Una success story tra Accademia ed Istituzioni Pubbliche

...in 5 minuti!

Maria Teresa PAZIENZA, Armando STELLATO,
Manuel Fiorelli, Tiziano Lorenzetti, Andrea Turbati,

http://art.uniroma2.it/

Roma, ELRC, 15 marzo 2016
Il nostro laboratorio di ricerca

Realized as a University Campus distributed over a wide area in the SE of Rome

We are located at the Faculty of Engineering, Information Engineering Building
The ART Group

Prof. Maria Teresa Pazienza
Full Professor, head of the ART group

Prof. Fabio Massimo Zanzotto
Associate Professor

Prof. Roberto Basili
Associate Professor

Armando Stellato
Researcher

(PhD students, research associates/fellows/assistants...)

Paolo Annesi
M.S., PhD Student

Danilo Croce
Eng., M.S., PhD Student

Diego De Cao
Eng., M.S., PhD

Manuel Fiorelli
M.S., PhD student

Cristina Giannone
Eng., M.S., PhD

Andrea Turbati
Eng., M.S., PhD
The ART Group

Areas of Interest

Natural Language Processing
- Robust Parsing
- Information Extraction
- Semantic Role Labeling
- Textual Entailment
- ML for Natural Language

Machine Learning
- ML for Natural Language
- Computational Language Learning
- Ontology Learning
- Classification, Multimedia IR
- Music IR

Knowledge Engineering
- Ontologies and Knowledge Modeling
- Semantic Web Technologies
- Knowledge Acquisition
- Knowledge Sharing
- Knowledge-based Systems
ART Technology Asset (KE Unit)

Semantic Turkey / VocBench*

Flagship KM&A Platform

Ontology Linguistic Enrichment Tool

ALE

Automatic Linguistic Enrichment

Linguistic Watermark

Library and Ontologies for Accessing Heterogeneous Linguistic Resources

CODA

Computer-aided Ontology Development Architecture

Wrappers for different LRs

Java Multi WordNet Library

Semantic Turkey extensions for Knowledge Acquisition

…e molto altro ancora

* VocBench in collaboration with FAO
The ART Group is actively contributing to both the research and end users communities with the software outcomes of its applied research

Knowledge Management and Acquisition

**Semantic Turkey**: a Knowledge Management and Acquisition Platform for the Semantic Web

**VocBench**: a collaborative editing environment for SKOS thesauri and concept schemes (developed in collaboration with [FAO](#))

**CODA**: an architecture used to project the data produced by one or more UIMA annotators onto an ontology

**Sheet2RDF**: a platform for acquisition and transformation of datasheets into RDF (also available as a [Semantic Turkey](#) plugin, and soon for [VocBench](#) as well)

**OntoLing**: a tool for Linguistic Enrichment of ontologies (available as extension for different ontology development platforms)

**Owl ART API**: api for accessing and managing RDF. Provide various implementations for several existing triple store

Natural Language Processing / Language Engineering

The **Chaos System**: A Natural Language Robust Processor

**OBI-WAN**: Ontology Based WrApper iNduction

**Linguistic Watermark**: an abstraction framework for uniformly accessing heterogeneous linguistic resources (thesauri, dictionaries, wordnets...):

**JMWNL**: Java (Multi)WordNet Library: a set of extensions to the [JWNL](#) library allowing access to WordNets for different languages (currently supports [EuroWordNet](#))

**LIME**: a vocabulary and an associated Java API for linguistic metadata about linguistically grounded RDF datasets and linguistic resources published as Linked Data

24/03/2016
Capitolo IV

2006 In una conferenza lontana lontana…

incontriamo alcuni rappresentati della FAO

si parla di tesauri, risorse online,

multilinguismo…
Interesse per nostre risorse per arricchimento linguistico
Linguistic Metadata for Concepts documentation

- Search linguistic expressions inside the LR
- Explore semantic relationships which characterize the LR
- ...and linguistic relationships

Integration between ontology and linguistic resource: search ontology terms inside the linguistic resource

Linguistic Enrichment of the Ontology

Ontology concepts bear a greater linguistic expressivity: this helps in identifying similarities with other conceptualizations.

Assist ontology creation by extracting portions of knowledge from the LR

Concepts documentation for...
Attività e Collaborazioni

2009: collaborazione allo sviluppo di VocBench, un sistema collaborativo, interno alla FAO, per lo sviluppo di tesauri multilingui


2012: inizio di una stretta collaborazione per il progetto VocBench 2.0. Un sistema general purpose completamente compliant agli standard W3C SKOS/SKOSXL per la rappresentazione di tesauri, basato su Semantic Turkey - framework per la gestione di ontologie, tesauri e dati RDF- sviluppato a Tor Vergata

2012: progetto europeo SemaGrow, per lo sviluppo di tecnologie a supporto di federazioni di linked data partners. Supporto parziale allo sviluppo di VocBench 2

2016: ……..
AGROVOC is a *controlled vocabulary* covering all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations. It is published by FAO and edited by a community of experts.

AGROVOC consists of over **32,000 concepts available in 23 languages**: Arabic, Chinese, Czech, English, French, German, Hindi, Hungarian, Italian, Japanese, Korean, Lao, Malay, Persian, Polish, Portuguese, Russian, Slovak, Spanish, Telugu, Thai, Turkish and Ukrainian.

You can use AGROVOC to look up the common name of a plant in a language that you do not master, or to find relations between a commodity and the crop from which it is produced. Your library can use AGROVOC to index its documents, or you can use it from inside your content management system (e.g., Drupal) to organize your documents or web site. You can also use AGROVOC as an hub to access many other vocabularies available on the web.

To date, AGROVOC is used by researchers, librarians and information managers for indexing, retrieving and organizing data in agricultural information systems and Web pages. Currently, AGROVOC is an SKOS-XL concept scheme and a Linked Open Data (LOD) set aligned with 16 other multilingual knowledge organization systems related to agriculture.

You may download AGROVOC, access its Web Services or SPARQL endpoint.
Agrovoc Oggi (1)

- **Agrovoc [1]**: Linked Open Dataset a 5 stelle
  - SKOS & SKOSXL
  - > 32.000 concetti
  - sviluppato collaborativamente a livello mondiale (tramite VocBench 2) e validato da FAO

- **Disponibile**: 
  - tramite HTTP dereferenciation (dati RDF o pagine Web)
  - tramite SPARQL endpoint ()

- **Metadati**
  - presto anche LIME description

- **Multilingue**
  - 25 lingue (22 fortemente rappresentate, 3 in via di sviluppo)
Agrovoc Oggi (2)

• **Uso**
  – utilizzato come vocabolario per l’indicizzazione di basi documentali
  – come riferimento per dimensioni discrete in ipercubi per dati statistici
  – repository più noto basato su Agrovoc: AGRIS [1]
    • data agglomeration da 150 provider, provenienti da 100 diversi paesi
    • 5 milioni di riferimenti bibliografici in ambito agriculturale, più di 130 milioni di triple RDF
    • spidered da google per popolare Google Scholar

• **Dissemination**


• **COD** (Computer-aided Ontology Development)
  – All processes for enriching ontology content through exploitation of external resources, by using (semi)automatic approaches.

• **CODA**: COD Architecture
  – An Architecture for systems for Computer-aided Ontology Development
  – A Platform supporting development of such systems
• ProjEction of Annotations Rule Language

• A language for transforming information extracted and stored as UIMA annotations into RDF Triples
A use case for CODA: AGROVOC

Natural Language

Documents

AGROVOC SKOS

http://aims.fao.org/standards/agrovoc/concept-scheme

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>c_3556</td>
<td><a href="http://aims.fao.org/aos/agronontology#isPestOf">http://aims.fao.org/aos/agronontology#isPestOf</a></td>
<td>c_1972</td>
</tr>
<tr>
<td>c_1972</td>
<td><a href="http://aims.fao.org/aos/agronontology#hasPest">http://aims.fao.org/aos/agronontology#hasPest</a></td>
<td>c_3556</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject</th>
<th>Predicate</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>hemiptera</td>
<td><a href="http://aims.fao.org/aos/agronontology#isPestOf">http://aims.fao.org/aos/agronontology#isPestOf</a></td>
<td>crop</td>
</tr>
<tr>
<td>crop</td>
<td><a href="http://aims.fao.org/aos/agronontology#hasPest">http://aims.fao.org/aos/agronontology#hasPest</a></td>
<td>hemiptera</td>
</tr>
</tbody>
</table>
Automatic Annotation and Knowledge Acquisition

Semantic Turkey

AgroIE plugin

Insects

Plants

Relations

AGROVOC Thesaurus (Small cut)
Showcase - Ontology Enrichment (1)

Adding Concepts
Showcase - Ontology Enrichment (2)
VocBench 2.0

VocBench is a **web-based, multilingual, editing and workflow tool that manages thesauri, authority lists and glossaries using SKOS-XL**.

Designed to meet the needs of semantic web and linked data environments, VocBench provides tools and functionalities that facilitate both **collaborative editing and multilingual terminology**. It also includes administration and group management features that permit flexible roles for maintenance, validation and publication.

VocBench 2 is developed by a collaboration between the **Food and Agriculture Organization of the United Nations (FAO)** and the **ART Group** of the **University of Rome 'Tor Vergata'**.

A growing **user community** has formed around VocBench, mainly made of **public organizations** which were looking for open source solutions for maintaining their thesauri, code lists and authority resources.

Born over the success of its first homonymous predecessor, which was developed for internal use in FAO, VocBench 2 represents a major shift in both design and outlook. Based on a **service-oriented OSGi architecture**, it is designed to allow external developers to add functionality via plug-ins without disturbing the core. This allows the current partnership to avoid the administrative burden of managing a full-blown open source community, while at the same time maintaining core functionalities and making the code public so that others can extend it.

Version 2 also features other important improvements such as native RDF support, support for different triple-stores, looser coupling between modules and more.

VocBench business and data access layers are handled by **Semantic Turkey**, an open-source platform for Knowledge Acquisition and Management realized by the **ART Research Group** at the University of Rome - Tor Vergata.
## VocBench 2.0

**Welcome to VocBench 2.0**

VocBench 2.0 is a tool for managing and visualizing ontologies. It provides a user-friendly interface for creating, editing, and sharing ontologies.

### Main Features
- **Concepts Management:** Add, edit, and manage concepts.
- **Properties:** Define and manage properties.
- **Schemes:** Link to external schemes.
- **Validation:** Validate the concepts and properties.
- **SPARQL:** Query the ontology.
- **ICV:** Import, Convert, Validate the ontology.

### Example:

<table>
<thead>
<tr>
<th>Language</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>coordination of EMU policies (Preferred)</td>
</tr>
<tr>
<td>Spanish</td>
<td>coordinación de políticas UEM (Preferred)</td>
</tr>
<tr>
<td>French</td>
<td>coordination des politiques UEM (Preferred)</td>
</tr>
<tr>
<td>Portuguese</td>
<td>coordenação das políticas UEM (Preferred)</td>
</tr>
</tbody>
</table>

**Legend:**
- **Proposed:** Proposed for consideration.
- **Published:** Published in a formal document.
- **Revise:** Under review.
- **Depreciated:** No longer recommended.

**License:** © FAO & ART Group, 2015
VocBench 2.0: multilinguismo

...and multilingual UI: (currently: English, Spanish, Dutch, Thai)

multilingual editing

...and visualization
VocBench Facts (1)

Utilizzato da Enti di rilevanza nazionale e internazionale (tra cui)

- Food and Agriculture Organization (FAO) > AGROVOC, Biotechnology, Land and Water, FAO Topics
- EU Publications Office > EUROVOC
- Italian Senate > TESEO
- United Nations Environment Programme (UNEP) > InforMEA, in the context of MEA, the Information and Knowledge Management Initiative
- European Environment Agency (EEA) > GEMET
- Harvard University > Unified Astronomy Thesaurus (UAT)
- EC Parliament Library
- Institut National de la Recherche Agronomique (INRA) > in the context of AnaEE France project (ecosystems and biodiversity thesaurus) and Délégation à l'information scientifique et technique (DIST)
- CABI
- United Nations Convention to Combat Desertification (UNCCD)
- Scottish Government > Gov metadata
- Columbia University > IEDA Thesaurus
VocBench Facts (2)

Pubblicazioni di riferimento

- Manuel Fiorelli, Maria Teresa Pazienza and Armando Stellato *Semantic Turkey goes SKOS: Managing Knowledge Organization Systems*, ISemantics 2012 Graz, Austria, 5-7 September, 2012

**Futuro: VocBench 3.0**

- completamente basato sul sistema Semantic Turkey
- consentirà di sviluppare collaborativamente sia tesauri che ontologie
- supporto avanzato per ontology evolution & validation
Contributi Principali Tor Vergata al Multilinguismo

• **OntoLing** (2006) [1]

• **Linguistic Watermark** [2]: una suite di librerie e ontologie per l’uso di risorse lessicali e per il loro interfacciamento con risorse ontologiche

• il già citato [3]: **VocBench** (2009 on)

• **OntoLex**: partecipazione al Community Group W3C Ontology-Lexica (OntoLex) e contributi (portati da esperienza Linguistic Watermark) al modulo principale per l’interfacciamento ontologie e lessici

• **Lime (Linguistic Metadata)** [4]: modulo per i metadati di OntoLex sviluppato principalmente da noi e con contributi di altri membri della comunità scientifica

---


