Hackathon project «Reference data for Anime and Manga: Semantic Linking and Publishing of Diverse Data-Sets»

Team lead: Victor Agroskin, vic5784@gmail.com
(MSK, UTC+4)

27th of February 2014
Ontology Summit 2014
A problem and an answer

• Ontology-based integration of engineering data should be transparent for engineers and to logicians. What does it mean? To understand simultaneously plant models, equipment, P&ID diagrams, and classes, properties and inverse properties, inheritance.
  • Few people know both. Even fewer among them use the same terms for objects they probably share.
  • Engineering datasets come from industrial proprietary CAD systems, ontology tools are open-source university apps.
  • Engineering data is guarded behind corporate firewall, ontology research in public.

• We need non-industrial example of engineering ontology work to test and publicize approaches, method and tools. What can it be?
Anime and manga production as engineering

• Studio, distribution and fandom as Enabling system and “inside-anime world” as System-of-interest – complexly related.

• Derivative works as Product lines and Custom-tuned models: what is our typical system?

• Complex data with high variety:
  • Unstructured texts, video, picture datasets and also relational production and fandom data.
  • Blurred border between structured and unstructured data (lessons from IBM Watson: ontologies help to answer questions but aren’t a primary source of information)
Datasets, datasets,

Databases, datasets,

**Datasets, datasets,**
A) The Reference Data

• We hold firm ontology commitments, grounded in ISO 15926 4D extensional ontology (self-education reading sequence: http://levenchuk.com/2012/10/01/iso-15926-self-education-sequence/).
• We want to be a part of federated Reference Data Libraries (RDL).
• We prefer to link to PCA reference data wherever possible: (http://rds.posccaesar.org/downloads/PCA-RDL.owl.zip or http://posccaesar.org/endpoint/sparql).
• We will try to develop domain ontology (classes and templates – n-ary relations) rich enough to map a sensible selection of anime-manga online resources.

Results: published anime-manga domain RDL (OWL file and probable SPARQL endpoint)
b) The Semantization of Datasets

• We will develop patterns for anime-manga semantic data structures (bases on classes and templates of domain RDL).

• We will program adapters to access selected datasets and map native data model to developed patterns.

• We will run adapters and create semantic data from anime-manga sources in ISO 15926-8 RDF.

• Datasets:
  • http://anidb.net (API: http://wiki.anidb.net/w/API)
  • http://myanimelist.net (API: http://myanimelist.net/modules.php?go=api)
  • IMDB (how to get it: http://blog.teamtreehouse.com/coding-a-dynamic-imdb-webapp-using-my-movie-api)
  • Wikipedia
  • …

• Documentation for .15926 Editor patterns and mapping: http://techinvestlab.ru/V4
c) The Linked Data

• We will attempt to use Flask web framework (http://flask.pocoo.org/) to publish our semantic data in human readable form.

• We will use the same patterns used for mapping to guide human representation.

• We will attempt to enhance Linked Data pages with more semantic mark-up (schema.org?).

Result: published human readable Linked Data pages for anime-manga semantic data.
Whom do we need?

• Otaku (anime and manga fans) to understand data and engineers to build ontology.
• Data modelers to develop patterns.
• Python programmers to access native APIs of online resources and encode mappings.
• Web designers to develop concept pages.
• Semantic markup (schema.org anyone?) specialists to develop and embed semantic markup.
Thank you

Contacts:
Team lead: Victor Agroskin, vic5784@gmail.com
.15926 Editor 1.4 freeware:
http://techinvestlab.ru/dot15926Editor
(documentation:
http://techinvestlab.ru/15926EditorDocumentation)