
Carla Geovana do N. Macário
Advisor: Claudia Bauzer Medeiros

September, 2008
Outline

- Introduction
- Objective
- Related Work
- Proposed Solution
- Expected Contributions
- Publications
Introduction

Agriculture corresponds to 25% of Brazil's GNP
Introduction

Strategic questions:

- When will be the best time to start harvesting coffee in this area?
- What is the expected yield of my sugar cane farm?

Answers to these questions involve a combined use of soil, climate, maps and temperature data, among others.
Introduction

Web
Introduction

Web

orange?
Introduction
Introduction

- The existing solutions do not focus on agricultural domain
- They mainly consider textual data
- Novel solutions must be found to support adequate discovery, retrieval and management of geospatial data on the Web, with agriculture in mind
Semantic Annotation

Describes a content and what it does, by means of formal concepts (e.g. using entities of an ontology)

It is represented by a set of metadata that provide a reference to each annotated entity by its unique Web identifier (e.g. URI).

**In this work:** a semantic annotation is a set of one or more metadata, with each one describing a given digital content using ontologies terms.
Objective

Provide a mechanism for semantic annotation of geospatial, distributed and heterogeneous data available on the Web -- satellite images, graphs, maps, etc – geared towards the generation of strategic information for agriculture.
Objective

The framework will support:

- creation, validation and management of semantic annotations of geospatial data;
- discovery, access and effective search for data in agricultural contexts.

The framework will be part of WebMAPS project
Traditional Approach

- Semantic Annotation of Digital Content
  - Annotate a content using the terms of a domain ontology
Traditional Approach

- Semantic Annotation of Digital Content
  - Annotate a content using the terms of a domain ontology
Traditional Approach

• Semantic Annotation of Digital Content
  – Annotate a content using the terms of a domain ontology
# Non Spatial Annotation Mechanisms

<table>
<thead>
<tr>
<th>Tool</th>
<th>Format</th>
<th>Ontology</th>
<th>Storage</th>
<th>Automated</th>
<th>Annotation Method</th>
<th>Annotated data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Embrapa Information Agency</td>
<td>XML, using Dublin Core metadata</td>
<td>no</td>
<td>Relational database</td>
<td>no</td>
<td>Manual, using natural language</td>
<td>Textual Web pages, videos, images and documents</td>
</tr>
<tr>
<td>Amaya</td>
<td>XML, RDF</td>
<td>no</td>
<td>Local files</td>
<td>yes, but very limited</td>
<td>Based on given parameters</td>
<td>Textual Web pages</td>
</tr>
<tr>
<td>Kim</td>
<td>RDF, OWL</td>
<td>yes</td>
<td>Local files or in an annotation server</td>
<td>yes</td>
<td>String matching and ML</td>
<td>Textual Web pages</td>
</tr>
<tr>
<td>AKTive Media</td>
<td>RDF</td>
<td>yes</td>
<td>Local files</td>
<td>yes</td>
<td>ML(induction), with continuous manual training</td>
<td>Textual Web pages and images</td>
</tr>
<tr>
<td>CREAM</td>
<td>RDF, OWL</td>
<td>yes</td>
<td>Local files or in an annotation server</td>
<td>yes, with supervised learned</td>
<td>ML(induction) manual training</td>
<td>Textual Web pages, videos and images</td>
</tr>
</tbody>
</table>
The use of ontologies is not enough: it should also consider the **spatial component**, that one that associates objects and events to localities.
Spatial Annotation Mechanisms

Are there hotels in the Campinas region?
Spatial Annotation Mechanisms

Focus mainly in textual data, where the spatial component is explicit
## Spatial Annotation Mechanisms

<table>
<thead>
<tr>
<th>Tool</th>
<th>Format</th>
<th>Ontology</th>
<th>Automated</th>
<th>Annotation Method</th>
<th>Annotated data</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-Culture</td>
<td>RDF, OWL, using VRA metadata</td>
<td>yes</td>
<td>no</td>
<td>Manual, using a structured schema</td>
<td>Images of painting</td>
</tr>
<tr>
<td>OnLocus</td>
<td>XML</td>
<td>yes</td>
<td>yes</td>
<td>geospatial evidences (addresses)</td>
<td>Textual Web pages</td>
</tr>
<tr>
<td>SPIRIT</td>
<td></td>
<td>yes</td>
<td>yes</td>
<td>geospatial evidences</td>
<td>Textual Web pages</td>
</tr>
<tr>
<td>Geodata Annotation</td>
<td>XML, using ISO 19115 metadata</td>
<td>yes</td>
<td>yes</td>
<td>Spatial methods, string matching</td>
<td>Geographic data</td>
</tr>
</tbody>
</table>
Annotation of Geospatial Data for Agriculture

A lot of strategic data for agriculture can not be annotated using the traditional approach.
Annotation of Geospatial Data for Agriculture

A lot of strategic data for agriculture can not be annotated using the traditional approach.

E.g. NDVI graph, which indicates the biomass conditions of a vegetation area
Annotation of Geospatial Data for Agriculture

A lot of strategic data for agriculture can not be annotated using the traditional approach.

How to annotate it?
The Proposed Solution

To explore the feature of each kind of content, focusing in agriculture
The Proposed Solution

To explore the feature of each kind of content, focusing in agriculture
WebMAPS

Client layer

Service layer

Data layer

Service layer

User applications

ontology service

annotation service

workflow service

catalog service

textual data service

géoprocessing services

Data layer

Semantic data

Ontologies

Annotations

Olther data

workflows

Textual data

Geospatial data

Temporal Series

Image Descriptors

Satellite images, maps

User applications

Client layer

25/46
WebMAPS

Client layer

Service layer

Data layer

- ontology Service
- Annotation Service
- workflow service
- Catalog Service
- textual data service
- geoprocessing services

- Semantic data
  - Ontologies
  - Annotations
- Other data
  - workflows
  - Textual data
- Geospatial data
  - Temporal Series
  - Image Descriptors
  - Satellite images, maps

User applications

IC/UNICAMP
The Proposed Annotation Service

1. Analysis
2. Generation of annotations
3. Semantic Catalog

External data sources

Aondê

Content
Metadata
Provenance
Geospatial evidence

Creation process
The Proposed Annotation Service

1. Analysis
2. Generation of annotations
3. Semantic Catalog

External data sources

Geospatial evidence

Creation process

Metadata

Provenance
The Proposed Annotation Service

1. Analysis
2. Generation of annotations
3. Semantic Catalog

External data sources

Geospatial evidence

Content

Metadata

Provenance

Creation process

Aondê

Semantic Catalog
Workflow for content Annotation

- Creation process
- Content
- Metadata
- Provenance
- Geospatial evidence

Annotation schema → Link to ontology → Annotations
Workflow for content Annotation

- creation process
- content
- metadata
- provenance
- Geospatial evidence

Annotation schema

Link to ontology

Annotations

<Extended_information>
<crop />
<quality_of_production />
<harvesting_period />
</Extended_information>

<Identification.Information>
<Spatial_Data_Organization.Information>
<Indirect_Spatial_Reference>
<System />
<county />
<state />
<country />
</Indirect_Spatial_Reference>
</Spatial_Data_Organization.Information>
</Identification.Information>

<Spatial_Reference.Information>
<Horizontal_Coordinate_System_Definition>
<Geographic>
<Latitude_Resolution />
<Longitude_Resolution />
</Geographic>
</Horizontal_Coordinate_System_Definition>
</Spatial_Reference.Information>
Workflow for content Annotation

- Creation process
- Content
- Metadata
- Provenance
- Geospatial evidence

Annotation schema → Link to ontology → Annotations

getSimilarCurves → getTheirQuality
getProductName
IdentifyCropStages
getCoordinates → getCountyName

FAOSTAT → generateAnnotation
IBGE

Embrapa
Workflow for content Annotation

creation process
content
metadata
provenance
Geospatial evidence

Annotation schema
Link to ontology
Annotations

<Identification_Information>
<Description>
<Abstract "NDVI_graph"/>
</Description>
ANNOTATION_SCHEMA

<Native_Data_Set_Environment "MODIS Vegetation"/>
<Keywords>
<Theme>
<Theme_Keyword_Thesaurus "ISO 19155 Topic Category"/>
<Theme_Keyword>
farming "agriculture"
</Theme_Keyword>
</Keywords>
<Extended_information>
crop "sugar cane"
<quality_of_production "high"/>
<harvesting_period "11"/>
</Extended_information>

<Spatial_Data_Organization_Information>
<Indirect_Spatial_Reference>
<System "IBGE"/>
<county "Piracicaba"/>
<state "Sao Paulo"/>
<country "Brazil"/>
</Indirect_Spatial_Reference>
</Spatial_Data_Organization_Information>

<Spatial_Reference_Information>
<Horizontal_Coordinate_System_Definition>
<Geographic>
<Latitude_Resolution "-22:43:31"/>
<Longitude_Resolution "-47:38:57"/>
</Geographic>
</Horizontal_Coordinate_System_Definition>
</Spatial_Reference_Information>

<Extended_information>
crop "sugar cane"
<quality_of_production "high"/>
<harvesting_period "11"/>
</Extended_information>

<Indirect_Spatial_Reference>
<System "IBGE"/>
<county "Piracicaba"/>
<state "Sao Paulo"/>
<country "Brazil"/>
</Indirect_Spatial_Reference>

<Indirect_Spatial_Reference>
<System "IBGE"/>
<county "Piracicaba"/>
<state "Sao Paulo"/>
<country "Brazil"/>
</Indirect_Spatial_Reference>
The Proposed Annotation Service

Analysis (1) → generation of annotations (2) → Semantic Catalog (3)

External data sources

Geospatial evidence

content

metadata

provenance

creation process
Proposed Solution

What is the expected yield of my sugar cane farm?
Proposed Solution

What is the expected yield of my sugar cane farm?

1)
Proposed Solution

What is the expected yield of my sugar cane farm?

2)
Proposed Solution

What is the expected yield of my sugar cane farm?

3)
Proposed Solution

What is the **expected yield** of my sugar cane farm?

4) Yield = 80 ton/ha
Methodology

1. Analysis and comparison of existing annotation tools ✓
2. Design of workflows, annotation schemas and services – in course
3. Implementation – in course
4. Validation
Implementation

YAWL Workflow management system
Implementation

SPOT sensor images for South America

The Power of Expressiveness

<table>
<thead>
<tr>
<th>Administrate</th>
<th>Workflow Specifications</th>
<th>Available Work</th>
<th>Checked Out Work</th>
</tr>
</thead>
</table>

- Produce Annotation
- County Name XML
- Indirect Spatial Reference
  - System: IBGE
  - County: Piracicaba
  - State: Sao Paulo
  - Country: Brazil
- Product XML: sugar cane

Submit | Suspend | Save | Cancel | Refresh

* - required | ? - help
Expected Contributions

- Proposal of an annotation mechanism directed to the agricultural context
- Specification of scientific processes describing the generation of semantic annotations
Expected Contributions

- The framework, which considers the following requirements:
  - annotation of any kind of agricultural geospatial data;
  - discovery of geospatial data in the Web, their fusion and/or integration;
  - end user validation of the annotations generated;
  - automate the annotation process as much as possible.
Publications


Presentation and publication of this work in ACM GIS Ph.D. Dissertation Showcase and in Brazilian WTDDDB
Acknowledgments

- Profa. Claudia B. Medeiros
- CNPTIA/Embrapa
- LIS
- IC/Unicamp
- CNPq, FAPESP