

【概要】

ComeJisyo は、医療施設で使われる用語を集めた、分かち書き用の辞書です。

分かち書きのツールには、形態素解析器 Mecab を使っています。

今回、ComeJisyoUtf8-3 を公開しました。全てではありませんが、ComeJisyoUtf8-2 で見つかったヨミガナ誤りと品詞誤りを訂正しています。

Mecab0.996 でコンパイルした dic ファイルと、csv ファイルを公開します。

登録語は 118,404 語です。半角カタカナ、全角英数字、シフト JIS 依存文字を含む語は含まれません。

但し「Ⅱ型糖尿病」等、用語に含まれるローマ数字は登録しています。

【登録語数】 118,404 語

【文字コード】 UTF-8 BOM 付き

公開しているのは以下の2つです。

◆ コンパイル済のユーザ辞書 ComeJisyoUtf8-3.dic (ipadic 形式)

動作検証：Windows10

利用方法は、適当な場所にフォルダを作成し、ComeJisyoUtf8-3.dic をコピーして下さい。

例えば、c : ¥Users¥namae¥Mecab¥userdic など

次に、¥Mecab¥dic¥ipadic の中にある dicrc に ComeJisyoUtf8-3.dic を格納した場所を追加して下さい。(Mecab をインストールする際、文字コードは Utf-8 を指定して下さい。)

例) userdic=C : ¥Users¥namae¥Mecab¥userdic¥ComeJisyoUtf8-3.dic
を追加して下さい。

また、文字コードの設定は、config-charset = utf8 として下さい。

◆ csv 形式の ComeJisyo

文字コードは Utf8 (BOM 付き)です。

辞書の形式は、Ipadic 辞書に合わせています。

【属性】ヨミガナの後ろの◆以降が ComeJisyo 固有の属性です。

◆：英語表記：出現情報：文書頻度：MEDIS 標準マスター索引出現情報：通番

◆： ComeJisyo 固有の属性との区切りを意味します。

英語表記： 看護師国家試験問題に併記されている英語、および、京都大学 金子周司先生より提供頂いたライフサイエンス辞書にある英語を掲載しています。

出現情報：「看」「助」「栄」は、看護師国家試験、助産師国家試験、そして管理栄養士国家試験 8 年分 (2013 年～2019 年度) の問題文に出現した用語であることを示しています。「教」は、看護師及び管理栄養士養成校で使われている教科書 57 冊の索引に使われている語であることを示しています。

文書頻度：実際の医療記録データ 3 種類、看護師教育で使われる看護記録データ、GSK2012-D 模擬診療録テキスト・データの 5 種類における文書頻度（0～5）

MEDIS 標準マスター索引出現情報：4 種類の MEDIS 標準マスターの索引と一致した用語であることを示しています。

病：病名マスターの索引

処：手術・処置マスター


看：看護観察マスター

所：症状・所見マスター

通 番： ComeJisyoV1 からの用語に付加した通し番号です。

【解析精度について】

今回の ComeJisyoUtf8-3 は、mecab 推奨の Ipadic のコストを基に設定しています。
ComeJisyoUtf8-3 の登録語の内、記号（“・” “／” “+” “2” “.” 等）やローマ数字が含まれる登録語は可分割されます。

これらを含めた合成語を抽出したい場合は、合成語生成ツール  「GoMusubi」をご利用下さい。

【利用する際に】

本辞書を利用して行なった研究等の成果を公表する場合は、その旨を明記してください。

【医療情報をご提供頂いている医療施設】

千葉大学医学部附属病院

聖路加国際病院

九州大学附属病院

佐賀大学医学部附属病院（教育用 看護記録データ）

【助成一覧】

ComeJisyo は次の助成金による援助で行われています。

1. 2004 年～2017 年 西南女学院大学 学内共同研究費
2. 2009～2011 年度 科学研究費・基盤研究（B）21300099 コメディカル実践用語辞書データベースの作成
3. 2018～2021 年度 科学研究費・基盤研究（B）18H03499 語形成および意味的情報を付加した実践医療用語辞書の構築

【作成者】

ComeJisyo プロジェクト代表 相良かおる （西南女学院大学）

sagara-k@users.sourceforge.jp