この資料は、クリエイティブ・コモンズ 表示 4.0 国際ライセンスの下に提供されています。

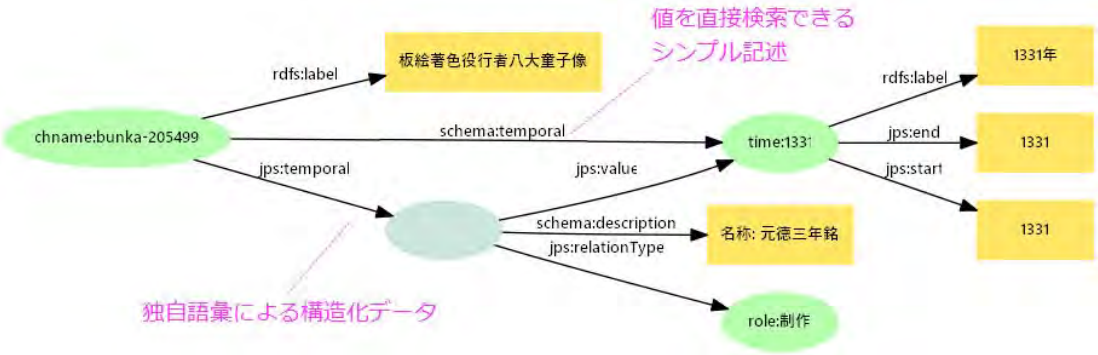
利活用スキーマの利活用

1. 利活用スキーマの基本モデル
2. 「いつ」でつなぐ
3. 「どこ」でつなぐ
4. 「だれ」でつなぐ
5. 「なに」でつなぐ
6. ジャパンサーチ内の新しい関係を見出す
7. クエリ結果を応用したコンテンツ
8. データとしての利活用スキーマRDF
9. ジャパンサーチを超えて
10. 参照先

ジャパンサーチ連携促進説明会
2019-07-17 神崎正英

1 利活用スキーマの基本モデル

- **使いやすさと詳細さの両立**
  - 「いつ」「どこ」「なに」「だれ」に絞った正規化
  - schema.org語彙によるシンプル検索
  - 独自語彙による役割区分と元データ保存



- **アクセス情報とソース情報**
  - 記述対象へのアクセス情報を集約
  - 提供データ（つなぎ役）に関する情報

2 「いつ」でつなぐ

■ 時間表記の正規化と実体化

- 「元徳三年銘」→ `time:1331`（文字列ではなく時間実体）
- 時代や曖昧表記を時間範囲として実体化
- 時間実体を利用して、異なる記述のアイテムを検索。例：鎌倉時代の絵画
  - 「鎌倉時代」だけでなく、その時代範囲の西暦、和暦記述も検索できる。

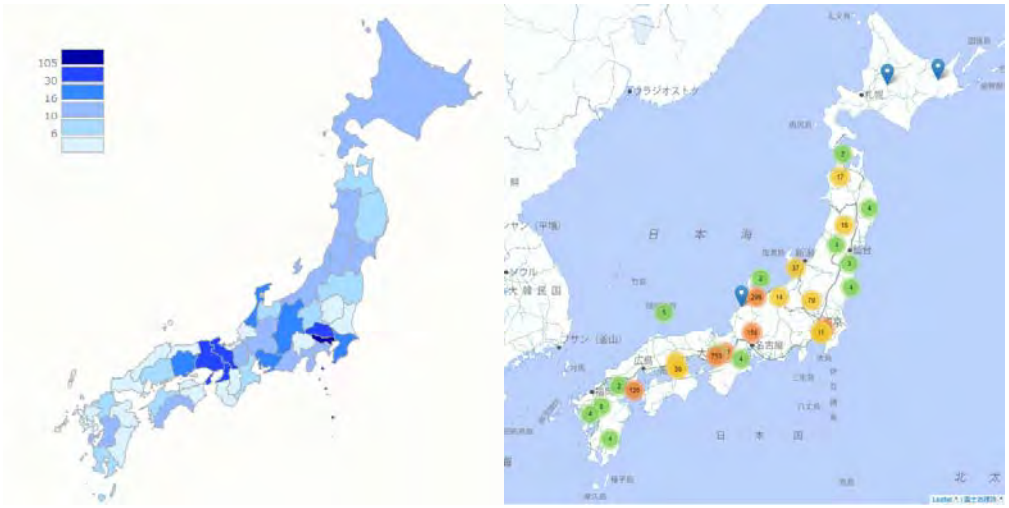
```
SELECT ?s ?start ?end WHERE {
    {time:鎌倉時代 jps:start ?st ; jps:end ?et}
    ?s a type:絵画 ;
        schema:temporal [jps:start ?start; jps:end ?end ] .
    FILTER (?start >= ?st && ?end <= ?et)
}
```

- さらに年表上への表示[1]も

3 「どこ」でつなぐ

■ 場所の正規化

- 場所記述は粒度がさまざま（国単位～街区住所）。旧国名や市町村合併→現在使わない地名
- 都道府県を単位に正規化（海外は国単位）。
- 正規化の結果、検索結果を都道府県別に集計して地図を色分け[2]できる
- さらに歴史地名データ[3]などを用い、可能なものは緯度経度を追加



4 「だれ」でつなぐ

■ 別名とさまざまな表記

- 例えば歌川国芳は一勇斉國芳とも号し、本名の井草孫三郎でも記される
- 「筆」「画」などの接尾辞も含めるとジャパンサーチで歌川国芳は37通りの記述
- マッピング辞書を用いてchname:歌川国芳に正規化、さらにWikidata、DBpediaなどのLODハブともリンク

■ LODハブとMLA統合検索

- SPARQLエンドポイントを提供するMLAの多くが、それぞれの作者典拠をLODにリンク
- 共通のLODリンクを利用した統合検索[4]が可能
- 作品記述のRDFモデルは各機関それぞれ異なるので、クエリは個別設計が必要
  - 次の例はWikidataへの統合クエリ

```
?wid owl:sameAs chname:歌川国芳 ;
    rdfs:isDefinedBy <http://www.wikidata.org/>.
SERVICE <https://query.wikidata.org/sparql> {
    ?uri wdt:P170 ?wid ; rdfs:label ?label .
        FILTER (lang(?label)="en")
    OPTIONAL{?uri wdt:P18 ?image}
}
```

5 「なに」でつなぐ

■ タイプと階層検索

- タイプは元データの「種別」などから生成 → データセットによって異なる
- クラスを階層化：「版画」「水彩」は「絵画」の一種
- クラス階層を利用して例えば神奈川県の絵画＆サブクラスを検索（**プロパティパス**）

```
?type rdfs:subClassOf* type:絵画
```

■ 主題とキーワード

- 分類記号、件名標目、自由キーワードをURI化し、いずれも`schema:about`
  - それぞれ名前空間URIは異なる
- 分類記号はNDCの階層とNDLSHの関係を利用して`rdfs:seeAlso`による関連付けとラベルを内部的に設定
  - クエリを簡単にするため、`broader`、`related`を区別せず、全て`seeAlso`でリンク
- 関連付けを使って例えば伝説に関係するものを検索

```
?s schema:about ?subj .
?subj rdfs:seeAlso*/rdfs:label "伝説"
```

- これらはプロパティパスを用いた一種の推論（元のグラフに直接記述されていない関係の利用）でもある

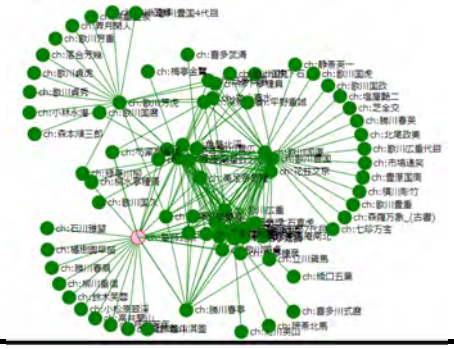
6 ジャパンサーチ内の新しい関係を見出す

■ 作者関連グラフの推論

- アイテムにはしばしば複数寄与者 → これらの寄与者の間には**何らかのつながり**がある。

```
SELECT ?someone (count(?s) as ?count) WHERE {
    ?s jps:agential
        [jps:relationType/skos:broader? role:制作;
            jps:value/owl:sameAs? chname:歌川広重 ],
        [jps:relationType/skos:broader? role:制作; jps:value ?someone ]
    FILTER(?someone != chname:歌川広重)
} GROUP BY ?someone
```

- つながりのある寄与者を集めることで、作者関連グラフ[5]が得られる。
  - CONSRUCT句で関係を表すRDFグラフ生成もできるがここではつながりの数も考慮して集計数を取得し、アプリケーション側で動的にグラフ作成



7 クエリ結果を応用したコンテンツ

■ 参考文献リストからIIIFコレクション生成

- NDLデジタルコレクションを活用した『まいボコ』[6]（通称）の参考文献リストを一括検索

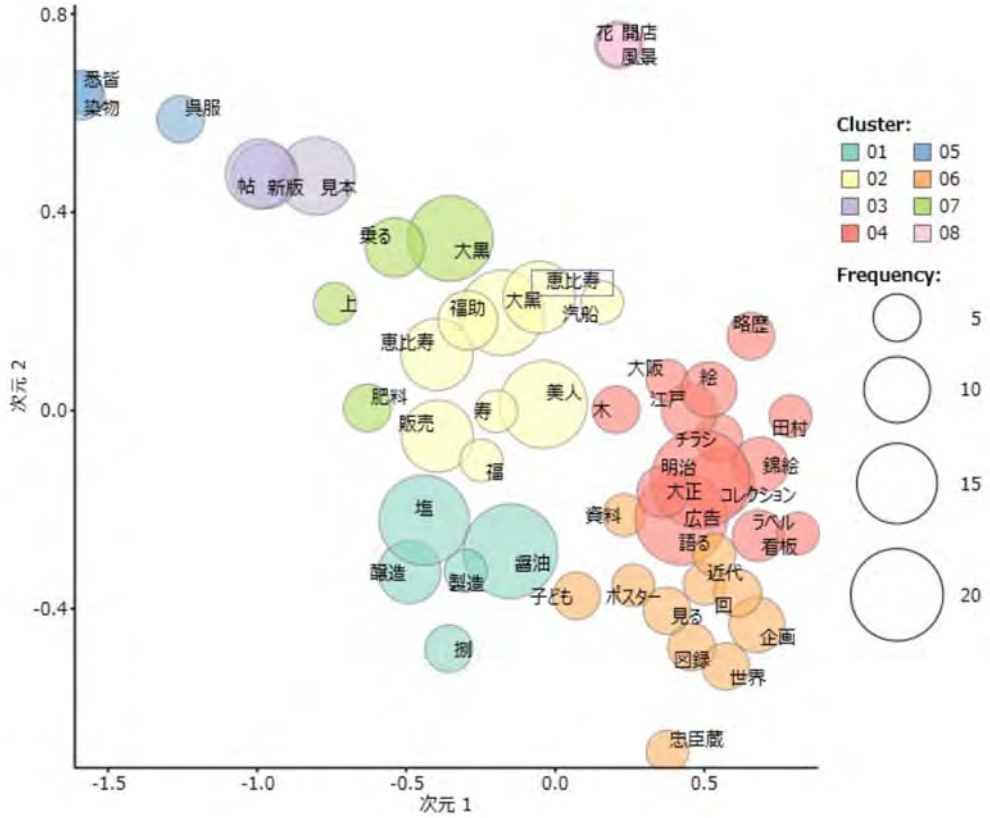
```
SELECT * WHERE{
    VALUES ?title {"豪傑荒尾龍之助" "剣道名人那珂一石斎" ...}
    ?s rdfs:label ?title ;
        schema:creator ?creator ;
        jps:accessInfo/schema:provider chname:NDLデジタルコレクション .
    OPTIONAL {?s schema:image ?image }
}
```

  - 150近くあるので、全て一括ではクエリURIが長すぎてエラー → 章ごとに実施
  - 文献リストとJPSの表記が少しでも違うと完全一致しないため、打率は6〜8割
- 不足分を手動検索し、文献を閲覧できるIIIFコレクションを生成[7]

8 データとしての利活用スキーマRDF

■ メタデータのテキスト分析

- キーワードをタイトルに含むアイテムを年別集計しグラフ化（例：猛烈,根性,感性,努力[8]）
- タイトル抽出 → KH Coder[9]で分析。図は「引札」を含むタイトルの多次元尺度法プロット



9 ジャパンサーチを超えて

■ 利活用モデルをオープンデータに適用

- 簡単な定義を用意すれば、CSV/TSVデータからジャパンサーチ型RDFを生成[10]できる
- オープンデータとして提供されている美術館データを変換。例：メトロポリタン美術館[11]、クリーブランド美術館館[12]
- 国立国会図書館の日本関係外国語図書の書誌情報（試行版）[13]

■ Wikidataからジャパンサーチを使う

- ジャパンサーチのSPARQLエンドポイントに対して、外部から統合クエリを送ることも可能
- オックスフォード大ボドリアン図書館の招聘ウィキメディアンであるポールター氏のブログ記事[14]でWikidataからのクエリ紹介
- 例：ジャパンサーチにあるイギリスの彫刻家の作品をWikidataから検索[15]

■ 利用者による情報の追加

- ジャパンサーチを核として、ユーザが更に情報を追加できればより豊かなものになり得る
- Wikiのような形？ → ジャパンサーチ自身がホストするのは、現状では難しいか
- 外部から注釈を加える（例：相撲絵の場合[16]） → 注釈を集約する第三者サービスをぜひ！

10 参照先

■ 参照したリソース

1. アイテムの年表表示, (非公式サポート)
   <https://www.kanzaki.com/works/ld/jpsearch/timeline>
2. アイテム分布の地図グラフ, (非公式サポート)
   <https://www.kanzaki.com/works/ld/jpsearch/jpmapchart>
3. 歴史地名データ, 人間文化研究機構
   <https://www.nihu.jp/ja/publication/source_map>
4. MLAの統合検索＝歌川国芳の例, (非公式サポート)
   <https://www.kanzaki.com/works/ld/jpsearch/fedquery?who=歌川国芳>
5. 作者の関連グラフ, (非公式サポート)
   <https://www.kanzaki.com/works/ld/jpsearch/agent-relation>
6. 「舞姫」の主人公を…ボコボコにする最高の小説…紹介する本, by 山下泰平, 2019-04, 柏書房
   <http://www.kashiwashobo.co.jp/book/b452226.html>
7. 『まいボコ』参考文献のIIIFコレクション
   <https://www.kanzaki.com/works/2016/pub/image-annotator?u=/works/2019/annot/myboco-collection.json>
8. キーワードの年別出現数グラフの例, (非公式サポート)
   <https://www.kanzaki.com/works/ld/jpsearch/kwitem-frequency?kw=猛烈,根性,感性,努力>
9. KH Coder: 計量テキスト分析・テキストマイニングのためのフリーソフトウェア
   <https://khcoder.net/>
10. CSVからジャパンサーチ型RDFを生成する, (非公式サポート)
    <https://www.kanzaki.com/works/ld/jpsearch/jps-generator>
11. メトロポリタン美術館RDF：ジャパンサーチモデル版, (非公式サポート)
    <https://www.kanzaki.com/works/ld/jpsearch/metmuseum>
12. クリーブランド美術館RDF：ジャパンサーチモデル版, (非公式サポート)
    <https://www.kanzaki.com/works/ld/jpsearch/clevelandma>
13. 日本関係外国語図書の書誌情報：ジャパンサーチモデル版, (非公式サポート)
    <https://www.kanzaki.com/works/ld/jpsearch/books-on-japan>
14. Welcome Japan Search to the web of Linked Open Data, by Martin Poulter, 2019-05-30
    <http://blogs.bodleian.ox.ac.uk/digital/2019/05/30/welcome-japan-search-to-the-web-of-linked-open-data/>
15. ジャパンサーチにあるイギリスの彫刻家の作品, Wikidataクエリ
    <https://w.wiki/4cW>
16. 相撲絵の場合：ジャパンサーチとWeb Annotation, (非公式サポート)
    <https://www.kanzaki.com/works/ld/jpsearch/annot-sumo>