

● ● ● | デジタル図書館研究室

研究室紹介  
2015/4/14 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？



- 「図書館」の定義を辞書で引くと...
  - 「図書・記録その他の資料を集め、整理・保管して、利用者の閲覧に供するための施設。」(大辞泉)
  - 「資料」を「収集・整理・保管」し「提供」する「施設」
- 「デジタル」の資料を扱う図書館
  - デジタル情報を「収集・整理・保管」し「提供」
  - インターネット上のサービスであり、「施設」があるわけではない



● ● ● | 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



主な研究テーマ



- Web情報**
  - 文書内容との関連性に基づくWikification
- 各種メディア**
  - 図形群の意味や階層構造を用いたプレゼンスライド検索システム
- 古典史料**
  - 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間における同一作品の同定

5

● ● ● | Web情報

6

## 文書内容との関連性に基づく Wikification

- インターネットの普及により、Web上でニュース記事を読む機会が増加
- 本文の中で読者にとって知らない語が出てくる
  - 記事の理解にさしつかえる場合がある

理化学研究所発生・再生科学総合研究センターと先端医療センターは12日、iPS細胞（人工多能性幹細胞）から膜の細胞を、目の難病「加齢黄斑変性」の患者に移植する臨床手術を行ったと発表した。

人工多能性幹細胞？  
加齢黄斑変性？  
?

7

## Wikificationとは

- 文書内の重要語（キーワード）に、それを説明する Wikipedia記事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

- 本研究では、日本語だけでなく、他の言語の該当記事へのリンクも実現

- 外国語学習者の読解支援

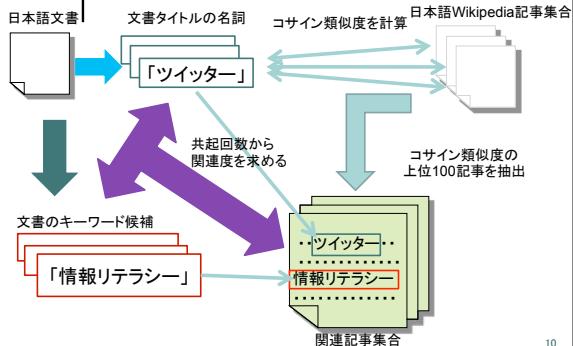
8

## 多言語 Wikipedia記事への言語横断 Wikification

文章中の用語と記事名が異なっていてもリンク可能

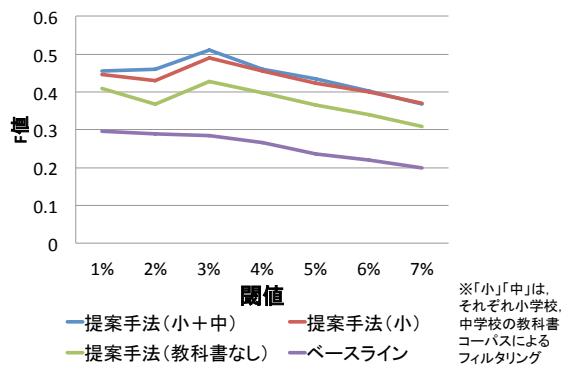


## キーワードと文書内容との関連度の計算



10

## キーワード抽出実験の結果



## 実際に抽出されたキーワード

記事名：「大統領権限で移民改革 500万人送還対象外に」  
赤字：アンケートによって選ばれたキーワード

順位	ベースライン	提案手法(小+中)
1	共和党	共和党
2	ホワイトハウス	民主党
3	中間選挙	ホワイトハウス
4	恩赦	大統領選
5	永住権	永住権
6	下院	市民権
7	野党	恩赦
8	上院	永住
9	オバマ	不法
10	井戸	オバマ

12

● ● ● | 各種メディア

13

● ● ● | 図形群の意味や階層構造を用いたプレゼンスライド検索システム

- プrezentationスライド
  - 視覚的にわかりやすいように図が多用
- 図の作成は手間がかかる
  - 図の再利用のニーズ
- 従来のスライド検索手法では、図自体の検索が困難




図の作成が面倒だなあ...  
以前作った図を再利用したい！

14

● ● ● | 形状でのグループ化

- 意味を持つと考えられる図形群をグループ化
  - 図形の形状や配置などの情報
  - 同領域で、同形状、近距離の場合
- グループ化された図形群の意味を推定する
  - 集合関係、重なり、包含関係、リスト(縦・横)

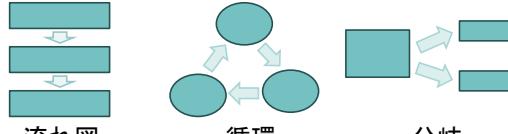


集合関係 重なり 包含関係 リスト(縦) 5

15

● ● ● | 矢印でのグループ化

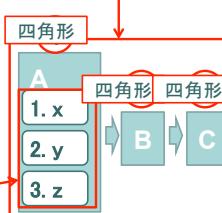
- 矢印图形が、どの図形群からどの図形群に流れることを示すか推定し、グループ化する
- グループ化された図形群の意味を推定する
  - 流れ図、循環、分岐、集中、拡散、相互関係



流れ図 循環 分岐 16

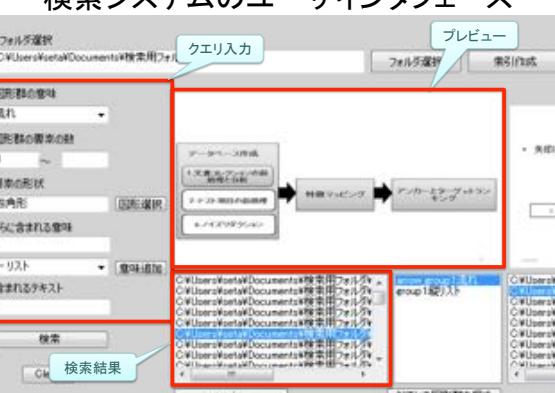
● ● ● | 入力クエリ

- 図形群の意味
  - 「流れ」
- 図形群の要素の数
  - 3(「A」, 「B」, 「C」)
- 図形群の要素の形状
  - 四角形
- 下の階層の図形群の意味
  - リスト(「1」, 「2」, 「3」)
- スライドに含まれるテキスト



17

検索システムのユーザインターフェース



18

**検索の例**

**入力ケイリ**  
図形群の意味: **重なり**  
図形群の要素の数: **4**  
図形群の要素の形状: **楕円**

候補	スコア
1号	100
2号	100
3号	95
4号	95
5号	85
6号	83
7号	80
8号	80
9号	80
10号	80

次に10件

**応用例:類似図形群の検索**

**↓類似図形群↓**

候補	スコア
1号	0
2号	25
3号	10
4号	16.666666
5号	17.516666
6号	19.166666
7号	19.166666
8号	19.166666
9号	19.166666
10号	21.666666

次に10件

**古典史料**

**古典史料テキストからの情報抽出**

役者評判記には、歌舞伎役者の演技や容姿に対する評価(レビュー)が書かれおり、役者名の変遷や、位付と評価文との関係など、さまざまな分析が可能

- 位付(役者の評価)
- 役者名
- 評価文

役者絵「古今俳優」(左)、役者評判記『役者大極舞』(右)  
立命館大学アート・リサーチセンター (<http://www.arc.ritsumei.ac.jp/index.html>) 所蔵

**機械学習による注釈候補の自動抽出**

**学習データ**  
人手で注釈が付けられた文書  
既存の注釈

**入力データ:**  
注釈がまったく付けられていない文書

**出力データ:**  
入力文書から、注釈の候補を自動的に抽出

既存の注釈: 東大寺: ..., 勝宝感神: ..., 姿婆: ..., 舎那: ...

規則抽出: 注釈文字列の出現規則を機械学習

出力: 当伽藍の本願、勝宝感神聖武天皇、俗に天帝と号するは是れなり。藤原宮に御宇せし文武天皇の太子なり。其の母は太皇后藤原宮子。是れ贈正一位太政大臣藤原朝臣不比等の女なり。

**異言語の浮世絵データベース間に  
おける同一作品の同定**

- 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

大英博物館  
オーストラリア国立工芸美術館  
江戸東京博物館  
ホノルル美術館  
メトロポリタン美術館

注: 本ページの画像はすべて  
メトロポリタン美術館  
([www.metmuseum.org](http://www.metmuseum.org))  
のものを使用

## 研究目的

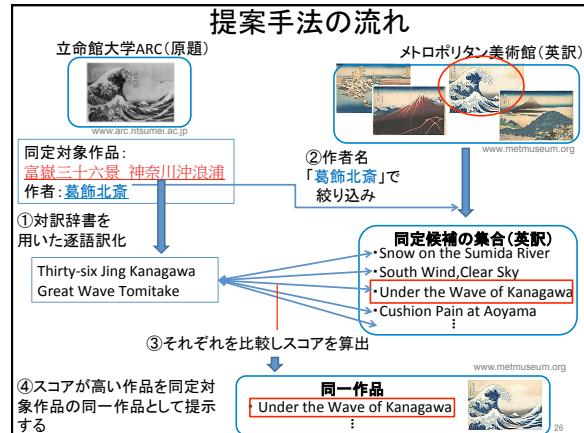
- 浮世絵研究者からの要望
  - データベース間で散在している同一作品を見つける
  - 同一作品のメタデータを比較することにより、誤りの発見・修正、より詳細な情報の取得などが可能となる

「富嶽三十六景 凱風快晴」  
絵師: 北斎  
立命館大学アート・リサーチセンター  
<http://www.arc.ritsumei.ac.jp/index.html> 所蔵

「South Wind, Clear Sky (Gaifu kaisei), also known as Red Fuji...」  
Artist: Katsushika Hokusai  
[www.metmuseum.org](http://www.metmuseum.org)

本研究では原題と英訳を用いた同一作品の同定手法を提案

25



## 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

27

## 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

28

## 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
  - 大学院修士課程5名
  - 学部生11名
- 共同研究
  - 本学MOT研究科、本学文学部など
- 研究成果(2014年度)
  - 国際会議6件、国内会議8件、受賞2件

29

## 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わ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@media.ritsumei.ac.jp](mailto:amaeda@media.ritsumei.ac.jp)

30