

# MIT MineR

## 事前準備マニュアル

2023 年 4 月 20 日

この資料は、Windows OS の PC を対象にし、事前準備の手順を説明するものである。

手順：

- [Java Runtime Environment \(JRE\)のインストール](#)
- [（日本語）形態素解析器 MeCab, JUMAN, 文節係り受け解析器 CaboCha のインストール](#)
- [（欧米語）形態素解析器 TreeTagger のインストール](#)
- [MTMineR 本体をダウンロードし、解凍する](#)

## 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 [https://www.java.com/ja/download/help/java\\_win64bit.html#Java%20for%2064-bit](https://www.java.com/ja/download/help/java_win64bit.html#Java%20for%2064-bit) からダウンロードできる。

## 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ち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。[このリンク](#)からダウンロードできる。

---

<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。

### 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://strawberryperl.com/>

手順 2. [TreeTagger](#) をダウンロードする.

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 に入れておくとも手間がかからない.

手順 4 : bin まで TreeTagger のパスを通す.

### 4. 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.