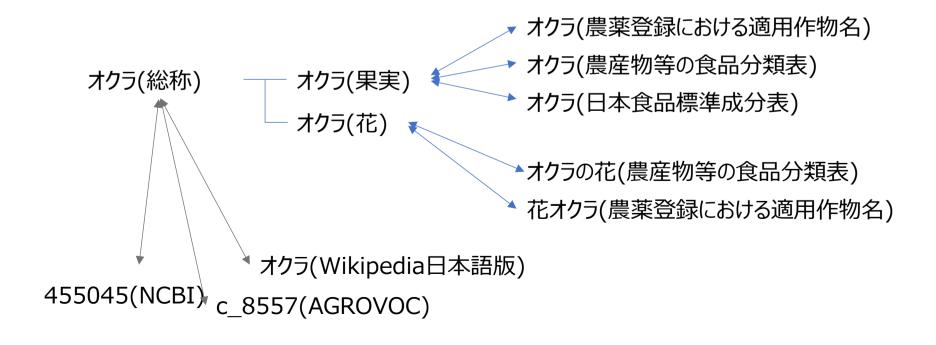
3. 日本食品標準成分表

食品番号	食品名	生物種の英名	学名
06032, 06033	オクラ	Okra	Abelmoschus esculentus

オクラの例

作物名(総称) 作物名(収穫部位1) (いわゆる植物名) 作物名(収穫部位2)

収穫部位、栽培方法によって定義された個別の農作物が総称の下位概念となる。





В	C	D	l E	F	G	Н		J	K	L	M	N	0	l P	Q	R	S	Т	U
	SIP_CVO							適用農作物						厚生労働省			■ 品の原料とな	る生物種の	英名・学名
	農作物名	科名	同意語	よみ	学名	英名	属性	大グループ	中グループ	作物名	作物名に含	カッコ書き		食品分類表	食品分類表	食品番号	食品名	生物種の英	学名
	オカノリ	アオイ科	フユアオイ	おかのり	Malva	Curly		名 野菜類	名	おかのり	まれる別名		部位) 茎葉	食具タ おかのり	食品分類 その他の野		() th/+	- /2	
階層	オクラ	アオイ科	アメリカネ	おくら	Hibiscus	Okra		野菜類		オクラ			果実	オクラ	菜 おくら	06032,	オクラ	Okra	Abelmosc
	オクラ(総	アオイ科	II アメリカネ	おくら(そ	Hibiscus			ST PRIAR		377			未失		40/0	06033	377	OKIA	hus
階層			リ (総称) 花オクラ	おくら(は	Hibiscus									オクラの花	その他の野				
階層	テンサイ	アカザ科	甜菜 サト	てんさい	Beta	Sugar beet		野菜類	-	花オクラ	サトウダイ		花	テンサイ	菜				
			ロ米Iソド ウダイコン トウヂシャ		udanie			-	-	てんさい	コン	採糖目的	根		てんさい その他の野	06261.		Swiss	Beta
		アカザ科	不察費 7	ふだんそう	Beta	Swiss		野菜類	-	ふだんそう	スイス		葉	フダン草	菜	06262	ふだんそう	chard	vulgaris
	フダンソウ	アカザ科	トウヂシャ 不 <u>終</u> 音 マ	ふだんそう	Beta	Swiss		野菜類	-	ふだんそう	チャード		葉	スイス	葉	00040		B 11 .	
	テープルビ		カエンサイ	てーぶるび・	Beta	Table beet		野菜類	-	かえんさい				ピート (カエンサイ)	その他の野 菜	06243, 06244	ピーツ	Red beet [Svn, Table	
	テーブルビ	アカザ科	カエンサイ	てーぶるび-	Beta	Table beet		野菜類	-	かえんさい	レッドピート		根	ピート (カ	その他の野菜	06243, 06244	ピーツ	Red beet [Syn. Table	Beta vulgaris
階層	トンプリ	アカザ科	ホウキギ	とんぶり	Brassica			野菜類	-	ほうきぎ	とんぶり		胞果	ほうきぐさ	その他の野菜				
階層	ホウキギ	アカザ科		ホウキグサ	Brassica	Summer		野菜類	-	ほうきぎ			胞果	ほうきぐさ	その他の野	06188	とんぶり	Summer	Bassia scoparia
階層	ホウキギ	アカザ科	ホウキグサ	ホウキグサ	Brassica	Summer		野菜類	-	ほうきぎ	ほうきぐさ		胞果	ほうきぐさ	その他の野				
	キノア	アカザ科		きのあ	Chenopodi	Quinoa		雑穀類	-	キノア			種子	キノア	その他の穀物	я			
	オカヒジキ	アカザ科	ミルナ オ	おかひじき	Salsola	Salt wort		野菜類	-	おかひじき			茎葉	おかひじき	その他の野		おかひじき	Japanese	Salsola
	オカヒジキ	アカザ科	カミル ミルナ オ	おかひじき	Salsola	Salt wort		野菜類	-	おかひじき	オカミル		茎葉		莱	06031		saltwort	komarovii
	オカヒジキ	アカザ科	カミル ミルナ オ	おかひじき	Salsola	Salt wort		野菜類	_	おかひじき	ミルナ		茎葉						
	ホウレンソ	アカザ科	カミル ほうれん草	ほうれんそ	Spinacia	Spinach		野菜類	_	ほうれんそ			茎葉	ほうれんそ	ほうれんそ		ほうれんそ	Spinach	Spinacia
要検討	ガーデンビ	アカザ科	まかいン		oleracea			野菜類	_	う かえんさい	ガーデン		根	3	3	269.	3		oleracea
要検討	クサヒジキ	アカザ科						野菜類	_	おかひじき	ピート クサヒジキ		茎葉						
	サポジラ	アカテツ科		さぼじら	Manilkara 2	Sapodilla		果樹類		サポジラ	,,,,,		果実	サポジラの:	その他の果	+			
	カニステル	アカテツ科	クダモノタ	かにすてる		Anistel Egg	fruit		_										
		アカテツ科		かにすてる		Anistel Egg		果樹類	-	カニステル			果実	<i>7</i> -717 11	その他の果?	*			
	カニステル	アカテツ科		かにすてる				果樹類	-		エッグフル・		果実						
			ソメモノメ			Anistel Egg		果樹類	-		クダモノタ	∀ ⊐	果実	25445					
		アカテツ科		みらくるふ		Miraculous	berry	果樹類	-	ミラクルフ	ルーツ		果実	ミラクルフ	その他の果	実			
要検討	アケビ	アケビ科		あけび	Akebia quinata	Five leaf													
要検討	アケビ(果			あけび(か	Akebia quinata			果樹類	-	あけび (果	実)		果実	あけび	その他の果	×.			
要検討	アケビ(茎		キノメ		auinata		キノメ;イ	野菜類	-	あけび(茎 葉)			茎葉(若 芽)	あけび(若 非)	その他の野菜				
要検討	アケビ(茎	アケビ科	キノメ	あけび(けい	Akebia		キノメ;イ	野菜類	-	あけび (茎 葉)	キノメ		茎葉(若 芽)	あけび(若 共)	その他の野 菜				
要検討	アケビ類	アケビ科		あけびるい	Akebia guinatal 4	Akebia													
要検討	ミツパアケ	アケビ科		みつばあける	Akebia trifolioto	Three leaf										07001, 07002	あけび (ミ ツパアケ	Akebia	Akebia trifoliata
	ムベ	アケビ科	トキワアケ	むべ		Japanese s	taunton vine	果樹類	-	むべ			果実	むべの実	その他の果		,,,,		S HVIIGU
	ホップ	アサ科	セイヨウカ	ほっぷ	Humulus	Нор		-	-	ホップ			雌花穂	ホップ	ホップ				
	ホップ	アサ科	セイヨウカ	ほっぷ	Humulus	Нор		-	-	ホップ	セイヨウカ		雌花穂						
	ワサビダイ	アプラナ科	ラハナソウ ホースラ	わさびだい。	Armoracia	Horseradi		野菜類	-	わさびだい	ラハナソウ		根	西洋わさび	西洋わさび				
			71000		rueticana	eh				٢4				(±-75					

- 名称の収集・同一性の確認

適用農作物一覧(農薬使用、農林水産省) 食品分類表(残留農薬、厚生労働省) 日本食品標準成分表(栄養、文部科学省) 品種リスト(農研機構)

(最も参照が多い) 適用農作物一覧を中心 に整理を行う



1.「農薬登録における適用作物名(農林水産省)」 農薬の登録申請に適用対象となる作物名の一覧(893語)

PDF ⇒ EXCEL 2 種, CSV, Turtle

2.「農産物等の食品分類表(厚生労働省)」 食品中に残留する農薬に関する情報(1,173語)

PDF ⇒ EXCEL, CSV, Turtle

3.「日本食品標準成分表(文部科学省)」 国民の健康、栄養管理のために管理する食品(625語)

EXCEL ⇒ EXCEL, CSV, Turtle

4. 「農業ITシステムで用いる農作物の名称に関する個別ガイドライン(内閣官房)」 (689語)

EXCEL ⇒ EXCEL, CSV



1.「農薬登録における適用作物名(農林水産省)」 PDF ⇒ EXCEL 2 種, CSV, Turtle

(別表1-1-2)

適用農作物のうち果樹類以外のもの(食品の用に供される農作物(特用作物及び 家畜の飼料の用に供される農作物を含む。): 作物残留試験成績を必要とするもの)

大グループ	中グループ	作物名	作物名に含まれる	備考
名	名		別名、地方名、品	
			種名等の例	
*		稲	水稲(移植、直	種子を収穫する
			播)、陸稲	もの
麦 類		えんばく	オートムギ、エン	
	2		麦、カラスムギ	
		大 麦	二条種、六条種、	
			はだか麦	
		小 麦		
		ライ麦	100	
雑 穀 類	いね科細粒	あわ		
	雑 穀 類	きび		
		ひえ		
	とうもろこ	とうもろこし(子		
	L	実)		
		未成熟とうもろこ	スィートコーン	種子(ある程度
		L		成熟した雌穂)
				を収穫するもの

特徴:

- ①農業における最も重要な語彙
- ② わかりやすい名称

問題点:

- ① 機械可読性ゼロ
- ② 加工のために必要な 体力、眼力、忍耐力
- ③ 老舗の継ぎ足しタレ
- ④ 今後の改善は遠い
- ⑤ 名称の一貫性がない



2.「農産物等の食品分類表(厚生労働省)」PDF ⇒ EXCEL, CSV, Turtle

農産物等の食品分類表

厚生労働省食品安全部 [2013.7版]

注1:いずれも生鮮食品としての分類である。

注2: 医薬品に該当するものは、同じ植物であっても、食品衛生法に規定される残留基準の

対象とはならない。

食品名	食品分類	
あおうり	しろうり	
青トウ<とうがらし>	その他のなす科野菜	
赤毛うり	その他のうり科野菜	
アカザ	その他の野菜	
赤または白フサスグリ	その他のベリー類果実	
アカメガシワ 葉・樹皮	その他の野菜	
アガリクス (カワリハラタケ)	その他のきのこ類	
アカワケギ (アカシャロット)	その他のゆり科野菜	
アキーアップル	その他の果実	
あけび	その他の果実	
あけび(若芽)	その他の野菜	
あさつき<チャイブ>	その他のハーブ	
アサの種子 (アサノミ)	その他のスパイス	
アサフェチダの根	その他のスパイス	
アサフェチダの根茎	その他のスパイス	
あさまこな (朝熊小菜)	その他のあぶらな科野菜	
あざみな (チリメンカラシ、ハゴロモカラシナ)	その他のあぶらな科野菜	
アシタバ (あしたば)	その他のせり科野菜	
味美菜 (あじみな)	その他のあぶらな科野菜	
アジョワンの種子	その他のスパイス	

特徴:

- ① 面白い
- ② 1よりコピペーが楽

問題点:

- ① 機械可読性"ほぼ"ゼロ
- ② 分類項目が曖昧
- ③ 人間には優しい名称
- ④ 謎が多い名称
- ⑤「その他」多すぎ



2. 「農産物等の食品分類表(厚生労働省)」の例

ケンポナシ(玄圃梨)(キグ(枳グ))の実	その他の果実
	<u> </u>
乾燥キマメ(インド名:Toor dal/アラハル豆)	その他の豆類
エシャロット (ベルギーシャロット) (注:日本型エシャロット (ラッキョウを若採りしたもの) は含まない。)	その他のハーブ
アマメシバ (通常の摂取方法に限る。)	その他の野菜
イリッペ・ナッツ (テンカワン) [イリッペ油、イリッペ/ ターの原料]	その他のナッツ類
花チンゲンサイ(チンゲンサイのナバナ的栽培。通常のチン	
ゲンサイ栽培方法と異なり、抽苔した花茎のみを利用する の)	も その他のあぶらな科野菜
野ブドウ (ノブドウ・ザトウエビ) の実 (本来は食べられい)	な その他の果実
· · · · · · · · · · · ·	and the state of t



3. 「日本食品標準成分表(文部科学省)」 PDF ⇒ EXCEL, CSV, Turtle

	A	В	С	D
1	2 食品の原料とな	る生物種の英名・学名		
2				
	本成分表に収載され	ている食品の原料となる生物種	の英名・学名を表18に示した	広く使われている複数の事
		、シノニムとしてSyn.の次に記		
3	(1-41%-0) 2 30 Electr	C D / L D C D C By II. O / C C BL	190 0720	
1		4 4 4 4 - W 5		
,	表18 食品の原料となる	生物種の英名・学名		
,	食品番号	食品名	生物種の英名	学名
	及即惟力	()内は生物種名	主物性の大有	7-41
	1 穀類		CEREALS	
,	01001	アマランサス	Amaranth	Amaranthus spp.
0	01002,01003	あわ	Foxtail millet	Setaria italica
	01004	えんぱく	Common oat	Avena sativa
1	01004	Z.VIZ.	[Syn. Oat]	Avena sativa
2	01005~010	おおむぎ	Barley	Hordeum vulgare
	01011	きび	Proso millet	Panicum miliaceum
3				7 United Manager
4	01012~079, 01144~150	こむぎ	Common wheat	Triticum aestivum
	01063~01064	(デュラムコムギ)	Durum wheat	Triticum turgidum
5		.,,		[Syn. Triticum durum]
6	01080~121, 01151~161	こめ (イネ)	Rice	Oryza sativa
7	01122~130	そば	Buckwheat	Fagopyrum esculentum
8	01131~137, 01162~165	とうもろこし	Corn [Syn. Maize]	Zea mays
9	01138	はとむぎ	Job's tears	Coix lacryma-jobi vat. ma-yuen
0	01139	ひえ	Japanese barnyard millet	Echinochloa esculenta
1	01140,01141	もろこし	Sorghum	Sorghum bicolor
2	01142,01143	ライむぎ	Rye	Secale cereale
3	2 いも及びでん粉類		POTATOES AND STARCHES	
4	02001,02041	きくいも	Jerusalem-artichoke	Helianthus tuberosus

特徴:

- ① 兄弟の中では一番データらしいデータ
- ② エクセル

問題点:

① 謎の改行

02034, 02033	イーコン	racon	Smattaninus sonenijotius
	(やまのいも類)	(Yams)	
02022~025	ながいも	Chinese yam	Dioscorea polystachya
02026	じねんじょ	Japanese yam	Dioscorea japonica
02027	だいじょ	White yam	Dioscorea alata
	(でん粉類)		
02028	キャッサバでん粉 (キャッサバ)	Cassava	Manihot esculenta



4. 「農業ITシステムで用いる農作物の名称に関する個別ガイドライン(内閣官房)」

A	В	С	D	E	F	G
農作物名ガイド	ライン					
	基本表		シソーラス	品標等		属性項目
大分類	中分類	小分類	27-3A	の様々	収穫部位	施工土机 日
*	*	水相		あいちのかおりSBL、養の神め、青系187号、あきげしき、あきさかり、あきたこまち、秋田清こまち、秋田63号、アキツホ、秋の詩、秋のきらめき、あきほなみ、あきまさり、あきろまん、アケボノ、朝の光、朝日、あさひの夢、彩雨月、あやひめ、あわみのり、淡雪こまち、イクヒカリ、いただき、一番星、祝、いわてっこ、うこん館、おいでまい、オオセト、おくほよれ、おてんとそだち、おぼろづき、鎖町、龍の、東庭鎖町、加賀ひかり、かけはし、風さやか、風鳴子、神の穂、神の舞、きたくりん、きたしずく、キヌヒカリ、きぬむすめ、吉優の意、京の輝き、きらら397、きらり宮崎、きらりん、ぎんおとめ、吟ぎんが、吟の夢、吟のさと、吟風、吟吹雪、金紋錦、くまさんの力、蔵の薬、京の輝き、きらら397、きらり宮崎、きらりん、ぎんおとめ、吟ぎんが、吟の夢、吟のさと、吟風、吟吹雪、金紋錦、くまさんの力、蔵り上カリ、五百万石、ゴロビカリ、西海134号、彩のがかやき、彩のきがなま、女の田によみ、彩のみのり、佐春線、こといぶき、技路単生、越淡顔、コシヒカリ、五百万石、ゴロビカリ、西海134号、彩のがかやき、彩のきがり、、たらかり、佐春線、こといぶき、大地の風、大地の屋、たかは蔵、ササニシキ、さとじまん、黒空の神き、さわかおり、さわびがり、しらかば錦、華星、干本館、空青180号、大地の風、大地の屋、たかは戯。たかねみのり、たきたて、玉葉、たんぼの夢、ちほのかおり、ちゅらひかり、井田一本寺、一が名の里、大地の風、大地の風、大地の屋、たかね鈴、たかな赤ろい、つぶみのり、つぶゆき、つぶゆたか、つやおとめ、一や版、露葉風、出羽きらり、出羽腹々、出羽の風、大心のこれへした。大使の詩、てんたかく、天のつぶ、天竜乙女、東北第194号、とがおとめ、土佐錦、とちぎ清14、とちぎの屋、ナドロキワセ、どまんない、宮の香、トヨニシキ、どんとこい、どんびしゃり、中手千本、なごりゆき、なすひかり、なつもが、夏の突み、ナツヒカリ、ななつぼし、雨間そだち、にこまる、日本晴、ねばりゆき、農林22号、熊春20号、北いり、はいほう、夏の突み、ナッヒカリ、なないほに、かいたいき線、ひだほまれ、ひとごこち、ひとめぼれ、ヒノヒカリ、兵庫本館、兵庫を登り、かえの表が、赤っくりた。からにしき、秦山館、まルナークィーン、ミルキーサマー、むつほまれ、柔の君、めんこいな、森のくまきん、山形95号、山形面104号、山田錦、やまの七、秦山館、まれたい木、東山館、まれたいり、中のおはた、からを入着、本の者、参しがよりまった。中のまで、中のまで、からまり、たっといまがりり、からをで、春一秋、からおう、たりまれた、カラがモチ、カバヤモチ、カグヤモチ、風の子もち、ありもち、加野の子も、からがよりまった。までは野り、とみちから、白山もち、きつよりまち、カレドナ・カバ・カー・デ・カがらもち、ドナコモチ、たまひがり、もち美人、モチミノリ、もちむすめ、モリモリモチ、ヤシロモチ、ヤシロモチ、たちのもち、ストピリ・ボースでは、ちゅうには、カラのよりは、大きのは、大きのは、からは、大きのは、からは、大きのは、大きのは、大きのは、大きのは、大きのは、大きのは、大きのは、大きの	種子	普通、直播栽培水 乾田 種屬 表稿、 「 「 「 「 「 「 「 「 「 「 「 「 「 「 「 「 「 「 「
4		陸框			種子	301
-		その他				2
麦類	麦類	えん麦	オートムギ、エン麦、カラスムギ		種子	

特徴:

① 面白い

問題点:

- ① 他の語彙を意識しているためこだわりがない
- ② エクセル形式で公開している意味がない
- ③ 伝家の宝刀「その他 |

EXCEL ⇒ EXCEL, CSV Turtleの提供が不可能

感想

PDFをCSVにしてみると、、

- ① コピペーが綺麗にできない
- ② 意味不明の階層構造が多い
- ③ 濁音が分離される
- ④ 文字コードが色々
- ⑤ txtにして簡単なスクリプトでクリーニング

PDFをxlsx、csvに綺麗に再現してみたら、、

- ① 担当部署の人に感謝される
- ② 文字化けの苦情が多い、、、、

PDFをturtleにしてみたら、、

- ① 脚注など、どうしたらいいかわからない情報が結構ある
- ② skos:scopeNoteは本当に便利
- ③ 公開されたらこっちもちょっと恥ずかしい

感想

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <a href="http://www.w3.org/2000/01/rdf-schema">http://www.w3.org/2000/01/rdf-schema</a>>.
@prefix skos: <a href="http://www.w3.org/2004/02/skos/core#">http://www.w3.org/2004/02/skos/core#>.
@prefix agrovoc: <a href="http://aims.fao.org/aos/agrovoc/">http://aims.fao.org/aos/agrovoc/</a>.
<a href="http://cavoc.org/cvo/ns/2/C11">http://cavoc.org/cvo/ns/2/C11</a> a skos:Concept;
skos:scopeNote "アオイ科"@ja;
skos:altLabel ""@ja, "アメリカネリ"@ja, "陸蓮根"@ja,
                                                                                            オクラ(花), オクラ(果実)
skos:altLabel "Abelmoschus esculentus"@en;
skos:narrower <a href="http://cavoc.org/cvo/ns/2/C823">http://cavoc.org/cvo/ns/2/C12</a>;
skos:closeMatch <http://cavoc.org/cvo/ns/syokuhin bunrui/オクラ>:
skos:closeMatch <a href="http://cavoc.org/cvo/ns/syokuhin">http://cavoc.org/cvo/ns/syokuhin</a> genryou/オクラ>:
skos:closeMatch agrovoc:c_8557;
                                                                                                       他の資源へのリンク
skos:closeMatch <a href="http://ja.dbpedia.org/page/オクラ">http://ja.dbpedia.org/page/オクラ>;
skos:closeMatch <a href="https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=455045">https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=455045</a>;
<a href="http://purl.org/dc/terms/created">http://purl.org/dc/terms/created</a> "2017-10-
02T12:02:48Z"^^<a href="http://www.w3.org/2001/XMLSchema#dateTime">http://www.w3.org/2001/XMLSchema#dateTime</a>;
<a href="http://purl.org/dc/terms/modified">http://purl.org/dc/terms/modified</a> "2020-04-
```

08T16:07:59Z"^^http://www.w3.org/2001/XMLSchema#dateTime;

skos:prefLabel "Okra"@en,"오쿠라"@ko,"黃秋葵"@zh,"オクラ"@ja.



、、、を思いながら整理しました。



1 括弧を全角から半角に変換。

http://cavoc.org ではこれらの語彙を:

- 1. 公開されているPDFファイルをの構造をそのまま再現したもの、フラットな構造に変換したものの両方を提供
- 2. 半角表記での独立の濁点、半濁点を濁音、半濁音に処理
- 3. 全角記号を半角に変換
- 4. UTF-8に変換
- 5. URIの提供

http://cavoc.orgにて公開中!



W略的イノベーション創造プログラム Cross-ministerial Strategic Innovation Promotion Program

本研究(の一部)は、内閣府 戦略的イノベーション創造プログラム(SIP) 「次世代農林 水産業創造技術」(管理法人:農研機構 生物系特定産業技術 研究支援センター)によって実施されました。