Using DBpedia in Europeana Food and Drink

Vladimir Alexiev, Ontotext Corp

DBpedia meeting, 12 Feb 2016, The Hague
Europeana Food and Drink
EFD Semantic Application

How is it done? Tech part:
Food and Drink Classification
Using DBpedia for Enrichment
Using GeoNames for Parent Places
Europeana Food and Drink

Collects FD content, makes FD creative applications

http://foodanddrinkeurope.eu/
The Semantic demonstrator demonstrates the use of semantic technologies for classification and discovery of Europeana objects related to Food and Drink. It aims to provide semantic enrichment, i.e., extract references to Food and Drink topics from free text in object metadata to strengthen the relevance of database search results on the theme of Food and Drink.

The application does this by applying a Europeana Food and Drink Classification scheme to Cultural Heritage Objects, such as those stored in Europeana to yield more precise groupings, relationships and categorisations for database items.

Showcases enrichment, faceted semantic search

 Demo, Data, SPARQL. Homepage, Description, Deliverable
Here: Bulgarian recipes

But this is by dataProvider, not through semantic enrichment

Only a few BG enrichments: will add 300+ term→DBpedia mappings by regexp, as no BG NLP is planned
EFD Sem App Shots

Europeana Food & Drink

The Semantic Demonstrator shows the use of semantic technologies for classification and discovery of Europeana objects related to Food and Drink. Detailed description, data, SPARQL endpoint.

Selected filters: Food: Fishing, Place: Oceania

Fishing objects from Oceania.

Semantic enrichment of objects using Concept Extraction (EN only)

Like a Mini-Europeana, but adds 2 semantic hierarchical facets: FD topic and Place
Objects from Fratelli Alinari related to the Roman Empire and Beverages

Semantic (conceptual) search / browse, query expansion
Tech part

HOW IS IT DONE?
Food and Drink Classification

FD is such a broad topic: how to find classifications for it?

- Main goal: FD gazetteer to recognize FD topics in CH objects
- Studied a number of LOD datasets: USDA Standard Reference, FAO AgroVoc, EC EuroVoc, Getty AAT, UMBEL SuperTypes, Yago WordNet Domains, WiBi, DBTax, Wiktionary
- Concluded we should use Wikipedia Categories
- Use through DBpedia
- Wikidata has expunged Categories as useless / too unorganized
- To complement with Wikipedia lists (extend Extraction Framework), UMBEL, DBTax

Also do Place enrichment using DBpedia/GeoNames

EFD Classification: Deliverable, Slides
Wikipedia FD Categories

- Starting from dbc:Food_and_drink, you reach 887k cats, 26 levels deep, representing 80% of all categories
- Most are irrelevant to FD (left: level 2, right: level 5)
Pruning the FD Category Tree

Mix of approaches:

- Statistical analysis of the category network
- Manual curation (chopping out irrelevant branches: about 300)
- Evidence-based feedback (CHOs, UMBEL, DBTax)

Reduced cats by 98%, level from 16 to 5 (reduce semantic drift)

- Left relevant: 17.5k cats, 221k articles, ~900k labels


- Extended version (LNCS 9398): preprint, published
Using DBpedia, Editing Wikipedia

→ DBpedia is a great integrated resource that we used for Categories (cat hierarchy, art→cat classification), Places, labels

→ When something is missing or wrong, don’t just cry! Add it to Wikipedia, e.g.
  • dbc:Bottles: added parent dbc:Drink_containers
  • dbr:Gourd: added category dbc:Bottles
  • “Muller” (vessel for making mulled beer or wine): added as alias (Redirect) of dbr:Mulled_wine
  • Made article dbr:Shepherd's_crook (cut & paste)

→ You learn a lot of curious facts in the process!
  • Did you know the Crook_and_flail are the insignia of pharaonic authority?
Using GeoNames for Parent Places

→ Surprisingly, DBpedia doesn't have good Parent Place info:
  • No statement that Bulgaria and France are part of Europe.
  • Have classes yago:MemberStatesOfTheEuropeanUnion, yago:EuropeanCountries, yago:EuropeanUnionMemberEconomies, but no relation to dbr:Europe

→ No uniform property. E.g. dbo:City has **dbo:region** and **dbo:country**, while dbo:Island (e.g. dbr:Andaman_Islands) has:
  • **dbo:archipelago** dbr:Andaman_and_Nicobar_Islands (physical parent)
  • **dbo:location** dbr:Bay_of_Bengal (physical parent)
  • **dbp:countryAdminDivisions** dbr:Andaman_and_Nicobar_Islands (admin parent)
  • **dbo:country** dbr:India (admin ancestor)
  • **dbo:capital** dbr:Port_Blair (admin child)
  • **dbo:majorIsland** dbr:North_Andaman_Island, dbr:South_Andaman_Island (physical child, partial)

→ So we decided to use GeoNames, which has uniform property gn:parentFeature
Using GeoNames for Parent Places

- GeoNames→Wikipedia/DBpedia links
  - Got from geonames.org, used script from Dbpedia
  - Coverage is about 65% (Wikidata has similar coverage)
  - Added 250 links (e.g. dbr:Sloane_Street
gn:parentFeature
dbr:Royal_Borough_of_Kensington_and_Chelsea)
- About 10 fixes to GeoNames on their site (e.g. North & South America were children of “America”…a small village)
- Loaded GeoNames & DBpedia in Ontotext GraphDB, used owl:sameAs optimization to merge (smush) the corresponding nodes
- Voila! A hierarchical semantic facet
  - Faceting is done by ElasticSearch, we use the GraphDB-ElasticSearch Connector
Thanks for your time!

Project co-funded by the European Union under the ICT Policy Support Programme