

● ● ● | デジタル図書館研究室

研究室紹介

2018/4/10 メディアプロジェクト演習2

前田 亮

1

「デジタル図書館」とは？

- デジタル情報を収集・整理・保管し、提供するサービス全般
 - 対象となる情報の種類は限定しない
 - ・ 電子書籍・新聞・雑誌、古典資料、芸術作品、Webページ、画像・映像・音源 etc.
- Web上の情報は、従来の図書館のように体系付けられていない
- しかし、人類の知識を蓄積・提供する一種の巨大なデジタル図書館と捉えることができる



2

● ● ● | デジタル情報の利点



- 「検索」ができる
- 情報の内容から自動的に「分類」できる
- 隠れた知識を「発見」(マイニング)できる
- 利用者に合った情報を「推薦」できる
- 様々な情報源を「統合利用」することができる
- これらの技術を「情報アクセス技術」と呼ぶ

3

研究の対象

Web情報

Twitter, ブログ,
レビュー, 電子掲
示板, etc.

情報検索

情報推薦

データマイニング

各種 メディア情報

楽曲, 動画, プレ
ゼン資料, etc.

図書館・ 人文情報

古典資料, 講演録,
etc.

4

研究テーマの例



・ Web情報

- 多言語 Wikipedia 記事への言語横断エンティティリンク
- Twitter のフォロー情報に基づくユーザ嗜好の推定
- Twitter を利用した漫画作品の評判分析
- 映画レビューからの印象推定に基づく映画推薦システム
- プロ野球の試合ダイジェストの自動生成

・ 図書館・人文情報

- 経営哲学に関する講演録を用いた質問応答システム
- 古典史料テキストからの情報抽出
- 異言語の浮世絵データベースからの同一作品の同定
- 「新古典籍総合目録データベース」のマルチリンガル化
- 篆書体による蔵書印の文字認識および検索

5

● ● ● | Web情報

6

多言語 Wikipedia 記事への言語横断エンティティリンク

- 文書中のキーワードから、該当する知識ベース記事 (Wikipediaなど) に自動的にリンクを付与する手法

日本語文書

日本語 Wikipedia 記事

7

言語横断エンティティリンクの例

英語版 Wikipedia の該当記事

ニュース記事

中国語版 Wikipedia の該当記事

日本語版 Wikipedia の該当記事

言語間リンクがなくても該当記事にリンク

Twitter のフォロー情報に基づくユーザ嗜好の推定

- 対象ユーザのフォロー情報および行動履歴から、ユーザの嗜好を予測し、タイムラインを並び替える

9

Twitterを利用した漫画作品の評判分析

作品に対するツイートを分析し、評価値を予測

10

映画レビューからの印象推定に基づく映画推薦システム

11

プロ野球の試合ダイジェストの自動生成

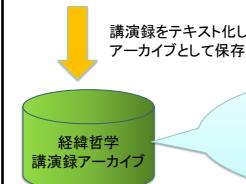
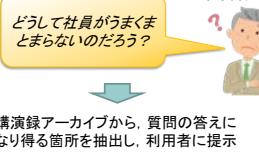
- 試合のデータからダイジェストの文章を自動生成
- 利用者の目的に応じて、動的に内容や分量を変更可能
 - 自分が応援している球団向けのダイジェスト
 - 媒体に応じて文章長を変更
 - スマートフォン、Webページ、新聞など

12

図書館・人文情報

13

経営哲学に関する講演録を用いた質問応答システム



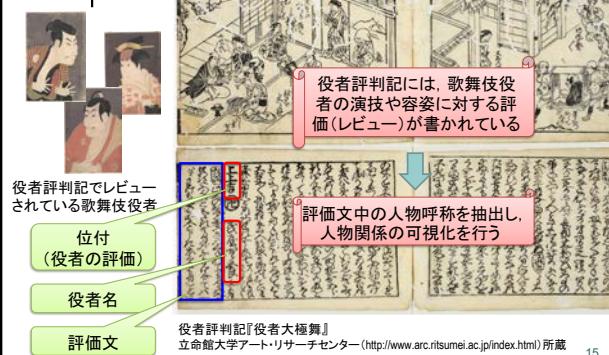
講演録アーカイブから、質問の答えになり得る箇所を抽出し、利用者に提示



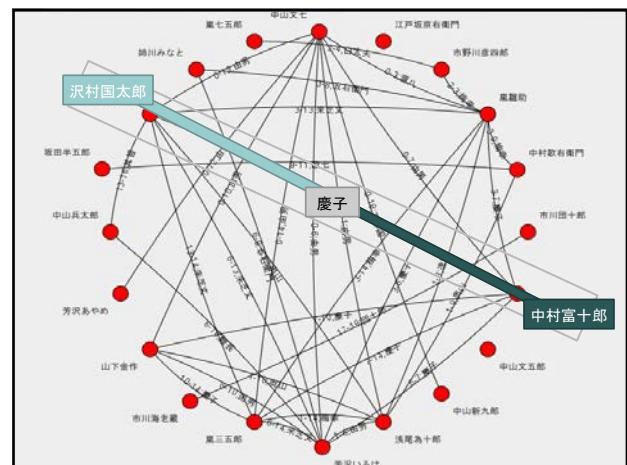
…同時に会社に住んでいる社員の考え方によって社風が決まつてくるわけです。ですから、いい考え方をする社員がいる場合には素晴らしい社風になっていきます。…

14

古典資料テキストから的人物情報の抽出と可視化

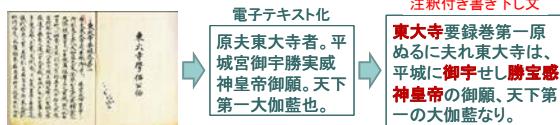


15



『東大寺要録』への注釈作業支援システム

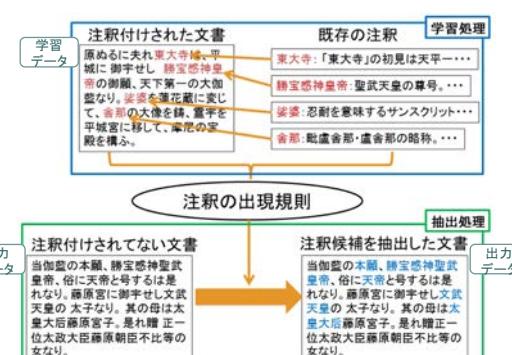
- 12世紀に成立したとされる歴史資料
- 東大寺の歴史と当時の状況などが記されている
- 全十巻のうち一巻(約2万字)が電子テキスト化
- 東大寺要録研究会が書き下しと注釈作業中



- 注釈作業には、複数の分野の専門知識が必要
- 専門分野によって知識に差がある

17

機械学習による注釈候補の自動抽出



● ● ●

異言語の浮世絵データベースからの同一作品の同定

- 浮世絵は、国内外の美術館・博物館で個々にデジタル化され公開
 - 同じ浮世絵作品が複数のデータベースに散在
 - データベースによってメタデータの言語・表記が異なる

19

国内外に散在する美術館・博物館の浮世絵情報に対して統合的なアクセスを実現

同作者や同画題等の関連情報へのリンクを動的に生成

利用者	江戸東京博物館	国立国会図書館	V&A ヴィクトリア&アルバート博物館
選択	葛飾北斎『東州江戻』	葛飾北斎『尾州不二見廻』	Suzuki Harunobu『Couple under umbrella in snow』
	喜川興宣『見返美人図』	葛飾北斎『朝靄快晴』	Lightnings below the summit』
	東洲斎写楽『大谷通次の江戸兵』	島居清信(2代目)『市川港老廻』	The Doll Ceremony』

● ● ●

「新古典籍総合目録データベース」のマルチリンガル化

- 国内外の日本研究者にとって基本的な古典資料の総合目録データベース
 - 英語版があるが、データは日本語のみ
 - 海外の利用者には使い難い
- ローマ字表記の自動分かち書き手法を開発中
 - 例:『喜多流装束付并ニ作物伝書』
kitaryūshōzokutsukerenarabini tsukurimonodensho
kitaryū shōzoku tsuke narabini tsukurimono densho

23

● ● ●

篆書体による蔵書印の文字認識および検索

- 篆書体:** 紀元前200年頃に作られた漢字の書体
 - 現代でも印鑑などに用いられる
- 蔵書印:** 所蔵を明示するために蔵書に捺した印影
 - 篆書体が多く使われる
- 当研究室で開発した「白川フォント」を学習データとし、機械学習による文字認識を行う

22

● ● ●

メディアプロジェクト演習2 作品制作のヒント(1)

- 「調査研究」**
 - 新しいWeb技術にはどのようなものがあり、それによって何が実現できるか
 - デジタル図書館システムとしてどのようなものが開発され、技術的課題は何か
- 「未来創造」**
 - 従来の図書館とデジタル図書館は、将来どのように共存していくべきか
 - デジタル情報を長期にわたって保存していくには、どのようにすればよいか

23

● ● ●

メディアプロジェクト演習2 作品制作のヒント(2)

- 「ソフトウェア」**
 - 簡単な情報検索システムの作成
 - メディア情報学実験2「情報検索」の発展
 - 検索結果のランキング・分類・可視化など
 - 簡単なデジタル図書館システムの作成
 - 各種WebサービスAPIの利用
 - 国立国会図書館サーチAPI
 - 図書館API(カーリル)
 - 楽天ブックス書籍検索API
 - 版元ドットコム・書誌情報API

24

デジタル図書館研究室について

- 現在のメンバー

- 前田亮 教授
- ビルゲサイハン・バトジャルガル 専門研究員
- 大学院 博士課程2名、修士課程8名
- 学部生14名

- 共同研究

- 本学東洋文字文化研究所、本学稻盛経営哲学研究センター、国文学研究資料館、国立歴史民俗博物館など

25

おわりに



- 当研究室では、**デジタル図書館・情報アクセス技術**に関する研究を幅広く行っている

- 研究室ホームページ

- <http://www.dl.is.ritsumei.ac.jp/>

- メールアドレス

- amaeda@is.ritsumei.ac.jp

26