Linked Data Driven Dynamic Web Services for Providing Multilingual Access to Diverse Japanese Humanities Databases

Biligsai khian Batjargal
Takeo Kuyama, Fuminori Kimura, and Akira Maeda

Digital Library Laboratory
Ritsumeikan University, Japan
A Federated Search System for Japanese Ukiyo-e Prints (FeSSU)

**Ukiyo-e:** Japanese traditional woodblock printing, a popular art during the 17th - 20th centuries
The FeSSU Concept

Accessing and searching across multiple resources simultaneously in real-time

A federated search system

### Ukiyo-e Databases

- **National Diet Library Digital Archive Portal**
  - approx. 10,000 Ukiyo-e prints

- **The Edo-Tokyo Museum**
  - approx. 6,100 Ukiyo-e prints

- **Ukiyo-e Database of Art Research Center, Ritsumekan University**
  - approx. 12,500 Ukiyo-e prints

- **The Museum of Fine Arts, Boston**
  - approx. 57,000 Ukiyo-e prints

- **The Victoria and Albert Museum**
  - approx. 38,000 Ukiyo-e prints

- **The British Museum**
  - approx. 15,000 Ukiyo-e prints

- **The Metropolitan Museum of Art, New York**
  - approx. 4,000 Ukiyo-e prints

- **The Boston Museum of Fine Arts**
  - approx. 57,000 Ukiyo-e prints

- **The Museum of Art and Archaeology, University of Oxford**
  - approx. 640 Ukiyo-e prints

- **The Art Institute of Chicago**
  - approx. 2,600 Ukiyo-e prints

- **The Library of Congress**
  - approx. 2,740 Ukiyo-e prints

- **Diverse data in Japanese**

- **Diverse Web databases (museums, online collections, etc)**
Diverse data

- In different languages
  - How to find information across languages
  - that not familiar with
  - Query in other language
- Deal with diverse metadata for searching
- Different representations or usage
  - Artist name: Danjûrô, Danjûrō, Danjuurou, Danjuro, Danjyurou
  - Ukiyo-e, Ukiyoe, Woodcuts, Color prints

Linked Open Data (LOD) resources

- Re-use of LOD in multilingual information retrieval applications
  - diverse resources in various languages
The Overall Research Goal

- **Linked Data** driven approach in providing integrated **multilingual access** to diverse Japanese databases **dynamically**
  - by **linking** and **re-using** LOD resources **dynamically**
  - **name authority**, **subject headings**, links to other **Linked Data resources**: DBpedia, etc
  - access to extra data **not only** in **Japanese** databases but also **multilingual** databases in other countries
  - **not depending on languages**

Conventional multilingual Web databases (museums, online collections, etc)
Access Across Databases

Database 1 (in Japanese)

Author: 葛飾北斎 (Katsushika Hokusai)
Title: 尾州不二見原 (Fujimi Fuji view field in the Owari province)

Database 2 (in Japanese)

Author: 葛飾北斎 (Katsushika Hokusai)
Title: 凱風快晴 (South Wind, Clear Sky)

Database 3 (in English)

Author: Katsushika Hokusai
Title: Lightning below the summit
Unique Features of our Approach

1. **Not collecting** data from each database in advance
   - Links are generated **dynamically** from the search results when a user performs a search
   - Always **up-to-date data**

2. **Not only monolingual** (e.g. Japanese) databases
   - but also **multilingual databases worldwide**

3. **Utilization of authority data**
   - showing **additional relevant information** for a given query input
Utilized Resources

- **Name authorities & Subject headings**
  - VIAF: The Virtual International Authority File
  - LCSH: Library of Congress Subject Headings
  - NDLSH: Japanese National Diet Library Subject Headings - Web NDL Authorities

- **Linked Open Data**
  - Databases of Ukiyo-e Prints
  - The thesaurus of Japanese Art
  - Japanese Museum LOD (LODAC)
  - Japanese DBpedia
  - Already available links
  - Newly generated links
Utilization of Authority Data (1)

Advantages

- Unifying **different notations** and synonyms
- **Distinguishing** individual names
  - even if **different names** are referring to the same person
- Obtaining **additional information**
  - different **notations** and **synonyms**
The Reasons

- Simple **string matching** is not viable
  - matching of **different representations**
- Databases in **different languages** could be **accessed** using the reading
  - **phonetic representation** of Japanese kanji
  - **A romanized representation (romaji)** in Western databases
- Could contain **links** to DBpedia as Linked Data
Connecting to LOD Resources

Authority data (NDL Authorities, VIAF)
- foaf:name
  - 安藤広重 (Hiroshige Ando)
- rdaGr2:dateOfBirth
  - “1797”

A search result

A returned record
- <Author>
  - 歌川広重 (Hiroshige Utagawa)
- <Title>
  - 琉球人来貢図巻 (Ryukyuan people are paying tribute)
- <Published year>
  - 1807

DBpedia (English)
- URI of “歌川広重”
- URI of Hiroshige Utagawa
- URI of dbpedia:Painting
- dbpedia-owl:field
- dbpprop:placeOfBirth
- URI of dbpedia:Edo

owl:sameAs

Japanese DBpedia
- URI of “歌川広重”

Other LOD resources in Japanese
- URI of ‘a work of art’
- “叙情的風景版画 (Lyrical Landscape prints)”

Display additional relevant information

User
An Example of Cross-language Access (1)

Query: “浮世絵” (ukiyo-e)

Web NDL Authorities

Synonym (xl:altLabel)

江戸絵 (Edo-e)

錦絵 (Nishiki-e)

横浜絵 (Yokahama-e)

LCSH

Hashirae

Pillar Prints

Ukiyo-e

Ukiyoe

Query: “浮世絵”

Japanese National Diet Library Subject Headings (Contents are mainly in Japanese)

Library of Congress Subject Headings (Contents are mainly in English)
An Example of Cross-language Access (2)

View of Mount Fuji from Satta Point in the Suruga Bay
Artist: **Utagawa Hiroshige**

Japanese National Diet Library Subject Headings
**歌川, 広重 1世, 1797-1858**

<table>
<thead>
<tr>
<th>ID</th>
<th>00222450</th>
</tr>
</thead>
<tbody>
<tr>
<td>標目 (shareLabel)</td>
<td><strong>歌川, 広重 1世, 1797-1858 (ウタガワ, ヒロシゲ, 1797-1858)</strong></td>
</tr>
<tr>
<td>別名 (shareLabel)</td>
<td><strong>安藤, 広重 (アンドウ, ヒロシゲ); 一立巌, 広重 (イチリュウサイ, ヒロシゲ); 安藤 (姓), 一立巌 (姓)</strong></td>
</tr>
</tbody>
</table>

an alias **Ando Hiroshige**
an alias **Ichiryusai Hiroshige**

Library of Congress Subject Headings
**Andō, Hiroshige, 1797-1858**

- **Variants**
  - Hiroshige I, 1797-1858
  - Utagawa, Hiroshige, 1797-1858
  - Ichiryūsai Hiroshige, 1797-1858
  - Ryūsai, 1797-1858
  - Ichiryūsai Hiroshige, 1797-1858
  - Ichiro Hiroshige, 1797-1858
  - Khirosige, 1797-1858
  - Andō, Tokitarō, 1797-1858
  - Andō, Tokutarō, 1797-1858
  - Andō, Jūemon, 1797-1858
  - Yūsai, 1797-1858
  - Andō, Tokubei, 1797-1858
  - Utagawa, Hiroshige, 1797-1858
  - 安藤広重, 1797-1858
  - 安藤広重, 1797-1858
  - 安藤広重, 1797-1858
  - 歌川広重, 1797-1858
  - 歌川広重, 1797-1858

The Virtual International Authority File

**Andō, Hiroshige, 1797-1858**

VIAF ID: 19678928 (Personal)
Permalink: http://viaf.org/viaf/19678928

4xx's: Alternate Name Forms (119)

- 400 0 _a Ando Hiroshige _d 1797-1856
- 400 1 _a Ando Hiroshige
- 400 1 _a Ando Hiroshige
- 400 1 _a Ando Hiroshige, _d 1797-1856
- 400 1 _a Ando Hiroshige
- 400 0 _a Andō Hiroshige _d 1797-1856
A user performs a search. When an author name is chosen, the system searches for Linked Data and it displays additional relevant information.
| Katsushika, Beikasai, active 19th century |
| Katsushika, Genryūsai, active 19th century |
| Katsushika, Hokusai, active 19th century |
| Katsushika, Taito, 19th cent. |
| Katsushika, Taito, active 19th century |
| Katsushika 戴斗, 19th cent. |
| Katsushika 戴斗, active 19th century |
| Doo mounting Katsushika, active 19th century |
| Der Mensch, der vom Zeichnen besessen ist |
| Der vom Malen Besessene |
| Gakyō Rōjin Manji |
| Gakyōjin |
| Gwakyōjin |
| Hishikawa, Sōri |
| Hokusai* |
| Hokusai |
| Hokusai Tetsuzō |
| Hokusai Tokimasa |
| Hokusai Tokitarō |
| Hokusai, Katsushika |
| Hokusai, Katsushika |
| Hokusai, Katsushika ( |
| Hokusai, Katsushika Japanese printmaker and pai |
| Hokusai, Tokimasa |
| Hokusai, Tokitarō |
| Hokusai. |
| Tokitarō Kakō |
| Zen Hokusai |
| Zenhokusai |
| Zenhokusai, Manjirōjin |
| Kanusika, Hokusai |
| Hokusai |
| Katsukawa, Shunrou |
| Katsushika, Hokusai |
| Kyo wax moth Zinnman page |
| Zenhokusai |
| Zenhokusai, Manjirōjin |
| Tokitarou, crater |
| Hokusai |
| Hokusai, Tokimasa |
| Nakajima Tameiichi |
| Before Hokusai |
| Hokusai, swastika old man before |
| KATSUKAWA spring Akira |
| KATSUKAWA, spring Akira |
| Hokusai |
| Hokusai Tokitaro |
| Hokusai Tetsu Kura |
| Hokusai, Dragon government |
| Yes climate |
| Sori |
| 戴斗 |
| Tokitaro |
| Tokitaro, possible climate |
| Tameiichi |
| Kondo Ichitaro |
| Tetsuzo |
| http://dbpedia.org/resource/Edo |
| http://dbpedia.org/resource/Edo |
| http://ja.dbpedia.org/resource/ Edo |
| http://dbpedia.org/resource/Hiroshige |
| http://ja.dbpedia.org/resource/Utagawa Hiroshige |
| http://dbpedia.org/resource/Kan% C5% 8D_Masanobu |
| http://dbpedia.org/resource/Katsukawa_Shunsh% C5% 8D |
| http://ja.dbpedia.org/resource/KATSUKAWA spring chapter |
| Hokusai |
| KATSUSHIKA, Hokusai |
| Katsushika Hokusai |
| Hokusai |
| Katsushika Hokusai |
| Katsushika Hokusai |
| 1849-05-10 |
| 1760-10-31 |
| Katsushika Hokusai |
| http://ja.dbpedia.org/resource/ Impressionism |
| http://ja.dbpedia.org/resource/ Utagawa Kuniyoshi |
| http://ja.dbpedia.org/resource/ Utagawa Hiroshige |
| 1760-10-31 |
| http://ja.dbpedia.org/resource/ KATSUKAWA spring chapter |
| http://ja.dbpedia.org/resource/ Kitao Masami |
| 1849-05-10 |
| "36 Views Fugaku," "Hokusai Manga", "Ama and octopus" |
### Evaluation

**4088 randomly chosen records**

- **manually checks** whether the data retrieved using the aliases are **relevant** to the **artist** in the original record or not

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Number of chosen records</th>
<th>Number of correct records</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4088</td>
<td>3397</td>
<td>83.09%</td>
</tr>
<tr>
<td>Sole names (surname, pen name or honored title only)</td>
<td>2179</td>
<td>1492</td>
<td>68.47%</td>
</tr>
<tr>
<td>Full names only</td>
<td>1909</td>
<td>1905</td>
<td>99.79%</td>
</tr>
</tbody>
</table>

**FeSSU**

- **Authority data**
- **Randomly chosen records in the search results**
- **Manually checking the matching records**
- **Newly returned records**

**Search results**

**Search by aliases**

**Precision =** \[rac{\text{Number of correct records}}{\text{Number of chosen records}}\]
Summary

Linked Data driven dynamic approach

- Providing multilingual access to multiple Japanese Humanities Databases
  - Real-time parallel search on diverse databases
  - a single query input in a single user-interface
  - Useful for humanities researchers

- Utilizing Linked Data Resources + Authority Data
  - multilingual access to multiple databases
    - by generating links between databases dynamically
    - across languages, not in her/his mother tongue
  - Showing additional relevant data from other databases
    - generating links dynamically from the search results of multiple databases using Linked Data resources
Further improvements

- **Decent precision - 99.79%**
  - for finding additional information using the aliases
  - full names retrieval

- **68.47% precision**
  - short string query e.g. sole names (surname, pen name or honored title)

- **Retrieval in other languages**
  - English, Japanese, and more
  - Europeana, the Pushkin State Museum of Fine Arts (Russian), etc

- **Humanist-oriented evaluation**
  - Their demand
Thank you for your attention!
Questions or Comments?