

資源としての文化

第11回

福田一史

大阪国際工科専門職大学

<https://scrapbox.io/fukudakz/資源としての文化>



図. 講義ウェブサイトリンク (QRコード)
※LMSの資料のページにもリンクがあります

授業計画

回次	タイトル
1	ガイダンス・文化資源の定義
2-3	文化資源研究の系譜と基礎概念
4-6	デジタルアーカイブ
7-9	メタデータ
10-12	文化資源データの分析
13-14	事例研究
15	総括

コンテンツ

1. テキストマイニングによる文化資源データ分析（続き）

テキストマイニング による文化資源データ分析

前回の続き

演習（テキストマイニング）のスケジュール

- 第10回講義
 - テーマ設定・分析ファイル作成
- 第11回講義
 - KH Coderインストール・分析・テーマ見直し・分析用ファイルの再作成
- 第12回講義
 - 再分析・レポート（課題）作成と提出

データ分析演習（準備編 1）

- 分析のテーマを決める。
 - 自分が興味を抱く／取組中のテーマやトピックに関連する、デジタル形式でテキストを取得可能な（a.k.a. コピペできる）データを設定する。

データ分析演習（準備編 2）

- 分析対象とする**日本語テキストデータ**を収集し、分析用ファイルを作る。
 - 文字数基準：10,000字以上、上限なし（PCのスペックによる）
 - 入力可能ファイル形式：**テキストファイル**（.txt）、CSVファイル（.csv）、MS Excelファイル（.xlsx）
 - この講義ではテキストファイル形式の作成を推奨します
 - 分析しやすいように段落や属性などの外部変数を記録する
 - 「話」や「章」などの見出しで文書を分割する
 - 3～10件に分割するイメージだと分析しやすい
 - [KH Coderを使って分析するためには、どのようにデータを準備すればよいですか？](#)

テキストファイルの保存と見出しの付け方

- テキストファイルとは、文字データだけで構成されたファイル。「メモ帳」や「[サクラエディタ](#)」「vs Code」「[EmEditor](#)」などのアプリケーションで作成できる。
- KH Coderでは、**見出し**を特定の形式で記録することで、章や節とその内容を構造化することが可能
 - 右は見出しによる構造化の事例 ([本文](#))

```
<h1>一 午後の授業</h1>  
「ではみなさん、さういふふうに川だと云はれたり、乳の流れたあとだと云はれたりしてゐた、このぼんやりと白いものが何かご承知ですか。」.....
```

```
<h1>二 活版所</h1>  
ジヨバンニが学校の門を出るとき、同じ組の七八人は家へ歸らずカムパネルラをまん中にして校庭の隅の櫻の木のところ集まつてゐました。... (以下続く) ...
```

```
<h1>三 家</h1>  
ジヨバンニが勢よく歸つて來たのは、ある裏町の小さな家でした。...
```

KH Coderによるテキストマイニングの実践

- 前回の講義で策定したテーマと作成した分析用ファイル (.txt) を用いて分析を行う
- まずは以下の手順でダウンロードから。
 1. <https://kncoder.net/> へアクセス
 2. [KH Coder 3 \(最新版\) ダウンロード](#) へアクセス
 3. Windowsの「フリー版 : [kncoder-3b03a.exe](#)」からダウンロード



KH Coder 3 最新版のダウンロード

[KH Coder]

Windows

◆フリー版 : [kncoder-3b03a.exe](#) (2021 05/03)

無料でご利用いただけますが、サポートや保証はございません。はじめてKH Coderに触れる方は[チュートリアル](#)をご一読下さい。

◆サポート付きパッケージ : [株式会社SCREEN AS様より販売中](#)

職場で教室で、安心してお使いいただける企業様発のパッケージです。便利な自動インストール、セキュリティ対策、セミナー受講券、ご質問受付から、万一のバグ対応まで。

Mac

自動設定ソフトウェア [有償サポート]

下記のソースコードなら無料でご利用いただけますが、こちらの方が比較的手軽です。

ソースコード (Linux / Mac / Windows共通)

[GitHub Repositoryにて公開中](#)

KH Coderチュートリアル

- 「KH Coder 3 チュートリアル」
 - <https://kncoder.net/tutorial.html>
 - PDF版 https://kncoder.net/kokoro_slides.pdf
- 上記のKH Coderの作者によるチュートリアルにそって、機能の概要を把握する

演習

- チュートリアルで示される手順に従い、新しいプロジェクトとして、**前回作成した分析用ファイル**をKH Coderで読み込む。
- 前処理を実行し、分析を実施する。
 - **抽出語リスト**や、**共起ネットワーク**、**対応分析**（テキストを見出しで分割している場合）、**自己組織化マップ**などといったKH Coderが提供する分析ツールで、データを分析し、結果をもとに解釈を行ってみる。
 - 各分析は、データの表示単位などを細かく設定することで、違った形でデータが表出します。想定するものと近いかたちでデータが表示された場合は、このチューニングによって論点を明確化することができるかも。
 - 想定していた通りのことがわかったか？ 想定していたことと違う発見があったか？

演習

- 思っていたような結果が全然データに現れない場合もある。
- そのような場合は、テーマが悪いか、データが悪いか。
- 必要に応じて、テーマの練り直しや、分析用データ（テキストファイル）を再度作成して、分析を行ってください。
- また、抽出語がグループ化できる場合は、対応分析などにより、ブロックごとのテーマやトピックの変遷や比較を行うことができる。

演習

- 次回、より詳細な再分析を行った後に、演習用ファイル (.docx) でデータ分析の結果を記述してもらいます。
- そこでは、テーマに掲げた論点について、何がわかったかを議論してください。
- LMSで提出予定
- **！！来週の提出物は課題です！！**
 - (提出しないと単位取れない)