Providing Bilingual Access to Multiple Japanese Humanities Databases: Text Retrieval Using English and Japanese Queries

Bilgsaikhan Batjargal
Akira Maeda, and Ryo Akama
Ritsumeikan University, Japan
2015.12.01
Background
The overall goal
The proposed Federated Search System
  - The main features
  - Current implementation/application
  - Query translation approach
  - Resources that can be utilized
  - Examples of Cross-language access
  - Connecting to Linked Data resources
  - Integrating the search results
Summary
Many **Japanese collections** and resources in libraries, museums, and research institutes are **digitized**

### Art Research Center, Ritsumeikan University
- Kyoto-based Japanese **arts and cultures**
- Digital archives of intangible culture and temporal arts
- **New interactive image culture**
  - Since **1998**, based in Kyoto
- **Over 40 databases**
  - Ukiyo-e, books, kabuki, shunga etc

1257.6m², 3-storey facilities
The overall goal

- To provide integrated **multilingual parallel access** to diverse Japanese databases **dynamically**

- Apply to the databases at the Art Research Center of the Ritsumeikan University
  - Digitized of Japanese Art and Cultures

**Conventional Web databases** (museums, online collections, etc)

- **On-the-fly**

**Linked Data + Web resources** (Authority data, DBpedia)

2015 DADH
① **Federated search** (aka *distributed information retrieval*)
- Accessing or searching across multiple resources simultaneously in one search box in real-time
- Search/Retrieve Web service (SRW), Search/Retrieve via URL (SRU) technologies

② **Metadata mapping**
- A simple metadata schema similar to Dublin Core Metadata Element Set (The Dublin Core Metadata Initiative. 2010)
  - `<title>`, `<author>`, `<subject>`, `<date>`, `<description>`, `<publisher>`, `<type>`, `<identifier>`, `<source>`, `<coverage>` and `<rights>`
- merging, unifying, normalizing returned results into a single list
The main features of the proposed system

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 databases
   - but also **multilingual databases worldwide**

3. Utilization of authority data and **Linked Data**
   - showing **additional relevant information** for a given query input
Current implementation (1)

- ARC Ukiyo-e Portal Database
- ARC Early Japanese Book Database
- ARC Japanese modern book database

The proposed Federated Search System

ARC DBs

- ARC Ukiyo-e Portal Database
- ARC Early Japanese Book Database
- ARC Japanese modern book database

FileMaker driven
Current implementation (2) - DBs

- **ARC Ukiyo-e Portal Database**
  - approx. 28,000 prints

- **Art Research Center (ARC), Ritsumeikan University**

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

- [http://www.dh-jac.net/db/nishikie-e/search_portal.php](http://www.dh-jac.net/db/nishikie-e/search_portal.php)
Current implementation (3) - DBs

- ARC Early Japanese Book Database
  - approx. 16,000 books
  - Art Research Center (ARC), Ritsumeikan University
  - Browsing rare books and early Japanese books
    - Page by page

- http://www.dh-jac.net/db1/books-e/search_portal.php
Current implementation (4) - DBs

- ARC Japanese modern book database
  - Approx. 400 books
- Art Research Center (ARC), Ritsumeikan University
- Browsing books since the Meiji Era - 1868
  - Page by page

- http://www.dh-jac.net/db1/mbooks/search.php
Query translation approach

- Cross-language information retrieval (CLIR)
  - A dictionary based query translation techniques
- Send original query and its translation into the same database simultaneously
  - Bilingual queries over Japanese databases
  - Merge the retrieved results

An user query

Original query

Query translation

Translated query

A database

Multilingual/bilingual search results

Merge the search results
Resources that can be utilized

- 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

![Diagram showing interconnected databases and relationships between VIAF, LCSH, NDLSH, Japanese Museum LOD (LODAC), DBpedia, The thesaurus of Japanese Art, and Japanese DBpedia.](image)

- Already available links
- Newly generated links
An example of Cross-language access using Authority Data

Artist: **Utagawa Hiroshige**

Search by SPARQL

Japanese National Diet Library Subject Headings

歌川, 広重 1世, 1797-1858

Library of Congress Subject Headings

Andō, Hiroshige, 1797-1858

The Virtual International Authority File

Andō, Hiroshige, 1797-1858

Hiroshige, 1797-1858

Utagawa, Hiroshige, 1797-1858

Andō, Hiroshige

Ichiryusai Hiroshige

Utagawa Hiroshige

An example of Cross-language access using Authority Data

2015 DADH
An example of Cross-language access using Subject Headings

**Query:** “浮世絵” (Ukiyo-e)

- Web NDL Authorities
  - Synonym (xl:altLabel)
- LCSH
  - Ukiyo
  - Hashirae
  - Pillar Prints
  - Ukiyo-e

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

Library of Congress Subject Headings
(Contents are mainly in English)
Hiroshige

Utagawa Hiroshige (Japanese: 歌川 広重), also Andō Hiroshige (Japanese: 安藤 広重; 1797 – 12 October 1858) was a Japanese ukiyo-e artist, considered the last great master of that tradition.

Wikipedia
Connecting to Linked Data resources

Authority data (NDL Authorities, VIAF)
foaf:name
歌川広重 (Hiroshige Ando)
drGr2:dateOfBirth “1797”

owl:sameAs
URI of dbpedia:Painting
URI of dbpedia:Edu

DBpedia(English)
URI of "歌川広重"

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

owl:sameAs
URI of Hiroshige Utagawa

Japanese DBpedia
URI of “歌川広重”

Other LOD resources in Japanese
URI of ‘a work of art’

Display additional relevant information

2015 DADH
A user performs a search.

When an author name is chosen, the system searches for Linked Data and it displays additional relevant information.
Integrating the search results (1)

- Snippets with the original links and thumbnails
- Translation or reading (romaji) in English page view
- Kanji or reading (hiragana) in Japanese

2015 DADH
Integrating the search results (2)
61. **Whale hunting at the island of Goto in Hizen.** Utagawa Hiroshige., 1859.

**Title transcription:** Hizen gotō kujiraryō no zu  
**Authors:** Utagawa Hiroshige;  
**Original records:** [Link] [Link]

**Information:** Whalers armed with harpoons hunting whales  
- Title and other descriptive information compiled by Nichibunken-sponsored Edo print specialists in 2005-06  
- Format: vertical Oban Nishiki  
- From the series: Shokokumireisho hyakkei: One hundred famous views of the country  
**Edition:** 1859 [Link], [Link]


**Title transcription:** Yamato yoshinoyama  
**Authors:** Utagawa Hiroshige;  
**Original records:** [Link] [Link]  

63. **Choshi Bay in Shimosa Province; Famous Places in the Provinces (Woodblock print).** Utagawa Hiroshige., 1854-1858.

**Title transcription:** Choshi Bay in Shimosa Province; Famous Places in the Provinces (Woodblock print)  
**Authors:** Utagawa Hiroshige;  
**Original records:** [Link]
1. ひぜん ごとう くじらりょう のす。うたがわ ひろしげ。1859.
   - Title transcription: ひぜん ごとう くじらりょう のす
   - Authors: うたがわ ひろしげ
   - Original records: [Link] [Link]

   Information: Whalers armed with harpoons hunting whales
   - Title and other descriptive information compiled by Nichibunken-sponsored Edo print specialists in 2005-06
   - Format: vertical Oban Nishiki
   - From the series: Shokokumreisho hokkei : One hundred famous views of the country

   Edition: 1859; [Link], [Link];

2. やまと よしのやま。うたがわ ひろしげ。1859。
   - Title transcription: やまと よしのやま
   - Authors: うたがわ ひろしげ
   - Original records: [Link] [Link]

3. Choshi Bay in Shimosa Province; Famous Places in the Provinces (Woodblock print)。うたがわ ひろしげ。1854-1858。
   - Title: Choshi Bay in Shimosa Province; Famous Places in the Provinces (Woodblock print)
   - Authors: うたがわ ひろしげ
   - Original records: [Link]
Summary

Federated Search System for Japanese Humanities Databases

Useful to find information from multiple resources
- real-time parallel search on diverse databases
- a single query input in a single user-interface
- saves users' access time to many databases

Utilizing Linked Data Resources + Authority Data
- multilingual access to multiple databases
  - 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

Instantaneous integration of the results
- aggregating/integrating the retrieved results
- multilingual retrieved results (English, Japanese, ++)
- original links and files
  - Snippets with the original links and thumbnails

2015 DADH
Thank you for your attention!

Questions or Comments?