

Rによるテキストマイニング入門(第2版) 正誤表

本書の内容に以下の誤りがございました。お詫びして訂正いたします。

お手持ちの本の「刷数」とこの表の「該当刷数」が一致する箇所をご参照ください。お手持ちの本の「刷数」の調べ方は[こちら](#)

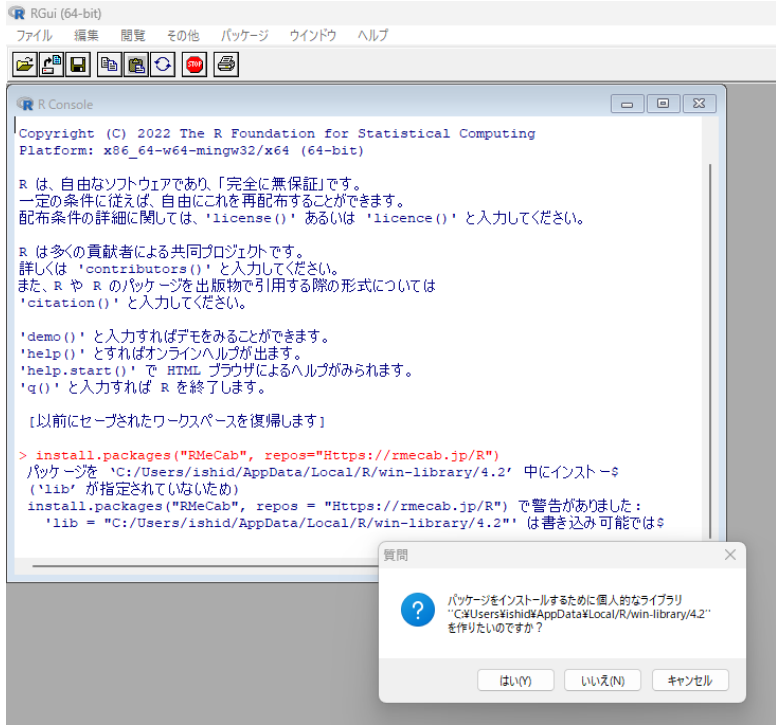
(2023年4月28日更新)

該当刷数	頁	行数など	誤	正
1,2,3,4,5	10	1~5行目	右の文章に差し替え	<p>なお Windows.zip という圧縮ファイルでは、本書で利用するファイルすべてについて、文字コードを日本語 Windows 標準の Shift-Jis に変更している。実は、Windows 版 R-4.2.0 バージョンでは文字コードが UTF-8 が標準となっている。OSとしての日本語 Windows の文字コードはいまも Shift-Jis(CP932)であるが、今後、R で日本語をファイル进行分析の場合は、対象となるファイルの文字コードを UTF-8 に変換する必要がある。筆者の GitHub に公開している Windows.zip ファイルは、Windows 版 R-4.1 までを利用しているユーザー向けのアーカイブである。本書を手にとっている読者が、もしも旧版の R を利用しているのであれば、これをアンインストールし、最新の R を導入し直すことを推奨したい。</p>
1,2,3,4,5	11	2.1 節 6行目の下	右の文章を挿入	<ol style="list-style-type: none"> 1. エクスプローラーを開いて(C:)ドライブ直下に work のような半角の名前のフォルダを用意しておく。 2. Windows の「スタート→設定」で「環境変数」と打ち込んで「環境変数を編集」を探す。 3. ユーザー環境変数の「新規(N)...」を押し、変数名として R_LIBS_USER, 変数値として C:\work を入力して、OK を押して設定を保存する (大文字小文字, アンダーバーなどの綴りを間違えないように注意)。
1,2,3,4,5	11	最下行	…, Download R-3.*.* for …	…, Download R-4.*.* for …
1,2,3,4,5	12	6~7行目	一方, Mac の場合は[Download R for (Mac) OS X] をクリックし, 次のページで R-3.*.*.pkg を選ぶ. こちらもインストールはダブルクリックするだけである.	一方, Mac の場合, [Download R for macOS] をクリックして進むが, インストーラとして Apple silicon arm64 用と Intel 用の2種類があるので注意が必要である.
1,2,3,4,5	12	2.2 節 2行目	(https://www.rstudio.com/)	(https://posit.co/)

該当刷数	頁	行数など	誤	正
1,2,3,4,5	13	図 2.3	右の図に差し替え	 <p>The screenshot shows the RStudio download page. A callout box with a black border and white background points to the 'Free' download button. The text inside the callout box reads 'Free の[DOWNLOAD]をクリック'. The page shows two options: 'RStudio Desktop Open Source Edition (AGPL v3)' for 'Free' and 'RStudio Desktop Pro' for '\$995 Per Year'.</p>
1,2,3,4,5	15	2.4.1 節	右の文章に差し替え	<p>Windows に MeCab をインストールする場合，開発者の工藤拓氏のサイト (https://taku910.github.io/mecab/) で配布しているインストーラではなく，ボランティアが独自に Windows 64bit 用にビルドを行ったバイナリ版を利用する方法を紹介する．また，Mac あるいは Linux では，ソースファイルから自身でビルドする方法を推奨する．</p>

該当刷数	頁	行数など	誤	正
1,2,3,4,5	15	2.4.2 節	右の文章に差し替え	<p>公式サイトには 32bit 版のインストーラが公開されているが、現在、ほとんどの Windows マシンは 64bit となっている。また R のほうも、Windows 版 R-4.2.0 バージョンからは 64bit 版のみがインストールされる。そこで、本書では https://github.com/ikegami-yukino/mecab/releases から 64bit 版の MeCab インストーラである <code>mecab-64-0.996.2.exe</code> をダウンロードしてインストールする方法を推奨する。非公式の MeCab バイナリではあるが、これをインストールすることで何らかの不具合が起こることはないと思われる。ただし、インストールにあたっては自己責任で行っていただきたい。なお、すでに 32bit 版の MeCab をインストールしている場合は、これをアンインストールしたうえで、改めて 64bit 版 MeCab を導入されたい。Mac や Linux でインストールに利用するソース版と異なり、Windows 用の MeCab インストーラには本体と辞書がバンドルされている。</p> <p>インストールは基本的に、インストーラをダブルクリックして [OK] を押していただくが、ただし、途中で日本語の文字コードを選択するダイアログが表示される。デフォルトでは Shift-Jis が選択されているが、これを UTF-8 に変更する (図 2.5 を参照)。日本語 Windows のデフォルトの文字コードは Shift-Jis(CP932)なのだが、Windows 版 R-4.2.0 バージョンでは文字コードが UTF-8 が標準となっているのである。したがって、今後 Windows 版 R で日本語テキストを扱う場合、それらのファイルの文字コードも UTF-8 に変換していく必要があることに注意してほしい。</p>
1,2,3,4,5	16	図 2.5	右の図に差し替え	 <p>図 2.5 Windows 版 MeCab のインストール</p>

該当刷数	頁	行数など	誤	正
1	16	2.4.3 4行目	…インストールしてほしい.	…インストールしてほしい (なお Homebrew あるいは MacPorts で導入した MeCab で RMeCab が落ちる理由は、辞書のパスが認識できないためである).
1,2,3,4	16	5~7行目	… (なお Homebrew…認識できないためである).	削除
1,2	16~ 17	2.4.3 節の 7~20 行目 (p.16 下から 5 行目~ p.17 9 行目)	右に差替え	<p>MeCab をソースからインストールするためには Mac に開発環境を導入する必要がある. Mac OS X では無償で開発環境が提供されている. まず[アプリケーション]フォルダにある[ユーティリティ]フォルダを開く. ここに[ターミナル]というソフトウェアがあるので, これを起動して次のように半角で入力して Enter キーを押す.</p> <pre>xcode-select --install</pre> <p>すると, ダイアログが表示されるので[インストール]を押すと, ライセンスへの同意を求められたあと, 自動的にダウンロードとインストールが始まる. これで MeCab をインストールするためのツールは整う. ただし, Mac OS X のバージョンによってはソースファイルからのインストールが制限されていることがある. この場合, インストール先のフォルダに書き込み権限を与える必要がある. これ以外にも, ユーザーの利用環境に固有の問題が生じることもあるだろう. 対処方法がわからないエラーに遭遇した場合は筆者のサポートサイトの掲示板 (https://rmecab.jp/new/forums/) に質問を投稿してみてください.</p>
1,2,3,4,5	17	6~8行目	対処方法がわからないエラーに遭遇した場合は筆者のサポートサイトの掲示板(https://rmecab.jp/new/forums/)に質問を投稿してみてください.	対処方法がわからないエラーに遭遇した場合は, MeCab の GitHub サイト (https://github.com/IshidaMotohiro/RMeCab)に Issue として投稿してほしい.
1,2,3,4,5	17	9~10行目	なお, インストールにつまづいたら筆者のサポートサイト…を参照されたい.	削除
1,2,3,4	17	10行目	https://sites.google.com/site/rmecab/home/install	https://rmecab.jp/new/
1,2,3,4,5	18	下から 5~1行目	なお, 実行例は Windows 64 bit の場合である. …防ぐことができる.	削除

該当刷数	頁	行数など	誤	正
1,2,3,4,5	19	1~2行目	<pre>> C:¥Program Files (x86) ¥MeCab¥bin¥mecab.exe ¥ c:¥work¥test.txt > c:¥work¥res.txt</pre>	<pre>> C:¥Program Files ¥MeCab¥bin¥mecab.exe ¥ C:¥work¥test.txt > c:¥work¥res.txt</pre>
1	20	2.6 5行目	<pre>install.packages("RMeCab", repos = "http://rmecab.jp/R")</pre>	<pre>install.packages("RMeCab", repos= "http://rmecab.jp/R") #windows の場合 install.packages("RMeCab", repos= "http://rmecab.jp/R", type = "source")#Mac/Linux の場合</pre>
1,2,3,4,5	20	2.6 節 1つ目の 網かけの後	右の文と図を挿入	<p>Windows で追加パッケージをインストールする場合、図 2.6 のように確認を求められることがある。これは、p.11 で説明した環境設定を行っていない場合に表示されるのだが、「はい」を押して進んでほしい。</p>  <p>図 2.6 パッケージインストール先の指定</p>

該当刷数	頁	行数など	誤	正
1,2,3,4,5	21	下から 8行目	…Cドライブの data フォルダ内 (以下“C:¥data”と表記する) に保存する.	…Cドライブの data フォルダに保存する。ただし Windows の場合、文字コードを UTF-8 に変更しておく必要がある。
1,2,3,4,5	21	下から 4行目	…, C:¥Program Files (x86)¥MeCab である.	…, C:¥Program Files¥MeCab である.
1,2,3,4,5	22	網掛けの中	右のように修正	<pre>> cd C:¥Program Files¥MeCab¥bin C:¥Program Files¥MeCab¥bin > mecab-dict-index.exe ¥ -d "C:¥Program Files¥MeCab¥dic¥ipadic"¥ -u C:¥data¥motohiro.dic -f utf-8 -t utf-8 ¥ C:¥data¥motohiro.csv</pre>
1,2,3,4,5	22	11行目	C : ¥Program Files (x86)¥MeCab¥etc フォルダ内に…	C : ¥Program Files¥MeCab¥etc フォルダ内に…
1,2,3,4	23	最終行の後	(最終行の後に右を追加)	なお, MeCab のデフォルトの辞書は登録語彙がやや古い. 一方, 新語や造語を毎週追加している拡張辞書に neologd(https://github.com/neologd/mecab-ipadic-neologd)がある. RMeCab でも利用することができる. 導入については https://rmecab.jp/new/ などを参照されたい.
1,2,3,4,5	33	11行目	<code>dat <- read.csv("data/classes.csv",fileEncoding = "CP932")</code>	<code>dat <- read.csv("data/classes_cp932.csv",fileEncoding = "CP932")</code>
1	33	下から 10行目	…文字コードについては後で説明するが, 重要なのは…	…文字コードで重要なのは…
1,2,3,4,5	33	下から 8行目	CSV形式,あるいは Excel で作成されたファイルのほとんどは Windows で	Excel で作成された CSV ファイルは Windows で
1,2,3,4,5	33	下から 3~1行目	ただし, Windows を使っている場合はもともと CP932 の設定になっているので, 上記のコードの <code>fileEncoding="CP932"</code> の部分は (直前のコマも含めて) 不要である.	ただし, 先にも説明したように Windows 版 R では UTF-8 が標準の文字コードになっている. Windows マシンの Excel で CSV ファイルを作成すると自動的に文字コードが CP932 となるので, <code>fileEncoding</code> の指定が必要である.
1,2,3,4,5	45	下から 7行目	<code>%>%</code> という記号をパイプ演算子という.	<code>%>%</code> という記号をパイプ演算子という. ちなみに, R-4.1.0 からは <code> ></code> もパイプ演算子として利用できる.
1	61	表 4.1 2行目	<code>"\\d"</code>	<code>"\d"</code>
1	65	下から 11行目	<code>alice %>% inspect</code>	<code>alice %>% inspect()</code>

該当刷数	頁	行数など	誤	正
1,2	109	下から 9行目	次に, ネットワークグラフを作成するために igraph パッケージを導入する.	次に, ネットワークグラフを作成するために igraph パッケージを導入する. なお Mac では別途 XQuartz (https://www.xquartz.org/) を導入する必要がある.
1,2,3,4	112	1つ目の 網掛けの 1行目	<code>okinawa <- read.csv("data/H18koe.csv")</code>	<code>okinawa<-read.csv("data/H18koe.csv", stringsAsFactors = TRUE)</code>
1	150	11行目	<code># 列名に形態素を設定</code>	<code># 行名に形態素を設定</code>
1,2	167	11.2 節 4~6行目	ただし, 2017年1月 … 実行できない.	ただし, 2018年5月の段階で rtweet パッケージはまだ部分的に実装の拡張が進行中であるため, 本書では twitteR パッケージを利用する.
1,2	168	11.4 節 1行目	ここでは, 総務省のアカウント…	ここでは, 文部科学省のアカウント…
1,2,3,4	179	11行目	11行目の下に右の文献を追加	・石田基広(2020):『実践 Rによるテキストマイニング』森北出版