Japanese Visual Media Graph

Providing researchers with data from enthusiast communities

Senan Kiryakos and Magnus Pfeffer
Outline

- Project background
- Data integration workflow
- Lessons learned
Japanese Visual Media Graph Project

Starting point

- Growing interest in research on Japanese visual media (manga, anime, games...)
- Lack of bibliographic or other database-like central resources suitable for research
- Active online enthusiast communities that discuss and catalogue specific subsets of the Japanese visual media domain (e.g. single mediums, characters, whole franchises)

Project goal

- Integrate the enthusiast data into a central knowledge graph tailored to the needs of researchers
# Japanese Visual Media Graph Project

Example enthusiast communities & statistics

<table>
<thead>
<tr>
<th>Visual Novel Database</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Novel Works</td>
<td>Unique Releases</td>
<td>Characters</td>
<td>Descriptive Work Tags</td>
<td>Descriptive Character Tags</td>
</tr>
<tr>
<td>28,000</td>
<td>73,000</td>
<td>91,000</td>
<td>2600</td>
<td>2800</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AnimeClick</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anime</td>
<td>Manga</td>
<td>Characters</td>
<td>Related Works</td>
<td>Authors</td>
</tr>
<tr>
<td>9400</td>
<td>11,000</td>
<td>102,000</td>
<td>15,000</td>
<td>28,000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Anime Characters Database</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Characters</td>
<td>Works</td>
<td>Character Relations</td>
<td>Descriptive Character Tags</td>
<td>Descriptive Work Tags</td>
</tr>
<tr>
<td>101,000</td>
<td>11,000</td>
<td>17,000</td>
<td>3800</td>
<td>1100</td>
</tr>
</tbody>
</table>
Japanese Visual Media Graph Project
RDF

- All information is turned into *triples* of the form “Entity - Property - Value”
  - This replaces the implicit “record” or “table row” context with explicit statements
  - URIs are used to denote entities and properties
  - Values can be literals or URIs

- Statements are independent of each other
  - Statements can easily be accumulated
  - There can be multiple values for the same property of the same entity
  - Contradictory statements are possible

- RDF-Schema and OWL for formal ontologies
  - Class hierarchies and type definitions
  - Domain and range for properties

- Named graphs
  - RDF databases (triple stores) support named graphs to separate groups of statements and attach additional information to these groups
Data Integration

For each data source

● Analyse the data model used by the source
● Formalize the model as an OWL ontology
● Mint URIs for entities and convert the data into RDF
● Ingest data into a separate graph

Anytime / Ongoing

● Additional analysis of the domain and existing data models from the sources
● Define relevant entities and their relationships
● Create attribute sets to describe entities and relationships
● Formalize the model as an OWL ontology

Finally

● Match entities from multiple sources and merge information
## Workflow example

Data from source, tables

<table>
<thead>
<tr>
<th>Visual Novel table</th>
<th>id</th>
<th>title</th>
<th>original</th>
<th>pid</th>
<th>l_wp</th>
<th>desc</th>
<th>length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7014</td>
<td>Dangan Ronpa: Kibou no Gakuen to Zetsubou no Koukousei</td>
<td>ダンガンロンパ 希望の学園と絶望の高校生</td>
<td>2807</td>
<td>Danganronpa:_Trigger_Happy_Havoc</td>
<td>Hope’s Peak Academy is a government-funded private high school recognized all across the nation as the cradle of hope and prosperity…</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Producer table</th>
<th>id</th>
<th>name</th>
<th>original</th>
<th>type</th>
<th>l_wp</th>
<th>subsidiary</th>
<th>lang</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2807</td>
<td>Spike Chunsoft</td>
<td>スパイク・チュンソフト</td>
<td>co</td>
<td>Spike_Chunsoft</td>
<td>5917</td>
<td>ja</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Character table</th>
<th>id</th>
<th>name</th>
<th>original</th>
<th>gender</th>
<th>vid</th>
<th>desc</th>
<th>role</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10195</td>
<td>Hagakure Yasuhiro</td>
<td>葉隠 康比呂</td>
<td>m</td>
<td>7014</td>
<td>Yasuhiro apparently has an inherent ability to predict the future through the use of his own intuition…</td>
<td>primary</td>
</tr>
</tbody>
</table>
Workflow example

RDF data

@prefix vndb: <http://mediagraph.link/vndb/ont> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://mediagraph.link/vndb/vn/7014> a vndb:VisualNovel ;
 rdfs:label "Dangan Ronpa: Hope’s Peak Academy is a government-funded private high school recognized all across the nation as the cradle of hope and prosperity..."
 vndb:length <http://mediagraph.link/vndb/ont/length/3> ;
 vndb:original "ダンガンロンパ 希望の学園と絶望の高校生" ;
 vndb:producedBy <http://mediagraph.link/vndb/producer/1761> ;
 vndb:title "Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei" ;

<http://mediagraph.link/vndb/producer/2807> a vndb:Producer ;
 rdfs:label "Spike Chunsoft" ;
 vndb:lang "ja" ;
 vndb:name "Spike Chunsoft" ;
 vndb:original "スパイク・チュンソフト" ;
 vndb:producerSub <http://mediagraph.link/vndb/producer/5917> ;
 vndb:type "co" ;
 vndb:vndbLink <http://vndb.org/p2807> .

<http://mediagraph.link/html/vndb/character/10195> a vndb:Character ;
 rdfs:label "Hagakure Yasuhiro" ;
 vndb:desc "Yasuhiro apparently has an inherent ability to predict the future through the use of his own intuition..." ;
 vndb:gender "m" ;
 vndb:name "Hagakure Yasuhiro" ;
 vndb:original "葉隠 康比呂" ;
 vndb:primaryIn <http://mediagraph.link/vndb/vns/7014> ;
 vndb:role "primary" .
Workflow example

Ontology information (excerpt)

```reason
@prefix vndb: <http://mediagraph.link/vndb/ont> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

vndb:producedBy a owl:ObjectProperty ;
   rdfs:label "Produced by"@en ;
   owl:inverseOf vndb:produced ;
   rdfs:domain vndb:VisualNovel ;
   rdfs:range vndb:Contributor .

vndb:title a owl:DatatypeProperty ;
   rdfs:label "Title"@en ;
   rdfs:comment "Base title property in VNDB, typically English."@en ;
   rdfs:domain vndb:VisualNovel ;
   rdfs:range rdfs:Literal .

vndb:desc a owl:DatatypeProperty ;
   rdfs:label "Description"@en ;
   vndb:comment "A short description of the character. While it is advised to keep the spoilers to a
   minimum, you can use the usual formatting codes such as [spoiler] to include spoilers while hiding them
   by default. When using a description from external sources, check out the general editing guidelines for
   quoting rules. [Source https://vndb.org/d12 ]"@en ;
   rdfs:domain vndb:Character ;
   rdfs:range rdfs:Literal ;
   owl:equivalentProperty vndb:description .
```
Custom web frontend

- Shows all stored information for a given entity URI
- Uses SPARQL query
  - All triples with URI as subject (i.e. the entity attributes and values)
  - All triples with URI as object (i.e. all links to the entity)
  - All named graphs that contain the above triples
  - All label information on all URIs in the result set

- URIs in the project namespace are resolved by the web frontend
  - Exploration of linked entities
  - Exploration of ontologies
# Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>● Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei vndb</td>
</tr>
<tr>
<td></td>
<td>● ダンガンロンパ 希望の学園と絶望の高校生 vndb</td>
</tr>
<tr>
<td>Description</td>
<td>Hope's Peak Academy is a government-funded private high school...</td>
</tr>
<tr>
<td></td>
<td>across the nation as the cradle of hope and prosperity...</td>
</tr>
<tr>
<td>Original</td>
<td>ダンガンロンパ 希望の学園と絶望の高校生 vndb</td>
</tr>
<tr>
<td>Produced by</td>
<td>Spike Chunsoft vndb</td>
</tr>
<tr>
<td>Title</td>
<td>Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei Visual Novel vndb</td>
</tr>
<tr>
<td>VNDB Link</td>
<td><a href="https://vndb.org/v7014">https://vndb.org/v7014</a> vndb</td>
</tr>
<tr>
<td>is Primary Role of</td>
<td>Hagakure Yasuhiro vndb</td>
</tr>
</tbody>
</table>
Merged entities

- Mint a new URI representing the merged entity and add property-value pairs from all sources
  - Allows for exploration and queries
  - Original data stays in source context
  - Merged data in a separate context
<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td>Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei <strong>vndb</strong></td>
</tr>
<tr>
<td></td>
<td>ダンガンロンパ希望の学園と絶望の高校生 <strong>vndb</strong></td>
</tr>
<tr>
<td></td>
<td>ダンガンロンパ希望の学園と絶望の高校生 <strong>acdb</strong></td>
</tr>
<tr>
<td>Content Rating</td>
<td>M / Mature <strong>acdb</strong></td>
</tr>
<tr>
<td>Date</td>
<td>2010/11/25 <strong>acdb</strong></td>
</tr>
<tr>
<td>Description</td>
<td>Hope's Peak Academy is a government-funded private high school recognized all across the nation as the cradle of hope and prosperity... <strong>vndb</strong></td>
</tr>
<tr>
<td>Genre</td>
<td>visual novel <strong>wikidata</strong></td>
</tr>
<tr>
<td></td>
<td>school anime and manga <strong>wikidata</strong></td>
</tr>
<tr>
<td>Media Type</td>
<td>Video Game <strong>acdb</strong></td>
</tr>
<tr>
<td>Narrative Location</td>
<td>Japan <strong>wikidata</strong></td>
</tr>
<tr>
<td>Original</td>
<td>Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei <strong>vndb</strong></td>
</tr>
<tr>
<td></td>
<td>ダンガンロンパ希望の学園と絶望の高校生 <strong>vndb</strong></td>
</tr>
<tr>
<td></td>
<td>ダンガンロンパ希望の学園と絶望の高校生 <strong>acdb</strong></td>
</tr>
<tr>
<td>Produced by</td>
<td>Spike Chunsoft <strong>vndb</strong></td>
</tr>
<tr>
<td>Synopsis</td>
<td>Danganronpa: Trigger Happy Havoc is a visual novel adventure game developed and published as the first game in the Danganronpa series. The game was originally released in Japan... <strong>wikidata</strong></td>
</tr>
<tr>
<td>Title</td>
<td>Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei <strong>vndb</strong></td>
</tr>
<tr>
<td>type</td>
<td>Visual Novel <strong>vndb</strong></td>
</tr>
<tr>
<td>VNDB Link</td>
<td><a href="https://vndb.org/v7014">https://vndb.org/v7014</a> <strong>vndb</strong></td>
</tr>
</tbody>
</table>

Dangan Ronpa Kibou no Gakuen to Zetsubou no Koukousei
Further integration: Common data model

- Limits of merged entities
  - Properties are described in multiple ontologies
  - Similar information is conveyed in different properties
  - Inconsistent naming conventions
  - Queries across sources are complex

- Merging properties process
  - Identify properties that describe the same attribute for a given entity type
  - Create a new property URI and map the information to that
    - Conflicting information can be left as-is (representing the community dissent) or manually resolved (for factual information that should represent a ground truth)
  - Can be limited to a core property set at first and extended as needed
Key properties

● Continuous integration
  ○ Additional data does not interfere with the existing information
  ○ Additional sources and additional properties from existing sources can be added at any time

● Accessibility and Explorability
  ○ Frontend shows a human-readable representation of entities, attributes and relations
  ○ Linked entities can be explored interactively

● Transparency
  ○ Each statement has a named graph associated, which can be further described
  ○ Source-specific ontologies preserve the naming and conventions of the enthusiast communities

● Expandability
  ○ Frontend can be used to connect web-based analysis functions
  ○ Without fixed data model, the scope of the project can be adjusted
Lessons learned

● Fan communities in the Japanese visual media domain
  ○ Collect and curate a wealth of data
  ○ Have a unique viewpoint on the domain that influences the data model
  ○ Are mostly open to collaboration with research projects
  ○ Often already make data available for download or using APIs

● Data integration
  ○ Technical problems were limited and mostly easy to overcome
  ○ Data licensing issues are more difficult than we imagined
    ■ Licence incompatibilities triggered by “share-alike” clauses

● Researcher interaction
  ○ Data-driven research has a lot of potential
  ○ Knowledge graphs are a hot topic in digital humanities
  ○ Similar problems and solutions in different domains
Thank you for listening

Project homepage and blog: https://jvmg.iuk.hdm-stuttgart.de/

Project funded by