

翻訳データ

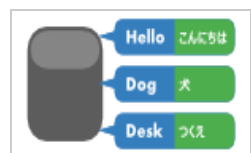
翻訳データについて

翻訳データとは自動翻訳や各ツールで利用する、テキストデータの集まりのことです。

種別

翻訳データには、以下の種別が存在します。

用語



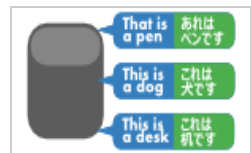
用語集です。

翻訳時に訳語を指定する場合などに利用します。

(例)

hello - こんにちは

対訳



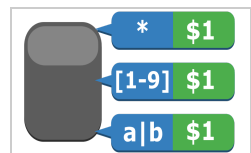
対訳集です。

翻訳時に原文に対応した訳文で翻訳したり、各ツールで例文として利用したりします。

(例)

This is a pen. - これはペンです。

正規表現



正規表現集です。

翻訳時に正規表現を使い、文字を変換させる場合などに利用します。

(例)

* - \$1

データ構造の親と子について

データ構造

翻訳データは必ず「親子関係のデータ構造」を持っています。

(例)用語集

用語集 - 用語

翻訳データを登録する場合、まずは親データを作成した後に、その親データに対して子データを登録していきます。

(例)

「用語集 A」を登録後、「用語集 A」に対して「hello - こんにちは」を登録する。

翻訳データの一覧画面について

翻訳データ一覧画面の構成について説明します。

名前	登録数	言語		対訳提供	操作	状態
		AnyLang	↔			
 用語集 A (1)	0 (2) 日本語	→	英語 (3)	許可 (4)	   (5)	■ (6)

項目	説明
(1) 名前	「翻訳データ」の名前を表示します。
(2) 登録数	登録されている「子 翻訳データ」の数を表示します。
(3) 言語方向	「翻訳データ」の言語方向を表示します。
(4) 対訳提供	「翻訳データ」の対訳提供を表示します。
(5) 操作	「翻訳データ」を操作するボタンです 編集・ダウンロード・削除を行えます
(6) 状態	ファイルを使用した登録の処理状態を表示します。

親 翻訳データの登録

はじめに

はじめに親データを登録し、その親データに対して子データを登録していきます。

※以下、種別「用語」で説明します。

親データ登録

親データを登録します。

※すでに親データを登録している場合は省略可能です。

(1)メニューから[翻訳データ > 用語集]を選択します。



(2)[新規登録]ボタンを押します。



(3)フォームの各項目を入力します。

用語集登録

 ヘルプ

名前

言語方向

日本語 ➡ 英語 ▼



対訳提供



不許可



許可

「許可」にすると、情報通信研究機構（以下、当機構という。）が当データを自動翻訳技術の研究開発の目的に使用することを許諾します。「不許可」にすると、当機構は当データを使用しません。

OK

Cancel

項目

説明

名前

用語集の名前

言語方向

用語集の言語方向

対訳提供

「許可」にすると、情報通信研究機構（以下、当機構という。）が当データを自動翻訳技術の研究開発の目的に使用することを許諾します。「不許可」にすると、当機構は当データを使用しません。

(4) 全て入力し終わったら[OK]ボタンを押します。

登録が成功すると「処理が完了しました。」と表示され、[戻る]ボタンを押すと画面に登録した用語集が表示されます。

親 翻訳データの編集

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ>用語集]を選択します。



(2)[編集]ボタンを押します。

	↔ AnyLang	対訳提供	操作	状態
➡	フランス語	許可	  	—
➡	ベトナム語	許可	  	—
➡	ポルトガル語	許可	  	—
➡	ミャンマー語	許可	  	—
➡	中国語(簡体字)	許可	  	—

(3)以降は登録時と同じ手順となります。

親 翻訳データの削除

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ>用語集]を選択します。



(2)[削除]ボタンを押します。

AnyLang	対訳提供	操作
フランス語	許可	  
ベトナム語	許可	  
ポルトガル語	許可	  
ミャンマー語	許可	  
中国語(簡体字)	許可	  
中国語(繁体字)	許可	  

(3)削除する項目を選択後、[一括消去]ボタンを押します。

 新規登録  一括削除		
 1 2 3		
名前	登録数	言語
英語 - フランス語	0	英語
英語 - ベトナム語	0	英語
英語 - ポルトガル語	0	英語
英語 - ミャンマー語	0	英語

(4)確認ダイアログが表示されますので、[OK]ボタンを押します。

削除しますか？

OK

キャンセル

削除が成功すると、成功画面が表示されます。

⚠ 注意

子翻訳データも削除されます。

一度削除すると、元に戻すことはできません。

子 翻訳データの登録

はじめに

親データに対して子データを登録していきます。

※以下、種別「用語」で説明します。

子データ登録

子データの登録方法は、一つずつ登録する方法と、ファイルで一括して登録する方法があります。

一つずつ登録

(1)メニューから[翻訳データ > 用語集]を選択します。



(2)用語を登録する用語集を押します。

用語集

対訳集

+

新規登録

名前

用語集 A

(3)入力フォームの「原テキスト」と「訳テキスト」を入力します。

用語集

対訳集

ヘルプ

用語集 A (日本語 → 英語)

登録数 0件

+

ファイルから登録

Q

原テキスト	訳テキスト	操作
		+ 登録

(4)全て入力し終わったら[登録]ボタンを押します。

登録が成功すると、画面に登録した用語が表示されます。

ファイル一括登録

(1)メニューから[翻訳データ > 用語集]を選択します。



(2)用語を登録する用語集を押します。

用語集

対訳集

+

新規登録

名前

📁

用語集 A

(3)[ファイルから登録]ボタンを押します。

用語集

対訳集

ヘルプ

用語集 A (日本語 ➡ 英語) 登録数 0件

+

ファイルから登録

Q

原テキスト	訳テキスト	操作
		+ 登録

(4)ファイル入力フォームが表示されますので、ファイルを選択します。

用語集

対訳集

ヘルプ

用語集 A (日本語 ➡ 英語) 登録数 0件

+

ファイルから登録

ファイル1

Browse...

No file selected.

0%

Up to 1000 MB

※対応ファイル拡張子

tsv

tmx

※文字コードはUTF-8のみ

登録が成功すると、「処理中」状態になります。

※登録には時間が掛かる場合があります。

子 翻訳データの編集

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ>用語集]を選択します



(2)対象用語が登録されている用語集を押します。



(3)対象用語の「原テキスト」か「訳テキスト」を押します。(押すと表示が入力フォームに変わります。)

原テキスト	訳テキスト	操作
		+ 登録
ペン	pen	✕ 削除

(4)「原テキスト」、「訳テキスト」を編集します。

原テキスト	訳テキスト	操作
		+ 登録
ペン	pen	更新

(5)全て入力し終わったら[更新]ボタンを押します。

成功すると、表示が編集した用語に変わります。

子 翻訳データの削除

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ > 用語集]を選択します。



(2)対象用語が登録されている用語集を押します。



(3)対象用語の[削除]ボタンを押します。

原テキスト	訳テキスト	操作
		<button>+ 登録</button>
これは、ペンです。	This is a pen.	<button>✕ 削除</button>

成功すると、画面が削除されます。

翻訳データの検索

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ > 用語集]を選択します。



(2)検索対象の用語集を押します。



(3)用語集名の下の検索フォームに、検索キーワードを入力して[検索]ボタンを押します。



翻訳データのダウンロード

※以下、種別「用語」で説明します。

(1)メニューから[翻訳データ>用語集]を選択します。



(2)[ダウンロード]ボタンを押し、tsv, tmx, csvから形式を選択します。

ミャンマー語	許可			
中国語(簡体字)	許可			
中国語(繁体字)	許可			
日本語	許可			
韓国語	許可			
中国語(簡体字)	許可			
中国語(繁体字)	許可			

成功すると、zip形式で圧縮されたファイルのダウンロードが開始されます。

正規表現集のデータ形式について

正規表現集のデータ形式について説明します。

「正規表現パターン」には、対象文字列の置換場所を指定するための、正規表現を入力します。

「置換パターン」には、「正規表現」で指定した置換場所の置換文字列を入力します。

正規表現パターン

(1)基本的な正規表現

正規表現	説明
.	改行を除くすべての文字
[]	角括弧に含まれるいずれか 1 文字
*	直前の文字を0回以上繰り返す
+	直前の文字を1回以上繰り返す
?	直前の文字を0回または1回繰り返す
{n}	直前の文字をn回繰り返す
{n,}	直前の文字をn回以上繰り返す
{n,m}	直前の文字をn回からm回まで繰り返す
	いずれかの条件 (OR条件) として使用
()	文字を 1 つにグループ化
-	文字の範囲の指定
^	角括弧の中の否定

(2)定義済みの正規表現

正規表現	説明
\t	タブ
\n	改行
\r	復帰
\d	すべての数字([0-9]と同意)
\D	すべての数字以外([^0-9]と同意)
\s	垂直タブ以外のすべての空白文字([\t\r\n]と同意)
\S	すべての非空白文字([^\t\r\n]と同意)
\w	すべての半角英数字とアンダースコア([a-zA-Z_0-9]と同意)
\W	すべての半角英数字とアンダースコア以外([^a-zA-Z_0-9]と同意)

(3)特定の位置関係の正規表現

正規表現	説明
^	直後の文字が行の先頭にある場合
\$	直前の文字が行の 末尾 にある場合

置換パターン

(1)置換文字記号

置換文字記号	説明
\$1 ~ \$9	一致した文字列の 1 ~ 9 番目に対応する文字列に置換