

資源としての文化

第3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)

コンテンツ

1. 文化資源の分析（つづき）
2. 文化資源の再利用

文化資源の分析

デジタル技術を用いて文化資源を分析する

テキストマイニング

- テキスト（文章・文字データ・文献など）を対象とするデータマイニング。
- 文章を単語や文節で区切り（形態素解析）、それら単語や文節の出現頻度や共起関係などを計量的に分析する。自然言語処理技術の一種。
- KH Coder
 - テキストマイニングソフトの一つ。
<http://khcoder.net/>

#	抽出語	品詞/活用	頻度
1	先生	名詞	595
2	K	タグ	411
3	奥さん	名詞	388
4	思う	動詞	296
5	父	名詞C	269
6	自分	名詞	264
7	見る	動詞	225
8	聞く	動詞	219
9	出る	動詞	185
10	人	名詞C	182
11	母	名詞C	170
12	お嬢さん	名詞	168
13	前	副詞可能	163
14	帰る	動詞	155
15	今	副詞可能	139
16	顔	名詞C	135
17	来る	動詞	131
18	考える	動詞	130
19	言葉	名詞	126
20	眼	名詞C	123

図. 夏目漱石「こころ」の抽出語リスト
<http://khcoder.net/> スクリーンショット集より

Voyant Tools

- <https://voyant-tools.org/>
- “Voyant Tools is a web-based text reading and analysis environment. It is a scholarly project that is designed to facilitate reading and interpretive practices for digital humanities students and scholars as well as for the general public.” – About – Voyant Tools Help.
<https://voyant-tools.org/docs/#!/guide/about> (accessed 2021-04-15)



ゲームタイトルの計量的分析

- 1980年代前半から2000年代前半までのビデオゲームのタイトルについての計量的分析。
- ゲームタイトルは経時的に長大化傾向。また、トレンドの変化が影響する。
 - 福田一史 et al.. 2017. “家庭用ゲームソフトのタイトルに関する研究：DBを活用した文字数・文字種の観点からみたマクロ的分析。” アート・リサーチ 17 (November): 29-44.
<http://doi.org/10.34382/00007765>.

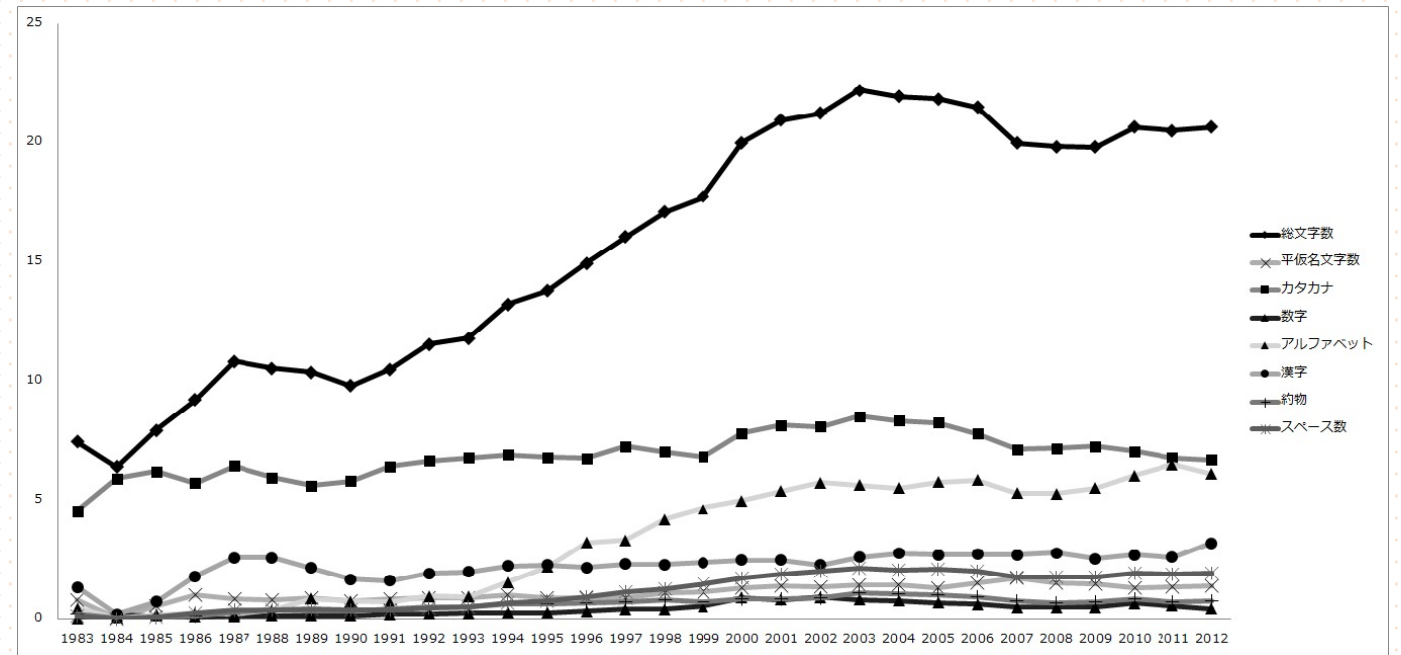


図. 文字種ごと平均文字数の推移

```

version="1.0" encoding="UTF-8"?>
model href="tei_all_ja_20191012_4.rnc" type="application/relax-ng-compact-syntax"
xmlns="http://www.tei-c.org/ns/1.0" xmlns:eaj="http://www.example.org/ns/ejaTEI"
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>走れメロス</title>
      <author>太宰治</author>
    </titleStmt>
    <publicationStmt>
      <distributor>青空文庫</distributor>
      <authority>金川一之</authority>
      <authority>高橋美奈子</authority>
      <date when="2011-01-17"> 2011年1月17日</date>
    </publicationStmt>
    <sourceDesc>
      <bibl>
        <author>太宰治</author>
        <title>走れメロス</title>
        <publisher>筑摩書房</publisher> 「太宰治全集3」 ちくま文庫、 <date when="
          >1988 (昭和63) 年10月25日</date>初版発行 <date when="
          >1998 (平成10) 年6月15日</date>第2刷 </bibl>
      </sourceDesc>
    </fileDesc>
    <revisionDesc>
      <list>
        <item>
          <date when="2011-01-17">2011年1月17日</date>修正 </item>
        <item> 入力<persName>金川一之</persName> 校正<persName>高橋美奈子</persName>
          <date when="2000-12-04"> 2000年12月4日</date>作成 </item>
        </list>
      </revisionDesc>
    </teiHeader>
    <text>
      <front>
        <p>
          <listPerson>
            <person xml:id="メロス">
              <persName>メロス</persName>
              <occupation>牧人</occupation>
            </person>
          </listPerson>
        </p>
      </front>
    </text>
  </teiHeader>

```

TEI

- Text Encoding Initiatives
- 人文学資料の適切なマークアップ（XML）のための仕様・標準、それを構築するためのプロジェクト。
- The TEI Guidelines
 - <https://www.tei-c.org/release/doc/tei-p5-doc/ja/html/index.html>
- 日本語向けTEIガイドライン（作成中・試行版）
 - https://github.com/TEI-EAJ/jp_guidelines/wiki

図. マークアップサンプルファイル「走れメロス」.
https://www.dhii.jp/nagasaki/dazai_all_20191012.xml

XML: Extensive Markup Language

- 任意の目的に用いるための汎用性が高いマークアップ言語。
- 比較的人間にとって読みやすいデータ形式。
- APIのレスポンスや、MS Word、TEIなどのテキスト文書の構造化などに用いられる。
- 参照：“Extensible Markup Language (XML) 1.1 (Second Edition)”。
<https://www.w3.org/TR/xml11/>.

IIIF: International Image Interoperability Framework

- <https://iiif.io/>
- 「画像へのアクセスを標準化し相互運用性を確保するための国際的なコミュニティ活動」
 - IIIFを用いた高品質／高精細の画像公開と利用事例 | 人文学オープンデータ共同利用センター. <http://codh.rois.ac.jp/iiif/>
- もしくはそれらにより定義されたAPIの仕様
 - [IIIF Image API](#)
 - [IIIF Presentation API](#)
 - など
- Community#Participating Institutions <https://iiif.io/community/>



図. IIIFロゴ

API

- Application Programming Interface
- プログラムコンポーネント同士を接続するためのインターフェイス
- バックエンドの一部の機能の公開や共有
- ワールド・ワイド・ウェブを介するAPIは「**ウェブAPI**」と呼ばれ、HTTP要求とJSONなどによる応答で構成される。

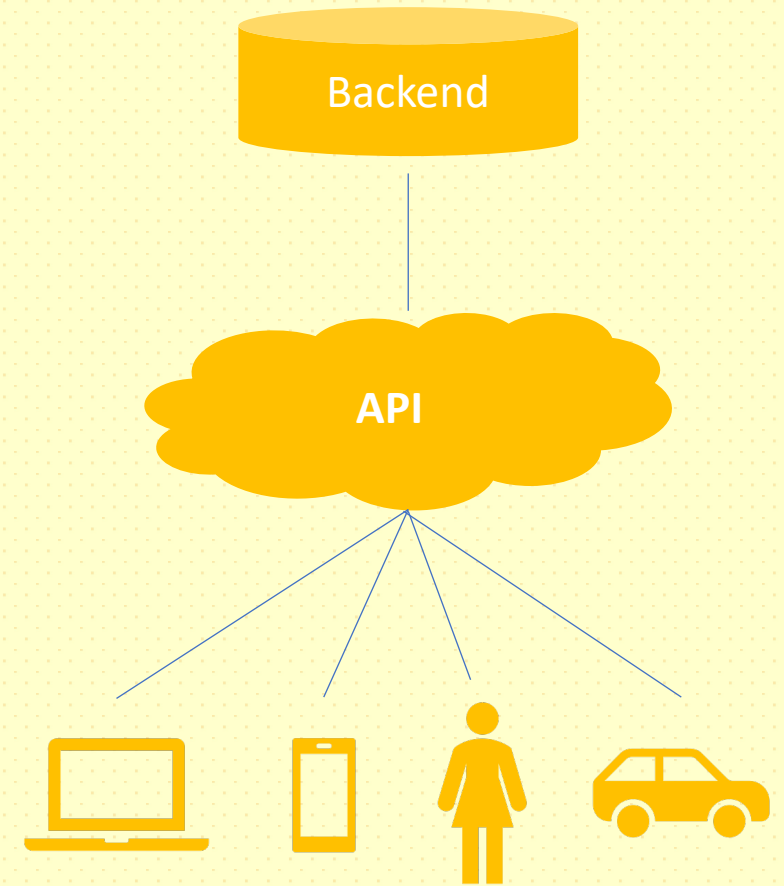


図. APIイメージ

API

- 「Web 2.0」 (2005～) の基礎要素
- **APIファースト**
 - ウェブサービス開発において、まずAPIの設計や開発を進めてから、ビューや他のシステムの接続などの開発を進めるという開発法
- APIを提供することで、それを軸とするエコシステム (≒**APIエコノミー**) が形成される。それにより、サービスの価値向上が生じる。
 - **ネットワーク外部性**が影響

APIの例

- 小売 : Amazon MWS API
- 金融 : Moneyforward, Freee, 各銀行API
- 気象 : Open Weather Map API, Dark Sky API
- 地理 : Google Maps API, 電子国土web
- SNS : Twitter API, Facebook API
- 図書・文化資源 : NDL, OpenBD, Japan Search, CiNii

JSON (JavaScript Object Notation)

- データ記述言語。
- **機械可読性の高い軽量のデータ交換フォーマット。**
 - JSONの紹介. <https://www.json.org/json-ja.html>.
- 1) 名前と値 (Key-Value) の組み合わせ、2) 配列 (Array) 、 の構造で記録される。
- ウェブ上のデータ交換のための形式における、標準的形式 (デファクトスタンダード) として広く用いられる。

Gettyと「どうぶつの森」

Getty

Animal Crossing Art Generator

Museum Research Institute Conservation Institute Foundation Support Us

Visit What's On Explore Art Resources Our Work About 



 Animal Crossing Art Generator <https://experiments.getty.edu/ac-art-generator#iiifloader> (accessed 2020-06-29)

機械学習

- KuroNetくずし字認識ビューア
- 人文学オープンデータ共同利用センターが提供する、ディープラーニングを用いたくずし字のOCRサービス。
 - <http://codh.rois.ac.jp/kuronet/>

集合知・クラウドソーシング

- 「みんなで翻刻」
- <https://honkoku.org/>
- 「すでに『みんなで翻刻』では、5000人の人々により600万文字もの史料が翻刻されています。」
 - <https://honkoku.org/>
 - 加納靖之. 2017. “みんなで翻刻--これまでとこれから.” リポート笠間, no. 63 (November): 53–56.
<http://hdl.handle.net/2433/228169>.
- くずし字学習支援アプリ「KuLA」
 - ゲーミフィケーションの方法論を採用。



みんなで翻刻

MINNA DE HONKOKU



くずし字、
読み解く
をしよう！

JOIN
参加する

※現在新バージョンの試験運用中です。
安定稼働するまで旧バージョンも併せてご利用ください。

Twitter

Bookmark LINEで送る

DHの方向性

- 既存の文化研究の**ブラックボックス化を危惧**し、デジタル技術を用いて**横串を指す**／**透明化**をすすめる。
- **オープン化**や**標準化**により、文化資源へのアクセシビリティを向上させ、持続的に提供可能とし、**知識基盤**を確立する。
 - 特にオープンな標準や、コミュニティで維持するオープンソースを指向する。
 - **方法論の共有地**（Methodological Commons）。
- アクセシビリティが向上することにより、これまで想定していなかった**新しい方法論**や**知見**が生まれつつある。

文化資源の二次利用

二次利用を支援する制度

背景

- デジタルアーカイブの発展により文化資源へのアクセシビリティが向上したとしても、デジタルコンテンツが**どのような条件下で利用できるか**を明示しなくては、それらの活用が進まない。
- そのようななか、海外ではクリエイティブ・コモンズ・ライセンスやパブリックドメインツールなどといった**枠組み・ツール**が普及しつつある。

パブリックドメイン

- 著作権保護期間の終了や、著作者の権利放棄などにより、著作物は自由に利用ができるようになる。パブリックドメイン (Public Domain: PD) とは、そのような状態の著作物、またそれらの群。
- “パブリックドメイン系ツール 使用ガイド”。
<https://resources.creativecommons.org/downloads/ccguides/publicdomain-ja-flat.pdf>.



図. “CC0” マーク (左) と “Public Domain Mark” (右)

パブリックドメイン

- 「著作権の保護期間は、原則として著作者の生存年間及びその死後70年間です。」
 - 著作者の権利の発生及び保護期間について – 文化庁.
<https://www.bunka.go.jp/seisaku/chosakuken/seidokaisetsu/gaiyo/hogokikan.html>.

著作物の種類	保護期間
無名・変名（周知の変名は除く）の著作物	公表後70年（死後70年経過が明らかであれば、その時点まで）
団体名義の著作物	公表後70年（創作後70年以内に公表されなかったときは、創作後70年）
映画の著作物	公表後70年（創作後70年以内に公表されなかったときは、創作後70年）

表. 著作権の保護期間の例外

青空文庫

- インターネットの電子図書館
- 「青空文庫は、誰にでもアクセスできる自由な電子本を、図書館のようにインターネット上に集めようとする活動です。」
 - “青空文庫早わかり”.
https://www.aozora.gr.jp/guide/aozora_bunko_hayawakari.html.
- パブリックドメインの小説などの著作物をデータ化し、公開するプロジェクト

クリエイティブ・コモンズ・ライセンス

- Creative Commons Licenses (CC Licenses)
- “Creative Commons is a nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity to address the world’s pressing challenges.”
 - What We Do – Creative Commons. <https://creativecommons.org/about/>

クリエイティブ・コモンズ・ライセンス

- インターネットにおける**著作物をシェアするための標準的ルール**として広く利用されている。
 - 例：Wikipediaのコンテンツ, デジタルアーカイブのコンテンツ
- クリエイティブコモンズは、アメリカで2001年に非営利団体として設立。ローレンス・レッシグら、アメリカのインターネット法や知的所有権問題の専門家などが発起人として関与。
 - “Laws that choke creativity | Larry Lessig – YouTube”. <https://www.youtube.com/watch?v=7Q25-S7jzgs>.
 - “Free to Mouse!: Lawrence Lessig on Disney, copyrights, and the creative commons”. <https://www.austinchronicle.com/screens/2002-03-08/84942/>

クリエイティブ・コモンズ・ライセンス

- “クリエイティブ・コモンズ・ライセンスとは”。
<https://creativecommons.jp/licenses/>. (accessed 2021-04-20)
- 作品を利用するための4つの条件（表示、非営利、改変禁止、継承）の組み合わせで6つのCCライセンスが存在する。



図. CCライセンスの例

<https://creativecommons.org/about/ccllicenses/>

Rights Statements

- CCは権利者による二次利用のための表示であるが、それに対して Rights Statementsは著作権のステータスや二次利用条件を表示するためのツールとして開発された。
- Europeana, DPLA, クリエイティブ・コモンズらにより、クリエイティブ・コモンズ・ライセンスと、パブリックドメイン・ツールでは不足する部分をフォローするためのツールとして開発された。
 - RightsStatements.org. <https://rightsstatements.org/en/>.
- 参考：数藤雅彦. 2020. “Rights Statementsと日本における権利表記の動向.” カレントアウェアネス, no. 343 (March): 19–23. <https://current.ndl.go.jp/ca1973>.

デジタルアーカイブの望ましい二次利用条件表示

1. 国際的に普及しているパブリック・ドメイン・ツール及びCCライセンス。特に、CC0、CCBYを強く推奨する。
 2. Rights Statementsからは、著作権あり、著作権あり - 教育目的の利用可、著作権なし - 他の法的制限あり、著作権なし - 契約による制限あり、著作権未評価のマーク。
 3. 日本独自の表示としては、文化庁長官裁定制度を利用した著作物であることがわかるマーク(著作権未決定-日本孤児著作物)。
- デジタルアーカイブジャパン実務者検討委員会.“ (抜粋) デジタルアーカイブにおける望ましい二次利用条件表示の在り方について (2019年版) ”。
https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/jitumusya/2018/nijiriyou2019.pdf.

デジタルアーカイブの望ましい二次利用条件表示

データ種別	自らが著作権を保有するものの二次利用条件
メタデータ	CC0
サムネイル／プレビュー	CC0, CC BY, PDM
デジタルコンテンツ	CC0, CC BY, PDM

表. 公的機関のもの又は公的助成により生成されたデータの二次利用条件

出所：デジタルアーカイブの連携に関する関係省庁等連絡会・実務者協議会。「デジタルアーカイブの構築・共有・活用ガイドライン」.

http://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_kyougikai/guideline.pdf. (accessed 2021-04-20)

デジタルアーカイブ：権利表記の事例

- メトロポリタン美術館 <https://www.metmuseum.org/>
 - ロサンゼルス・カウンティ美術館 <https://collections.lacma.org/>
 - J・ポール・Getty美術館 <http://search.getty.edu/gateway/landing>
 - 大英図書館 <https://www.flickr.com/photos/britishlibrary/albums>
- 参考：“【2020年版】パブリックドメインで無料！世界の名画550万枚をダウンロードできる美術館サイト19個まとめ”. <https://photoshopvip.net/111868>.

孤児作品（Orphan Works）

- 著作者の死亡、著作者である団体の廃止や、著作権の譲渡などといった原因で、権利保持者がわからない作品が多く存在する。場合によってはいつからPD入りするかわからない場合もあるなどといったこともあり、国際的な課題となっている。
- 国内では、裁定制度の見直し（2016年）により、権利者搜索の要件が緩和され、保証金制度が運用されるようになった（第67条）。
 - 著作権法 | e-Gov法令検索. <https://elaws.e-gov.go.jp/document?lawid=345AC0000000048>.
 - 著作権者不明等の場合の裁定制度 | 文化庁. https://www.bunka.go.jp/seisaku/chosakuken/seidokaisetsu/chosakukensha_fumei/.
 - 任天堂社による裁定制度利用の事例. <https://www.nintendo.co.jp/clvj/index.html>.