

MecabEditorについて

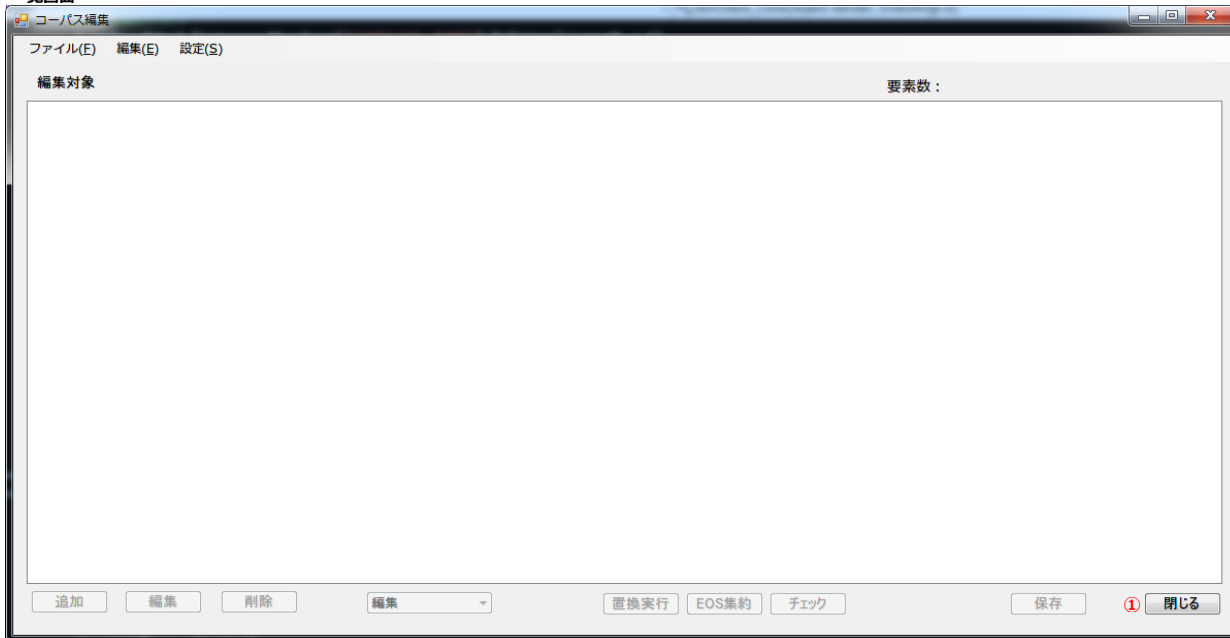
■動作開発環境

Windows7 SP1 64bit
Microsoft Visual C# 2012
Microsoft .Net Framework 4.5
MySQL 5.5
MySQL.Data 6.9.9.0
mecab 0.996

■機能概要

Mecabのipadicのコーパス学習用に使用します。
コーパスファイルと学習用CSVファイルを編集し、新しいコーパスファイルと学習用CSVファイルを編集することが目的です。

一覧画面



機能概要

メインの画面です。初めに学習用CSVファイルを読み込むようダイアログが表示されます。

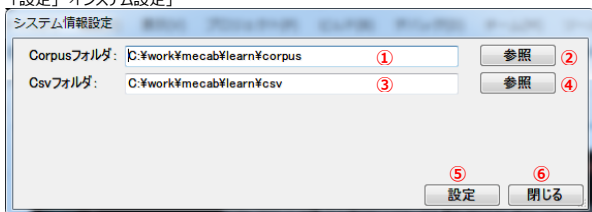
機能説明

- ① アプリケーションを終了します

■事前準備

システム設定画面

「設定」→「システム設定」



機能概要

コーパスファイルとCSVファイルの格納フォルダの場所を指定します。

機能説明

- ① コーパスファイルを格納してあるフォルダを指定します。
- ② コーパスファイルを格納してあるフォルダを指定するダイアログを開きます。
- ③ 学習用CSVファイルを格納してあるフォルダを指定します。
- ④ 学習用CSVファイルを格納してあるフォルダを指定するダイアログを開きます。
- ⑤ 画面上の内容で設定を更新します。
- ⑥ 修正内容を取り消して戻ります。

形態要素置換情報設定画面

「設定」→「置換情報設定」

機能概要

一括変換時の変換情報を登録します。

機能説明

- 登録されている形態要素置換情報のリストが表示されます。上から順にチェックされている情報が適用されます。形態要素置換情報格納ファイルパスはConst.xml中のMORPHEME_INFO_FILE_PATHで指定されます。
- 形態要素置換情報を追加するため編集モードにします。
- 選択された形態要素置換情報を編集するため編集モードにします。
- 選択された形態要素置換情報を削除します。
- 選択された形態要素置換情報(の優先順位)を一つ上げます。
- 選択された形態要素置換情報(の優先順位)を一つ下げます。
- 編集した内容を形態要素置換情報格納ファイルに反映します。
- 編集した内容を取り消して戻ります。

「設定」→「置換情報設定」→編集モード

機能概要

置換前の形態要素リストと、置換後の形態要素リストを登録します。

機能説明

- 形態要素置換情報名を設定します。
- 置換前形態要素リストに形態要素を追加します。
- 選択された形態要素を編集します。
- 選択された形態要素を削除します。
- 選択された形態要素の優先順位を上げます。
- 選択された形態要素の優先順位を下げます。
- コーパスファイル中に置換前形態要素リストで指定されたパターンが続く場合、チェックするとパターンが続く限り置換し続けます。
- 形態要素編集画面を出さずに自動的に置換完了とさせます。チェックされない場合は形態要素編集画面が表示されます。
- 表層系を指定して置換する場合チェックします。
- 表層系ファイルを適用する置換前形態要素リスト中のIndexを指定します。0から始まります。上記の場合ですと、「名詞－接尾－助数詞」に適用されます。
- CRLF区切りで表層系のパターンを指定します。表層系に該当するパターンが含まれると置換対象になります。
- 置換後形態要素リストに形態要素を追加します。
- 選択された形態要素を編集します。
- 選択された形態要素を削除します。
- 登録した内容で形態要素置換情報を確定し、形態要素置換情報リストの末尾に追加あるいは該当情報を修正します。

形態要素編集画面

「設定」→「置換情報設定」→編集モード→形態要素編集

形態要素編集

表層系: ①
左文脈ID: ②
右文脈ID: ③
コスト: ④
品詞: 名詞 ⑤
品詞詳細1: 数 ⑥
品詞詳細2: ⑦
品詞詳細3: ⑧
活用形: ⑨
活用型: ⑩
基本型: ⑪
読み: ⑫
発音: ⑬

解析結果一覧:

表層系	品詞	品詞詳細1	品詞詳細2	品詞詳細3	活用形	活用型	基本形	読み	発音
-----	----	-------	-------	-------	-----	-----	-----	----	----

解析(K) 学習済(L) 選択(T) ⑭追加(A) ⑮設定(I) 保存(S) 閉じる(C)

機能概要

形態要素の情報を編集します。置換のための情報なので、全ての項目は使用しません。

機能説明

- ① 使用しません。
 - ② 使用しません。
 - ③ 使用しません。
 - ④ 使用しません。
 - ⑤ 品詞を設定します。
 - ⑥ 品詞詳細 1 を設定します。
 - ⑦ 品詞詳細 2 を設定します。
 - ⑧ 品詞詳細 3 を設定します。
 - ⑨ 活用形を設定します。
 - ⑩ 活用型を設定します。
 - ⑪ 基本型を設定します。
 - ⑫ 読みを設定します。
 - ⑬ 発音を設定します。
 - ⑭ 編集した内容を登録して、画面を閉じます。
 - ⑮ 編集した内容を破棄して、画面を閉じます。
- ※⑤～⑬はどれか一つは入力が必要です。

■形態要素編集

一覧画面

「ファイル」→「コーパス読み込み」→編集モード(初期値)

コーパス編集

ファイル(E) 編集(E) 設定(S)

編集対象 C:\work\mecab\learn\corpus\work33 3.txt ① コーパス個数:1085 ②

表層系	品詞	品詞...	品詞...	品詞...	活用形	活用型	基本形	読み	発音
本日	記号	一般	*	*	*	*	*	本日	ホンジツ
24時	名詞	副詞...	*	*	*	*	*	ホンジツ	ホンジツ
から	数詞	時間	*	*	*	*	*	カラ	カラ
ラジオ関西	助詞	格助詞	一般	*	*	*	*	ラジオカンサイ	ラジオカンサイ
「	名詞	固有...	一般	*	*	*	*	「	「
なんでもカウンセリング～You&...	記号	括弧開	*	*	*	*	*	なんでもカウンセリング～ユー...	ナンデモカウンセリング～ユー...
」	名詞	固有...	番組名	*	*	*	*	」	」
に	記号	括弧閉	*	*	*	*	*	に	に
出演	助詞	格助詞	一般	*	*	*	*	出演	シュツエン
し	名詞	サ変...	*	*	*	*	*	し	シ
ます	動詞	自立	*	*	サ変・...	連用形	する	ます	マス
。	助動詞	*	*	*	特殊・...	基本形	ます	。	。
EOS	記号	句点	*	*	*	*	*	。	。
『	記号	括弧開	*	*	*	*	*	『	『
リニア	名詞	一般	*	*	*	*	*	リニア	リニア
、	記号	読点	*	*	*	*	*	、	、
時速	名詞	一般	*	*	*	*	*	時速	ジソク
603キロメートル	数詞	単位	*	*	*	*	*	603キロメートル	ロツビヤクサンキロメートル
を	助詞	格助詞	一般	*	*	*	*	を	ヲ
記録	名詞	サ変...	*	*	*	*	*	記録	キロク
記号	記号	空白	*	*	*	*	*	キロク	キロク
世界最高速	名詞	一般	*	*	*	*	*	サカイサイフウノハ	サカイサイフウノハ

③ 追加 ④ 編集 ⑤ 削除 ⑥ 編集 ⑦ 置換実行 ⑧ EOS集約 ⑨ チェック ⑩ 保存 ⑪ 閉じる

機能概要

形態要素リストを表示します。本画面から編集用画面を呼び出して編集します。

機能説明

- ① 現在編集中のファイルのフルパスです。
- ② 現在のコーパスの個数です。
- ③ コーパス編集画面を呼び出し、コーパスを追加します。
- ④ コーパス編集画面を呼び出し、選択されたコーパスを編集します。
- ⑤ 選択されたコーパスを削除します。
- ⑥ 編集モードと集計モードを切り替えます。
- ⑦ 一括置換を実行します。
- ⑧ EOSが複数連続する場合に集約します。Mecabの学習時にエラーが発生するためです。
- ⑨ 異常なコーパスがないかチェックします。
- ⑩ コーパスファイルと学習用CSVファイルを保存します。
- ⑪ アプリケーションを終了します。

「ファイル」→「コーパス読み込み」→集計モード

コーパス編集

ファイル(E) 編集(E) 設定(S)

編集対象 C:\work\mecab\learn\corpus\work33 3.txt コーパス個数:350

表層系	品詞	品詞...	品詞...	品詞...	出現...
本日	記号	一般	*	*	1
24時	名詞	副詞...	*	*	1
から	数詞	時間	*	*	1
ラジオ関西	助詞	格助詞	一般	*	13
「	名詞	固有...	一般	*	1
なんでもカウンセリング～You&...	記号	括弧開	*	*	19
」	名詞	固有...	番組名	*	1
に	記号	括弧閉	*	*	19
出演	助詞	格助詞	一般	*	40
し	名詞	サ変...	*	*	1
ます	動詞	自立	*	*	15
。	助動詞	*	*	*	20
EOS	記号	句点	*	*	39
『	記号	括弧開	*	*	1
リニア	名詞	一般	*	*	2
、	記号	読点	*	*	68
時速	名詞	一般	*	*	9
603キロメートル	数詞	単位	*	*	2
を	助詞	格助詞	一般	*	21
記録	名詞	サ変...	*	*	3
記号	記号	空白	*	*	26
世界最高速	名詞	一般	*	*	2
更新	名詞	サ変...	*	*	3
タタ	名詞	組織	*	*	0

追加 編集 削除 集計 置換実行 EOS集約 チェック 保存 閉じる

形態要素編集画面

「ファイル」→「コーパス読み込み」→編集モード(初期値)→追加 or 編集

形態要素編集

表層系: 603キロメートル

左文脈ID: 1000

右文脈ID: 1000

コスト: 1000

品詞: 数詞

品詞詳細1: 単位

品詞詳細2: *

品詞詳細3: *

活用形: *

活用型: *

基本型: 603キロメートル

読み: ロッピャクサンキロメートル

発音: ロッピャクサンキロメートル

解析結果一覧:

表層系	品詞	品詞詳	品詞詳	品詞詳	活用形	活用型	基本形	読み	発
5メートル	数詞	単位	*	*	*	*	5メートル	ゴメートル	ゴ
5回	数詞	単位	*	*	*	*	5回	ゴカイ	ゴ
5人	数詞	単位	*	*	*	*	5人	ゴニン	ゴ
5人	数詞	単位	*	*	*	*	5人	ゴニン	ゴ
5兆円	数詞	単位	*	*	*	*	*		
6%	数詞	単位	*	*	*	*	6%	ロクパーセント	ロ
6, 500t	数詞	単位	*	*	*	*	6, 500t	ロクセンゴヒャクトン	ロ
6. 1キロ	数詞	単位	*	*	*	*	*		
6.3m2	数詞	単位	*	*	*	*	6.3m2	ロクテンサンヘーベ	ロ
6. 3平米	数詞	単位	*	*	*	*	6. 3平米	ロクテンサンヘイベイ	ロ
60%	数詞	単位	*	*	*	*	*		
600キロ	数詞	単位	*	*	*	*	600キロ	ロツピャクキロ	ロ
600キロメートル	数詞	単位	*	*	*	*	600キロメートル	ロツピャクキロメートル	ロ
600キロメートル台	数詞	単位	*	*	*	*	600キロメートル台	ロツピャクキロメートルダイ	ロ
600キロ超	数詞	単位	*	*	*	*	600キロ超	ロツピャクキロチョウ	ロ
600メートル	数詞	単位	*	*	*	*	*		
603キロメートル	数詞	単位	*	*	*	*	603キロメートル	ロツピャクサンキロメートル	ロ

① 解析(K)

② 学習済(L)

③ 選択(T)

④ 追加(A)

⑤ 中断(Q)

⑥ 設定(I)

⑦ 保存(S)

⑧ 閉じる(C)

⑨ ☐ 以後変換しない

機能概要

形態要素の編集をします。

機能説明

- ① ipadicの辞書を使って解析した結果が一覧に表示されます
- ② 学習用CSVに追加されている形態要素を使って解析した結果が一覧に表示されます
- ③ 解析結果一覧から1件選んだものを選択状態にします。
- ④ 編集した内容を新しい形態要素として追加して選択状態にします。
かつ、学習用CSVファイル(学習用CSVファイル格納フォルダ+Const.xml中のLEARN_CSV_FILEタグの値)に上書きします。
- ⑤ 一括置換の最中に途中で中断する際に使用します。
- ⑥ 選択状態の形態要素の内容で編集対象の形態要素を上書きし、一覧画面に戻ります。
- ⑦ 選択状態の形態要素の内容で編集対象の形態要素を上書きし、一覧画面に戻ります。
かつ、学習用CSVファイル(学習用CSVファイル格納フォルダ+Const.xml中のLEARN_CSV_FILEタグの値)に上書きします。
- ⑧ 編集した内容を破棄して、一覧画面に戻ります。
- ⑨ 一括置換の最中に以降は置換対象外とする際に使用します。