ChaMame ユーザーズマニュアル

2018.3.9 Version 1.0

概要

平文形式のテキストファイルを、MeCab形式やCaboCha形式に変換するGUIツールです。句読点等による改行処理や、コマンドラインからMeCabやCaboChaを呼び出す操作を自動的に行います。文字コードは自動判別します。

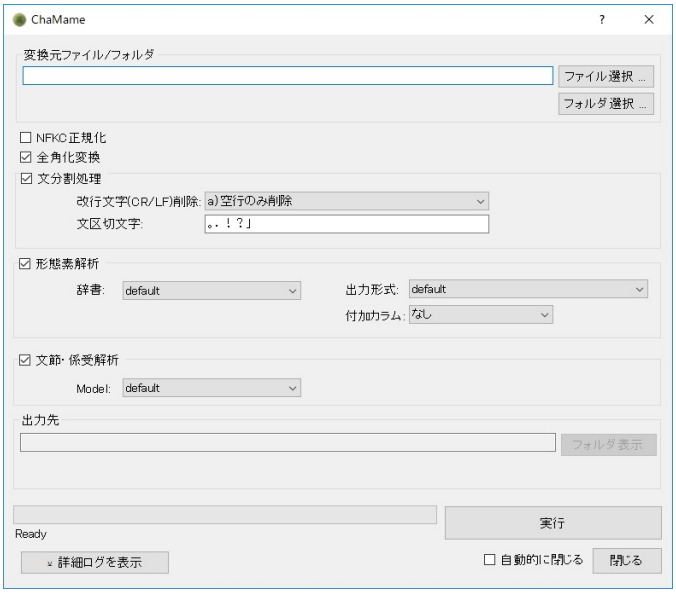
本ツールが出力する.mecabや.cabochaファイルの文字コードは、入力ファイルやMeCab辞書等の文字コードに関わらず、常にUTF-8となります。

インストール

インストーラファイルChaMameSetup.msiをエクスプローラー等から実行してください。インストーラは64/32bit共通です。

使い方

スタートメニューまたはデスクトップから、ChaMameアイコン をダブルクリックして実行します。



入力を設定する

- 単一のテキストファイルを指定する場合：

  ファイル選択 ボタンを押して、ダイアログからテキストファイルを選択するか、または形態素解析したいテキストファイルをChaMameウィンドウ上にドラッグ＆ドロップします。

- あるフォルダ内のすべてのテキストファイルを指定する場合

  フォルダ選択 ボタンを押して、フォルダ選択ダイアログからフォルダを選択するか、、またはそのフォルダをChaMameウィンドウ上にドラッグ＆ドロップします。

前処理オプションを指定する

- NFKC正規化

  NFKC正規化 チェックボックスをONにすると、UnicodeのNFKC正規化(Normalization Form Compatibility Composition)を行います。特定文字の分解・再合成により、表記の揺れをできるだけ少なくするためのオプションです。※ 変換の実装はWindows APIのNormalizeStringを呼び出しており、その動作仕様に従います。

- 全角化変換

  全角化変換 チェックボックスをONにすると、文書中のすべての半角文字（アルファベットや半角カナ）を全角に変換します。本機能はデフォルトでONになっています。
  ※ この機能は英語OSでは利用できません。

- 文分割処理

形態素解析処理は、1文が1行を構成していることが前提条件であるため、元ファイルがそうなっていない場合、**文分割処理** チェックボックスをONにしたうえで、どのような文分割・整形処理を行うかを指定します。

改行文字削除 の指定では、不要な改行文字を削除する方式を指定できます。

- 空行のみ削除：改行文字が2つ以上連続で現れた場合、それを1つの改行に置き換えます。
- 単一改行を空白に置き換え：連続しない改行文字を空白に置き換えます。文の途中に改行があるとき、行をスペースをはさんで連結する場合に使用します（英文など）。
- 単一改行を削除：連続しない改行文字を削除します。文の途中に改行があるとき、行をスペースをはさまずに連結する場合に使用します（和文など）。
- すべて削除：すべての改行文字(CR, LF, CR+LF)を削除します。

文区切文字 欄には、文区切文字となる文字をスペースなしで並べて指定します。文区切文字とは、その出現の直後に改行を挿入するような文字の集合です。デフォルト値は、`。. ！？」`です。

これらの文分割処理は基本的なものであり、これらの設定だけで確実に1文1行になる整形が常にできるわけではありません。

形態素解析(Mecab)オプションを指定する

辞書

Mecabのインストールフォルダ下の`dic`フォルダに存在する辞書サブフォルダ(`dicrc`ファイルを含むフォルダ)がリストされ、Mecabの`-d`オプションでどの辞書を使用するかを選択できます。

これ以外に、下記の選択肢があります。

- `default`：`-d`オプションを指定しません。
- `UniDic`：UniDicインストールフォルダ(通常は`C:/Program Files (x86)/unidic/`)の下にあるUniDic-Mecab辞書を使用して解析します。このオプションは互換性のために残されています。

出力形式

上で指定した辞書フォルダにあるdicrcに定義されている出力形式がリストされますので、必要な形式を選択します。この形式はMecabの`-O`オプションに相当します。

`default`を選ぶと、`-O`オプションを付加しません。

付加カラム

分類語彙表に従って、形態素解析された単語それぞれが分類される語彙IDやラベルをMecabの第3カラム(Tab区切りで)に出力します。分類語彙表データは、ChaMameインストールフォルダの下の`BunruiNo_LemmaID.txt`にあります。

このオプションを利用する（付加カラムなし以外を指定する）ためには、Mecab出力の最終カラム(CSV区切りで)がLemmaIDとなっている必要があります。(`unidic22`出力形式など)
また、次の文節・係り受け解析をONにした場合は、付加カラムはcabochaの出力に付加されます。

文節・係り受け解析(Cabocha)オプションを指定する

文節・係り受け解析チェックボックスをONにすると、Mecab出力をそのままCabocha解析器に渡します。この場合は出力は`.mecab`と`.cabocha`の両方となります。

別途Cabochaがインストールされていることが必要です。

処理を実行する

以上の設定を行ったうえで、**実行** ボタンをクリックすると処理が開始されます。

解析結果は、**出力先** 欄に示されるフォルダ（変換元パスに`/out`を付加したフォルダ。なければ生成します）に出力されます。**フォルダ表示** ボタンでフォルダの内容がExplorerによって表示されます。出力先の`out`フォルダは設定によって変更することはできません。

自動的に閉じる チェックボックスをONにしていると、処理終了後にアプリケーションを自動的に閉じます。

その他

起動オプション

ファイル名を指定して起動すると、そのファイルを入力として画面が開きます。それ以外の起動オプションやGUIなしで使用する手段は提供していません。

GUI表示言語について

基本的に、OSの言語（ロケール）に従って、GUI画面は日本語表示または英語表示が自動的に切り替わります。ただし、ChaKi.NETをインストールしている場合に限り、ChaKi.NETにおける言語設定を優先して使用するようになっています。

詳細ログを表示する

詳細ログを表示 ボタンをクリックするとウィンドウが下方向に広がり、デバッグ用途のログを見ることができます。ログの内容は、不具合がある場合などに、作成者サイドで原因を特定する助けになることがあります。