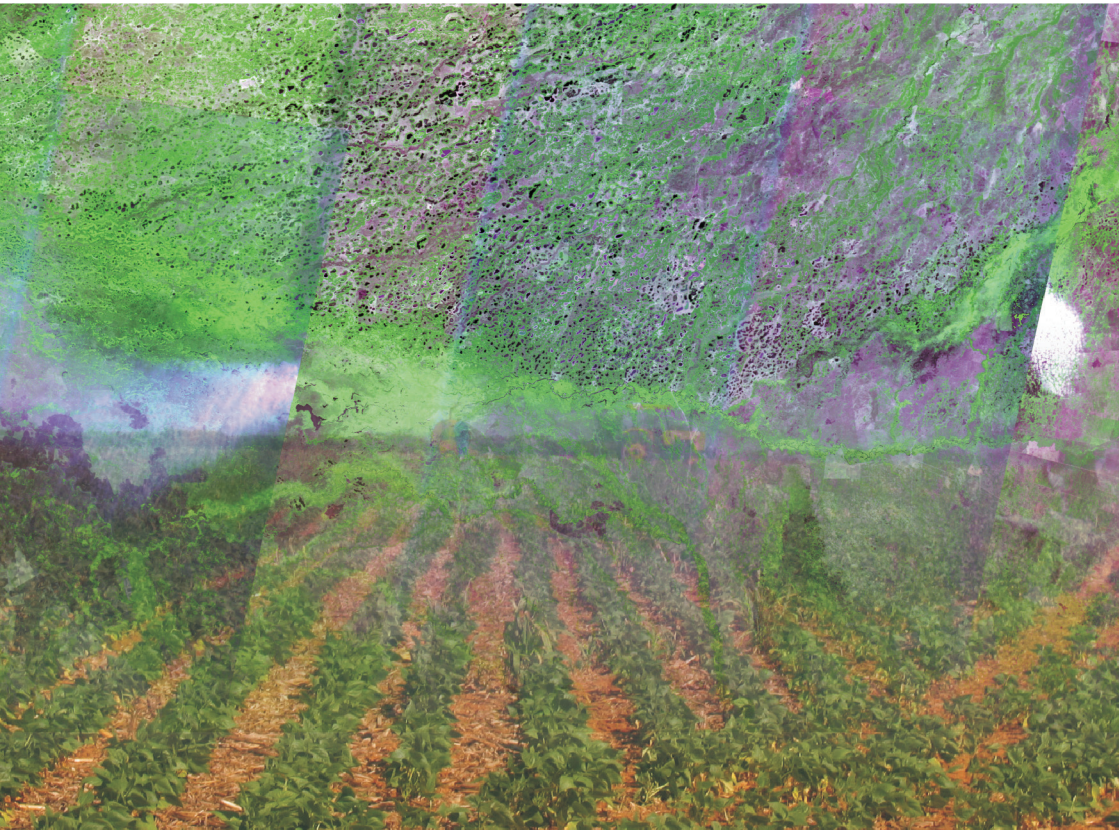


**Anotação semântica de dados  
geoespaciais para a agricultura**



*Empresa Brasileira de Pesquisa Agropecuária  
Embrapa Informática Agropecuária  
Ministério da Agricultura, Pecuária e Abastecimento*

# ***Boletim de Pesquisa e Desenvolvimento 25***

## **Anotação semântica de dados geoespaciais para a agricultura**

*Carla Geovana Nascimento Macário*

## **Embrapa Informática Agropecuária**

Av. André Tosello, 209 - Barão Geraldo  
Caixa Postal 6041 - 13083-886 - Campinas, SP  
Fone: (19) 3211-5700 - Fax: (19) 3211-5754  
www.cnptia.embrapa.br  
sac@cnptia.embrapa.br

### **Comitê de Publicações**

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa, Stanley Robson de Medeiros Oliveira, Maria Goretti Gurgel Praxedes, Adriana Farah Gonzalez, Neide Makiko Furukawa*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo, Paula Regina Kuser Falcão*

Supervisor editorial: *Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica: *Neide Makiko Furukawa*

Fotos da capa: *Embrapa Informática Agropecuária*

Secretária: *Carla Cristiane Osawa*

### **1ª edição on-line 2010**

#### **Todos os direitos reservados.**

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

#### **Dados Internacionais de Catalogação na Publicação (CIP) Embrapa Informática Agropecuária**

---

Macário, Carla Geovana Nascimento.

Anotação Semântica de Dados Geoespaciais para a Agricultura / Carla Geovana Nascimento Macário. - Campinas : Embrapa Informática Agropecuária, 2010.

32 p. : il. - (Boletim de pesquisa e desenvolvimento / Embrapa Informática Agropecuária, ISSN 1677-9266 ; 25).

1. Dados geoespaciais. 2. Anotação semântica. 3. Workflow científicos. I. Título. II. Série.

005.1015118 CDD 21 ed.

---

© Embrapa 2010

# Sumário

<b>Resumo</b> .....	5
<b>Abstract</b> .....	6
<b>1 Introdução e motivação</b> .....	7
<b>2 Aspectos de pesquisa envolvidos</b> .....	8
2.1 Interoperabilidade de processos geoespaciais .....	9
2.2 Heterogeneidade de dados geoespaciais .....	10
2.3 Padrões de metadados e ontologias geográficos .....	10
2.4 Anotação semântica .....	11
2.5 Automação do processo de anotação .....	12
2.6 <i>Workflows</i> científicos .....	13
<b>3 Material e métodos</b> .....	13
3.1 Metodologia adotada .....	14
3.2 Desenvolvimento do trabalho .....	14
<b>4 Resultados e discussão</b> .....	16
4.1 Anotações semânticas .....	16
4.2 Visão geral do <i>framework</i> .....	18
4.3 Arquitetura do <i>framework</i> .....	20
4.4 Detalhes de implementação .....	21

<b>5 Validação</b> .....	22
5.1 Estudo de caso - planejamento agrícola no Brasil .....	23
5.2 Configurando o <i>framework</i> .....	23
5.2.1 Definindo um esquema de anotação .....	23
5.2.2 Selecionando ontologias.....	24
5.2.3 Projetando e desenvolvendo os <i>workflows</i> .....	24
5.3 Produzindo as anotações - a execução do <i>workflow</i> .....	25
<b>6 Conclusões</b> .....	28
<b>7 Agradecimentos</b> .....	28
<b>8 Referências</b> .....	29

# Anotação semântica de dados geoespaciais para a agricultura

---

*Carla Geovana Nascimento Macário*

## Resumo

Dados geoespaciais são base para sistemas de decisão em vários domínios. Para serem usados esses dados precisam ser analisados e interpretados por especialistas. Essas interpretações geralmente não são armazenadas ou correspondem apenas a alguma informação textual e em linguagem própria, gravadas em arquivos técnicos. A ausência de soluções eficientes para armazená-las leva a problemas como retrabalho e dificuldades de compartilhamento de informação. Este trabalho apresenta uma solução para esse problema que baseia-se no uso de anotações semânticas, uma abordagem que promove um entendimento comum dos conceitos usados. Com a adoção de *workflows* científicos e também de um esquema de metadados e de ontologias bem conhecidos, foi especificado e parcialmente implementado um *framework* para anotação semântica de dados geoespaciais, focando na solução de problemas em agricultura.

**Termos para indexação:** Dados geoespaciais; anotação semântica, *workflow* científicos

---

<sup>1</sup> Doutora em Ciência da Computação, Pesquisadora da Embrapa Informática Agropecuária, [carla@cnptia.embrapa.br](mailto:carla@cnptia.embrapa.br)

# Semantic annotation of geospatial data for agriculture

---

## Abstract

Geospatial data are a basis for decision making in a wide range of domains. To be used, such data have to be analyzed and interpreted. Usually, these interpretations are not stored, and when this happens, they correspond to descriptive text, which is stored in technical files. The absence of solutions to efficiently store them leads to problems such as rework and difficulties in information sharing. This work presents a solution for these problems based on semantic annotations, an approach for a common understanding of concepts being used. We propose the use of scientific workflows to describe the annotation process for each kind of data, and also the adoption of well known metadata schema and ontologies. This framework was partially validated, through an implementation to produce annotations for applications in agriculture.

**Index terms:** Geospatial data; semantic annotation; scientific workflows

# 1 Introdução e motivação

O termo *dado geoespacial* refere-se a todos os tipos de dados sobre objetos e fenômenos do mundo que têm características espaciais e que referenciam alguma localidade na superfície da Terra. Esses dados constituem a base para sistemas de decisão usados em vários domínios -- por exemplo, em sistemas de transporte, planejamento agrícola ou ambiental. Em especial na agricultura são úteis para responder questões como “o que, onde, quando, e como plantar”. Sendo a agricultura uma atividade de destaque no Brasil, cujos ganhos correspondem à aproximadamente 30% do Produto Interno Bruto do país, melhorias no acesso e no uso desse tipo de dado permitiriam mais eficiência no planejamento e na previsão de culturas. Como consequência, teríamos o aumento nos valores obtidos com a produção.

Para serem usados, esses dados precisam ser analisados e interpretados, atividades muitas vezes trabalhosas e dependentes do contexto e do domínio de uso. Essas interpretações, quando armazenadas, geralmente correspondem a alguma informação textual, em linguagem própria e gravadas em arquivos técnicos. A ausência de soluções eficientes para armazenar essas interpretações leva a problemas como retrabalho e dificuldades de compartilhamento de informação.

Uma solução para esses problemas baseia-se no uso de anotações, que no contexto deste trabalho são definidas como dados que descrevem dados. Entretanto, a simples adoção de anotações não é suficiente, já que cada especialista ou empresa pode usar linguagem ou métodos de descrição próprios, criando barreiras para o entendimento da informação. Para reduzir esse problema, uma opção é usar, na descrição, termos de ontologias, como forma de prover semântica. A utilização de ontologias dá origem ao uso das chamadas *anotações semânticas*, uma abordagem que promove um entendimento comum dos conceitos usados, garantindo a interoperabilidade semântica entre produtores e consumidores da informação. Soluções existentes focam na anotação manual de dados - atividade trabalhosa e sujeita a erros. Além disso, consideram principalmente dados textuais. No entanto, dados geoespaciais compreendem também informação não textual, como mapas e imagens de satélites. Assim, novas aborda-



gens devem ser propostas para descrever esse tipo de dados, da maneira mais automática possível.

Este trabalho contribuiu para resolver esse desafio. Seu principal objetivo era prover um mecanismo para anotação semântica de dados geoespaciais, como forma alternativa de armazenamento da interpretação associada a esse tipo de dados, aplicando a solução em problemas em agricultura. Anotações semânticas facilitam a recuperação de dados anotados, aumentando o seu reuso e apoiando especialistas na aplicação de métodos de planejamento e previsão de culturas.

O trabalho desenvolvido combina pesquisa em anotação semântica, metadados geoespaciais e uso de *workflows* científicos, de maneira a fornecer um ambiente de anotação de dados geoespaciais genérico, com facilidades de extensão. O restante deste texto está organizado da seguinte forma: a Seção 2 discute alguns tópicos de pesquisa associados ao trabalho e a Seção 3 descreve a metodologia utilizada no seu desenvolvimento; a Seção 4 apresenta os resultados obtidos com o trabalho, descrevendo a especificação do mecanismo de anotação proposto, aspectos de implementação e validação. Por fim, a Seção 5 apresenta as contribuições e extensões do trabalho.

## 2 Aspectos de pesquisa envolvidos

Uma anotação é representada por um conjunto de metadados que provê uma referência para cada entidade anotada por seu identificador único, como uma URI. O uso de termos de uma ontologia na descrição do conteúdo é uma das abordagens para promover interoperabilidade (UREN et al, 2006). O aumento na qualidade da informação recuperada é um dos benefícios da adoção de anotações.

A anotação semântica de dados geoespaciais combina vários aspectos de pesquisa em computação. Os principais desafios encontrados envolvem questões em anotação semântica, especificação de *workflows* científicos e manipulação de dados heterogêneos, destacando-se: adoção de um padrão de metadados, que seja suficiente para descrever os diferentes tipos

de dados a serem anotados e, ao mesmo tempo, facilite a interoperabilidade; definição de conjunto mínimo de requisitos para a busca semântica de dados geoespaciais; definição de conjunto de características básicas que uma ferramenta de anotação deve apresentar; adoção de tecnologia capaz de prover a anotação semântica para diferentes tipos de dados, sejam eles textuais ou não.

A solução adotada envolveu todos esses aspectos, definindo o uso de *workflows* científicos como ferramenta básica para o processo de anotação. Esses pontos são analisados em seguida, sempre com o foco na agricultura, nosso estudo de caso.

## 2.1 Interoperabilidade de processos geoespaciais

Interoperabilidade de processos geoespaciais requer o manuseio apropriado de dados. A proposta desse trabalho lida com a anotação desses dados, o que requer o acesso a diferentes fontes de dados, como dados de temperatura, clima e produtividade de culturas. Em 1999, especialistas da área de Descoberta de Conhecimento Geográfico conduziram um encontro para identificar prioridades de pesquisa nessa área. Uma das questões foi a incorporação, às aplicações geográficas, de conhecimento descoberto por meio de consultas à base de dados distribuídas. (MILLER; HAN, 2000).

Dez anos depois isso ainda é uma questão em aberto. A Web Semântica para informação geográfica, denominada por (EGENHOFER, 2002) de Web Semântica Geoespacial, é uma forma de processar requerimentos envolvendo diferentes tipos de informação geoespacial. Segundo o autor, ela captura e analisa a informação, agrupando-a segundo critérios que extrapolam seu significado sintático. Todo esse processo requer o desenvolvimento de múltiplas ontologias espaciais e de domínio, sua representação numa forma compreensível e processável pelos computadores, o processamento de consultas considerando essas ontologias e a avaliação dos resultados baseados na semântica requerida.

Infelizmente, a Web Semântica ainda é uma realidade distante e, embora vários esforços venham sendo desenvolvidos, ainda há muito a ser feito. Vocabulários consensuais e ontologias são difíceis de se estabelecer

e manter. Além disso, a maioria dos mecanismos de buscas continuam restritos a texto, e a consulta semântica a outros tipos de mídia colocam inúmeros desafios para o desenvolvimento efetivo da Web Semântica (CARVALHO et al., 2006).

## **2.2 Heterogeneidade de dados geoespaciais**

Dados geoespaciais podem ser de diferentes tipos. Como especificar um mecanismo que permita a anotação desses diferentes tipos de dados e que seja genérico o suficiente para lidar com as questões de heterogeneidade? Foi testado e analisado um conjunto de ferramentas de anotação de dados, algumas das ferramentas consideram o componente espacial (BORGES, 2006; HOLLINK, 2006; JONES et. al. 2004; KLIEN, 2007). Percebe-se que a maioria considera, principalmente, dados textuais, baseando o processo de anotação em métodos de aprendizagem de máquina. Como a identificação de anotações é baseada em casamento de strings, o uso de ontologias é essencial para a desambiguação e também a identificação correta de evidências espaciais. Entretanto, se esse conteúdo é uma imagem ou vídeo, o uso de ontologias não é suficiente para a anotação automática: o conteúdo tem de ser anotado manualmente. Com exceção de (KLIEN, 2007), as ferramentas analisadas não consideram outros tipos de dados como imagens de satélite, mapas e gráficos.

## **2.3 Padrões de metadados e ontologias geográficos**

A produção de anotação semântica para dados geoespaciais requer o acesso a diferentes fontes de dados, como informação sobre clima, zoneamento agrícola e produtividade de culturas. Aliado a isso, há o problema de heterogeneidade dos dados. Uma abordagem para lidar com essas dificuldades é a adoção de padrões de metadados e de ontologias, visando prover um conjunto de terminologias e de definições para documentação e troca de dados. Exemplos de padrões de metadados bem estabelecidos e usados na descrição de dados geográficos são o ISO 19115 Metadata Standard (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2008) e o FGDC geospatial metadata standard (FEDERAL GEOGRAPHIC DATA COMMITTEE, 1998). A descrição de dados geralmente é disponibili-

zada em portais geoespaciais, permitindo que usuários realizem consultas e recuperem dados (NOGUERAS-ISO et. al., 2005) .

Qual o melhor esquema de anotação? Quais requisitos devem ser considerados nessa escolha? Esse esquema pode ser adotado diretamente em aplicações agrícolas? Questões como essas direcionaram a adoção do padrão FGDC, o qual foi estendido para acomodar necessidades específicas e agricultura.

Enquanto padrões proveem uma solução estruturada para o problema de heterogeneidade, ontologias consideram semântica. Uma ontologia é uma especificação explícita e formal de elementos de um domínio e os relacionamentos entre eles (GRUBER, 1993). Um vocabulário comum definido para um domínio facilita o compartilhamento de informação, mecanismos de busca e reúso (BECK; PINTO, 2002). Spaccapietra et. al. (2004) consideram dois tipos principais de ontologias:

- ontologias descritivas, que são aquelas que enriquecem a descrição da semântica de conceitos, associando a cada conceito uma descrição estruturada de suas propriedades; essas ontologias compartilham com o esquema conceitual da base de dados o esforço de modelar alguns domínios e atividades;
- ontologias espaço-temporais, que são aquelas que levam em conta características espaço-temporal das informações (elementos e relacionamentos espaciais e elementos e relacionamentos temporais). A dificuldade no seu desenvolvimento é a ausência de um modelo apropriado que seja capaz de lidar com espaço e tempo em nível ontológico, propiciando um mecanismo de busca adequado.

Ontologias vêm sendo usadas na computação para a integração de banco de dados e organização de conhecimento em banco de dados. Entretanto, o desenvolvimento e a adoção de ontologias comuns ainda são desafios (SHADBOLT et. al., 2006). Portanto, este trabalho focou na adoção de ontologias existentes e bem conhecidas.

## 2.4 Anotação semântica

Anotar significa adicionar notas, comentar. Em computação, uma anotação é usada para descrever um recurso (comumente textual) e o que ele faz,

usando conceitos formais (por exemplo, termos de uma ontologia). Uma anotação é representada como um conjunto de metadados que proveem uma referência a cada entidade anotada por meio de um identificador único da web, como um URI. Uma maneira de prover interoperabilidade é usar as entidades de uma ontologia de domínio como conceitos (UREN et. al., 2006). Por exemplo, uma anotação pode relacionar a palavra *laranja* que ocorre num texto a uma ontologia que indica que essa palavra está relacionada ao conceito abstrato fruta e não ao conceito cor. Isso ajuda a remover ambiguidade do significado. O aumento na qualidade da informação recuperada e a interoperabilidade são benefícios da adoção de anotações.

No caso de aplicações geográficas, anotações devem considerar também o componente espacial, já que informação geográfica associa objetos e eventos a lugares, por meio de um vocabulário de nomes de lugares e objetos geográficos, relacionamentos espaciais e padrões. Desta forma, o processo de anotação geográfica deve ser baseado em evidências geográficas -- aquelas que conduzem a um lugar ou fenômeno geográfico.

Como anotações semânticas podem ser de difícil leitura, decidiu-se manter descrições em linguagem natural para descrever as unidades de anotação, sendo isso parte da pesquisa desenvolvida.

## 2.5 Automação do processo de anotação

Parte da eficiência do processo de anotação vem da forma como ele é realizado. Reeve e Han (2005) destacaram que há dois métodos principais de anotação: baseado em padrões e baseado em aprendizagem de máquina. Sistemas baseados em padrões são aqueles que tem um conjunto inicial de entidades definidas -- manualmente ou não. Essas entidades são entendidas como padrões a serem encontrados no conteúdo. Caso novas entidades sejam descobertas, elas se tornam novos padrões. O processo continua recursivamente até que não existam mais entidades a serem descobertas ou que o usuário finalize o processo.

Sistemas baseados em aprendizagem de máquina se concentram no projeto e desenvolvimento dos algoritmos que permitam aos computa-

dores “aprender”. Esses sistemas utilizam dois métodos: probabilidade e indução. O primeiro usa modelos estatísticos para indicar localização de entidades em um texto -- por exemplo, identificar componentes de endereços (número, prédio, cidade). Os métodos de indução, por sua vez, extraem regras e padrões de conjuntos de dados, reusando-os em processos subsequentes de anotação.

O processo de anotação deve ser o mais automático possível, uma vez que o processo manual pode ser lento e sujeito a erros. Isso permanece como um desafio que vem sendo abordado por inúmeros projetos de pesquisa (GREENBERG et. al., 2006). Entretanto, a maioria deles propõem mecanismos que consideram apenas a anotação de recursos textuais, não levando em conta outros conteúdos como mapas e imagens de satélite. Mecanismos para anotar esses dados são poucos, tendo sido uma das motivações deste trabalho.

## 2.6 Workflows científicos

*Workflows* científicos emergiram como um paradigma para representação de gerenciamento de computações científicas distribuídas. Tais *workflows* capturam as transformações individuais de dados e etapas de análise, bem como os mecanismos para realizar etapas em um ambiente distribuído (WORKSHOP ON THE CHALLENGES OF SCIENTIFIC WORKFLOWS, 2006). A sua adoção para gerenciar o processo de anotação aparece como uma boa opção. Entretanto, isso leva a novos desafios, dada a complexidade demandada por essas anotações.

## 3 Material e métodos

Essa seção discute alguns tópicos de pesquisa associados ao trabalho, focando na anotação semântica de dados geoespaciais para a agricultura. Tendo como base esses desafios, bem como os objetivos do trabalho, é descrita sucintamente a metodologia utilizada no seu desenvolvimento

### 3.1 Metodologia adotada

Para atingir os objetivos propostos, atacando os desafios científicos existentes, o trabalho foi desenvolvido considerando, então, os seguintes aspectos:

- identificação de requisitos para o mecanismo de anotação;
- identificação de características que descrevem cada tipo de fonte de dados;
- combinação de ferramentas visando à descrição automática das características identificadas;
- automação, o máximo possível, do processo de automação;
- descrição de cada passo desse processo;
- armazenamento e gerenciamento do processo de anotação.

### 3.2 Desenvolvimento do trabalho

Com o objetivo de atacar as questões científicas apresentadas na seção 3, o trabalho teve início com a avaliação de portais e catálogos bem conhecidos que publicam dados geoespaciais, como o FAO e GOS. Deste trabalho derivaram-se as principais características e requisitos para acesso a dados. Detalhes deste trabalho podem ser obtidos em (MACÁRIO; MEDEIROS, 2009b).

O passo seguinte foi o teste de um grupo de ferramentas de anotação bem conhecidas, como a AKTiveMedia (CHAKRAVARTHY et. al., 2006) e a CREAM (HANDSCHUH; STAAB, 2002), que apresentam métodos para anotação semântica de recursos visuais. Isso mostrou o que deve ser apoiado por um mecanismo de anotação automática, como ele funciona e como deveria considerar a informação espacial. Nesse estudo percebeu-se que algumas das ferramentas não automatizam totalmente o processo de anotação, muitas delas funcionando como ferramentas de ajuda para o processo de anotação. Durante esse teste, estávamos interessados na avaliação do processo de anotação de fontes de dados não textuais. Entretanto, encontramos apenas ferramentas que anotavam dados textuais. A única exceção foi em (KLIEN 2007), que propõe um modelo teórico para anotação semântica de mapas, baseado em operações espaciais. A avaliação completa das ferramentas pode ser vista em (MACÁRIO; MEDEIROS, 2009a).

O resultado desse passo foi a identificação de requisitos importantes para o *framework* de anotação, como a necessidade de estabelecer um esquema de metadados, armazenamento e formato de anotações, e métodos de anotação. Isso permitiu a especificação da arquitetura de um *framework* para apoiar o mecanismo. Uma vez que o *framework*, diferente das ferramentas testadas, deveria considerar vários tipos de dados, decidimos usar *workflows* científicos para especificar cada processo de anotação.

A validação da proposta teve início com a anotação de um gráfico de NDVI<sup>2</sup>. Com a entrevista de especialistas, identificamos as seguintes características para essa fonte de dados: nome e localidade de uma cultura, valores de produtividade histórica e esperada para a cultura, e fases de produção de uma cultura. Também descrevemos como obter e produzir essa informação. De posse disso, o passo seguinte foi buscar fontes de dados e ferramentas a serem usados. Decidimos usar as fontes de dados de produtividade e de localidade do Instituto Brasileiro de Geografia e Estatística (IBGE), imagens de satélite do sensor Moderate Resolution Imaging Spectroradiometer (MODIS) e uma ferramenta que executa consulta baseada em similaridade de séries temporais (MARIOTTE et. al., 2007). Em seguida, foi necessário projetar um *workflow* descrevendo como combinar o acesso às fontes de dados selecionados, com o objetivo de produzir as anotações desejadas. Outra questão foi a definição de um esquema de metadados a ser usado. Embora tenha sido adotado o padrão FGDC (FEDERAL GEOGRAPHIC DATA COMMITTEE, 1998), percebeu-se que ele não era suficiente pra prover informações necessárias para a agricultura. Assim, foi proposta uma extensão para o padrão, focando em campos adicionais.

O passo seguinte foi a implementação da arquitetura proposta. Para a execução dos *workflows* de anotação, escolheu-se usar o ambiente YAWL (VAN DER ALAST; HOFSTEDE, 2005). Isso apresentou alguns desafios, o que nos levou a algumas decisões de implementação. Tais desafios e sua solução estão descritos em (MACÁRIO et. al., 2009a).

---

<sup>2</sup> *Normalized Difference Vegetation Index* – um valor computado de pixels de imagens de satélite, os quais indicam a quantidade de biomassa numa região.



Durante esse processo identificou-se a necessidade de uma fase de configuração, indicando quais termos de ontologias deveriam ser considerados durante o processo de anotação. Também foi necessário decidir sobre o armazenamento de anotações. Essa parte foi implementada por Sousa (2010), em seu trabalho de mestrado.

Finalmente o mecanismo de anotação e o *framework* teve de ser validado considerando outros tipos de dados. Decidimos pela anotação de imagens de sensoriamento remoto, usadas para identificar áreas de cultura em uma dada área geográfica (MACÁRIO et. al., 2010).

## 4 Resultados e discussão

Essa seção apresenta, de maneira geral, a solução desenvolvida para o problema identificado, considerando os pontos de pesquisa ressaltados na seção 2. Inicialmente, descrevemos a aplicação da metodologia proposta com alguns dos resultados obtidos. Em seguida, definimos anotações semânticas no contexto deste trabalho, descrevemos o processo de anotação e finalmente apresentamos a arquitetura adotada e detalhes da sua implementação.

### 4.1 Anotações semânticas

Uma anotação semântica combina conceitos de metadados e ontologias: metadados são campos que são preenchidos com termos de uma ontologia, sendo estes usados para descrever esses campos. Definimos anotações semânticas como se segue (MACÁRIO et. al., 2009):

- **Unidades de Anotação:** uma *unidade de anotação a* é uma tripla  $\langle s, m, v \rangle$ , em que  $s$  é o objeto sendo descrito,  $m$  é o rótulo do campo de metadado e  $v$  é o seu valor ou descrição.
- **Anotação:** uma *anotação A* é um conjunto de uma ou mais unidades de anotação.
- **Unidades de Anotação Semântica:** uma *unidade de anotação semântica sa* é uma tripla  $\langle s, m, o \rangle$ , em que  $s$  é o objeto sendo descrito,  $m$  é

o rótulo do campo de metadado e o é um termo de uma ontologia de domínio.

- **Anotação Semântica:** uma *anotação semântica SA* é um conjunto de uma ou mais unidades de anotação semântica.
- **Esquema de Anotação e Conteúdo:** uma anotação/anotação semântica tem um *esquema* e um *conteúdo*. O esquema é a sua estrutura, especificada por campos de metadados. O conteúdo corresponde aos valores desses campos.

Enquanto unidades de anotação descrevem dados usando linguagem natural, unidades de anotação semântica usam termos de ontologias e podem ser processados por uma máquina. As unidades de anotação são especificadas usando tuplas, de maneira similar à estrutura RDF. Isso ajuda o subsequente armazenamento e reuso. Usuários, entretanto, manipulam as anotações em formatos mais amigáveis.

Como um exemplo, as Figuras 4.1 e 4.2 respectivamente ilustram uma unidade de anotação e a unidade semântica correspondente. Destacamos que a solução apresentada é baseada no armazenamento de ambas anotações, a primeira para consumo do usuário (em linguagem natural), enquanto que a segunda é processável por máquina.

```
-<metadata>
-<extendinfo>
-<location>
  <county>Monte Santo de Minas</county>
  <state>Minas Gerais</state>
  <country>Brazil</country>
</location>
-<product>
  <name>Arabica Coffee</name>
</product>
</extendinfo>
</metadata>
```

**Figura 4.1.** Anotação gerada para uma imagem de sensoriamento remoto.

```

<fgdc:formcont>
-
<rdf:Bag>
<rdf:li rdf:resource="http://sweet.jpl.nasa.gov/ontology/biosphere.
owl#Crop"/>
<rdf:li>Coffe crop</rdf:li>
</rdf:Bag>
</fgdc:formcont>
</fgdc:digtinfo>
</fgdc:digform>
-
<crop>
<rdf:Bag>
<rdf:li rdf:resource="http://www.lis.ic.unicamp.br/ont/agricZoning.
owl#Arabica"/>
<rdf:li>Arabica Coffee</rdf:li>
</rdf:Bag>
</crop>
</rdf:Description>

```

**Figura 4.2.** Parte de uma anotação semântica gerada para a mesma imagem de sensoriamento remoto.

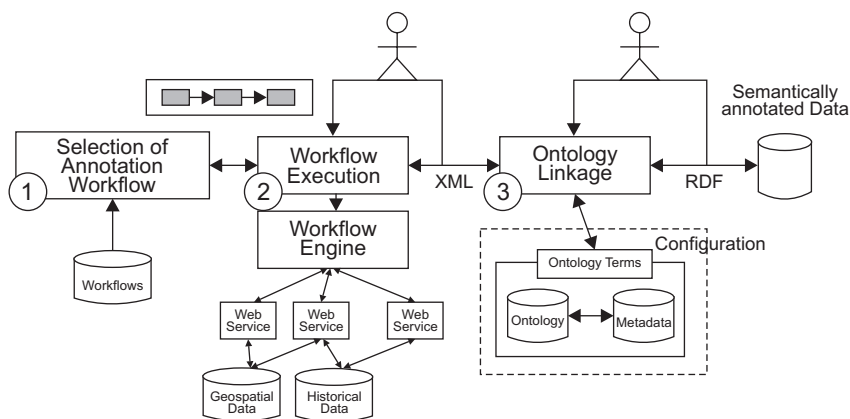
## 4.2 Visão geral do *framework*

A premissa básica deste trabalho é que informação geoespacial pode ser usada para acelerar o processo de anotação, aliviando a tarefa dos especialistas de domínio. Outra premissa é que, para vários tipos de dados geoespaciais, existem procedimentos centrais que podem ser especificados por esses especialistas. Tais procedimentos podem ser construídos para atingir demandas de anotação específicas de contexto.

Com essas premissas, nosso cenário de anotação é o seguinte: Primeiro, especialistas precisam definir esses procedimentos centrais de anotação para cada tipo diferente de dados geoespaciais (p.ex. mapas temáticos, imagens de satélite e séries temporais de dados de sensores). Cada um desses procedimentos é especificado e armazenado como um *workflow*. Assim, sempre que determinada fonte de dados precisa ser anotada, o *workflow* correspondente é executado, gerando uma anotação básica, que pode ser subsequentemente validada pelo especialista. Além disso, esses *workflows* podem ser especializados para atingir necessidades específicas (por exemplo, considerando uma certa cultura agrícola).

Embora sistemas especialistas sejam frequentemente usados em sistemas de anotação (KLIEN; LUTZ, 2005; REEVE; HAN, 2005), nem todos processos de anotação podem ser descritos por sistemas de decisão. Além disso, estamos lidando com fenômenos geográficos. Assim, optou-se pela utilização de *workflows* científicos para descrever o processo de anotação (FILETO et. al., 2003; TSALGATIDOU et. al., 2006). Cada *workflow* contém informação sobre o esquema de anotação que será usado durante o processo, as ontologias para descrever esses dados e as operações a serem executadas.

A Figura 4.3, reproduzida de (MACÁRIO et al., 2009), fornece uma visão geral do processo de anotação apoiado pelo *framework*, o qual tem três passos principais: seleção do *workflow* de anotação, execução do *workflow* e ligação às ontologias. O *workflow* orquestra a geração das unidades de anotação. No último passo, cada unidade de anotação é transformada em uma unidade de anotação semântica, trocando o conteúdo em linguagem natural por uma referência a um termo da ontologia associada. Usuários podem intervir para validar as anotações sendo produzidas.



**Figura 4.3.** A anotação de dados geoespaciais - principais passos.

Com um pouco mais de detalhe, o *framework* recebe como entrada um arquivo de dados geoespaciais para ser anotado, bem como dados de sua proveniência. O tipo do dado é identificado, indicando o *workflow* a ser executado. O *workflow* descreve o esquema da anotação e as opera-

ções a serem executadas para produzir o conteúdo da anotação. Durante esse processo, as unidades de anotação produzidas são apresentadas ao usuário para validação, que pode escolher um outro *workflow* ou definir um novo. Na etapa final, termos de ontologias são escolhidos para montar as anotações semânticas (ligando as unidades de anotação aos termos de ontologias). As anotações semânticas são armazenadas como triplas RDF em uma base de dados XML, de onde poderão ser recuperadas, por exemplo, usando declarações XQuery.

Aqui um dos principais desafios é a especificação do *workflow*, de tal forma que as anotações possam ser geradas. Isso requer não apenas a especificação do *workflow* em si, mas também de suas atividades - por exemplo, implementação como serviços web. Além disso, o esquema de anotação para cada fonte de dados geoespaciais deve ser definido. Isso é o que denominamos de passo de configuração e que é necessário para preparar o *framework* para anotar alguns tipos específicos de dados geoespaciais.

### 4.3 Arquitetura do *framework*

A arquitetura do *framework* é dividida em duas partes: (1) gerenciador de anotações, serviços de anotações e ligador às ontologias, e (2) camada de persistência, que inclui o gerenciador do banco de dados. Essa arquitetura básica foi projetada levando em conta questões de interoperabilidade e é ilustrada na Figura 4.4, reproduzida de (MACÁRIO et. al., 2009). Caixas brancas correspondem a módulos externos invocados pelo *framework*.

O gerenciador de Anotações gerencia a execução dos passos apresentados na Figura 4.4, funcionando como um controlador de eventos. Ele recebe requisições de anotação de dados, identifica o tipo do dado e cria a requisição para o *workflow* correspondente. Esse *workflow* será executado pelo Sistema de Gerenciamento de *Workflow* - *Workflow Management System* (WfMS) e uma vez que a anotação está pronta e validada, ela é encaminhada ao Ligador de Ontologias, para a associação com termos da ontologia. O Serviço de Anotação Corresponde aos serviços implementados e que são invocados pelo *workflow* de anotação para gerar o conteúdo desejado. O Gerenciador do Banco de Dados atua como um mediador, provendo interoperabilidade para as bases de dados em uso. Essas bases

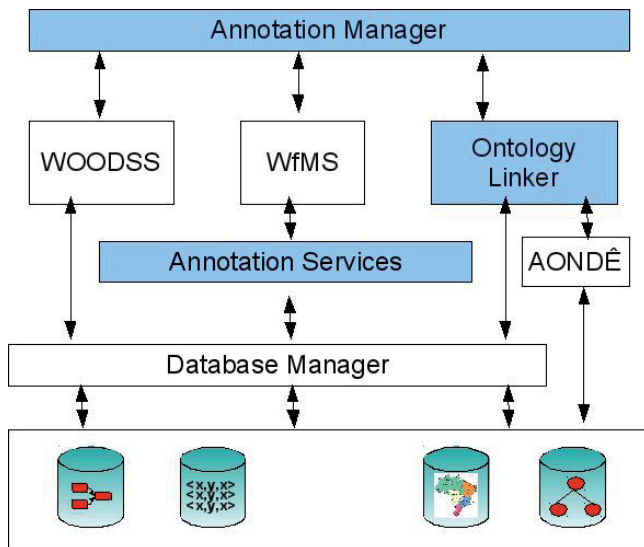


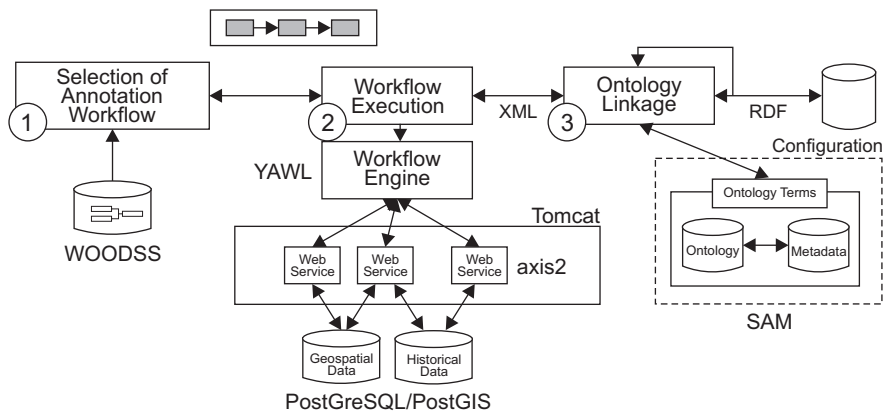
Figura 4.4. Arquitetura do *Framework*.

de dados armazenam os *workflows* de anotação, ontologias, dados geoespaciais anotados e outros dados geoespaciais que são usados por esses serviços (por exemplo, séries históricas de informação de produtividade de culturas ou séries temporais para uma localidade ou um fenômeno, como chuvas e temperatura).

#### 4.4 Detalhes de implementação

Implementamos um protótipo para validar o processo de anotação proposto. A Figura 4.5, fornece uma visão geral das tecnologias usadas nessa implementação. Usamos o sistema WOODSS (MEDEIROS et al., 2005) para a edição dos *workflows* de anotação. Esses *workflows* são traduzidos para o ambiente YAWL (VAN DER ALAST; TER HOFSTEDÉ, 2005), um ambiente escolhido para a sua execução. O YAWL foi escolhido por estar disponível para uso e por importar com facilidade o esquema de metadados do FGDC.

Cada tarefa no *workflow* de anotação é responsável por produzir uma ou mais unidades de anotação, por meio da invocação de um serviço web. Os



**Figura 4.5.** Visão geral da implementação.

serviços web foram implementados usando a linguagem java e também o *framework* Axis2, que gera automaticamente os *stubs* para os serviços. Esses serviços acessam dados geoespaciais e informação histórica armazenados em bases de dados sob controle do sistema de gerenciamento de banco de dados PostgreSQL e sua extensão para dados geográficos PostGIS, e executam operações específicas, como consultas espaciais, para gerar ou obter a informação desejada. O servidor Tomcat foi usado como um *container* para os serviços implementados.

As anotações produzidas são armazenadas como arquivos XML. Essas anotações são traduzidas em anotações semânticas durante a execução da etapa Ligação às Ontologias. A etapa de configuração do *framework* é executada usando a infraestrutura fornecida pelo Gerenciador de Anotação Semântica (SAM) (SOUSA, 2010), que provê meios para a indexação e recuperação de termos de uma ontologia por sua semântica. O SAM também executa a etapa de Ligação aos termos das Ontologias.

## 5 Validação

O *framework* implementado foi validado com a anotação de dois tipos de dados: um gráfico de NDVI (MACÁRIO; MEDEIROS, 2009c) e uma imagem de sensoriamento remoto (MACÁRIO et. al., 2010). Foram usados

dados de produtividade e de municípios do IBGE. O padrão de metadados FGDC foi escolhido por ser um padrão internacional e aberto. Nesta seção será apresentado o segundo desses casos, como ilustração do ambiente implementado. um estudo de caso que considera o uso desse *framework* de anotações para manipular com imagens de sensoriamento remoto para planejamento agrícola.

## 5.1 Estudo de caso - planejamento agrícola no Brasil

Vários fatores influenciam a estimativa e produção de culturas, como tipo de solo, relevo, clima e gerenciamento de práticas agrícolas. Tais fatores são também usados para zoneamento agrícola. Previsão e zoneamento são a base para definição de políticas para financiamento de atividades agrícolas.

Imagens de sensoriamento remoto são usadas intensivamente para planejamento e monitoramento agrícola e monitoramento de culturas, provendo uma base para a tomada de decisão. São usadas para identificar a extensão de uma plantação numa dada região, ou detectar sinais de problemas na saúde da planta (por exemplo, por ocorrência de pragas ou excesso de chuvas).

Especialistas agrícolas interpretam essas imagens para obter a informação desejada. Apresentamos o processo semi-automatizado para a geração de anotações de uma imagem de sensoriamento remoto para a região da cidade de Monte Santo de Minas, localizada em Minas Gerais, uma das regiões brasileiras com maior índice de produtividade de café. A Figura 5.2 (lado esquerdo) mostra uma imagem do satélite SPOT tirada em agosto de 2005.

## 5.2 Configurando o *framework*

### 5.2.1 Definindo um esquema de anotação

O interesse deste trabalho é o uso de dados geoespaciais para apoio à tomada de decisão em agricultura, portanto o esquema de metadados escolhido é baseado no padrão de metadados FGDC, que é genérico e aberto.



Estendemos esse padrão para prover campos adicionais para questões agrícolas, como informação sobre a produção de uma cultura. A execução do *workflows* produz informação para preencher cada um dos campos do esquema definido.

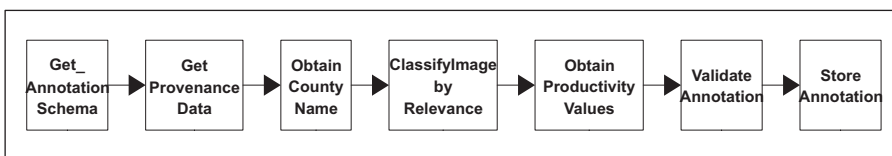
### 5.2.2 Selecionando ontologias

Especialistas são requisitados para escolher termos apropriados de ontologias que poderiam ser usados para produzir cada uma das unidades de anotação semântica, ou seja, para preencher os campos de metadados. Isso é executado pela ferramenta de extração de ontologias. Para cada ontologia escolhida, todos os termos são extraídos, rotulados e indexados em um dos repositórios do *framework*, para futura recuperação. Como no passo de ligação, os termos possíveis são associados a cada campo de anotação. Por exemplo, para o campo *crop* alguns termos possíveis são “www.lis.ic.unicamp.br/ont/agricZoning.owl#coffee” e “www.lis.ic.unicamp.br/ont/agricZoning.owl #sugarCane”, todos extraídos de POESIA (FILETO et. al., 2003), uma ontologia para zoneamento agrícola no Brasil.

### 5.2.3 Projetando e desenvolvendo os *workflows*

Especialistas precisam definir os processos de anotação, os quais são direcionados para cada tipo de fonte de dados geoespaciais. Por exemplo, para uma mesma região, existem diferentes processos de anotação para arquivos contendo imagens de satélite, fotos, séries temporais (ex. dados de precipitação) ou características de uma cultura.

A Figura 5.1 apresenta um *workflow* central para a anotação de uma imagem de sensoriamento remoto, descrevendo as principais tarefas a serem executadas.

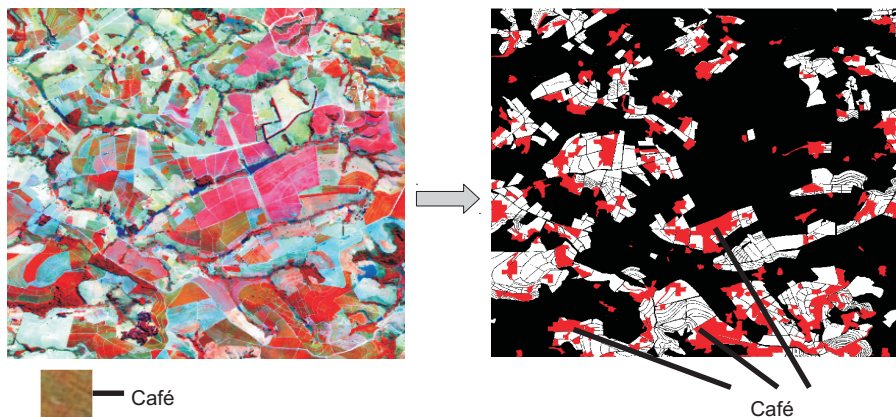


**Figura 5.1.** O workflow núcleo para a anotação de imagens de sensoriamento remoto

### 5.3 Produzindo as anotações - a execução do *workflow*

Neste trabalho adotamos o ambiente YAWL para execução de *workflow*. O primeiro passo, depois da recuperação do esquema de anotação, é obter informações de proveniência dos dados (por exemplo, nome do sensor, resolução espacial, data de aquisição e informação de coordenadas), que são extraídas do cabeçalho da imagem. A próxima atividade a ser realizada é obter o nome da cidade (“Obtain County Name” - ver Figura 5.1). Isso é obtido com a execução do serviço web CountyNameService, invocado pelo YAWL. A implementação desse serviço retorna o nome da cidade para um dado grupo de um ou mais pares de coordenadas geográficas. O serviço acessa uma fonte de dados do IBGE, que contém informação de todos os municípios, cidades e estados do Brasil. Se as coordenadas da imagem mostram que ela cobre mais de uma cidade, todos os nomes são retornados.

No passo seguinte, uma ferramenta de classificação de imagem é invocada. Essa ferramenta usa técnicas de processamento de imagem para identificar polígonos na imagem que casem com um certo padrão de entrada. Identificação e reconhecimento de um padrão de uma cultura em uma imagem de sensoriamento remoto é um processo difícil e por essa razão, uma ferramenta de apoio é uma importante ajuda para os especialistas (SANTOS et. al, 2009). Essa ferramenta recebe como entrada uma imagem padrão e uma imagem  $I$  a ser classificada. Como resultado provê uma nova imagem  $I'$ , composta por um conjunto de polígono identificados como contendo o padrão informado. Como ambas as imagens  $I$  e  $I'$  são georreferenciadas, é possível sobrepô-las e checar o resultado. A Figura 5.2, ilustra o processo de classificação. Tendo, como entrada, um padrão de café (quadrado pequeno na base esquerda inferior), a ferramenta divide a imagem em pequenas partes (considerando parâmetros como tamanho usuais de áreas para culturas específicas) e extrai suas características. Essas pequenas imagens são ordenadas de acordo com sua similaridades com as características do padrão de entrada. As partes mais similares são então convertidas em polígonos, os quais são apresentados como resultados para validação (Figura à direita). A ferramenta identifica corretamente mais de 66% dos talhões de café que já haviam sido identificados previamente em atividades de campo, o que é considerado um resultado excelente, quando comparado a métodos tradicionais de classificação (TUSK et al. 2003).



**Figura 5.2.** O processo de classificação de imagem

Fonte: Extraído de Santos et. al, 2009.

Esse processo de classificação pode ser executado para diferentes padrões de entrada para uma mesma imagem. No exemplo, além de café, há também áreas de cana-de-açúcar que podem ser identificadas usando a mesma ferramenta, tendo, como entrada, o padrão para cana-de-açúcar. Esse processo é dependente da espacialidade. Um determinado padrão de imagem pode ser associado a diferentes tipos de culturas, dependendo da região sendo examinada e da resolução da imagem.

A próxima atividade no *workflow* (Figura 5.1) é obter valores históricos de produtividade para essa cultura e região. A invocação do serviço *ProductivityService* retorna informação para uma certa cultura, para uma região específica em um ano. A informação retornada inclui valores de produtividade recuperados de uma outra base de dados do IBGE, que mantém dados oficiais de diferentes culturas agrupadas por região geográfica - macro e micro região, estado e município - e por ano. A Figura 5.3 ilustra um exemplo dessa informação histórica para diversos municípios. Ela indica que Monte Santo de Minas produziu 963 kg/ha de café para o ano de 2005 - ano em que essa imagem foi tirada.

Uma vez que a anotação é validada, as unidades de anotação são transformadas em unidades semânticas usando termos de ontologias selecionados durante a fase de configuração. A Figura 5.4, apresenta parte dessas

Unidades da Federação, Mesorregiões, Microrregiões e Municípios produtores	Área destinada à colheita (ha)	Área colhida (ha)	Quantidade produzida (t)	Rendimento médio (kg/ha)
São Sebastião do Paraíso	87 185	87 185	69 810	800
Juruáia	5 280	5 280	3 986	754
Monte Belo	4 500	4 500	3 321	738
Monte Santo de Minas	8 960	8 960	8 637	963
Muzambinho	7 000	7 000	4 802	686

**Figura 5.3.** Valores históricos de produtividade para café em Monte Santo de Minas.

```

<crop>
<rdf:Bag>
<rdf:li rdf:resource="http://www.lis.ic.unicamp.br/ont/agricZoning.
owl#Coffee"/>
<rdf:li> Coffee</rdf:li>
</rdf:Bag>
</crop>

```

**Figura 5.4.** Unidade de anotação semântica para imagem de sensoriamento remoto da região de Monte Santo de Minas.

anotações, correspondendo à extensão do esquema para considerar questões da agricultura. Por exemplo, a imagem contém talhões de café, que são identificadas pelo par <crop>, <recurso rdf>. Para mais detalhes no esquema de anotação adotado, unidades de anotação e ontologias usadas (MACÁRIO et. al, 2009).

Por fim, os especialistas devem validar as anotações semânticas criadas. Com base nessas anotações, um especialista do governo brasileiro pode confirmar a extensão espacial de uma cultura e computar valores de produtividade. Um outro uso importante é a possibilidade de identificação de ocorrência de doenças, o que podem impactar em pagamento de seguros. Como ganho adicional, nossas anotações, devido às descrições semânticas, podem aumentar o número de documentos relevantes recuperados numa operação de busca (fator *recall*).

## 6 Conclusões

Este trabalho descreveu a especificação e a implementação parcial de um *framework* para anotação semântica de dados geoespaciais. Embora o trabalho desenvolvido tenha focado o domínio agrícola, o *framework* mostrou ser genérico, podendo ser facilmente estendido e adotado por outros domínios que possuam necessidades de anotação de dados geoespaciais, sempre focando o contexto.

Como principais contribuições deste trabalho, além do mecanismo de anotação proposto tem-se: (i) identificação de um conjunto de características que catálogos geoespaciais devem prover para apoiar a busca semântica; (ii) identificação de requisitos para a anotação semântica de dados geoespaciais; (iii) implementação parcial de um *framework* de anotação, para validação do mecanismo proposto; (iv) identificação e desafios no uso de *workflows* científicos para orquestrar o processo de anotação de dados geoespaciais e como lidar com eles; (v) identificação de desafios existentes no uso de *workflows* científicos para orquestrar o processo de anotação de dados geoespaciais e como lidar com eles.

As possibilidades de extensões do trabalho abrangem desde o mecanismo de anotações proposto e o esquema de metadados utilizado, até o *framework* implementado. Dentre elas destacam-se: outras extensões do esquema de metadados adotado; adaptação do *framework* para outros domínios; tradução automática de um *workflow* especificado no WOODSS para o YAWL; anotação de partes do dado; versionamento de *workflows* e definição de uma metodologia para configuração de uma anotação.

## 7 Agradecimentos

Este trabalho apresenta o resumo da tese de doutorado da autora, que agradece à Unicamp e ao Instituto de Computação pelo suporte para o seu desenvolvimento, e ao demais colegas envolvidos. Em especial, agradece à Profa. Claudia Bauzer Medeiros, pela orientação e a Jefersson dos Santos pelas imagens fornecidas, resultado de seu trabalho de mestrado.

O apoio financeiro recebido da Embrapa, FAPESP, Instituto Virtual FAPESP-Microsoft(projeto eFarms), CNPq e CAPES.

## 8 Referências

BECK, H.; PINTO, H. **Overview of approach, methodologies, standards, and tools for ontologies**. Draft Paper, The Agricultural Ontology Service, UN FAO, 2002. Disponível em: <[www.fao.org/agris/aos/Documents/BackgroundPaper.pdf](http://www.fao.org/agris/aos/Documents/BackgroundPaper.pdf)>. Acesso em: 19 nov. 2010.

BORGES, K. A. V. **Uso de uma Ontologia de lugar urbano para reconhecimento e extração de evidências geo-espaciais na web**. 2006. 181 f. Tese (Doutorado) - Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte. Disponível em: <<http://hdl.handle.net/1843/SLBS-6XYFYS>>. Acesso em: 19 nov. 2010.

CARVALHO, A. C. P. L. F.; BRAYNER, A.; LOUREIRO, A.; FURTADO, A. L. STAA, A. von; LUCENA, C. J. P. de; SOUZA, C. S. de; MEDEIROS, C. M. B.; LUCCHESI, C. L.; SILVA, E. S. e; WAGNER, F. R.; SIMON, I.; WAINER, J.; MALDONADO, J. C.; OLIVEIRA, J. P. M. de; RIBEIRO, L.; VELHO, L.; GONÇALVES, M.; ARANAUSKAS, M. C. C.; MATTOSO, M.; ZIVIANI, N.; NAVAU, P. O. A.; TORRES, R. da S.; ALMEIDA, V. A. F.; MEIRA JUNIOR, W.; KOHAYAKAWA, Y. **Grand challenges in computer science research in Brazil - 2006 - 2016**. São Paulo: Brazilian Computer Society, 2006. Workshop report.

CHAKRAVARTHY, A.; CIRAVEGNA, F.; LANFRANCHI, V. **AKTiveMedia**: cross-media document annotation and enrichment. In: INTERNATIONAL SEMANTIC WEB CONFERENCE, 5., 2006, Athens. **Proceedings**... Berlin; New York: Springer, 2006. (Lecture notes in computer science, 0302-9743; 4273). ISWC2006. Poster.

EGENHOFER, M. J. Toward the semantic geospatial web. In: ACM INTERNATIONAL SYMPOSIUM ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 10., 2002. **Proceedings**... New York: ACM, 2002, p. 1-4. GIS'02.

FEDERAL GEOGRAPHIC DATA COMMITTEE. FGDC. **Content Standard for Digital Geospatial Metadata**. Washington, D.C. June 1998.

FILETO, R.; LIU, L.; PU, C.; ASSAD, E. D.; MEDEIROS, C. B. POESIA: An ontological *workflow* approach for composing Web services in agriculture. The **VLDB Journal**, v. 12, n. 4, p. 352-367, 2003.

GREENBERG, J.; SPURGIN, K; CRYSTAL, A. Functionalities for automatic metadata generation applications: a survey of metadata experts' opinions. **International Journal of Metadata**, Semantics and Ontologies, v. 1, n. 1, p. 3-20, 2006.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, n. 2, p. 199-220, Apr. 1993.

HANDSCHUH, S.; STAAB, S.. Authoring and annotation of web pages in CREAM. In: INTERNATIONAL CONFERENCE ON WORLD WIDE WEB, 11., 2002, Honolulu, HI. **Proceedings...** New York: ACM Press, 2002. p. 462-473.

HOLLINK, L. **Semantic annotation for retrieval of visual resources**. 2006. 141 p. (PhD Thesis) - Vrije Universiteit Amsterdam, 2006. (SIKS Dissertation Series, n. 2006-24).

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. **ISO 19115:2003**: Geographic information - metadata. 2008. Disponível em: <<http://www.iso.org/iso/home.htm>>. Acesso em: 19 nov. 2010.

JONES, C. B.; ABDELMOTY, A. I.; FINCH, D.; FU, G.; VAID, S. The SPIRIT spatial search engine: architecture, ontologies and spatial indexing. In: INTERNATIONAL CONFERENCE, GISCIENCE, 3., 2004, Adelphi, MD, USA. **Geographic Information Science**: proceedings. Berlin; Heidelberg: Springer, 2004. p. 125-139. (Computer science).

KLIEN, E. A rule-based strategy for the semantic annotation of geodata. **Transactions in GIS**, v. 11, n. 3, p. 437-452, June 2007.

KLIEN, L.; LUTZ, M. The role of spatial relations in automating the semantic annotation of geodata. In CONFERENCE ON SPATIAL INFORMATION THEORY, 2005, Ellicottville, New York, **Proceedings...** Berlin; Heidelberg: Springer, 2005. p. 133-148. COSIT'05. (Lecture notes in computer science, v. 3693).

MACÁRIO, C. G. N.; MEDEIROS, C. B. A framework for semantic annotation of geospatial data for agriculture. **International Journal Metadata**, Semantics and Ontology. v. 4, n. 1/2, p. 118-132, June, 2009a. Special issue on Agricultural Metadata and Semantics.

MACÁRIO, C. G. do N.; MEDEIROS, C. B. The geospatial semantic web: are GIS catalogs prepared for this? In: INTERNATIONAL CONFERENCE ON WEB INFORMATION SYSTEMS AND TECHNOLOGIES, 5., 2009, Lisboa, Portugal. **Proceedings...** Lisboa: Insticc, 2009b. p. 335-340. Webist 2009.

MACÁRIO, C. G. N.; MEDEIROS, C. B. Specification of a framework for semantic annotation of geospatial data on the web. **Sigspatial**, Special v. 1, n. 1, p. 27-32, Jul. 2009c.

MACÁRIO, C. G. N., NAKAI, A. M., MEDEIROS, C. B.; MADEIRA, E. Using scientific workflows for semantic annotation of geospatial data: what are the challenges involved? **Journal of Universal Computer Science**, 2009.

MACÁRIO, C. G. N.; SANTOS, J. A. dos; C. B. MEDEIROS, C. B.; TORRES, R. da S. Annotating data to support decision-making: a case study. In: WORKSHOP ON GEOGRAPHIC INFORMATION RETRIEVAL, 6., Zurich. **Proceedings...** New York: ACM, 2010. p. 1-7. GIR '10.

MARIOTTE, L. E.; MEDEIROS, C. B.; LEE, I. Diagnosing similarity of oscillation trends in time series. In IEEE INTERNATIONAL CONFERENCE ON DATA MINING WORKSHOPS, 7., 2007, Omaha. **Proceedings...** Washington, DC: IEEE Computer Society, 2007. p. 643-648. ICDMW '07.

MEDEIROS, C. B.; PÉREZ-ALCAZAR, J., DIGIAMPIETRI, L., PASTORELLO JUNIOR, G. Z.; SANTANCHE, A.; TORRES, R. S.; MADEIRA, E; BACARIN, E. Woodss and the web: Annotating and reusing scientific workflows. **SIGMOD Record**, v. 34, n. 3, p. 18-23, Sept. 2005.

MILLER, H. J.; HAN, J. Discovering geographic knowledge in data rich environments: a report on a specialist meeting. **ACM SIGKDD Explorations Newsletter**, v. 1, n. 2, p. 105-107, Jan. 2000.

NOGUERAS-ISO, J.; ZARAZAGA-SORIA, F. J.; BÉJAR, R., ALVAREZ, P. J.; MURO-MEDRANO, P. R. OGC catalog services: a key element for the development of spatial data infrastructure. **Computers & Geosciences**, v. 31, n. 2, p.199-209, Marc. 2005.

REEVE, L.; HAN, H. Survey of semantic annotation platforms. In: ACM SYMPOSIUM ON APPLIED COMPUTING, 2005. **Proceedings...** New York: ACM, 2005. p. 1634-1638. SAC 2005.

SANTOS, J. A. dos; LAMPARELLI, R.; TORRES, R. da S. Using relevance feedback for classifying remote sensing images. In: SIMPOSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 14., 2009, Natal. **Anais...** São José dos Campos: Inpe, 2009. p. 7909-7916.

SPACCAPIETRA, S.; CULLOT, N.; PARENT, C.; VANGENOT, C. On spatial ontologies. In BRAZILIAN SYMPOSIUM ON GEOINFORMATICS, 6., 2004, Campos do Jordão. **Proceedings...** São José dos Campos: Inpe, 2004.

SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The semantic web revisited. **IEEE Intelligent Systems**, v. 21, n. 3, p. 96-101, 2006.



SOUSA, S. R. **Gerenciamento de anotações semânticas de dados na Web para aplicações agrícolas**. 2010. 64 f. Dissertação (Mestrado em Ciências da Computação), Instituto de Computação, Universidade Estadual de Campinas, Campinas.

TSALGATIDOU, A., ATHANASOPOULOS, G., PANTAZOGLU, M.; PAUTASSO, C.; HEINIS, T.; GRONMO, R.; HOFF, H.; BERRE, A.; GLITTUM, M.; TOPOUZIDOU, S. Developing scientific workflows from heterogeneous services. **SIGMOD Record**, v. 35, n. 2, p. 22-28, 2006.

TUSK, C.; KOPERSKI, K.; AKSOY, S. E.; MARCHISIO, G. Automated feature selection through relevance feedback. In: GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2003, Toulouse. **Proceedings...** Piscataway, NJ: IEEE, 2003. p. 3691-3693, v. 6. IGARSS 2003.

UREN, V.; CIMIANO, P.; IRIA, J.; HANDSCHUH, S.; VARGAS-VERA, M.; MOTTA, E.; CIRAVEGNA, F. Semantic annotation for knowledge management: requirements and a survey of the state of the art. *Web Semantics: science, services and agents on the World Wide Web*, v. 4, 1, p. 14-28, 2006.

VAN DER ALAST, W. M. P.; HOFSTEDE, A. H. M. ter. Yawl: yet another workflow language. **Information Systems**, v. 30, n. 4, p. 245-275, 2005.

WORKSHOP ON THE CHALLENGES OF SCIENTIFIC WORKFLOWS, 2006, Arlington. **Workshop...** Arlington: National Science Foundation, 2006.



---

**Informática Agropecuária**

Ministério da  
Agricultura, Pecuária  
e Abastecimento



CGPE 9140