

## ウェブ施設名辞書ファイル仕様

ウェブ施設名辞書は、ウィキペディア日本語版から施設の名称、よみ、住所、カテゴリ、緯度、経度を取得し、整形して収録した辞書である。施設の情報収集ならびに整形は自動で行われているため、辞書に記載されている情報には誤りが含まれる可能性があることに注意していただきたい。

### 1. ファイル形式

- ・ ファイル名 : [A] ウェブ施設名辞書.txt  
[B] ウェブ施設名辞書\_utf8.txt
- ・ ファイル形式 : CSV (カンマ区切りテキストファイル)
- ・ データ項目間の区切 : “,” (カンマ)
- ・ レコード間の区切 : CR+LF
- ・ 文字コード : [A] UTF-16LE エンコード(BOM あり)  
[B] UTF-8 エンコード(BOM なし)

※ ファイル[A][B]は文字コードが異なるだけで、内容は同一です。

### 2. レコード構成

各レコードは、以下の①～⑫の12のデータフィールドを持ち、各レコードの先頭から順に格納されている。

#### ① 施設名称

- 定義 : 施設の名称を示す。
- 文字種 : 全角の漢字、かな、カナ、英字、数字、記号、空白  
および“(、)” (半角の括弧)。
- 例 : 東京国立近代美術館

#### ② 施設名称よみ

- 定義 : 施設名称の読み仮名を示す。
- 文字種 : 全角のかな、記号、空白。
- 例 : とうきょうこくりつきんだいびじゅつかん

#### ③ 住所

- 定義 : 施設の住所表記を示す。  
“丁目”、“番”、“号”は用いず、“-”(全角ハイフン)で区切る。
- 文字種 : 全角の漢字、かな、カナ、英字、数字、記号、空白  
および“ ” (半角空白)。
- 例 : 東京都千代田区北の丸公園 3 - 1

#### ④ カテゴリ

- 定義 : 施設の種類を示す。

文字種 : 全角の漢字、かな、カナ、記号、空白  
および“/”(半角スラッシュ)。

例 : スポーツ施設/野球場

⑤ 施設名異称 1

⑥ 施設名異称 2

⑦ 施設名異称 3

⑧ 施設名異称 4

⑨ 施設名異称 5

定義 : 施設名の異称・略称などを示す。  
ウェブ施設名辞書では常に null 値である。

値 : null 値

⑩ 緯度

定義 : 施設所在地の緯度(北緯)を世界測地系で示す。

文字種 : 半角(1byte)数字、“.”(小数点)。  
整数部 2 桁+小数点+小数点以下 6 桁。  
小数点以下の末尾桁の“0”は省略する。  
緯度が不明のときは null 値とする。

例 : 35.690545

⑪ 経度

定義 : 施設所在地の経度(東経)を世界測地系で示す。

文字種 : 半角(1byte)数字、“.”(小数点)。  
整数部 3 桁+小数点+小数点以下 6 桁。  
小数点以下の末尾桁の“0”は省略する。  
経度が不明のときは null 値とする。

例 : 139.754642

⑫ 緯度・経度のソース

定義 : 緯度・経度の情報源を示す。

値 : 情報源に応じて 0-3 のいずれかの値をとる。  
3 (ウィキペディア)。  
2 (国土交通省の街区レベル位置参照情報)。  
1 (国土交通省の大字・町丁目レベル位置参照情報)。  
0 (緯度・経度が不明のとき)。

### 3. 収録内容

- ・ レコード件数 : 32419 件
- ・ 収録方法

- 施設の情報は2012年10月27日のウィキペディアデータベース(日本語版)<sup>1</sup>から取得した。
- ウィキペディアデータベース内にある施設のページから名称などの情報を抽出した。施設のページであるかの判定はおよそ以下のように行った。
  - ✧ ウィキペディアにおいて所在地の記載のあるページに付与されているカテゴリのうち、「XのY」(「日本の小学校」「神奈川県の商業施設」など)というパターンにマッチするもので、Yが施設としてふさわしいもののリストを人手で作成した。
  - ✧ ページが「XのY」というカテゴリにあり、かつYが前述の施設のリストにマッチするとき、施設のページであるとみなした。
  - ✧ ただし、「かつて存在した日本の百貨店」「2003年廃止の鉄道駅」など、施設としてふさわしくないカテゴリ名のパターンを人手で作成し、これらのパターンにマッチするカテゴリのページは対象外とした。
- ①施設名称では、施設がいくつかの種類に分かれているとき、半角の括弧で注釈をつけた。
 

例. 京王百貨店(新宿店)、京王百貨店(聖蹟桜ヶ丘店)
- ③住所は、可能な限り日本郵政株式会社が提供する郵便番号データ<sup>2</sup>での表記にあわせた。また、住所にビル名などが記載されている場合は、住所の番地の後に半角スペースで区切って記した。
 

例 1. 愛知県小牧市小牧3-5 5 5 ラピオ4階

例 2. 埼玉県さいたま市大宮区高鼻町4 大宮公園内
- ④カテゴリは、ウィキペディアのページに記載されているカテゴリであり、前述の「XのY」というカテゴリ名におけるYを掲載した。ひとつのページに複数のカテゴリがある場合、“/”で区切って記した。
- ⑩緯度、⑪経度は、ウィキペディアならびに国土交通省の「街区レベル位置参照情報」「大字・町丁目レベル位置参照情報」<sup>3</sup>を基に、以下の順序で付与した。
  - ✧ ウィキペディアのページに緯度・経度が記載されている場合、それを付与した。この場合、⑫緯度・経度のソースには値“3”を入れた。
  - ✧ 住所が国土交通省の街区レベル位置参照情報に存在する場合、その緯度、経度を付与した。この場合、⑫緯度・経度のソースには値“2”を入れた。街区レベル位置参照情報は、おおまかにいえば「A県B市C町D丁目E番F号」のうちEのレベルの緯度・経度であるため、緯度・経度が指し

<sup>1</sup> <http://ja.wikipedia.org/wiki/Wikipedia:データベースダウンロード>

<sup>2</sup> <http://www.post.japanpost.jp/zipcode/dl/kogaki-zip.html> より2012年9月16日に取得。

<sup>3</sup> <http://nlftp.mlit.go.jp/isj/>

示す地点が施設名の位置と正確に一致しない点に注意を要する。

- ◇ 住所が国土交通省の大字・町丁目レベル位置参照情報に存在する場合、その緯度・経度を付与した。この場合、⑫緯度・経度のソースには値“1”を入れた。大字・町丁目レベル位置参照情報は、おおまかにいえば「A 県 B 市 C 町 D 丁目 E 番 F 号」のうち D のレベルの住所代表点の緯度・経度であるため、緯度・経度が指し示す地点が施設名の位置と正確に一致しない点、またそのずれは街区レベル位置参照情報を用いた場合よりも大きい点に注意を要する。
- ◇ 上記のいずれの処理でも緯度・経度を付与できなかった場合は、緯度、経度の値はともに null 値とした。この場合、⑫緯度・経度のソースには値“0”を入れた。

・ 緯度・経度のソースの内訳

⑫緯度・経度のソース	施設名数
3 (ウィキペディア)	24859
2 (国土交通省の街区レベル位置参照情報)	5898
1 (国土交通省の大字・町丁目レベル位置参照情報)	1193
0 (緯度・経度なし)	469

・ カテゴリの内訳

末尾の付録を参照。

付録 カテゴリー一覧

- ・ 本辞書における施設のカテゴリおよびその出現頻度(上位 99 件)を以下に示す。

鉄道駅	9199	ホテル	240	市役所	62
公立小学校	3969	ユースホステル	209	仏教系中学校	59
公立高等学校	3783	女子中学校	203	人物記念館	58
公立中学校	3385	通信制高等学校	197	企業博物館	53
学校	2592	観光地	191	コンベンションセンター	52
神社	1514	理数に関する学科設置高等学校	172	文学館	51
私立高等学校	1388	野球場	165	水族館	51
医療機関	1330	バスケットボール競技施設	165	公園	50
寺	1193	サッカー競技施設	141	国立高等専門学校	50
小学校	958	バレーボール競技施設	140	国宝(建造物)	48
博物館	780	ボクシング会場	126	政治	44
ショッピングセンター	773	美術館	120	ラグビー競技施設	44
スポーツ施設	749	映画館	115	都道府県議会	44
警察署	733	カトリック系高等学校	115	郷土博物館	42
私立中学校	721	プロレス会場	115	植物園	40
寺院	715	信号場	113	コミュニティ放送局	39
高等学校	672	私立小学校	112	臨時乗降場	35
商業に関する学科設置高等学校	627	消防本部	111	競技場	34
定時制高等学校	621	仏教系高等学校	107	分譲マンション	32
重要文化財	602	劇場	107	信用金庫	32
工業に関する学科設置高等学校	526	コンサートホール	106	百貨店	31
中学校	524	福祉に関する学科設置高等学校	105	都道府県庁	30
商業施設	468	看護に関する学科設置高等学校	103	競馬場	28
ホール	397	カトリック系中学校	101	名勝	28
超高層ビル	339	科学館	99	超高層ホテル	27
総合学科設置高等学校	334	陸上競技場	96	灯台	26
女子高等学校	321	国際関係に関する学科設置高等学校	87	発電所	25
公立特別支援学校	313	プロテスタント系高等学校	82	放送送信所	24
図書館	302	仮乗降場	75	テーマパーク	24
ダム	297	歴史博物館	75	塔	23
農業に関する学科設置高等学校	287	専修学校	73	新宗教系高等学校	22
家庭に関する学科設置高等学校	261	動物園	64	アウトレットモール	22
屋内競技施設	251	研究所	64	文化施設	20