ASIt

A Linked Open Data Geolinguistic Web

Emanuele Di Buccio*, Giorgio Maria Di Nunzio*, Jacopo Garzonio**, Diego Pescarini* and Gianmaria Silvello*

* University of Padua
** Ca’Foscari University of Venice

LOD2014, 20th - 21st February 2014, Archivio Centrale dello Stato, Roma
Outline

- Background and motivations
- ASIt Enterprise
- ASIt and LOD
Background and Motivations 1/2

- Digital Geolinguistics
  - Mapping the geographical distribution of linguistic phenomena
- Linguistic Atlas
- Geographic Information Systems (GIS)
Background and Motivations 2/2

• Heterogeneity of geolinguistic projects limits the reusability of linguistic tools and data collections

• Design and distribution of language resources
  ‣ coverage and documentation support
  ‣ suitable format
  ‣ document the existence of a resource
ASIt Enterprise

• Long standing tradition of
  ‣ collecting linguistic data
  ‣ analyzing linguistic data

• Research on syntax of Italo-Romance varieties

• Several important lines of research in linguistics
  ‣ it allows comparison between closely related varieties
  ‣ formulation of hypotheses about the nature of cross-linguistic parametrization
An Ontology for Geolinguistic Resources
8 questionnaires

Each questionnaire contains a set of sentences in Italian

Il direttore è arrivato

El diretor l’è rivà

Each questionnaire has a set of translations (one for each dialect)

200 observed localities
240 dialects
Modeling Geolinguistic Data

• Geolinguistic concepts organized in three major areas:
  ‣ Geographic
  ‣ Derivation
  ‣ Tagging
Modeling Geolinguistic Data

• Geolinguistic concepts organized in three major areas:
  ‣ Geographic
  ‣ Derivation
  ‣ Tagging

• Comprehends classes and properties related to physical places
• Defines places where a dialect is spoken
Modeling Geolinguistic Data

- Geolinguistic concepts organized in three major areas:
  - Geographic
  - Derivation
  - Tagging

- People speaking a certain language
- The geographical area where they live
- Relationship among people

allows for the study of the correlation between social factors, education and knowledge of dialect
Modeling Geolinguistic Data

- Geolinguistic concepts organized in three major areas:
  - Geographic
  - Derivation
  - Tagging

Language specific classes
- Document
- Sentence
- Word
- Their relationships
# ASIt RDF Specification

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Namespace</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>asit</td>
<td><a href="http://purl.org/asit/terms/">http://purl.org/asit/terms/</a></td>
<td>ASIt vocabulary terms</td>
</tr>
<tr>
<td>dcterms</td>
<td><a href="http://purl.org/dc/terms/">http://purl.org/dc/terms/</a></td>
<td>Dublin Core terms</td>
</tr>
<tr>
<td>foaf</td>
<td><a href="http://xmlns.com/foaf/0.1/">http://xmlns.com/foaf/0.1/</a></td>
<td>Friend of a friend</td>
</tr>
<tr>
<td>geo</td>
<td><a href="http://www.w3.org/2003/01/geo/wgs84_pos#">http://www.w3.org/2003/01/geo/wgs84_pos#</a></td>
<td>WGS84 Geo Positioning</td>
</tr>
<tr>
<td>gn</td>
<td><a href="http://www.geonames.org/ontology#">http://www.geonames.org/ontology#</a></td>
<td>GeoNames Ontology</td>
</tr>
<tr>
<td>owl</td>
<td><a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a></td>
<td>OWL vocabulary terms</td>
</tr>
<tr>
<td>rdf</td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a></td>
<td>RDF vocabulary terms</td>
</tr>
<tr>
<td>rdfs</td>
<td><a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a></td>
<td>RDF Schema</td>
</tr>
</tbody>
</table>
Modeling Geolinguistic Data

- ASIt ontology: starting point for modeling and describe geolinguistic resources
  - provides general-purpose concepts and relationships
  - is extendable by adding more fine-grained classes
Exposing Geolinguistic Data
Access Control

RESTful Web Service

RDF DAO

RDF Store

Relational Database

Syncronization Service

Serialization Service

Mapping Service

ASIt System

Web Application

Users

(Web Service)

Linguistic Layer

RDF Layer

Datastore

Application Service

User Interface

Web Service

(Database)

Helps to deal with diverse information needs

Helps to deal with a curated dataset

Helps to deal with an evolving dataset

(modularity) helps to deal with diverse information needs

helps to deal with a curated dataset

helps to deal with an evolving dataset
Exploiting Geolinguistic Data
ASIt Geolinguistic App

- SPARQL Endpoint
  - http://purl.org/asit/rdf/sparql

- Graphical user interface to perform tag-based search
  - result list (each list item is a sentence)
  - dynamically generated maps

- Graphical user interface to navigate the RDF graph
### RDF Browser

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Namespace</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbpedia</td>
<td><a href="http://dbpedia.org/resource/">http://dbpedia.org/resource/</a></td>
</tr>
<tr>
<td>asit-province</td>
<td><a href="http://purl.org/asit/resource/Province">http://purl.org/asit/resource/Province</a></td>
</tr>
<tr>
<td>asit</td>
<td><a href="http://purl.org/asit/terms/">http://purl.org/asit/terms/</a></td>
</tr>
<tr>
<td>gn</td>
<td><a href="http://www.geonames.org/ontology#">http://www.geonames.org/ontology#</a></td>
</tr>
<tr>
<td>rdf</td>
<td><a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a></td>
</tr>
<tr>
<td>owl</td>
<td><a href="http://www.w3.org/2002/07/owl#">http://www.w3.org/2002/07/owl#</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>asit:Region</td>
</tr>
<tr>
<td>asit:hasProvince</td>
<td>asit-province:Verona</td>
</tr>
<tr>
<td></td>
<td>asit-province:Vicenza</td>
</tr>
<tr>
<td></td>
<td>asit-province:Belluno</td>
</tr>
<tr>
<td></td>
<td>asit-province:Treviso</td>
</tr>
<tr>
<td></td>
<td>asit-province:Venezia</td>
</tr>
<tr>
<td></td>
<td>asit-province:Padova</td>
</tr>
<tr>
<td></td>
<td>asit-province:Rovigo</td>
</tr>
<tr>
<td>owl:sameAs</td>
<td>dbpedia:Veneto</td>
</tr>
<tr>
<td>gn:officialName</td>
<td>Veneto</td>
</tr>
<tr>
<td>gn:geographicPartition</td>
<td>Nord-est</td>
</tr>
</tbody>
</table>
LOD Live
ASIt Geolinguistic App
Final Remarks

• Contribution
  ‣ LOD approach to increasing the level of interoperability of geolinguistic applications
Final Remarks

- **System design and implementation**
  - decoupling between system which manages the data and the one that provides services over those data
  - geolinguistic open dataset can be used by third-party projects to
    - enrich data
    - build up new services
BACKUP SLIDES
Evolving Dataset

- **Tags:**
  - Jan '11: 523
  - Jul '11: 526
  - Jan '12: 529
  - Jul '12: 532

- **Documents:**
  - Jan '11: 460
  - Jul '11: 475
  - Jan '12: 490
  - Jul '12: 505

- **Sentences:**
  - Jan '11: 520
  - Jul '11: 540
  - Jan '12: 550
  - Jul '12: 560

- **Tags on Sentences:**
  - Jan '11: 10
  - Jul '11: 20
  - Jan '12: 30
  - Jul '12: 40

- **Tags on Words:**
  - Jan '11: 5
  - Jul '11: 10
  - Jan '12: 15
  - Jul '12: 20
Setup of a new project
Retrieval and preparation of written texts, conduction of fieldwork, transcription of audio data, translation of sentences, DB population

Segmentation of sentences in words and constituents
Validation of editors’ work
Tagging of words, constituents and sentences
Validation of editors’ work

Access to the results of the project

Linguistic Analyses
Tags
Documents
Dialects
Linguistic Project

System design and model

Administrator
Supervisor
Editor
Informant
Visitor