Andreas Harth, Hannes Gassert, Ina O'Murchu, John Breslin, Stefan Decker

WikiOnt: An Ontology for Describing and Exchanging Wiki Articles

Wikimania, Frankfurt, 5th August 2005

Why Develop an Ontology for Wikis?

A specific wiki ontology can be built to integrate Wikipedia (and by extension other wiki-based sites) into the Semantic Web framework and to make Wikipedia machine-processable and -understandable

Overview

- Semantic Web / ontologies / metadata
- Main classes / properties in ontology
- Converting Wikipedia to metadata instances
- Exchange of and query interface to metadata instances
- Connecting Wikipedia to other Semantic Web data sources

0. Semantic Web in Brief

- What is the Semantic Web?
- *“An extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation”,* Sir Tim Berners-Lee et al., Scientific American, 2001
- Requires web pages to have **metadata** with underlying **ontologies**

Ontologies in Brief

- Ontologies are formal specifications of how to represent the entities (by **classes**) in a **specific domain** as well as the various interrelations among them (by **properties**), e.g.
- “Article” is a class, and “Stub” is a sub-class of Article
- “internalLink” property links Articles

Metadata in Brief

- While the ontology provides the schema or vocabulary, the main data lies in the actual **instances** that are described using the ontology, e.g.

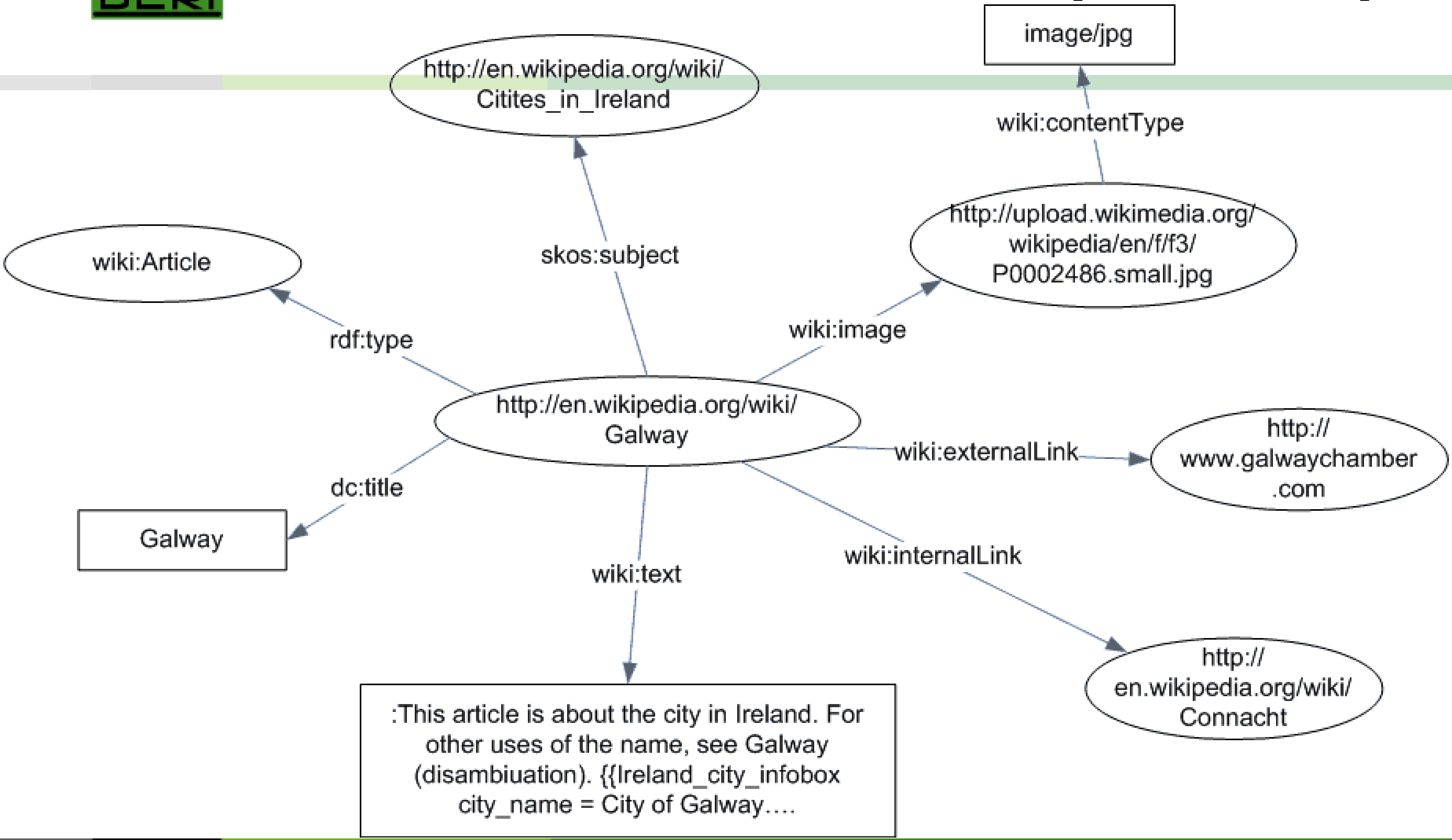
```
<Article rdf:about=”
  http://en.wikipedia.org/wiki/Galway”>
   <internalLink rdf:resource=”
     http://en.wikipedia.org/wiki/Connacht”>
...</Article>
```

1. Main Classes / Properties in Ontology

- We use OWL (Web Ontology Language)
- Classes: Article, Category, Stub, Image
- Properties: dc:title, text, dc:creator, dc:date, internalLink, externalLink, skos:subject, skos:narrower, redirectsTo, sioc:views*, img:height, img:width, contentType

* *SIOC is another ontology we've developed for connecting discussion forums, http://rdfs.org/sioc*

Example Instance (Graph)



Example Instance (RDF/N3)

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix wiki: <http://sw.deri.org/2005/04/wikipedia/index.rdf#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
<http://en.wikipedia.org/wiki/Galway> rdf:type wiki:Article ;
    dc:title "Galway" ;
    skos:subject <http://en.wikipedia.org/wiki/Cities_in_Ireland> ;
    wiki:image
      <http://upload.wikimedia.org/wikipedia/en/e/e3/Galway.jpg> ;
    wiki:externalLink <http://www.galwaychamber.com/> ;
    wiki:internalLink <http://en.wikipedia.org/wiki/Connacht> ;
    wiki:text ":This article is about the city in Ireland..." .
<http://upload.wikimedia.org/wikipedia/en/e/e3/Galway.jpg>
   wiki:contentType "image/jpg" .
```

2. Converting Wikipedia to Metadata Instances

- MySQL dump of *cur* / *categorylinks* tables
- class *WikipediaArticle*: extracts information from the text, the wiki markup and the database fields in an article
- class *N3WikipediaArticle*: produces an RDF/N3 representation of an article
- *QueryRunner*: Initiates conversion for one term / set of articles / entire encyclopedia

3. Exchange of and Query Interface to Instances

- Exchange of articles between different wiki systems by developing exporters and importers for the various systems
  - replace cut-and-paste methods currently used
  - eliminate need for SQL dumping and importing
- Allow querying of Wikipedia RDF metadata instances
  - see next slide

RESTful Web Service Interface

- Straightforward to provide a web service interface to converted Wikipedia RDF data
- Dataset stored in “YARS” RDF store
- Pose queries via HTTP to retrieve the pieces of Wikipedia information we want
- Since data in a machine-readable format, namely RDF, it is straightforward to read in and reuse the data in software programs

Sample Query for Text of Galway Wikipedia Article

```
@prefix : <http://sw.deri.org/2004/06/yars#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ql: <http://www.w3.org/2004/12/ql#> .
@prefix wiki: <http://sw.deri.org/2005/04/wikipedia/index.rdf#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .

<> ql:select {
       (?text) .
}; ql:where {
       ?x rdf:type wiki:Article .
       ?x dc:title "Galway" .
       ?x wiki:text ?text .
} .
```

4. Connecting Wikipedia to Other SW Data Sources

- Wikipedia URLs can become general URIs identifying concepts in the Semantic Web
- Infer connections between Wikipedia articles and ODP/DMOZ categories
- Use WordNet definitions and synonyms to fill in wiki stub pages
- RDF version of CIA World Factbook can be combined with RDF from Wikipedia

WikiOnt URLs

1. Ontology namespace:
   – http://sw.deri.org/2005/04/wikipedia/index.rdf
2. PHP scripts to convert Wikipedia to RDF:
   – http://sw.deri.org/2005/04/wikipedia/
3. RESTful WikiOnt RDF query web service:
   – http://sw.deri.org/2005/04/wikipedia/store
   – No slash at the end!

Conclusions / Questions

1. A proposed ontology (WikiOnt) to describe articles from the Wikipedia free encyclopedia using the Web Ontology Language (OWL)
2. PHP scripts developed to convert to RDF
3. RESTful web interface to query data
4. Can combine with other SW data sources

- Thank-you for your attention – questions?