SeCoGIS 2009

A Semantic Approach to Describe Geospatial Resources

Sidney Roberto de Sousa

Master degree in Computer Sciences
Institute of Computing - UNICAMP
This work is financially supported by FAPESP
Grant number: 08/51028-4

Agenda

- Introduction and Motivation
- Annotation
- Semantic Annotation
- Prototype
- Storage of semantic annotations
- Conclusions

Introduction and motivation

- Web: large repository of geospatial data
- Retrieval of these data requires special attention
- Solution: geographic metadata standards and geospatial information portals
- Search: use of keywords and metadata fields
- Problem: natural language text

Introduction and motivation

- Use of ontologies
- Represent knowledge about some domain of interest
- Geographic domain: terms and concepts to describe geospatial information
- e.g.: spatial references, geographic formats, etc.
- Improve retrieval of geospatial information

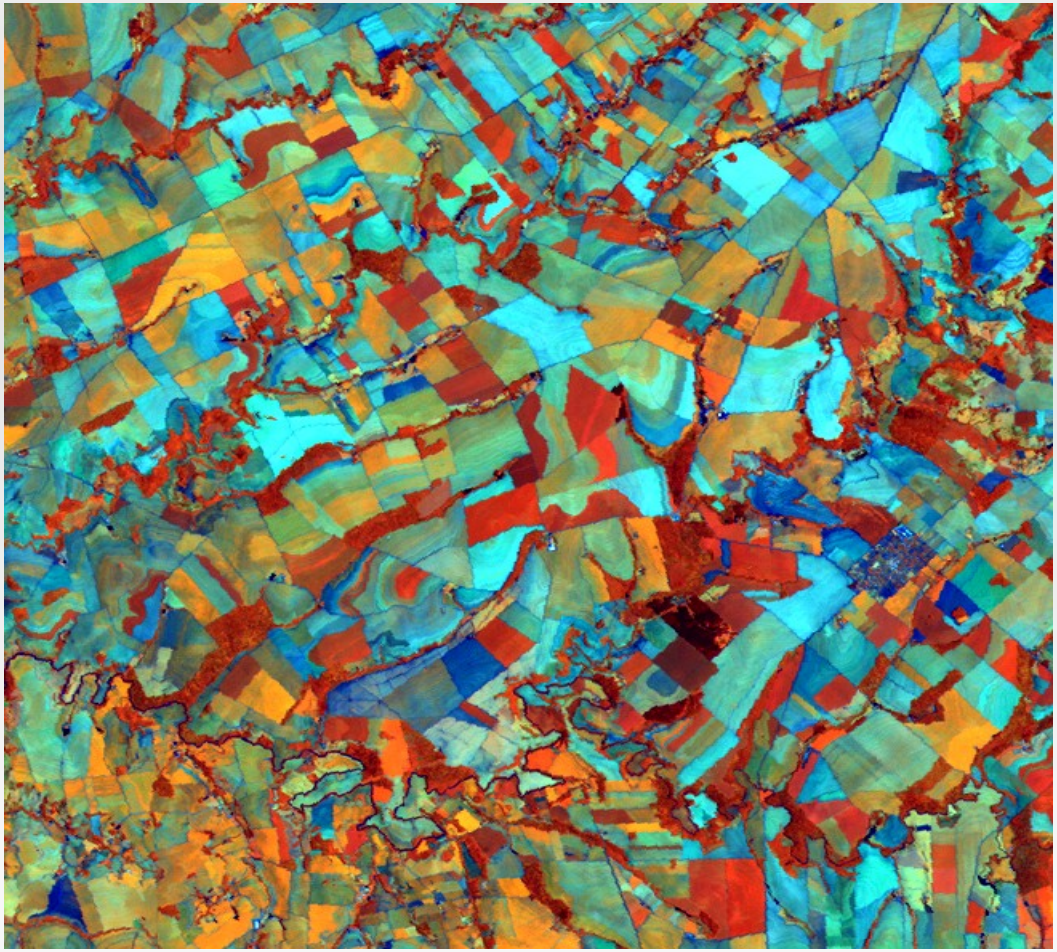
Proposal

- Use of semantic annotations for describing geospatial resources
- Semantic annotation: set of triples
- Triple: <*resource, metadata field, ontology term*>
- Study of storage mechanisms for semantic annotations

Proposal

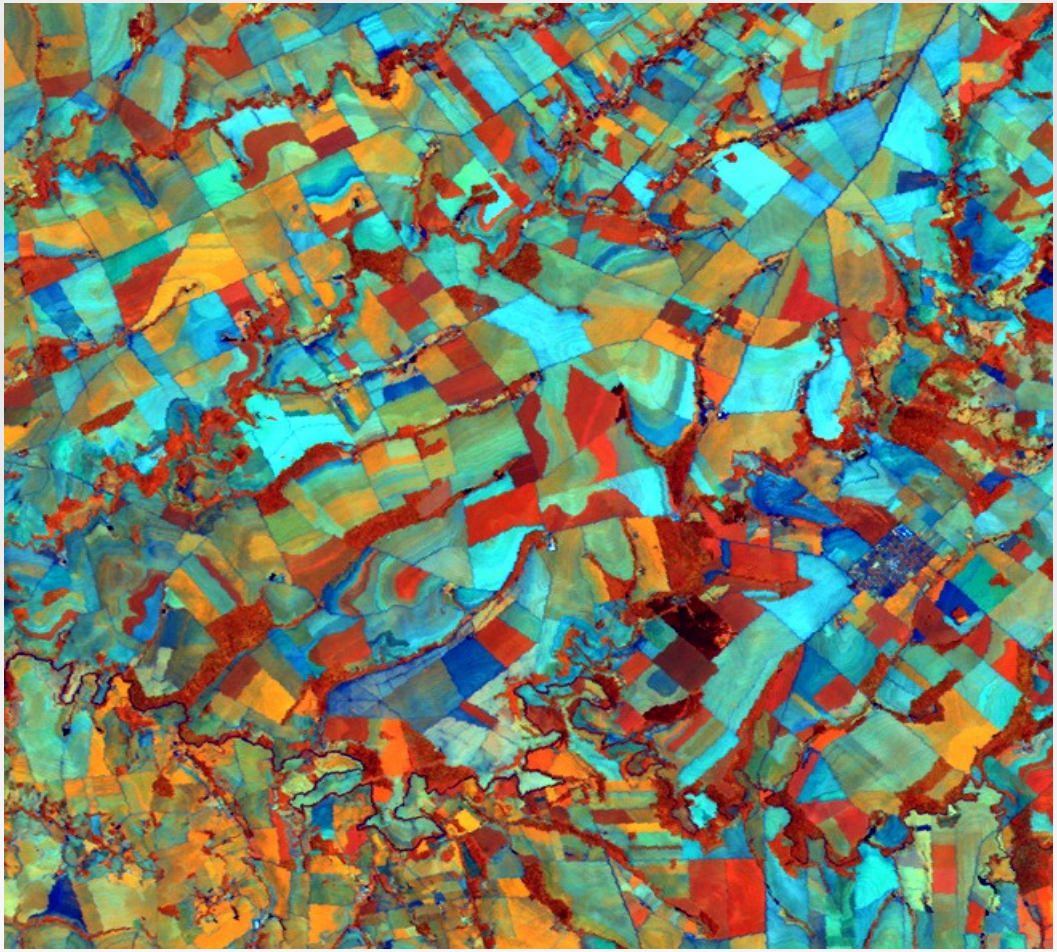
- eFarms project: multidisciplinary project combining research in Computer Science and Agricultural Sciences
- It attacks problems involving agricultural data management and low cost wireless data communication
- The prototype developed in this work will be part of the eFarms tools set

Annotation



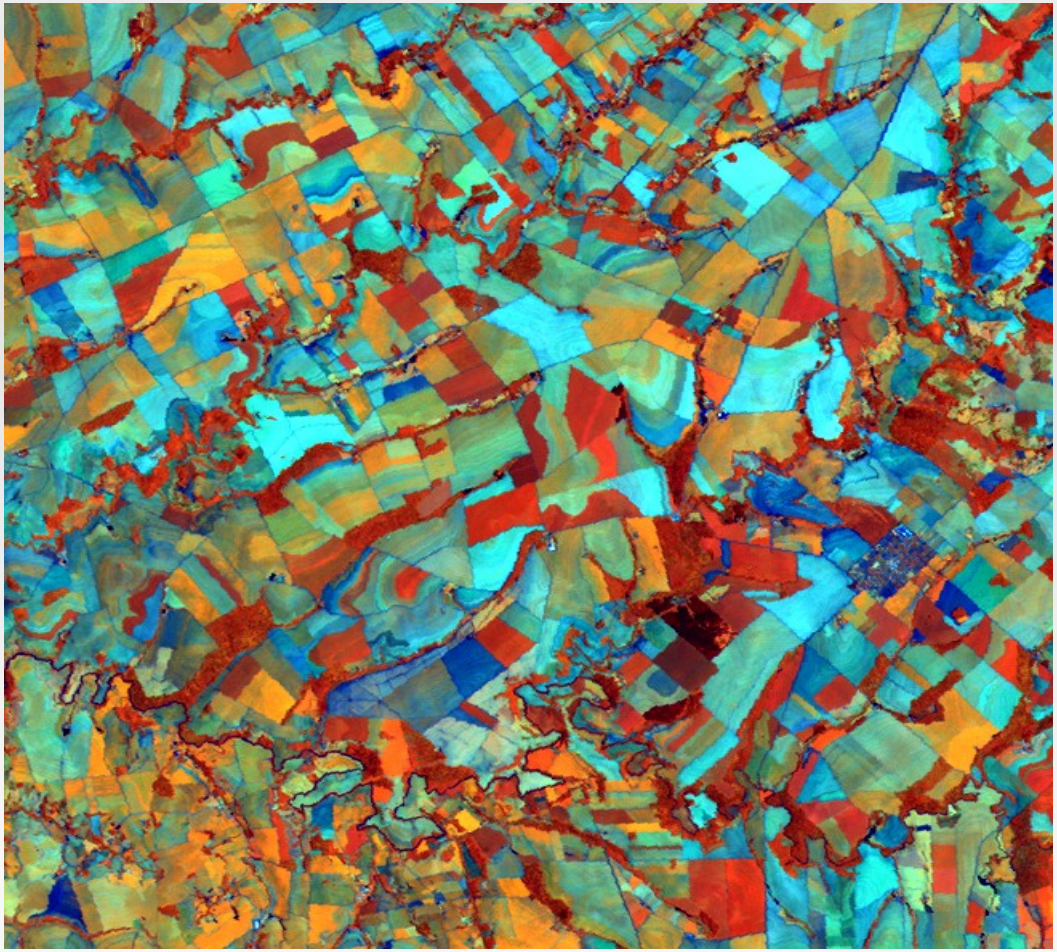
Remote sensing image for a carioca bean crop from Irecê city, BA, Brazil. Taken by Landsat satellite.

Annotation



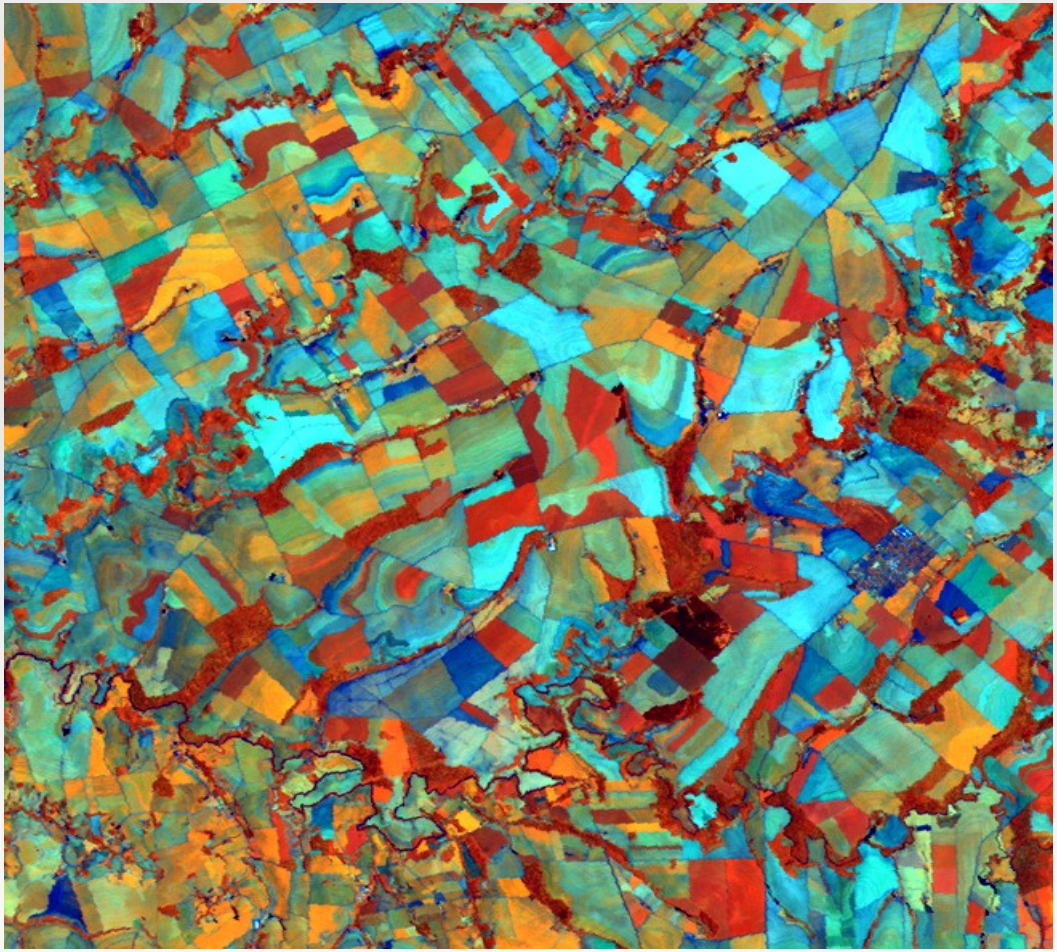
Remote sensing image for a carioca bean crop from Irecê city, BA, Brazil. Taken by Landsat satellite.

Metadata structured annotation



Format: Remote sensing image
City: Irecê
State: Bahia
Country: Brazil
Issue: arabica coffee crop
Source: Landsat satellite

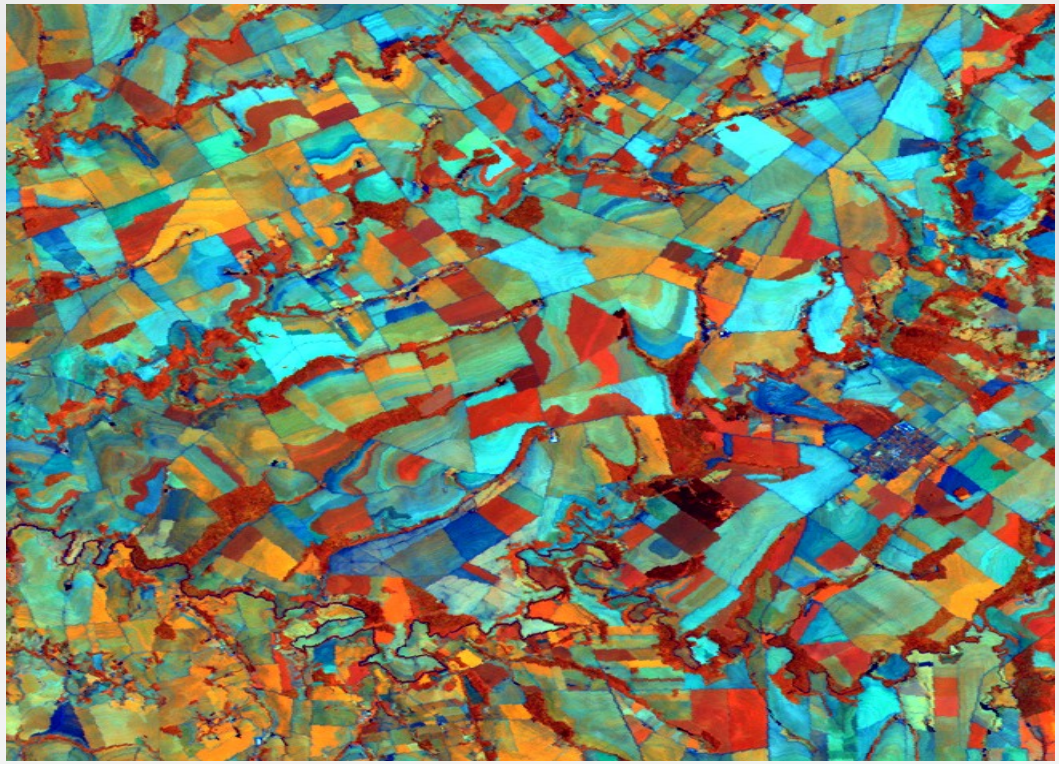
Using standard metadata



FGDC

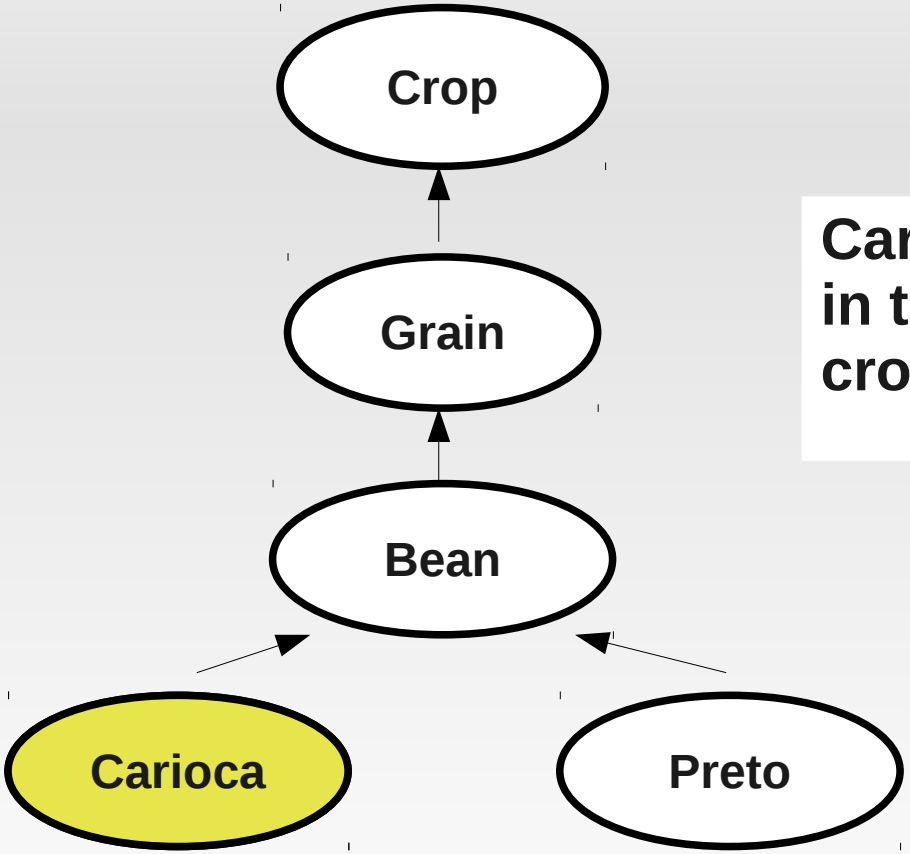
Geographic Format:
Imagem de sensoriamento remoto
City: Irecê
State: Bahia
Country: Brazil
Issue: carioca bean crop
Originator: Landsat satellite

It is possible to improve it!



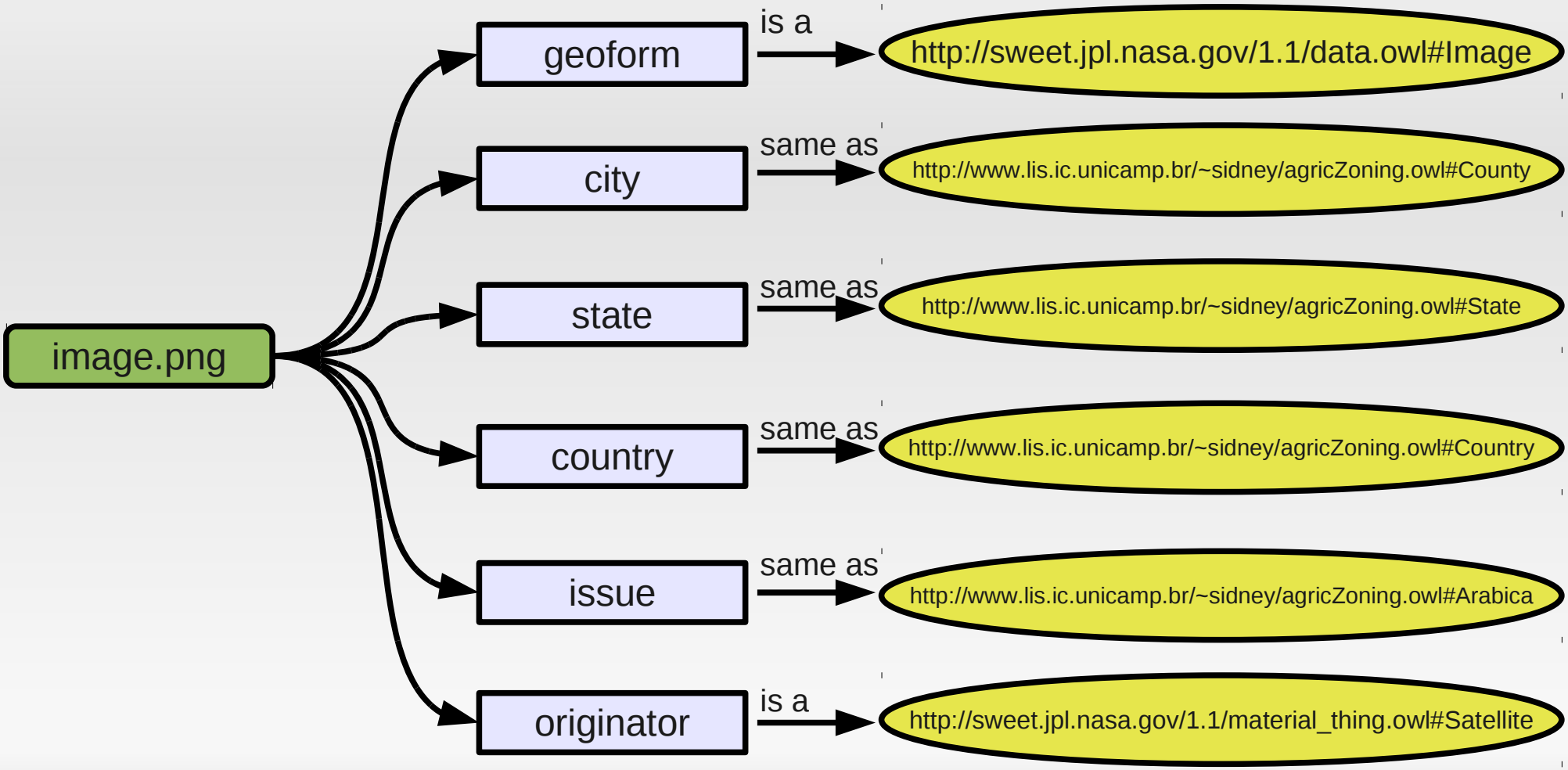
Issue:
http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Carioca

Why is it better?



Carioca is a kind of bean, that in turn is a kind of grain and a crop too.

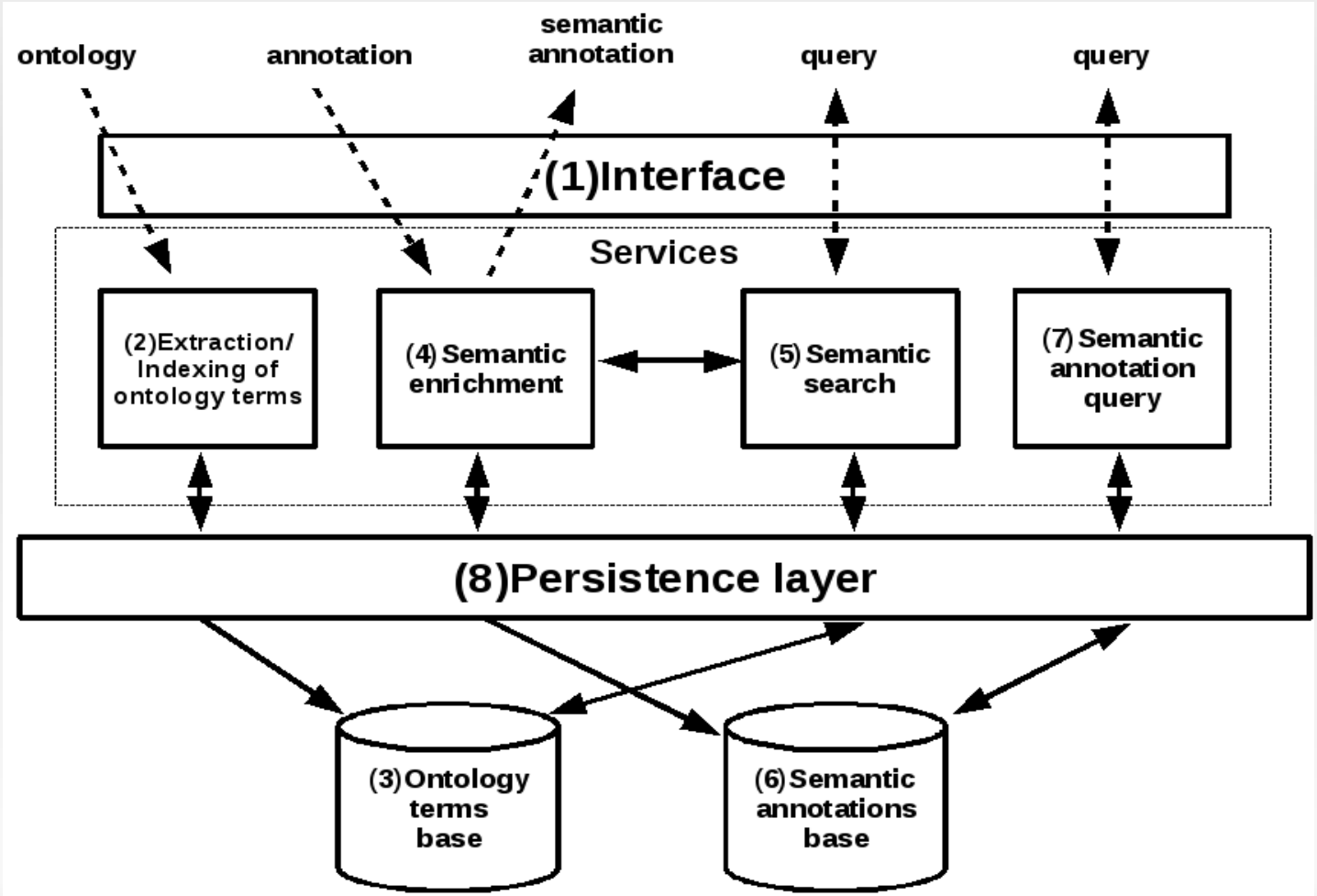
Semantic annotations



Prototype

- Tool for creating semantic annotations
- Case study: agricultural resources
- Metadata used: FGDC + agricultural extension
- Already implemented services (and interfaces):
  - Extraction and indexing of ontology terms
  - Semantic search
  - Creation of semantic annotations

Architecture



Extraction/Indexing of ontology terms

- It extracts and indexes ontology terms (classes)
- Lucene API → text indexing, storage and search
- Term → <termURI, {tags}>
- Tags → className + synonyms + superClassesNames

Example

- Ontology class: **Bean**
- Synonyms (WordNet): bean plant, dome, noodle, attic, bonce, noggin
- Ontology superclasses: Grain, Crop, Raw, AgriculturalProduct
- {http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Bean, {bean, bean plant, dome, noodle, attic, bonce, noggin, grain, crop, raw, agricultural, product}}

Semantic search

- Lucene API
- Search performed over tags
- API performs terms ranking according to tags occurrence
- Search service performs pos-ranking according to syntactic similarity

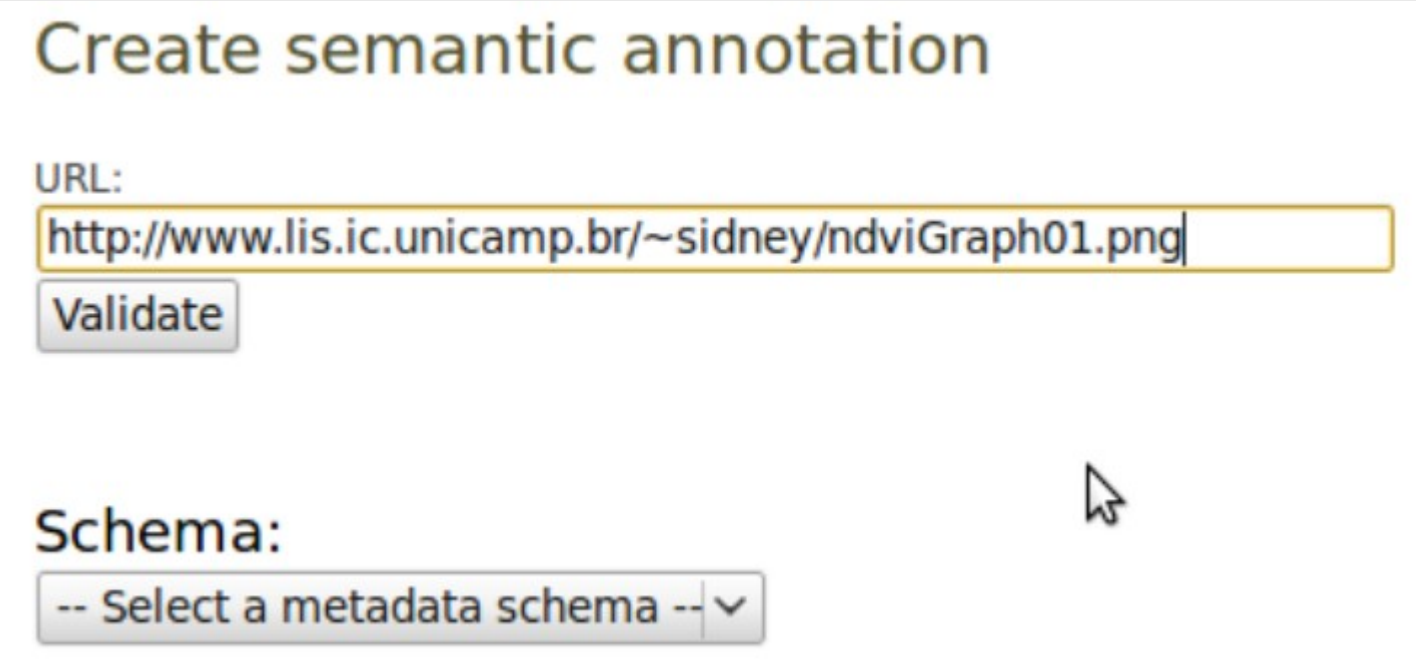
Search example

Search:
crop

Submit

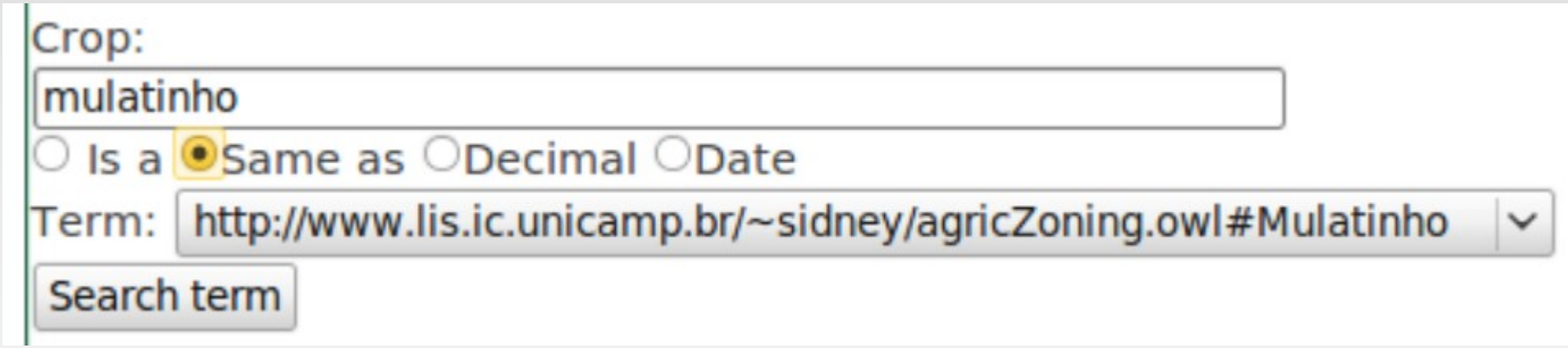
1 http://sweet.jpl.nasa.gov/2.0/biolPlant.owl#Crop
2 http://sweet.jpl.nasa.gov/1.1/biosphere.owl#Crop
3 http://morpheus.cs.umbc.edu/aks1/ontosem.owl#crop
4 http://sweet.jpl.nasa.gov/1.1/human_activities.owl#CropProduct
5 http://sweet.jpl.nasa.gov/1.1/property.owl#CropMoistureIndex
6 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#CropGroup
7 http://sweet.jpl.nasa.gov/1.1/human_activities.owl#CroppingSystem
8 http://sweet.jpl.nasa.gov/1.1/human_activities.owl#CropsProduction
9 http://sweet.jpl.nasa.gov/2.0/biolPlant.owl#CropMoistureIndex
10 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#KindOfCrop
11 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Grapes
12 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#SpecificCrop
13 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Mango
14 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Crop
15 http://www.owl-ontologies.com/unnamed.owl#CropProduction
16 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Carioca
17 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Cotton
18 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#OtherCrop
19 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Robusta
20 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Fruit
21 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Grain
22 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Arabica
23 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Carrot
24 http://www.lis.ic.unicamp.br/~sidney/agricZoning.owl#Cucumber

Creation of semantic annotations



Metadata schema: FGDC / FGDC + agricultural extension

Creation of semantic annotations



Storage of semantic annotations

- Service not yet implemented
- Semantic annotations are represented in RDF
- RDF → a framework, not a language
- It can be represented in various languages
- More human-understandable → Notation3
- More machine-processable → RDF/XML
- RDF storage depends on the language used for representing it

Native XML databases

- Storage of RDF/XML
- No mapping needed
- Query language: XQuery e XPath
- Query based on DOM trees, not triples
- RDF/XML is handled as native XML, not RDF

RDF databases

- Frameworks for handling RDF
- e.g.: Sesame and Jena
- In general, allow persistence in binary files or relational databases
- Query language: SPARQL, RDQL or proprietary languages
- Query based on RDF triples
- Allow serialization in various RDF languages

Relational databases

- Requires mapping from RDF to relational model
- Query language: SQL
- More complex implementation

Related work

Link to several research initiatives:

- semantic interoperability on GIS
- geospatial resources description
- metadata storage

Semantic interoperability

- Use of ontologies to help information integration
- Use of ontologies to facilitate retrieval of geospatial information
- Controled vocabulary into the Geospatial Semantic Web

Description of geospatial resources

- Use of RDF to catalog geospatial information
- Use of ontologies to improve the use of geographic metadata

Metadata storage

- Storage of metadata into XML databases
- Solutions for efficient RDF metadata storage

Conclusions

- Proposal of an approach based on RDF, metadata, and ontologies for describing geospatial resources → semantic annotations
- Semantic Web + geographic standards
- Inclusion of ontologies in information description → unique meaning / no ambiguities
- The study of solutions for storing semantic annotations has considered RDF databases as the best solution

Conclusions

- Use of Lucene API for storing, indexing, and searching ontology terms
- Interface for creating semantic annotations

Future work

- Service for storing semantic annotations (URL's from resources as primary keys?)
- Use of ontology axioms in the triples (not just *is a*, *same as*, *decimal*, or *date*)
- Improvement of the tagging mechanism:
  - synonyms restriction
  - use of specialization (not just generalization)

Acknowledgements

- FAPESP-Microsoft Research Virtual Institute (eFarms project)
- CAPES and CNPq
- Prof. Claudia Medeiros
- Laboratory of Information Systems (LIS)
- Institute of Computing (IC-UNICAMP)

SeCoGIS 2009

Thank you!

sidney@lis.ic.unicamp.br