

# テキストマイニング 2018 事 前準備マニュアル

同志社大学

データサイエンス研究室

2018 年 4 月 20 日

テキストマイニング・計量分析 2018 のセミナーでは 6/10（日）の講習会を設けており、当日はテキストマイニングのためのソフトウェア MTMineR を用いて演習を行う予定である。そのための事前準備として、演習参加者の方には Windows OS の PC を持参していただきたい。以下のソフトウェアをインストールしていただくことを願いたい。この資料は事前準備の手順を説明するものである。

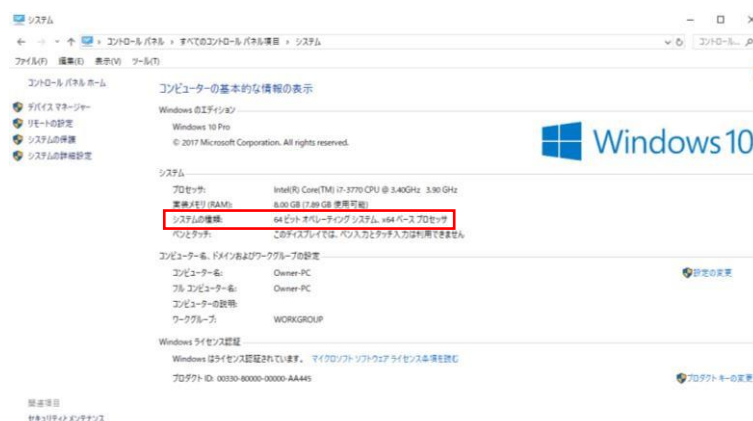
手順：

- ・ 個人 PC の bit 数の確認
- ・ 各自の PC の bit 数にマッチする Java Runtime Environment (JRE) のインストール
- ・ 形態素解析器 MeCab、JUMAN のインストール
- ・ 文節係り受け解析器 CaboCha のインストール
- ・ R 言語のインストール
- ・ MTMineR 本体をダウンロードし、解凍する。

## 0. 個人 PC の bit 数確認

これからインストールしていただく多くのソフトウェアは、32bit 版と 64bit 版で大別される。ソフトウェアを正しくインストールするためにまず次の手順を踏んで自分の PC の bit 数を確認する。以下は、Windows10 を例とした場合の説明である。

- ・ スタート→「W」欄→Windows システムツール→コントロールパネル
- ・ システムをクリック→システムの種類にて確認



## 1. JRE のインストール

MTMineR は Java 言語で開発され、それを動かすのに Java Runtime Environment (JRE) をインストールする必要がある。JRE は[このリンク](#)からダウンロードできる。JRE をダウンロードする際に必ず自分の PC の bit 数と同じものの JRE をダウンロードしてください。また MTMineR をダウンロードする際にも同じ bit のものにする必要がある。

## 2. 日本語形態素解析器 MeCab、JUMAN と係り受け解析器の CaboCha のインストール

テキストマイニングは、テキストを何らかの形で集計を行い、その集計データに対して統計処理を施す一連のプロセスを指す。日本語の文章は分かれ書きされていないため、集計を行うためにテキストを分割しなければならない。そのためのソフトウェアとして形態素解析器の MeCab や JUMAN、係り受け解析器の CaboCha などが挙げられる。

MeCab と JUMAN は、日本語文章を形態素<sup>1</sup>ごとに分割してくれるソフトウェアである。それぞれ[このリンク](#)または[このリンク](#)からダウンロードできる。自分の PC の bit 数と一致するものをダウンロードし、インストールする。

CaboCha は、日本語文章を文節<sup>2</sup>ごとに分割してくれるソフトウェアである。最新版の 0.69 は[このリンク](#)からダウンロードできる。

インストールが終わると、パソコンのシステムから実行ファイルを呼び出せるようにパスを通す。MeCab のパスの通し方を次に示す。JUMAN と CaboCha の操作手順は MeCab と同じである。

### パスの通し方

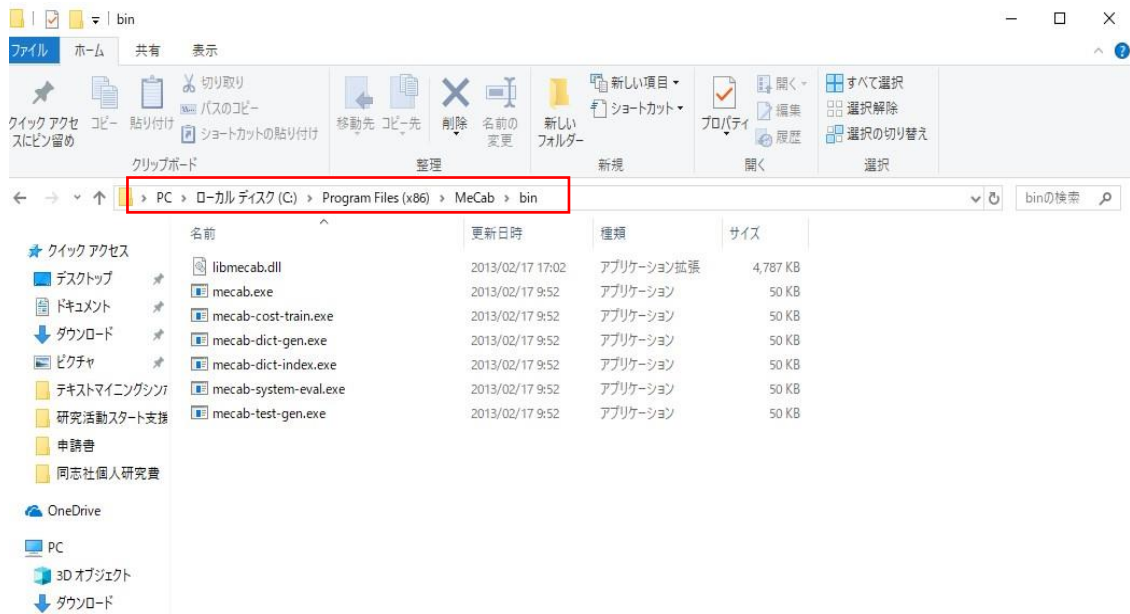
(Windows10 の場合)

- MeCab インストール先の bin フォルダ直下のルートをコピー

---

<sup>1</sup> 意味を表す最小単位である。

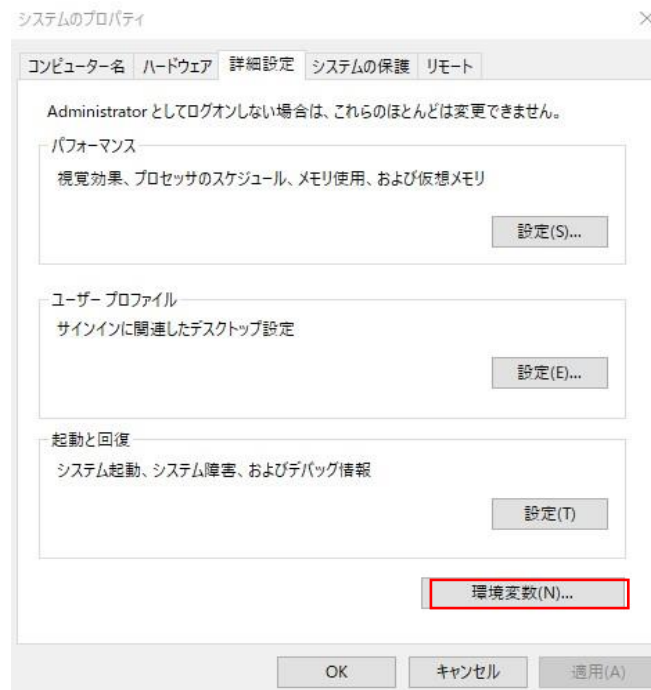
<sup>2</sup> 文を実際の言語として不自然でない程度に区切ったときに得られる最小単位。



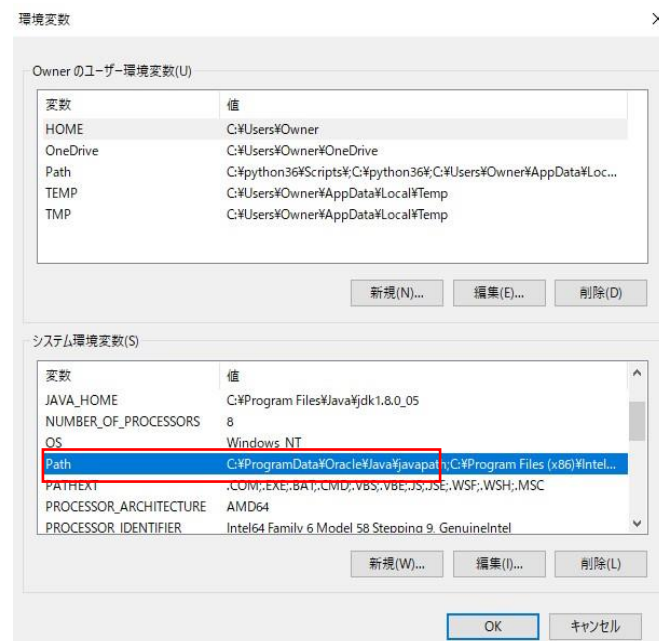
- スタート→「W」欄→Windows システムツール→コントロールパネル
- システム→システム詳細設定



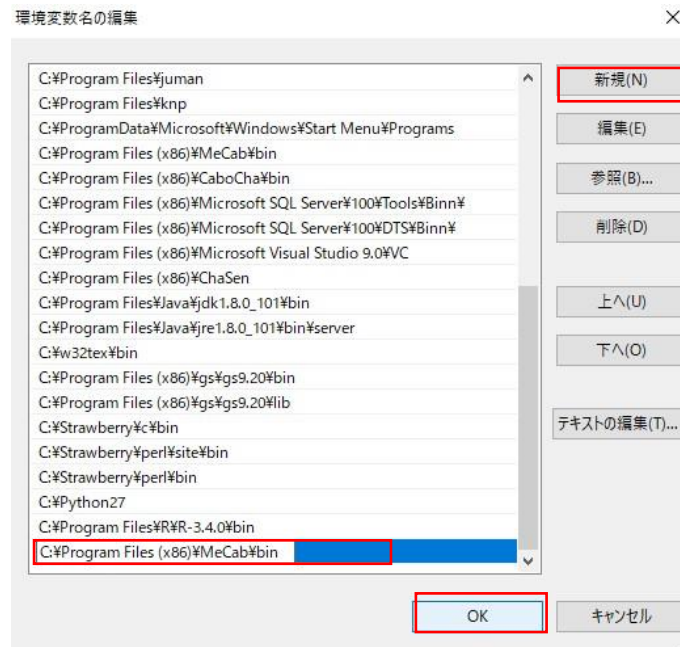
- ・環境変数 (N) をクリック



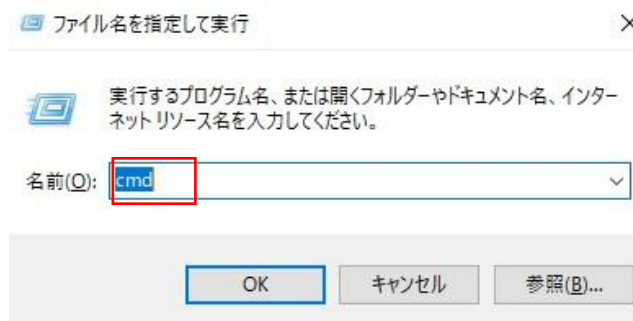
- ・システム環境変数 (S) 欄の Path を左ダブルクリック



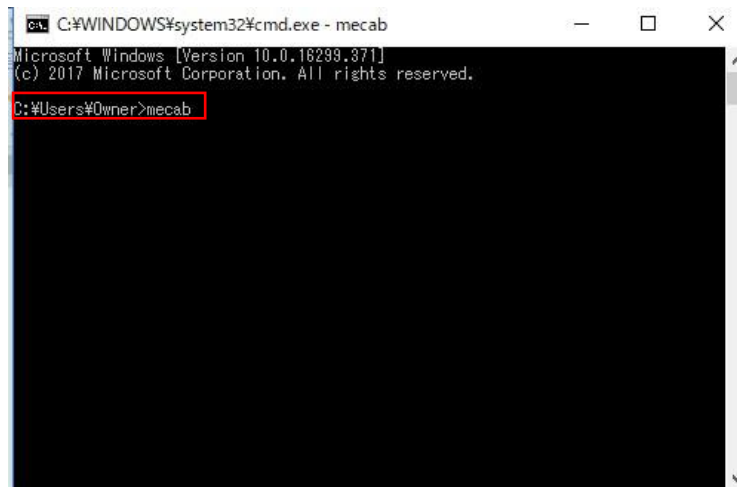
- ・「環境変数名編集」の新規をクリック
- ・ MeCab インストール先の bin フォルダ直下のルートをペスト→OK



- ・ window キー+R を押して、cmd を入力→OK



- ・ mecab を入力、エラーメッセージが出ていなければ成功



```
C:\WINDOWS\system32\cmd.exe - mecab
Microsoft Windows [Version 10.0.16299.371]
(c) 2017 Microsoft Corporation. All rights reserved.
C:\Users\owner>mecab
```

### 3. 英語、フランス語とドイツ語形態素解析器 TreeTagger をインストール

手順 1. TreeTagger では Perl プログラムを実行するため、まず Perl をインストールする。

<https://www.activestate.com/activeperl/downloads> bin まで Perl のパスを通す。例えば

C:\Perl64\bin。

手順 2. [TreeTagger](#) をダウンロードする。

TreeTagger のホームページ :

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

Windows version:

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger-windows-3.2.1.zip>

MacOS version:

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger-MacOSX-3.2.tar.gz>

手順 3. ダウンロードしたファイルを解凍し、ローカルに展開しておく。

C:\TreeTagger に入れておくと最も手間がかからない。

自分が好きなローカルに置いた場合は、解凍したファイルの中のエンジン部分の PATH を変更する必要がある。

例えば、C:\mydir\TreeTagger に解凍した場合 :

TreeTagger 中のフォルダ bin の中の tag-english.bat ファイルをテキスト

エディターで開き、第2行の TAGDIR=C:\TreeTagger を  
TAGDIR=C:\mydir\TreeTagger に変更し、上書き保存する。

手順4 : bin まで TreeTagger のパスを通す。

例えば C:\TreeTagger\bin 手順5. 言語ごとに必要となるパラメータファイルをダウンロードする。

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/#Windows>

- ◇ [English parameter file](#) (gzip compressed, UTF-8, [tagset documentation](#), trained on the Penn treebank)
- ◇ [French parameter file](#) (gzip compressed, UTF-8, [tagset documentation](#))
- ◇ [German parameter file](#) (gzip compressed, UTF-8, [tagset documentation](#)) 左のリンクはパラメータファイル、右のリンクは品詞体系などの説明である。手順6. ダウンロードしたファイルを解凍し、中のファイルを TreeTagger 中の lib フォルダに入れる。

英語の場合 : ダウンロードファイル名 english-par-linux-3.2-utf8.bin

解凍すると中にはファイル english-utf8.par ([このリンク](#)からもダウンロードできる) がある。フランス語、ドイツ語は上記の english がそれぞれ french, german に入れ替えただけである。しかし、French parameter file, German parameter file を解凍した後 french.par ([このリンク](#))、german.par ([このリンク](#)) のファイルになる。使えるようになるため、ファイル名を french-utf8.par、german-utf8.par に変更しないとイケない。

また、ダウンロードした english-utf8.par、french-utf8.par、german-utf8.par は名前を変更する必要ではなく、直接 TreeTagger の lib に入れてください。

#### 4. R のインストール

MTMineR では R 環境を呼び出しての統計処理を行っているため、R をインストールする必要がある。Windows のための R は[このリンク](#)からダウンロードできる。R にもパスを通す作業が必要であるため、MeCab のパスの通し方をご参照ください。



## 5. MTMineR 本体のインストール

MTMineR 本体を[このリンク](#)からダウンロードし、解凍しておく。