

この教材は『手を動かしながらやさしく学べるはじめてのAIデータサイエンスリテラシー』（技術評論社）の実践編として本書のv～viiページで紹介している発展的な実践教材です。本書で学んだことを活かして、発展的で面白いプロジェクトに挑戦できます。この教材はPDFで提供しています。必要に応じて印刷し、Wolframノートブックに自分でコードを入力して実行してください。（Wolframノートブックでの配布はしていません。）

なお、ここでは、本書独自のフォーマットを使い、プログラムの入力部分にオレンジの枠、出力部分にブルーの枠をつけてプログラムの部分をわかりやすく表示しています。新規でノートブックを開いた時にはこれらの色枠は付きません。

実践編5：文学作品のイメージに合うワードクラウドを作ろう

第6章の演習6-4で作成した「走れメロス」のワードクラウドを参考にして、文学作品のワードクラウドを作りましょう。本書サポートページで提供している文学作品のテキストファイルの中から好きな作品を選び、自分だけのオリジナルのワードクラウドを作ってください。

文学作品名とそのファイル名(全10作品)

サポートページにある文学作品は、以下の10作品です。
(著者が形態素解析したテキストファイルで提供しています)

- 「吾輩は猫である」（夏目漱石）：wagahai_morph.txt
- 「人間失格」（太宰治）：ningenShikkaku_morph.txt
- 「みだれ髪」（与謝野晶子）：midaregami_morph.txt
- 「源氏物語 桐壺」（紫式部、與謝野晶子訳）：genjiKiritsubo_morph.txt
- 「源氏物語 夕顔」（紫式部、與謝野晶子訳）：genjiYugao_morph.txt
- 「母を尋ねて三千里」（アミーチス）：hahawotazunete_morph.txt
- 「みにくいアヒルの子」（アンデルセン）：minikui_morph.txt
- 「白雪姫」（グリム）：shirayukihime_morph.txt
- 「ラプンツエル」（グリム）：tangled_morph.txt
- 「若草物語」（オルコット）：wakakusa_morph.txt

＊注意＊

本書サポートページでは、青空文庫（<https://www.aozora.gr.jp/>）から著作権フリーのテキストデータを著者が形態素解析し、空白で区切られた自立語の並びのテキストデータを用意しました。

この演習では、テキストデータファイルをWolfram言語にインポートして行っています。（外部ファイルをインポートするには、Import関数に外部ファイルを与えるため、Wolfram Cloudの有料アカウントが必要です。無料アカウントを利用している場合は、テキストエディタでファイルを開いてからテキスト部分を「コピー＆ペースト」でプログラムに直接テキストを挿入する方法があります。）

データファイルのインポートの方法については本書サポートページ、あるいはWolfram CloudのImport関数のオンラインドキュメントを参照して下さい。

ワードクラウドの作成手順

1. テキストデータの準備

本書サポートページからダウンロードできるSampleDataFilesフォルダの中にあるPractice_5_Dataフォルダには、以下の10作品の文学作品を形態素解析したテキストファイルが入っています。事前にPractice_5_DataフォルダごとWolfram Cloudにアップロードしておきましょう。アップロードの方法については本書サポートページを参照して下さい。

Practice_5_Dataフォルダの中から好きな作品のテキストファイルを選び、**Import**関数を使ってファイルを読み込みます。その後の作業のために、読み込んだテキストは変数に代入しましょう。

例えば、**Import**関数を使って、変数**wagahai**に「吾輩は猫である」のテキストファイル "wagahai_morph.txt"を変数**wagahai**に代入するプログラムコードは、以下のようになります。この演習では、演習を行うWolframノートブックがあるフォルダの下にPractice_5_Dataフォルダがあることを想定しています。

```
wagahai = Import[NotebookDirectory[] <> "Practice_5_Data/wagahai_morph.txt"];
```

[インポート] ノートブックのディレクトリ

2. ワードクラウドの作成

1.で準備したテキストを**WordCloud**関数に与え、ワードクラウドを作成します。1.で代入した変数**wagahai**のワードクラウドは、次のプログラムで出力させることができます。

```
WordCloud[wagahai]
```

[ワードクラウドの生成]

3. 応用

さらに、第6章の演習6-4を参考にしてワードクラウドの色や形を変え、作品のイメージに合うワードクラウドを作成していきましょう。

図1は、大学生による個性豊かなワードクラウドの作品例です。それぞれのワードクラウドがどの文学作品のものなのか、言葉はもちろん、そのイメージからもわかるのではないのでしょうか？

Wolfram言語の関数は、本書28ページのコラムで紹介した「オンラインドキュメント」で調べたり、例題を見たりすることができます。新しく出てきた関数は自分でドキュメントを調べて、いろいろ試してみてください。



図1 文学作品のイメージに合わせて作成したワードクラウドの例