

R

R.package

quanteda: Quantitative Analysis of Textual Data

quanteda v3 参照のこと

<https://rdr.io/cran/quanteda/>

- ・ quanteda を使った分析例 quanteda2
 - ・ quanteda を使った連語処理 quanteda3
-

■ コーパスを構築してコーパスデータとして扱う

- ・ tmp にテキストファイルが入っているとして

```
tmp.cor <- corpus(tmp)
```

■ コーパスデータの概要を見る

```
summary(tmp.cor)
```

■ コーパスデータの中身を見る

```
texts(tmp.cor)
```

■ KWIC 検索をする kwic()

```
kwic(tmp.cor, pattern = " 文字列 ")
```

■ トークンのリスト作成

```
tokens(tmp)
```

- ・ オプション

```
remove_numbers=T
```

- ・ オプション

```
remove_punct=T
```

- ・ 文単位

```
what="sentence"
```

- ・ 単語単位

what="word"

- 文字単位

what="char"

quanteda.corpora

<https://github.com/quanteda/quanteda.corpora>

- reference

<https://data.library.virginia.edu/a-beginners-guide-to-text-analysis-with-quanteda/>