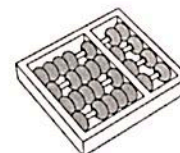


Lucas Oliveira Batista

**“Apoio ao Estudo de Correlações entre Séries
Temporais baseadas em Anotações Semânticas”**

CAMPINAS
2015



Universidade Estadual de Campinas
Instituto de Computação

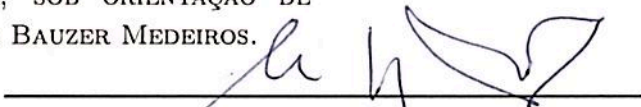
Lucas Oliveira Batista

“Apoio ao Estudo de Correlações entre Séries Temporais baseadas em Anotações Semânticas”

Orientador(a): Prof. Claudia Bauzer Medeiros

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Computação da Universidade Estadual de Campinas para obtenção do título de Mestre em Ciência da Computação.

ESTE EXEMPLAR CORRESPONDE À VERSÃO DA DISSERTAÇÃO APRESENTADA À BANCA EXAMINADORA POR LUCAS OLIVEIRA BATISTA, SOB ORIENTAÇÃO DE PROF. CLAUDIA BAUZER MEDEIROS.



Assinatura do Orientador(a)

CAMPINAS
2015

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

B32a Batista, Lucas Oliveira, 1991-
Apoio ao estudo de correlações entre séries temporais baseadas em
anotações semânticas / Lucas Oliveira Batista. – Campinas, SP : [s.n.], 2015.

Orientador: Claudia Maria Bauzer Medeiros.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de
Computação.

1. Ciência da computação. 2. Banco de dados - Busca. 3. Web semântica. I.
Medeiros, Claudia Maria Bauzer, 1954-. II. Universidade Estadual de Campinas.
Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Supporting the study of correlations among time series via semantic annotations

Palavras-chave em inglês:

Computer science
Database searching
Semantic Web

Área de concentração: Ciência da Computação

Titulação: Mestre em Ciência da Computação

Banca examinadora:

Claudia Maria Bauzer Medeiros [Orientador]
Ariadne Maria Brito Rizzoni Carvalho
Carla Geovana do Nascimento Macario

Data de defesa: 07-07-2015

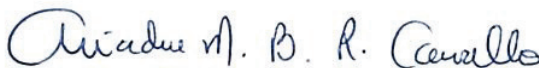
Programa de Pós-Graduação: Ciência da Computação

TERMO DE APROVAÇÃO

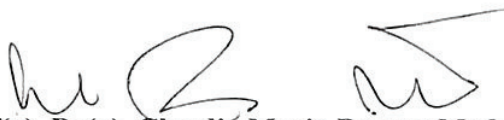
Defesa de Dissertação de Mestrado em Ciência da Computação, apresentada pelo(a) Mestrando(a) **Lucas Oliveira Batista**, aprovado(a) em **07 de julho de 2015**, pela Banca examinadora composta pelos Professores(as) Doutores(as):



Prof(a). Dr(a). **Carla Geovana do Nascimento Macario**
Titular



Prof(a). Dr(a). **Ariadne Maria Brito Rizzoni Carvalho**
Titular



Prof(a). Dr(a). **Claudia Maria Bauzer Medeiros**
Presidente

Apoio ao Estudo de Correlações entre Séries Temporais baseadas em Anotações Semânticas

Lucas Oliveira Batista¹

07 de julho de 2015

Banca Examinadora:

- Prof. Claudia Bauzer Medeiros (*Orientadora*)
- Prof. Dr. Ariadne Maria Brito Rizzoni Carvalho
Instituto de Computação - UNICAMP
- Dr. Carla Geovana do Nascimento Macario
Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA
- Prof. Dr. Fábio Luiz Usberti
Instituto de Computação - UNICAMP (*Suplente*)
- Dr. Alexandre Camargo Coutinho
Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA (*Suplente*)

¹Suporte Financeiro: Bolsa CAPES 2013–2014; Bolsa FAPESP (Processo 2014/07303-1) 2014–2015

Abstract

Time series are used in several knowledge domains, such as economics, meteorology and agriculture. In many situations, scientists often associate annotations to series during their analysis. They have moreover to search and correlate many types of time series in order to conduct their research. This process is hampered not only by the heterogeneity among the series, but also by the search for relevant series to compute a given correlation. The predominant methods to search series are based either in keyword (annotation) matching or in pattern matching. They do not support looking for additional information not always directly associated to series. Given this scenario, this dissertation proposes *TS³Annotation*, a framework based on the use of semantic annotations to support the study of correlations among time series. The main contributions of this work are: (1) a time series semantic annotation model; (2) and the *TS³Annotation* a framework that allows experts to create semantic annotations and uses these annotations as the basis for the search.

Resumo

Séries temporais são utilizadas em diversos domínios do conhecimento, por exemplo, economia, meteorologia e agricultura. Em várias situações, cientistas, muitas vezes, associam anotações a séries durante sua análise. Além disso, precisam buscar e correlacionar vários tipos de séries para estudar algum problema. Isto é dificultado não só pela heterogeneidade entre as séries, como também pela limitação dos mecanismos de busca por séries relevantes a uma correlação. As modalidades predominantes na busca por séries são baseadas ou em casamento de texto (anotações) ou em casamento de padrões. Não permitem buscas por séries que estejam relacionadas semanticamente. Diante deste cenário, esta dissertação propõe o *TS³ Annotation*, um *framework* que usa anotações semânticas como base para permitir o estudo de correlações entre séries. As principais contribuições desta dissertação são: (1) um modelo de anotação semântica para séries temporais; (2) e o *framework TS³ Annotation* que permite a especialistas anotar semanticamente séries, além de explorar o uso destas anotações como uma nova possibilidade na busca por séries temporais.

Agradecimentos

Eu gostaria de agradecer à professora Claudia Bauzer Medeiros por toda paciência, suporte, oportunidade e aprendizado a mim concedidos.

Agradeço também ao professor André Santanchè pela disponibilidade e apoio durante o desenvolvimento do trabalho.

Eu gostaria de agradecer aos meus pais, Batista e Nide, ao meu irmão, Mateus e a toda minha família pelo suporte e por sempre acreditar em meu potencial, me ajudando em todas as dificuldades que encontrei até o momento.

Eu gostaria de agradecer à Jacqueline Midlej por estar sempre ao meu lado, comemorando nossas vitórias e me apoiando nos momentos de tristeza.

Agradeço aos membros do LIS pelas conversas, companheirismo e pelas reuniões que contribuíram para o aperfeiçoamento desta pesquisa.

Agradeço a todos os meus amigos, em especial Anderson, Matheus, Ivo, Jaudete, Daniel, Jorge, Ive, Renato, Patrícia, Vitor e Lucas Miguel que me ajudaram a alcançar esta conquista.

Agradeço à EMBRAPA pela parceria e pelos dados disponibilizados para a validação desta pesquisa.

Agradeço às agências de fomento pelo suporte financeiro que, direta ou indiretamente, contribuíram para a realização desta pesquisa: FAPESP (Processo 2014/07303-1), FAPESP/Cepid em Ciência e Engenharia da Computação (2013/08293-7), Instituto Virtual de Pesquisa FAPESP-Microsoft (projeto NavScales), CNPq (projeto MuZOO), FAPESP-PRONEX (projeto eScience), INCT em Ciência Web e CAPES. As opiniões expressas neste trabalho não necessariamente refletem as opiniões das agência de fomento.

Agradeço aos membros da banca, Prof. Dr. Ariadne Maria Brito Rizzoni Carvalho, Dr. Carla Geovana do Nascimento Macario, Prof. Dr. Fábio Luiz Usberti (suplente) e Dr. Alexandre Camargo Coutinho (suplente) pela disponibilidade e contribuições dadas a esta pesquisa.

Sumário

Abstract	ix
Resumo	xi
Agradecimentos	xiii
1 Introdução	1
2 Conceitos Básicos e Trabalhos Correlatos	4
2.1 Anotações	4
2.2 Anotações Semânticas	8
2.3 Busca por séries temporais	10
2.4 Conclusões	13
3 TS³Annotation - <i>Framework</i> para Anotação Semântica e Busca por Séries Temporais	14
3.1 TSSAM - Modelo de Anotações Semânticas	14
3.2 Descrição da TSSAMOnto	21
3.3 Busca por Séries Temporais via Anotações Semânticas	26
3.4 Arquitetura do <i>TS³Annotation</i>	29
3.5 Conclusões	32
4 Detalhes de Implementação e Estudo de Caso	33
4.1 Detalhes de Implementação	33
4.2 Estudo de Caso	37
4.2.1 Dados utilizados	37
4.2.2 Testes e Validação	39
4.3 Conclusões	47

5	Conclusões e Trabalhos Futuros	48
5.1	Conclusões	48
5.2	Trabalhos Futuros	49
	Referências Bibliográficas	53

Lista de Tabelas

4.1	Conceitos anotados e a quantidade de anotações.	38
4.2	Resultados obtidos utilizando o parâmetro “Milho”, sem utilizar as regras de expansão do $TS^3 Annotation$	43
4.3	Resultados obtidos utilizando os parâmetros “Corn” e “Maize”, sem utilizar as regras de expansão do $TS^3 Annotation$	43
4.4	Resultados obtidos utilizando o parâmetro “Milho”, utilizando as regras de expansão do $TS^3 Annotation$	44

Lista de Figuras

2.1	Criação de anotação em diferentes tipos no <i>Boom Chameleon</i> , extraída de [39].	5
2.2	Exemplo de uso de dois protótipos do <i>BibPhone</i> , extraída de [23].	5
2.3	Criação de anotação semântica do Annotea, extraído de [17].	6
2.4	Criando anotações no TSPad, extraído de [33].	6
2.5	Visualização de um biosinal com o painel de anotações à esquerda e regiões anotadas destacadas em vermelho, extraído de [22].	7
2.6	Inserindo uma anotação para um intervalo de uma série temporal, extraído de [36].	8
2.7	Exemplo de anotação semântica de trajetória utilizando o modelo de episódio e viagem, extraído de [13].	9
2.8	Modelo de anotação semântica de séries temporais do Tatoo, extraído de [32].	10
2.9	Resultado de uma busca na ferramenta QuerySketch, extraído de [42].	11
2.10	Resultado de uma busca na ferramenta TimeSearcher, extraído de [14].	12
2.11	Resultado de uma busca na ferramenta proposta por Landesberger, Voss e Kohlhammer, extraído de [41].	12
3.1	Visão geral do <i>framework</i> proposto.	15
3.2	Modelo de anotação semântica de séries temporais proposto.	16
3.3	Características de uma série temporal.	17
3.4	Tela do primeiro protótipo para anotação semântica de séries temporais.	18
3.5	Exemplos de anotações semânticas de séries temporais utilizando o TSSAM.	20
3.6	Mais exemplos de anotações semânticas de séries temporais utilizando o TSSAM.	22
3.7	TSSAMOnto - Representação gráfica, não formal, da ontologia criada para definir os componentes do Time Serie Semantic Annotation Model.	25
3.8	Representação de uma série temporal utilizada pelo <i>TS³Annotation</i>	26
3.9	OntoEx - ilustração da instância de uma ontologia hipotética utilizada para anotações semânticas de séries temporais.	28

3.10	Representação das séries temporais anotadas com os conceitos da OntoEx.	29
3.11	Exemplo de arquivo de regras.	29
3.12	Representação de uma anotação semântica expandida.	30
3.13	Arquitetura do <i>TS³Annotation</i> .	30
4.1	Tecnologias utilizadas na implementação do <i>framework</i> .	34
4.2	Arquitetura básica de um reasoner. Extraída de [44].	36
4.3	Descrição da sintaxe de regras. Extraída de https://jena.apache.org/documentation/inference/ .	36
4.4	Instância de ontologia utilizada na criação das anotações semânticas.	39
4.5	Tela do protótipo para criação de anotações semânticas.	40
4.6	Anotação semântica em RDF/XML.	41
4.7	Anotação semântica versionada.	42
4.8	Regras utilizadas na busca por séries temporais.	44
4.9	Exemplos de séries temporais e suas anotações semânticas utilizadas no estudo de caso. Todas são recuperadas pela busca por “Milho”.	45
4.10	Representação gráfica da expansão de uma anotação semântica.	46
4.11	Tela utilizada na busca por séries temporais.	46
4.12	Conjunto de 10 séries temporais presentes no resultado de busca do <i>TS³Annotation</i> .	47
5.1	Tela inicial - proposta de melhoria de interface.	51
5.2	Tela de Login - considera a inclusão de usuários para controle de autorização.	51
5.3	Tela para a criação de uma conta.	51
5.4	Tela para a criação de anotações em séries temporais e visualização gráfica das anotações, associadas a uma série, com seu tempo inicial e final. A parte esquerda mostra uma inserção de anotação. À direita, todas as anotações da série ilustrada no centro da tela.	52

Capítulo 1

Introdução

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo [6], de forma genérica $\{< v_i, t_i >\}$, onde v_i é o conteúdo da observação e t_i uma marca de tempo. Séries são utilizadas em diversos domínios do conhecimento como economia (taxa mensal de desemprego), meteorologia (temperatura diária), agricultura (variação do índice de vegetação de determinada área) ou saúde (eletrocardiograma). Segundo Jensen e Snodgrass [16], um dado pode ser caracterizado como temporal quando apresenta mais de um estado ao longo do tempo. A análise de séries temporais pode fornecer aos especialistas a capacidade de identificar tendências e de prever possíveis cenários daquele dado temporal. Por exemplo, Koh *et al* [19] examinam a sazonalidade e luz do sol como influentes nos incidentes de tuberculose usando análise e correlação de séries temporais. Há dois tipos de séries utilizadas no trabalho: incidência de tuberculose em *Birmingham* e séries meteorológicas do período de luz solar na mesma cidade.

Séries temporais são geradas a partir de diferentes fontes de dados, como imagens de satélites ou sensores, sendo processadas e analisadas por uma grande variedade de pesquisadores. Em várias situações, cientistas precisam buscar e correlacionar vários tipos de séries para estudar algum problema. Isto é dificultado não só pela heterogeneidade entre séries, mas também pela falta de fatores em comum que possibilitem tal correlação e até mesmo a busca de séries relevantes a uma correlação. Além disso, em muitas aplicações (por exemplo, em agricultura, biodiversidade ou tráfego) o fator usado para a busca da série é a localização à qual ela está associada. Entretanto, encontrar séries temporais utilizando poucos parâmetros, como apenas a localização geográfica, pode não ser suficiente para obter as séries temporais mais adequadas ao problema em questão.

Uma busca eficiente por séries temporais facilita a tarefa de análise destas séries. Há algumas modalidades predominantes na busca por séries. Por exemplo, os sistemas existentes permitem que os usuários busquem séries temporais selecionando uma categoria (por exemplo, séries financeiras) e um período, ou usando uma *string* em linguagem

natural comparada ao título da série, ou mesmo via parâmetros que são comparados ao conteúdo das séries. Como se verá, o Capítulo 2 desta dissertação descreve em detalhes tais modalidades. Todavia, nenhuma dessas possibilidades permite a busca por séries temporais utilizando informações complementares nem sempre diretamente associadas às séries, por exemplo, utilizando termos técnicos ou conceitos de um domínio.

Durante a análise de séries temporais, especialistas muitas vezes anotam as séries criando e/ou analisando associações entre anotações e as séries temporais. Uma anotação pode ser definida como dados que descrevem dados [36]. Como destacado em Silva [36], anotações são utilizadas para prover informações adicionais que podem ser relevantes para a análise dos dados. Além disso, podem conter informações sobre a proveniência dos dados [1]. Anotações relacionadas a séries temporais são potencialmente feitas por diferentes pesquisadores ou grupos de pesquisa e são geralmente criadas sob a forma de texto, utilizando uma linguagem própria e sendo armazenadas em arquivos.

Outros sistemas de busca utilizam anotações em texto livre, que foram associadas às séries, para encontrá-las. Entretanto, em vários domínios, devido à ausência de um padrão, as anotações possuem grande heterogeneidade, dificultando a busca, o compartilhamento e a integração de dados [25]. O uso de anotações semânticas permite atacar este problema. Anotações semânticas podem ser definidas como a descrição de algum conteúdo digital de acordo com sua semântica [37], usando ontologias. O uso de anotações semânticas permite novas possibilidades de busca e facilita encontrar séries temporais relacionadas a um conteúdo, pois estarão associadas de forma indireta, pela semântica. Tal abordagem permite, por exemplo, que um pesquisador interessado em milho encontre: (1) séries temporais relacionadas a milho independente do termo de pesquisa utilizado, por exemplo, “corn” ou “milho”; (2) séries temporais associadas indiretamente a “milho”; e (3) outras séries temporais (por exemplo, séries de temperatura ou chuva) associadas a regiões onde o milho foi plantado mesmo que não anotadas diretamente como “milho”.

Dado este cenário de opções de busca, esta dissertação propõe um *framework* que permita apoiar o estudo da correlação entre séries temporais, facilitando a busca por séries a partir de suas anotações semânticas. A hipótese associada é que ao usar tais anotações para encontrar séries, há mais possibilidades dos pesquisadores tirarem proveito da semântica associada. O trabalho foi desenvolvido no LIS – *Laboratory of Information Systems* – do Instituto de Computação da UNICAMP; os dados utilizados são do domínio da agricultura e foram fornecidos pela EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária.

As principais contribuições deste trabalho são:

1. um modelo de anotação semântica para séries temporais;
2. e um *framework* que explora o uso destas anotações como uma nova possibilidade

na busca por séries temporais.

Esta pesquisa gerou as seguintes publicações:

- BATISTA, L. O. ; MEDEIROS, C. B. . Supporting the Study of Correlations between Time Series via Semantic Annotations. In: CSBC - VIII Brazilian eScience Workshop (BreSci), 2014, Brasília. Porto Alegre: Sociedade Brasileira de Computação, 2014. p. 379-382.
- BATISTA, L. O. ; MEDEIROS, C. B. . Searching Time Series via Semantic Annotations. In: SBBD - XIII Workshop de Teses e Dissertações em Banco de Dados (WTDBD), 2014, Curitiba. UFPR; PUC-PR, 2014. p. 1-6.

Além disso, gerou o seguinte poster:

- BATISTA, L. O. ; MEDEIROS, C. B. Supporting the Study of Correlations Among Time Series via Semantic Annotations. In: The 10th IEEE International Conference on e-Science, 2014, Guarujá.

O restante deste documento está organizado como segue. O Capítulo 2 define os conceitos básicos e os trabalhos correlatos que foram utilizados como base no desenvolvimento da dissertação. O Capítulo 3 descreve o modelo de anotação semântica, o algoritmo utilizado na busca por séries e a arquitetura do *framework*. O Capítulo 4 detalha as ferramentas utilizadas para construção do *framework* e os testes realizados para a validação desta dissertação. Por fim, as conclusões e trabalhos futuros são abordados no Capítulo 5.

Capítulo 2

Conceitos Básicos e Trabalhos Correlatos

Este capítulo apresenta os principais conceitos necessários para o desenvolvimento desta dissertação. Além disso, descreve alguns trabalhos correlatos ao problema abordado. A seção 2.1 apresenta o conceito de anotações e diferentes trabalhos onde anotações são utilizadas. A seção 2.2 apresenta a definição de anotações semânticas e trabalhos que as utilizam. A seção 2.3 apresenta trabalhos relacionados à busca por séries temporais e suas limitações. Por fim, a seção 2.4 descreve as conclusões deste capítulo.

2.1 Anotações

Anotações podem conter diferentes informações, em função do objeto anotado e de quem faz as anotações. Por exemplo, em dados de mídia, anotações comumente feitas pelo próprio *software* que gera a mídia são: a data de criação, tamanho, codificação ou formato [11]. Anotações podem ser usadas para expor uma opinião, prover informações adicionais, melhorar a recuperação de informação ou explicar determinado conteúdo. Por exemplo, o endereço de uma casa não pode ser descoberto a partir de sua foto digital, mas uma anotação da foto poderia conter esta informação [36]. Alternativamente, uma anotação com efeito semelhante poderia ser o georreferenciamento da foto.

As anotações podem ser criadas em diferentes formatos, por exemplo, em desenho, em áudio ou em texto. Uma ferramenta que usa anotações em desenho e áudio é a *Boom Chameleon* [39], cuja criação foi motivada pelo aumento do uso de dados em três dimensões (3D). A partir de um dispositivo especializado para visualizar ambientes 3D com tela *touchscreen* e um microfone acoplado, as anotações são feitas em formato de áudio ou capturando os gestos do usuário em três categorias: caneta, luz ou imagens. A Figura 2.1 ilustra o uso da ferramenta na criação de anotações em uma imagem 3D de

um veículo.

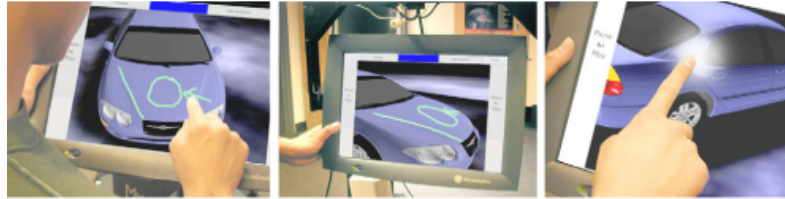


Figura 2.1: Criação de anotação em diferentes tipos no *Boom Chameleon*, extraída de [39].

Anotações no formato de áudio são feitas, por exemplo, pelas ferramentas BibPhone [23] e TableTrans [24]. O BibPhone é um protótipo que grava e reproduz anotações de áudio em livros. A ferramenta utiliza RFID (*Radio-Frequency IDentification*) e *Bluetooth* para estabelecer uma conexão entre o livro e sua anotação que está armazenada em um banco de dados central da biblioteca na qual o livro se encontra. A Figura 2.2 mostra dois protótipos do *BibPhone*. Cada um consiste de: um microprocessador, um módulo *Bluetooth*, uma bateria recarregável, um leitor RFID, um microfone, uma caixa de som e um botão utilizado para o controle da gravação do som.



Figura 2.2: Exemplo de uso de dois protótipos do *BibPhone*, extraída de [23].

A ferramenta TableTrans proposta por Ma *et al.* [24] serve para anotações colaborativas em formato de áudio, que são armazenadas em um banco de dados relacional. A TableTrans é responsável pelo compartilhamento de anotações entre usuários. Os dados ficam armazenados em um servidor compartilhado e a ferramenta, além de acessá-los, gerencia restrições das permissões de usuários sobre as anotações.

A grande maioria das anotações é feita sob forma de texto. Um exemplo de ferramenta que utiliza anotações textuais é o Annotea [17], que é um sistema *Web* de anotação para documentos da *Web*. As anotações são armazenadas em um servidor remoto no formato RDF (*Resource Description Framework*) e podem estar associadas a todo o documento ou parte dele (ver Figura 2.3).

O foco desta dissertação é em anotações de séries temporais. Há vários tipos de dados que se encaixam nesta categoria. Por exemplo, há ferramentas que realizam anotações

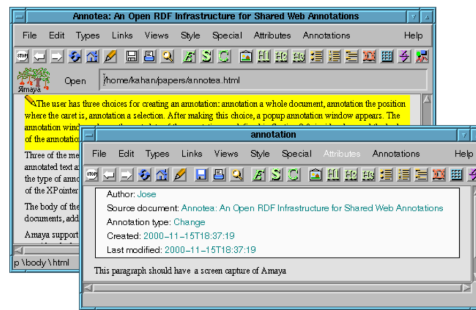


Figura 2.3: Criação de anotação semântica do Annotea, extraído de [17].

em trajetórias. De acordo com Yan *et al.* [43], uma trajetória consiste em um fluxo de pontos no espaço temporal (x, y, t) , onde x, y são coordenadas geográficas e t é um instante de tempo. O sistema *Easy Tracker* [9] utiliza o GPS (*Global Positioning System*) do *smartphone* para capturar os pontos espaço-temporais do movimento do usuário. Assim, o sistema permite que o usuário realize manualmente anotações textuais em partes de sua rota para descrever suas atividades, por exemplo, "caminhando para o trabalho".

Outros exemplos de trabalhos que realizam anotações em séries temporais são [33, 36, 22]. Pressly [33] desenvolveu um aplicativo distribuído para *tablets* e computadores chamado "TSPad". O sistema permite que os pesquisadores colaborarem em dados de séries temporais que surgem em sua pesquisa. É possível criar anotações em formato textual em partes das séries como, também, efetuar a busca por anotações (ver Figura 2.4). A ferramenta TSPad aceita diferentes tipos de dados de séries temporais, permite anotações nestes dados e sua visualização por outros. Além disso, o TSPad restringe as ações dos usuários sobre anotações de terceiros.

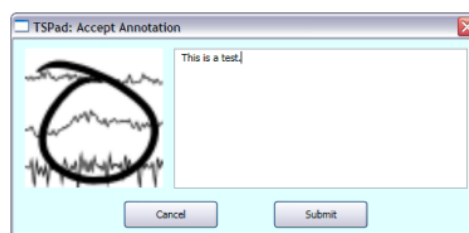


Figura 2.4: Criando anotações no TSPad, extraído de [33].

Ainda outro exemplo é o trabalho de Lourenço *et al.* [22], que desenvolveram uma plataforma para anotação e visualização de biosinais (por exemplo, eletrocardiograma - ver Figura 2.5). A ferramenta permite que usuários façam anotações manuais de séries temporais, importem anotações existentes em fontes externas e organizem as anotações em hierarquias a partir do uso de rótulos de nomes com barras (/). O sistema permite,

ainda, que o usuário visualize, corrija, elimine e navegue entre anotações.



Figura 2.5: Visualização de um biosinal com o painel de anotações à esquerda e regiões anotadas destacadas em vermelho, extraído de [22].

O *framework Serial Annotator* proposto por Silva [36] é outro exemplo de sistema de anotação de séries temporais e é a principal base para o nosso trabalho. O autor desenvolveu um aplicativo para dispositivos móveis voltado para a área de agricultura, possibilitando que especialistas da área façam anotações e as relacionem a intervalos de séries temporais (ver Figura 2.6). Tanto as séries como as anotações estão armazenadas em um banco de dados relacional. O *Serial Annotator* armazena separadamente a série temporal e suas anotações, o que diminui o espaço de armazenamento e possibilita associar uma anotação a várias séries temporais. Além disso, versiona anotações, fato demandado por especialistas (consultados pelo autor), que podem rever e restaurar estados anteriores de anotações.

Apesar de grande variedade de projetos de anotação, é possível observar algumas desvantagens destas últimas ferramentas. Em primeiro lugar, elas armazenam anotações em modo textual, com texto livre, o que dificulta a busca, o compartilhamento, a integração dos dados e a interpretação destes por máquinas. Além disso, não tratam da ambiguidade de informações. Estes problemas podem ser atacados com o uso de anotações semânticas, abordagem utilizada nesta dissertação.

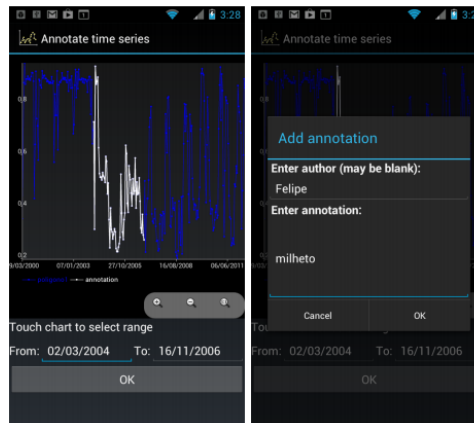


Figura 2.6: Inserindo uma anotação para um intervalo de uma série temporal, extraído de [36].

2.2 Anotações Semânticas

Neste trabalho, uma anotação semântica é um metadado específico que segue uma estrutura formal e possui seu significado especificado em uma ontologia [21], [38], [18]. De acordo com Oren *et al.* [31], existem três tipos de anotação: (a) informal - não são interpretadas por máquinas e não possuem uma linguagem formal; (b) formal - possuem uma estrutura formalmente definida, podendo assim ser interpretadas por máquinas; e (c) ontológica - possuem uma estrutura formal definida e utilizam apenas termos ontológicos que são socialmente aceitos e entendidos. Como se verá, nossa proposta considera anotações do tipo ontológicas.

Liao *et al.* [20] definem um modelo formal de anotações semânticas. O trabalho identifica os principais componentes para criar uma anotação semântica, a saber: o projeto ou a seleção de uma ontologia, o projeto de um modelo da estrutura da anotação semântica e o desenvolvimento da aplicação para atingir o propósito do usuário. O trabalho também faz um *survey* dos métodos de anotação semântica que são aplicados em diferentes propostas e domínios; por exemplo, no domínio de serviços *Web*, um modelo de anotação semântica foi utilizado para construir uma rede semântica que é explorada por um algoritmo que automaticamente encontra um componente de serviço para satisfazer um pedido.

Anotações semânticas também são tratadas por Macario [25] e Sousa [37]. A primeira explora anotações semânticas em dados geoespaciais, aplicando as soluções a problemas de agricultura. A principal vantagem do trabalho, para o problema abordado nesta proposta, é um *framework* genérico que provê um mecanismo de anotação semi-automática de dados. O segundo desenvolveu um sistema *Web* que cria e gerencia anotações semânticas. Além disso, transforma anotações em anotações semânticas utilizando ontologias do domínio

agrícola.

Outros trabalhos fazem anotação semântica de trajetórias, como [13, 43]. Guc *et al.* [13] desenvolveram uma ferramenta em que os usuários realizam anotações semânticas de suas trajetórias. Para isto, propõem um modelo de anotação semântica que possui dois elementos: episódio e viagem. Episódios dividem uma trajetória e semanticamente descrevem essas partes. Viagens são grupos de episódios que possuem uma semântica em um nível maior (ver Figura 2.7). A mesma ideia é implementada na ferramenta SeMi-Tri [43], que descreve uma trajetória como uma sequência de episódios semânticos. Para isso, o sistema automaticamente associa episódios a anotações semânticas referentes a locais (por exemplo, compras ou trabalho), meios de transporte (por exemplo, andando ou ônibus) para obter uma trajetória semanticamente estruturada. Um dos motivos de interpretar semanticamente o comportamento das pessoas é a utilização dessas informações em planejamento urbano, otimização do transporte público, dentre outros.

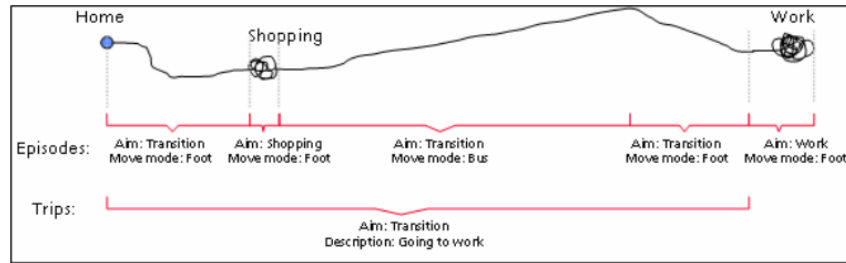


Figura 2.7: Exemplo de anotação semântica de trajetória utilizando o modelo de episódio e viagem, extraído de [13].

Já o framework Tatoo [32] tem por objetivo realizar anotações semânticas em recursos ambientais, possuindo um modelo para anotação semântica de séries temporais. Este modelo (ver Figura 2.8) segue o modelo proposto pela ontologia de observações e medidas cujo objetivo é representar os dados medidos por sensores.

Anotações e anotações semânticas são utilizadas para anotar diversos objetos (por exemplo, documentos, vídeos, etc) e são utilizadas em diferentes sistemas, cada um com seu propósito. Assim, outras ferramentas de anotação, suas comparações e detalhes podem ser encontrados em [37, 35, 40].

Como se verá, nosso trabalho se concentra em dar apoio a especialistas na especificação de anotações semânticas à semelhança de [37, 25], estendendo tais trabalhos no que se refere a possibilidade de busca.

aos métodos que utilizam para representar a série temporal, por exemplo, transformada discreta de Fourier ou polinômio de Chebyshev.

Em outro tipo de busca, a consulta é gráfica. Um exemplo é a ferramenta QuerySketch [42], uma aplicação para banco de dados financeiros que utiliza a distância Euclidiana para encontrar similaridade entre séries temporais. O QuerySketch permite que o usuário desenhe um gráfico a mão livre e, então, efetua a busca por séries temporais em que o preço é similar ao desenho. A Figura 2.9 ilustra o resultado de uma busca no QuerySketch. A série temporal desenhada pelo usuário está em negrito, e abaixo do desenho, estão alguns quadros contendo resultados da busca. Outra ferramenta que utiliza elementos gráficos para a busca por séries temporais é a TimeSearcher [14] que permite o desenho de “*timebox*” (uma janela de tempo) em séries temporais e utiliza os valores dos dados englobados pela “*timebox*” para efetuar a busca por outras séries temporais com os valores de dados semelhantes. A Figura 2.10 ilustra o resultado da busca por estoques com preços entre 70 e 190 em um banco de dados contendo 52 semanas de preços de estoques para 1430 estoques.

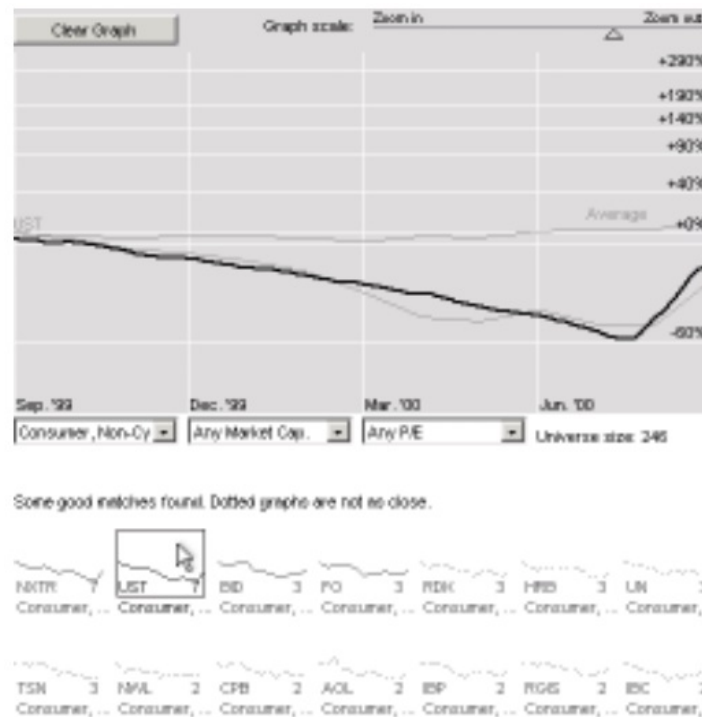


Figura 2.9: Resultado de uma busca na ferramenta QuerySketch, extraído de [42].

Landesberger, Voss e Kohlhammer [41] propõem uma ferramenta que busca séries temporais a partir de um texto em linguagem natural, por exemplo, “produto interno bruto na Alemanha”. A ferramenta busca em diferentes bancos de dados financeiros e

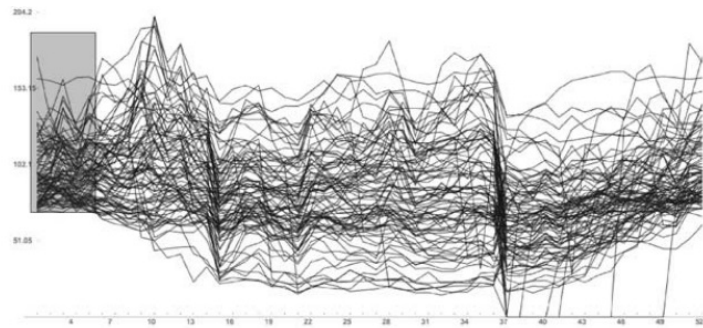


Figura 2.10: Resultado de uma busca na ferramenta TimeSearcher, extraído de [14].

retorna o resultado de maneira gráfica (ver Figura 2.11). Ainda outra ferramenta que busca séries a partir de parâmetros textuais é a TQuEST [3] que utiliza como parâmetro de busca limites definidos pelo usuário, por exemplo, busca por séries temporais em que a pressão está acima de um “limite 1” e ao mesmo tempo a temperatura está em “limite 2”.

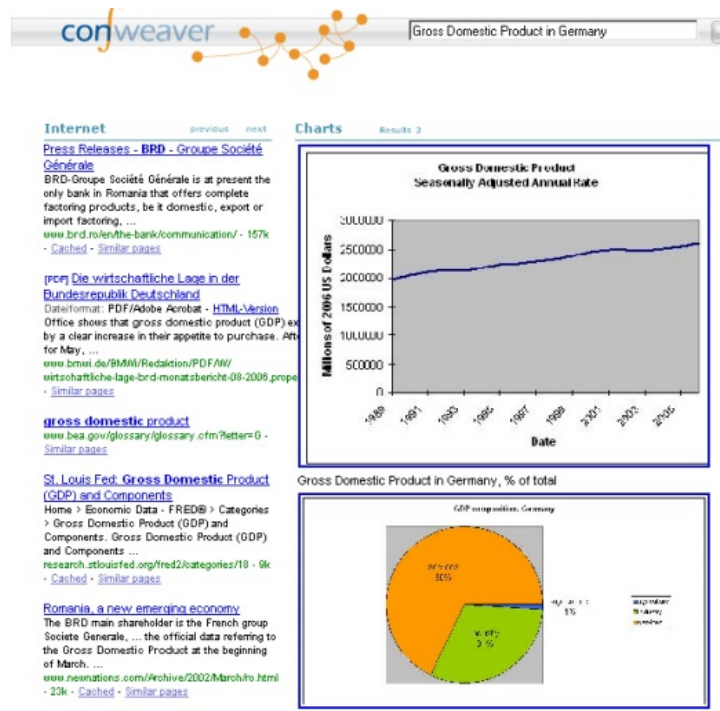


Figura 2.11: Resultado de uma busca na ferramenta proposta por Landesberger, Voss e Kohlhammer, extraído de [41].

Várias outras ferramentas possibilitam o uso de anotações textuais para realizar a busca por séries temporais [33, 36, 22]. Entretanto, esta abordagem recupera apenas as

séries temporais em que o usuário sabe exatamente a anotação que foi utilizada. Por exemplo, se uma série temporal que deve ser anotada como “Milho”, é anotada erroneamente como “Miloh”, ou em outra língua como “Corn”, não é possível recuperar esta série temporal se o termo utilizado na pesquisa for “Milho”. A abordagem desta dissertação utiliza anotações semânticas permitindo encontrar séries temporais associadas semanticamente ao termo utilizado na pesquisa. Além disso, por meio da navegação por ontologias é possível encontrar novos termos associados, por exemplo, a “Milho” e encontrar outras séries temporais anotadas com estes novos termos.

2.4 Conclusões

Este capítulo apresentou os principais conceitos básicos para esta dissertação e os trabalhos relacionados. O estudo de ferramentas que realizam anotações ou anotações semânticas em diversos tipos de dados foi necessário para entender e conhecer: porquê e como anotações são criadas, quais os modelos utilizados e quais abordagens estão disponíveis para criar anotações.

Capítulo 3

TS^3 Annotation - *Framework* para Anotação Semântica e Busca por Séries Temporais

Este capítulo apresenta o *framework* desenvolvido para auxiliar especialistas no estudo da correlação de séries temporais, denominado TS^3 Annotation (*Time Series Search and Semantic Annotation*). Este sistema possui duas funcionalidades principais: anotação semântica e a busca por séries temporais via anotações semânticas. A Figura 3.1 apresenta uma visão geral do TS^3 Annotation. Usuários realizam anotações semânticas de séries temporais seguindo o nosso modelo de anotação semântica e ontologias escolhidas para tal. Várias anotações podem ser associadas à mesma série (e a trechos de séries). A busca por séries é feita a partir de um conjunto de regras, um termo de busca e as anotações. O conjunto de regras é informado pelo próprio usuário e é utilizado para navegar nas ontologias, inferindo novas informações de interesse do usuário. Tais informações são utilizadas para expandir as anotações semânticas, possibilitando encontrar séries utilizando termos relacionados à anotação inicial. A seção 3.1 discute o modelo de anotação semântica criado denominado TSSAM. A seção 3.2 apresenta a ontologia TSSAMOnto criada para a definir o TSSAM. A seção 3.3 apresenta o algoritmo desenvolvido para a busca por séries temporais. A seção 3.4 mostra a arquitetura do sistema. Por fim, a seção 3.5 apresenta as conclusões do capítulo.

3.1 TSSAM - Modelo de Anotações Semânticas

Para alcançar o objetivo principal deste trabalho, primeiramente, definimos um Modelo de Anotação Semântica de Séries Temporais (o qual chamamos de TSSAM, do inglês “*Time Series Semantic Annotation Model*”). O modelo foi criado na linguagem RDF

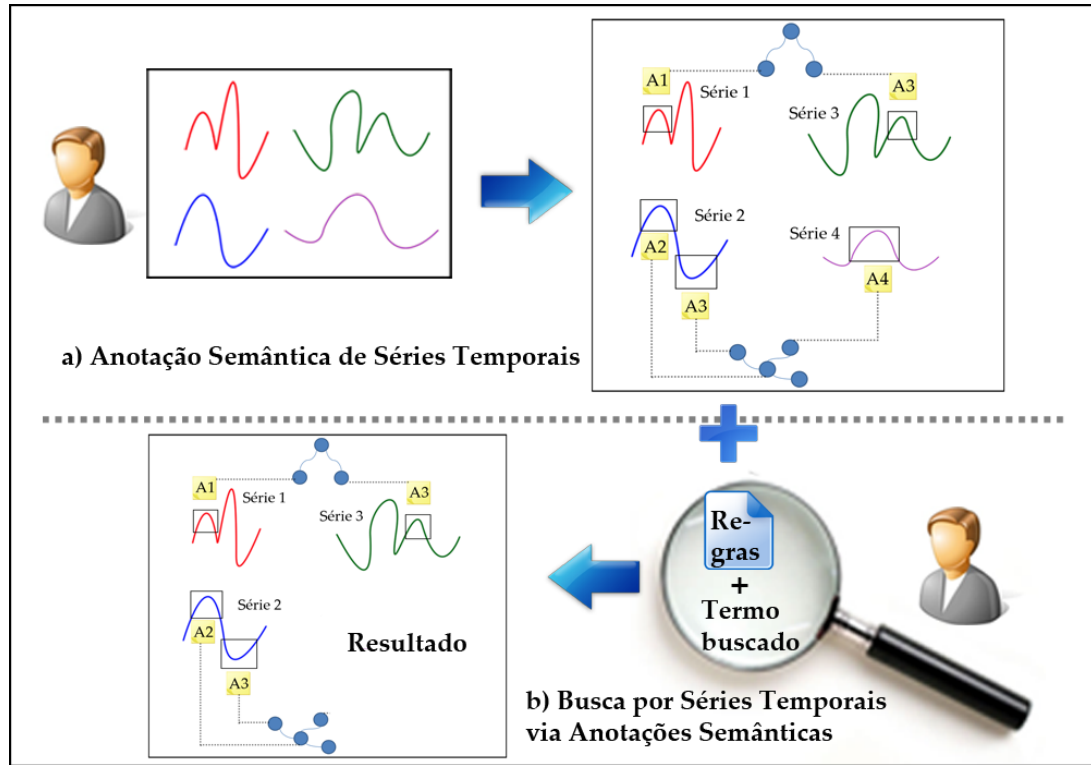


Figura 3.1: Visão geral do *framework* proposto.

(*Resource Description Framework*), pois esta linguagem é utilizada na *Web Semântica* como um mecanismo para descrever recursos, independente de seus domínios, de uma maneira compreensível por máquinas [44]. Além disso, é uma linguagem que é facilmente integrada com ontologias. Em RDF o conhecimento é representado a partir de triplas *Recurso – Propriedade – Valor*, onde valor pode ser um literal ou outro recurso.

A Figura 3.2 ilustra a base do TSSAM, na qual *Recursos* das triplas são as elipses (que representam URIs - *Unified Resource Identifier*), *Valores* são as elipses ou os retângulos (que representam literais) e as propriedades são as arestas que conectam recursos a valores. O nó central do modelo da Figura 3.2 contém o identificador da anotação semântica e em sua volta estão os demais componentes do TSSAM. O nome dos componentes foi especificado em inglês porque este modelo utiliza algumas propriedades que já estão definidas pelo Annotea [17], por exemplo, a propriedade “*annotates*” relaciona uma anotação a seu recurso e está definida em <http://www.w3.org/2000/10/annotation-ns#annotates>. As propriedades que estão antecidas por “*an*” estão definidas na ontologia do Annotea [17] <http://www.w3.org/2000/10/annotation-ns>; por “*to*” estão definidas na ontologia Time Ontology <http://www.w3.org/TR/owl-time/>; por fim, as outras propriedades são especificadas em uma ontologia própria chamada TSSAMOnto. O TSSAM

é extensível e pode ser assim modificado por usuários de acordo com suas necessidades.

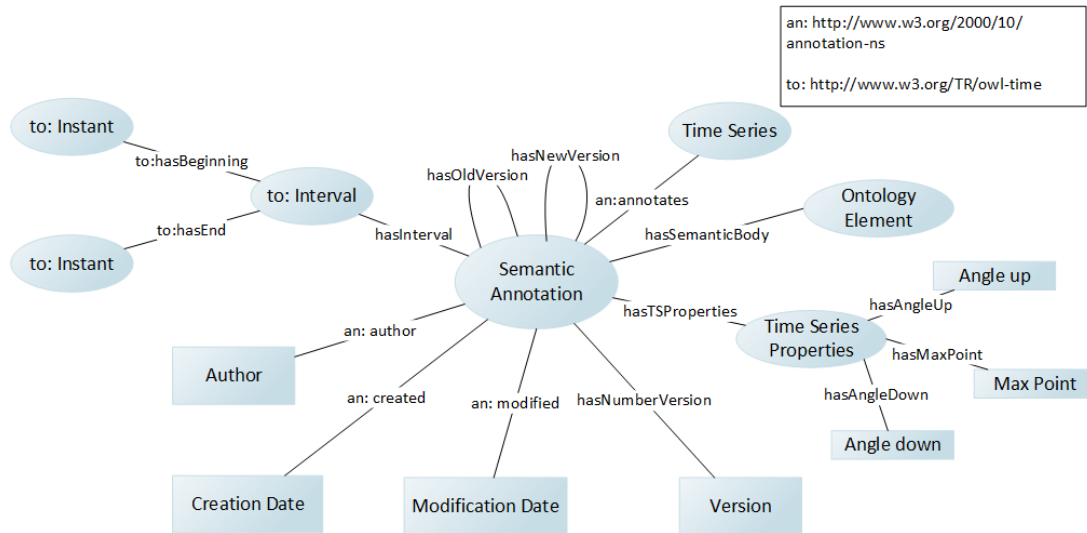


Figura 3.2: Modelo de anotação semântica de séries temporais proposto.

O TSSAM possui os seguintes componentes:

- **Semantic Annotation:** URI que identifica uma anotação semântica, é o componente central do TSSAM;
- **Time Series:** URI que identifica a série temporal à qual a anotação semântica está associada;
- **Interval:** URI que define a granularidade temporal da (sub)série anotada, isto é, uma anotação semântica pode estar associada a toda a série temporal ou apenas parte dela. Esta granularidade é definida por um intervalo de tempo limitado por seu início e seu fim. Esses elementos de tempo são definidos pela Time Ontology¹;
- **Author:** *string* que identifica a pessoa que criou a anotação semântica;
- **Creation Date:** valor que identifica a data em que a anotação semântica foi criada. Esta data não é alterada durante o versionamento da anotação;
- **Modification Date:** valor que identifica a data em que a anotação semântica foi modificada. Uma anotação, ao ser versionada, possui a sua data de modificação alterada;

¹<http://www.w3.org/TR/owl-time/>

- **Time Series Properties:** URI que identifica propriedades relacionadas com a parte anotada da série temporal, por exemplo, o ponto máximo, um dos atributos ilustrados; outras propriedades podem ser associadas, conforme a necessidade dos usuários. Essas propriedades podem ser utilizadas pelos especialistas durante a análise da série anotada para sanar eventuais dúvidas ou confirmar conceitos. Por exemplo, no contexto da agricultura, o ângulo de inclinação da curva auxilia especialistas a identificar se o padrão de uma série temporal representa uma plantação de cana-de-açúcar ou pasto, pois essas culturas possuem características similares. A Figura 3.3 ilustra as características específicas ilustradas na Figura 3.2, as quais são: o ângulo de inclinação inicial (letra “a”, *angle up*), o ângulo de inclinação final (letra “b”, *angle down*) e o ponto máximo (letra “c”, *max point*) de um trecho de uma série;

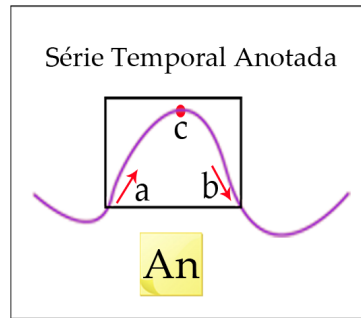


Figura 3.3: Características de uma série temporal.

- **Version:** um número inteiro que identifica a versão de uma anotação semântica. O versionamento possibilita a análise da mudança do conteúdo de uma anotação ao longo do tempo. Cada modificação da anotação gera uma nova anotação, com uma nova versão, com o seu conteúdo modificado, com a mesma data de criação e com uma nova data de modificação. As propriedades *hasNewVersion* e *hasOldVersion*, respectivamente, conectam uma anotação à sua nova versão e vice-versa;
- **Ontology Element:** URI que identifica os elementos das ontologias. Uma anotação pode ter um ou mais elementos de ontologias em seu conteúdo. Os elementos das ontologias são informados pelo usuário e são conectados a uma anotação pela propriedade *hasSemanticBody*.

A criação do TSSAM utilizou os passos descritos no trabalho de Liao *et al.* [20] (ver Capítulo 2). O primeiro passo foi criar a ontologia que define os componentes do TSSAM. Esta ontologia, descrita na seção 3.2, foi criada utilizando a ferramenta Protegè

3.5². O segundo passo foi a criação do modelo de anotação semântica com diversos componentes além daqueles que identificam a anotação semântica (Semantic Annotation) e fazem referência a conceitos em ontologias (Ontology Element). De acordo com Liao, estes são os dois componentes básicos de um modelo para anotação semântica. Por fim, o terceiro passo foi o desenvolvimento da aplicação.

Após a definição do TSSAM, um primeiro protótipo foi criado para permitir anotação semântica das séries. Essa primeira versão permite a criação de anotações semânticas e seu uso. Uma série pode ter mais de uma anotação. A Figura 3.4 mostra uma cópia de tela do protótipo. Para criar uma anotação, o usuário preenche um formulário informando a URI que remete ao conteúdo anotado (componente “Ontology Element” do TSSAM), a granularidade da anotação (que será associada ao “Interval” do TSSAM) e o autor (componente “Author” do TSSAM). Os outros componentes são automaticamente criados pelo sistema. No lado direito é possível observar uma representação gráfica das anotações para uma série e abaixo mais detalhes sobre as anotações. O Capítulo 4.2 apresenta mais detalhes sobre questões de implementação.

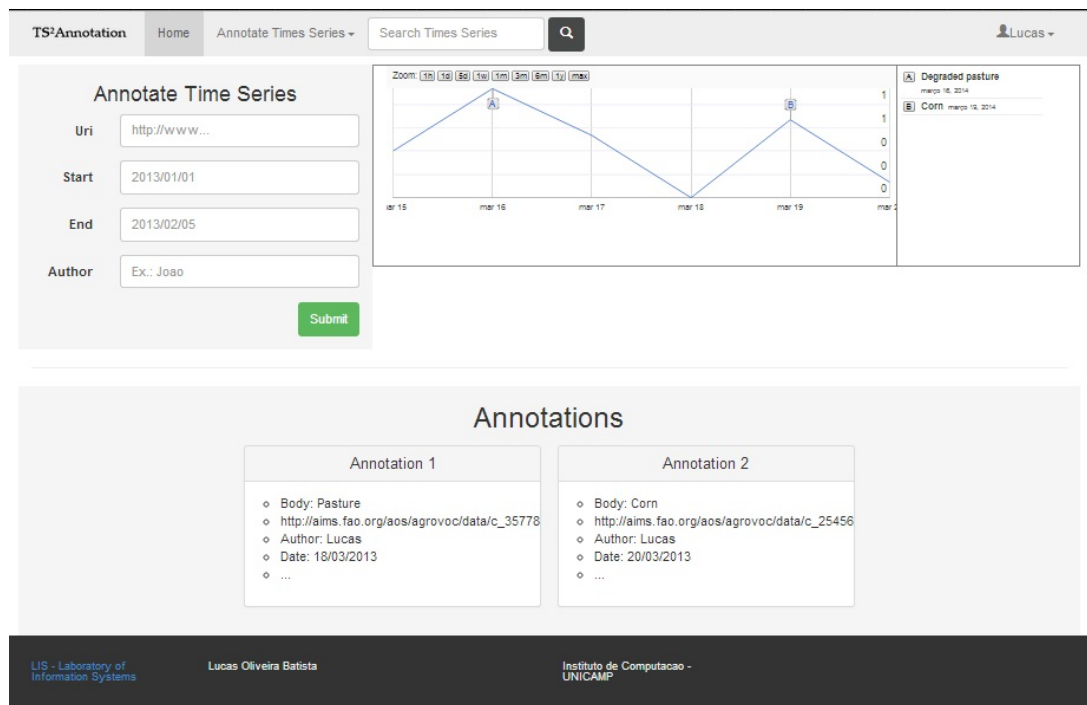


Figura 3.4: Tela do primeiro protótipo para anotação semântica de séries temporais.

Consideramos nosso modelo de anotação semântica formal e semi-ontológico, visto que não são utilizados apenas termos ontológicos em seus componentes. Os componentes

²<http://protege.stanford.edu/>

“Author”, “Creation Date”, “Modification Date” e “Version” não utilizam termos ontológicos. Por exemplo, o nome do autor é especificado utilizando texto em linguagem natural. Entretanto, os demais componentes, inclusive o conceito presente no conteúdo da anotação (“Ontology Element”), devem estar definidos em ontologias.

Nos exemplos a seguir, as anotações estão relacionadas a séries temporais NDVI (*Normalized Difference Vegetation Index*), que representam o nível de biomassa em uma determinada região. Esse tipo de série tem sua localização geográfica associada e pode ser utilizada, por exemplo, para detectar o tipo de cultura agrícola daquele local e naquele período de tempo, como ilustrado, dentre outros, pelo projeto MAPAGRI³ (Metodologia para Mapeamento da Agricultura Brasileira). Além disso, alguns componentes da anotação estão ocultos para melhor visualização; ao lado das URI’s está o nome dos componentes, para melhor entendimento.

As Figuras 3.5 e 3.6 ilustram exemplos, no contexto da agricultura, de anotações semânticas que poderiam ser criadas utilizando o TSSAM:

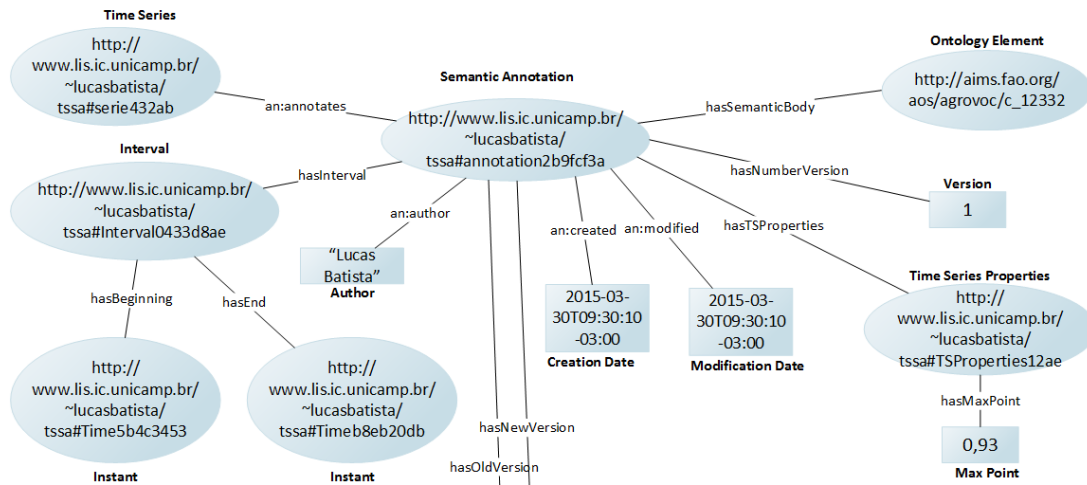
- A anotação “A” possui como identificador a URI <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#annotation2b9fcf3a> e foi criada por “Lucas Batista”. Esta anotação está associada à série temporal que possui o URI <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#serie432ab>. A anotação se refere ao intervalo representado pela URI <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#Interval0433d8ae>, possui tempo inicial e final <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#Time5b4c3453>, <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#Timeb8eb20db>, respectivamente. Além disso, a subsérie anotada possui como ponto máximo o valor de “0,93”, que está representado pela URI <http://www.lis.ic.unicamp.br/~lucasbatista/tssa#TSProperties12ae>.

Esta anotação possui duas versões, indicando alteração em seu conteúdo. O versionamento altera o conteúdo de uma anotação de acordo com as mudanças realizadas pelos usuários e atualiza a data de modificação da anotação. Assim, a anotação “A” foi criada e modificada no dia “2015-03-30”, possui o número de versão “1” e contém em seu conteúdo a URI http://aims.fao.org/aos/agrovoc/c_12332, que representa o conceito de “Milho” na ontologia da FAO/Agrovoc. Posteriormente, esta anotação foi versionada, o que gerou uma nova anotação semântica, com uma nova data de modificação “2015-04-01”, um novo autor “Mateus Batista”, um novo número de versão “2” e seu conteúdo agora está associado à URI http://aims.fao.org/aos/agrovoc/c_14477, que representa “Soja” no mesmo dicionário Agrovoc. Desta maneira, é possível manter o histórico das mudanças do conteúdo das anotações. Além disso, a anotação da versão “1” é co-

³<http://www.alice.cnptia.embrapa.br/handle/doc/954536>

nectada à anotação da versão “2” pela propriedade “hasNewVersion” e a anotação de número de versão “2” é associada a sua versão anterior pela propriedade “hasOldVersion”. A ontologia utilizada nesta anotação foi a AGROVOC (<http://aims.fao.org/standards/agrovoc/concept-scheme>);

a) Anotação A



Anotação A - Versão 2

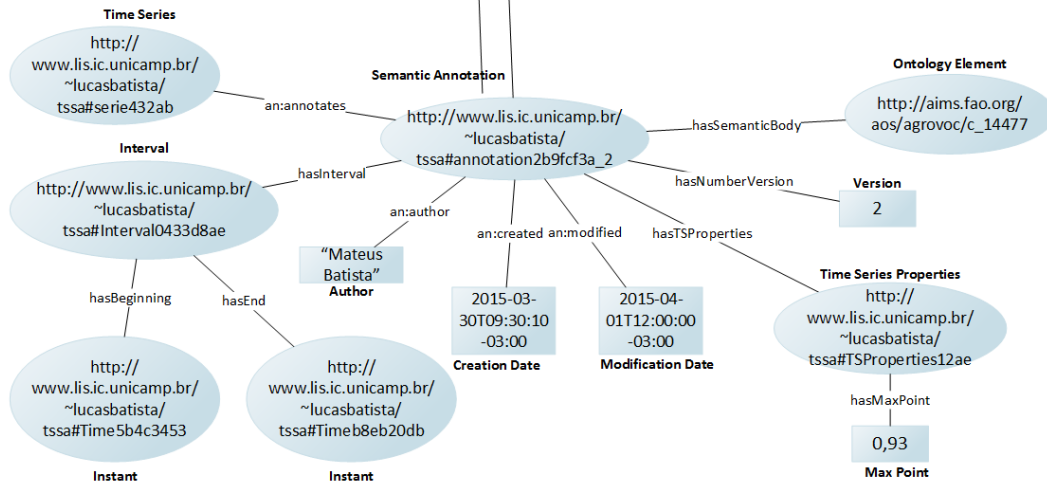


Figura 3.5: Exemplos de anotações semânticas de séries temporais utilizando o TSSAM.

Para facilitar a leitura, os demais exemplos não copiarão as URI's das figuras, que podem ser lidas de forma análoga.

- A anotação “B” tem como parte de seu conteúdo “Milho” no estado de “Desen-

volvimento de Sementes”. Para isto, dois conceitos foram associados à anotação: um para representar a planta, neste caso “Milho” e a outra para representar o estágio de desenvolvimento da planta, neste caso “Desenvolvimento de Semente”. Nesta anotação, duas ontologias foram utilizadas, a AGROVOC e a Plant Ontology (<http://www.plantontology.org/>). O identificador da anotação, a série temporal associada, a granularidade da anotação, o autor, a data de criação e modificação e as propriedades da subsérie anotada são similares ao que foi explicado na Anotação A;

- A anotação “C” tem como parte de seu conteúdo “Milho, Soja e Algodão”. Pode representar uma situação em que o especialista não sabe o tempo exato de plantio de cada cultura. A informação contida na anotação é que naquela região foram cultivadas esses três tipos de culturas agrícolas no tempo especificado pela anotação. Apenas a AGROVOC foi utilizada para representar “Milho, Soja e Algodão” com as URI’s correspondentes. O identificador da anotação, a série temporal associada, a granularidade da anotação, o autor, a data de criação e modificação e as propriedades da sub-série anotada são similares ao que foi explicado na Anotação A.

Os exemplos ilustram que o TSSAM permite que vários conceitos, definidos em ontologias, façam parte do conteúdo de uma anotação. Isto possibilita ao usuário flexibilidade em anotar diferentes informações complementares sobre uma série.

3.2 Descrição da TSSAMOnto

Uma ontologia possui classes e os relacionamentos entre essas classes podem ser expressos utilizando propriedades [44]. Em OWL (*Web Ontology Language*), essas propriedades podem ser: *object properties*, que relacionam uma classe a outra; ou *datatype properties*, que relacionam uma classe a um literal. O *domain* de uma propriedade representa as classes às quais a propriedade pode ser aplicada e o *range* limita o valor da propriedade.

A Figura 3.7 ilustra a ontologia TSSAMOnto, que foi criada como parte desta dissertação para definir os componentes do TSSAM. Na figura, as elipses representam as classes, os retângulos representam os literais e os hexágonos representam as propriedades da ontologia com seus respectivos *domains* (D) e *ranges* (R). A linguagem *RDF-schema* e as ontologias *Annotea*, *Time Ontology* e *Basic Geo*⁴ foram incorporadas à nossa ontologia, pois nosso modelo utiliza classes e propriedades já definidas por esta linguagem e ontologias. As classes ou as propriedades antecidas por *geo* são derivadas do *Basic Geo*; as antecidas por *to* são derivadas da *Time Ontology*; as antecidas por *rdfs* são derivadas

⁴<http://www.w3.org/2003/01/geo/>

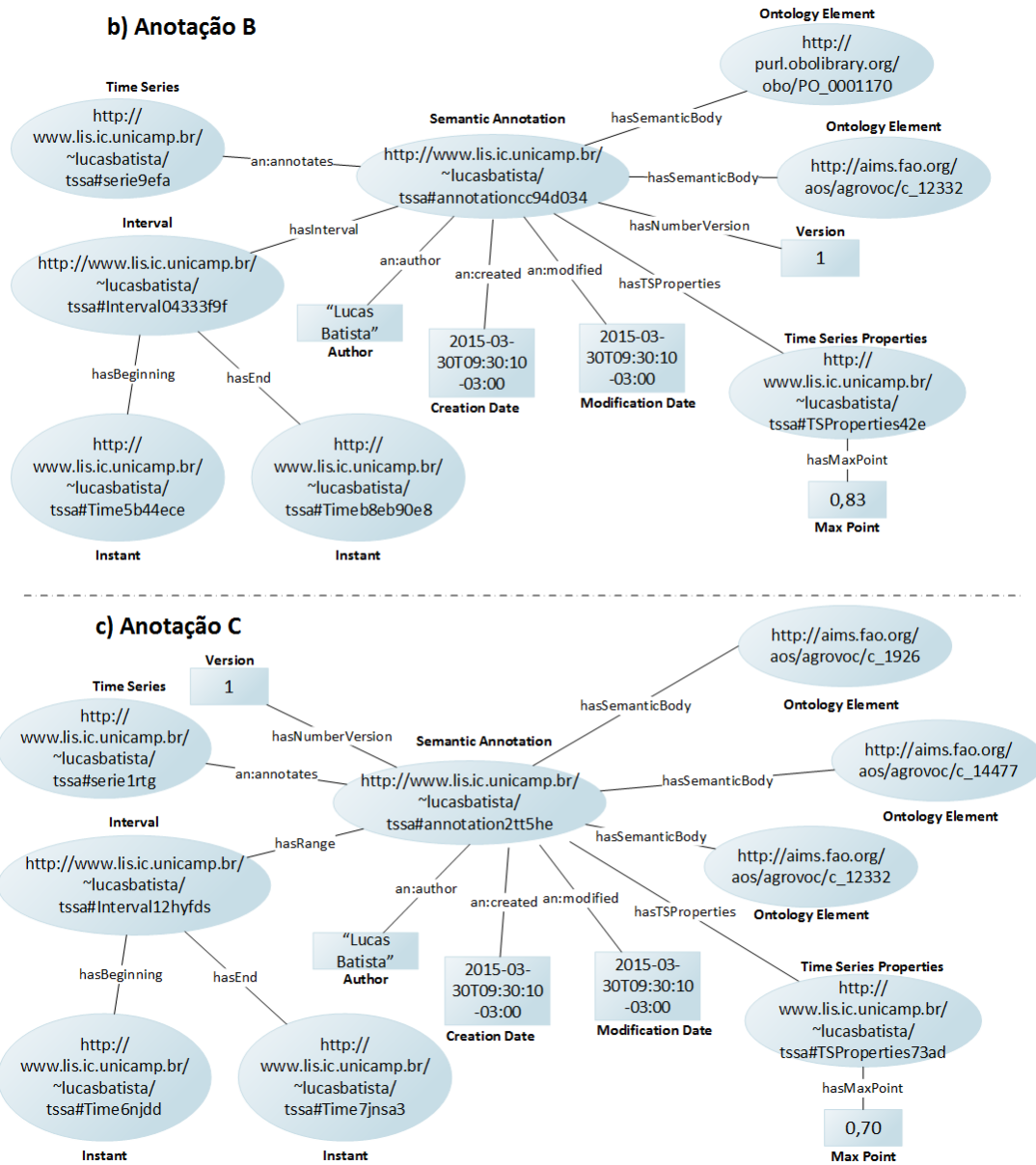


Figura 3.6: Mais exemplos de anotações semânticas de séries temporais utilizando o TS-SAM.

do RDF-schema⁵; e as antecedidas por *an* são derivadas do *Annotea*. As demais classes e propriedades são próprias da TSSAMOnto. As classes e propriedades da TSSAMOnto são:

⁵<http://www.w3.org/TR/rdf-schema/>

- **Classes**

- **Classe definida pelo Annotea (an):**

- * **an:Annotation** - Classe que identifica um recurso, por seu tipo, como sendo uma anotação;

- **Classes definidas pela Time Ontology (to):**

- * **to:Instant** - Classe que representa um determinado ponto no tempo;
 - * **to:Interval** - Classe que representa um período de tempo;

- **Classes definidas pela TSSAMOnto:**

- * **Semantic Annotation** - Classe que representa uma anotação semântica;
 - * **Time Series** - Classe que representa uma série temporal;
 - * **Time Series Properties** - Classe que representa características da parte anotada de uma série temporal.
 - * **Ontology Element** - Classe que representa os elementos das ontologias. Esses elementos podem ser as classes ou as propriedades das ontologias;

- **Propriedades**

- **Propriedades definidas pelo RDFS:**

- * **rdfs:subClassOf** - Propriedade que relaciona a classe *Semantic Annotation* como subclasse da classe *Annotation*, representando que as instâncias da primeira classe também são instâncias da segunda;

- **Propriedades definidas pela Time Ontology:**

- * **to:inXSDDateTime** - Propriedade que relaciona a classe *Instant* ao literal que representa a data e hora desta classe;
 - * **to:hasBeginning** - Propriedade que relaciona a classe *Interval* à classe *Instant* que representa o início de um intervalo;
 - * **to:hasEnd** - Propriedade que relaciona a classe *Interval* à classe *Instant* que representa o final de um intervalo;

- **Propriedades definidas pelo Annotea:**

- * **an:annotates** - Propriedade que relaciona a classe *Semantic Annotation* à classe *TimeSeries*, representando a associação de uma anotação semântica com a série temporal a que esta anotação se aplica;
 - * **an:author** - Propriedade que relaciona a classe *Semantic Annotation* ao literal que representa a pessoa que criou uma anotação;

- * **an:created** - Propriedade que relaciona a classe *Semantic Annotation* ao literal que representa a data e hora da criação de uma anotação;
- * **an:modified** - Propriedade que relaciona a classe *Semantic Annotation* ao literal que representa a data e hora em que uma anotação foi modificada;
- **Propriedades definidas pela Basic Geo:**
 - * **geo:lat** - Propriedade que relaciona a classe *Time Series* ao literal que representa a latitude de uma série temporal;
 - * **geo:long** - Propriedade que relaciona a classe *Time Series* ao literal que representa a longitude de uma série temporal;
- **Propriedades definidas pela TSSAMOnto:**
 - * **tsBegin** - Propriedade que relaciona a classe *Time Series* à classe *Instant* que representa o tempo inicial de uma série temporal;
 - * **tsEnd** - Propriedade que relaciona a classe *Time Series* à classe *Instant* que representa o tempo final de uma série temporal;
 - * **hasInterval** - Propriedade que relaciona a classe *Semantic Annotation* à classe *Interval*, que representa o intervalo da série temporal que está associado a uma anotação;
 - * **hasNumberVersion** - Propriedade que relaciona a classe *Semantic Annotation* ao literal que representa a versão da anotação;
 - * **hasSemanticBody** - Propriedade que relaciona a classe *Semantic Annotation* à classe *Ontology Element*, que representa um elemento de uma ontologia;
 - * **hasTSPproperties** - Propriedade que relaciona a classe *Semantic Annotation* à classe *Time Series Properties* que representa características da parte anotada da série temporal;
 - * **hasAngleUp** - Propriedade que relaciona a classe *Time Series Properties* ao literal que representa o ângulo de inclinação inicial do trecho da (sub)série anotada;
 - * **hasAngleDown** - Propriedade que relaciona a classe *Time Series Properties* ao literal que representa o ângulo de inclinação final do trecho da (sub)série anotada;
 - * **hasMaxPoint** - Propriedade que relaciona a classe *Time Series Properties* ao literal que representa o ponto máximo do trecho da (sub)série anotada;

As propriedades *hasAngleUp*, *hasAngleDown* e *hasMaxPoint* exemplificam particularidades associadas a anotações de usuários. Em outras palavras,

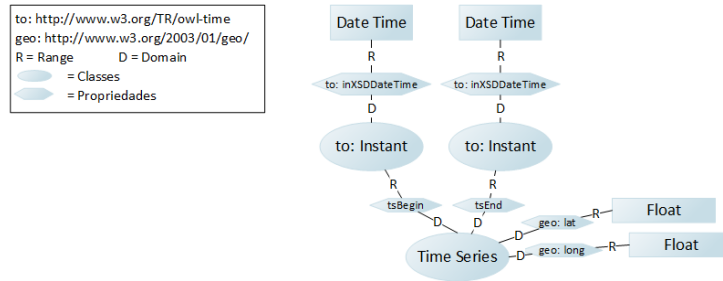


Figura 3.8: Representação de uma série temporal utilizada pelo *TS³ Annotation*.

3.3 Busca por Séries Temporais via Anotações Semânticas

O *TS³ Annotation* tem como objetivo principal auxiliar especialistas na busca por séries temporais, visando apoiá-los na correlação de séries temporais. Para isto, permite que os especialistas busquem séries a partir das anotações semânticas e encontrem séries temporais semanticamente relacionadas.

O Algoritmo 1 percorre o conjunto S de séries e busca aquelas que têm alguma anotação com conceito igual ou derivável semanticamente ao termo buscado. A função *Match* verifica a igualdade de conceitos. Há dois casos principais: Caso 1: A anotação está diretamente associada a uma série. Dada uma série $S_i \in S$, o bloco de linhas 6 a 11 verifica se alguma anotação $AS_j \in S_i$ tem conceito $C_k \in AS_j$ com valor igual ao procurado. Se tiver, o algoritmo interrompe a busca naquela série, inclui a série anotada ao *Resultado* e vai para S_{i+1} .

Caso 2: A anotação pode ser deduzida a partir das anotações de uma série. Caso não encontre igualdade (no caso 1), aplica as regras nas anotações $AS_j \in S_i$ (bloco de linhas 12 e 13) gerando um conjunto de termos expandidos que são semanticamente relacionados ao conteúdo das anotações. Uma regra é definida de forma *antecedente* \rightarrow *consequente*, onde antecedente e consequente são listas de triplas RDF (recurso, propriedade, valor/recurso). Regras são aplicadas, usando um *reasoner*, para obter novas informações deriváveis a partir da anotação armazenada. As vantagens em utilizar regras são: permitem a inferência de outras informações; podem ser ativadas automaticamente quando novas informações são adicionadas ao modelo; e podem ser combinadas possibilitando uma descoberta de informações mais abrangentes.

O bloco de linhas 14 a 18 verifica se os termos expandidos possuem conceito com valor igual ao procurado. Se sim, o algoritmo insere a série temporal anotada ao *Resultado* e vai para S_{j+1} . Caso contrário, a série temporal não satisfaz as condições de busca e o algoritmo vai para S_{j+1} .

Terminado este processo para cada conceito e anotações das instâncias de séries temporais, o conjunto *Resultado* é retornado ao usuário contendo todas as instâncias de séries temporais com anotações semânticas associadas que possuem algum relacionamento com o termo buscado.

Algoritmo 1: Busca por séries temporais via anotações semânticas

Entrada: Termo buscado, conjunto de regras, ontologias envolvidas, base de séries S, base de anotações AS

Saída: Conjunto de séries e anotações que atendem à busca

```

1  início
2  Resultado <- {};
3  para cada instância de série  $S_i \in S$  faça
4      para cada Anotação  $AS_j \in S_i$  faça
5          Encontrou <- Falso;
6          para cada Conceito  $C_k \in AS_j$  e  $Encontrou = Falso$  faça
7              se  $Match(\text{termo buscado}, C_k)$  então
8                  Encontrou <- Verdadeiro;
9                  Resultado <- Resultado  $\cup S_i$ ;
10             fim
11         fim
12         para cada Conceito  $C_k \in AS_j$  e  $Encontrou = Falso$  faça
13             Aplica regras de expansão das anotações, gerando um conjunto de
14             termos expandidos semanticamente relacionados ao conteúdo da
15             anotação;
16             para cada termo expandido e  $Encontrou = Falso$  faça
17                 se  $Match(\text{termo buscado}, \text{termo expandido})$  então
18                     Encontrou <- Verdadeiro;
19                     Resultado <- Resultado  $\cup S_i$ ;
20                 fim
21             fim
22         fim
23     retorna Resultado;
24 fim

```

Para melhor compreensão, consideramos o seguinte exemplo. Suponha que a ontologia OntoEx (ver Figura 3.9) foi utilizada por especialistas para realizar anotações semânticas

durante a análise de séries temporais. Nesta Figura, os conceitos da ontologia estão representados por Conceito1, Conceito2, Conceito3, Conceito4, Conceito5 e Conceito6; os relacionamentos entre esses conceitos estão representados por p1, p2, p3, p4, p5, p6 e p7; por fim, X, Y e Z representam literais. A Figura 3.10 mostra uma representação gráfica das diferentes séries temporais anotadas com estes conceitos e armazenadas no *TS³Annotation*. Apenas o conteúdo do corpo da anotação está sendo representado, os demais foram ocultados. A Figura também mostra que uma série pode ter mais de uma anotação. Assim, por exemplo, a série 2 está anotada com “Conceito2” e “Conceito6”.

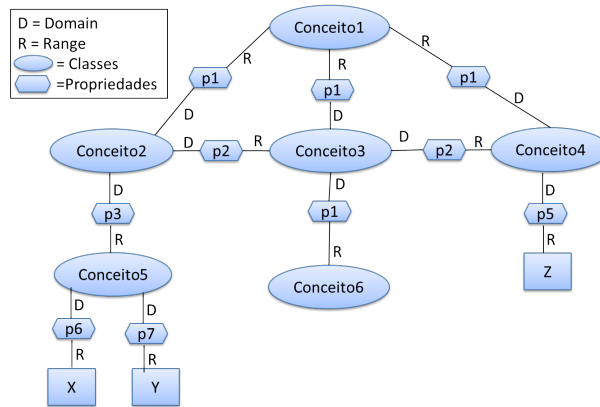


Figura 3.9: OntoEx - ilustração da instância de uma ontologia hipotética utilizada para anotações semânticas de séries temporais.

Um especialista deseja buscar séries temporais utilizando o *TS³Annotation*. Suponha que ele informa ao sistema “Conceito4, Regras.txt, OntoEx” que representam, respectivamente, o termo que deseja buscar, o arquivo de regras que será utilizado e a ontologia envolvida.

A Figura 3.11 mostra o arquivo de regras “Regras.txt” utilizado como entrada no sistema. A regra 1 permite derivar que anotações associadas a “Conceito1” também devem ser associadas a “Conceito2”, “Conceito3” e “Conceito4” via propriedade “p1” e vice versa. A regra 2 expressa uma transitividade, indicando que anotações associadas a “Conceito2” devem ser associadas a “Conceito4” via propriedade “p2” e vice versa.

Ao iniciar a busca por séries temporais o sistema, seguindo o Algoritmo 1, inicializa o conjunto *Resultado* como vazio e percorre o conjunto de séries da seguinte forma. A anotação da instância Série 1 não está diretamente associada a “Conceito4”. Entretanto, utilizando a regra 1, pode-se inferir que “Conceito1” está relacionado a “Conceito4” via “p1”. Assim, a instância Série 1 é adicionada ao *Resultado*. A Figura 3.12 mostra que Série 1 tem uma anotação estendida com “Conceito4”. Analogamente, a instância Série 2 (pela regra 2) e a instância Série 4 (por comparação direta) vão fazer parte do resultado. A instância Série 3 não satisfaz as condições de entrada.

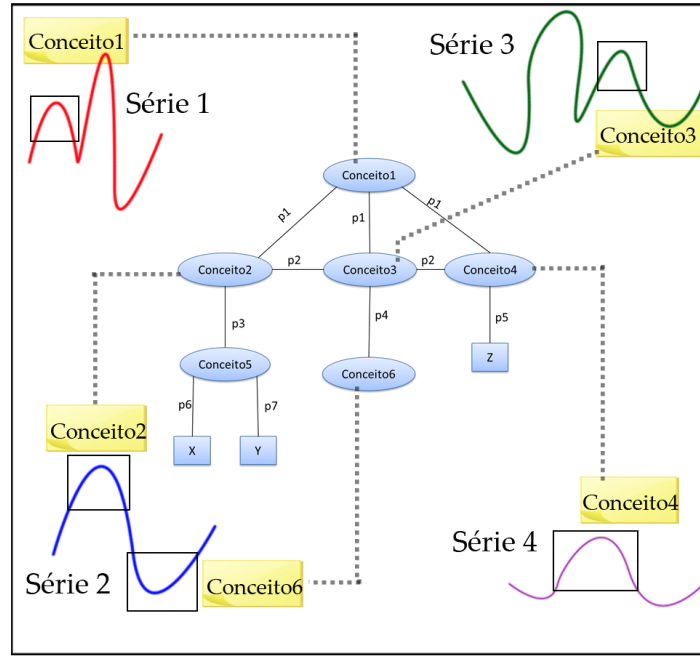


Figura 3.10: Representação das séries temporais anotadas com os conceitos da OntoEx.

```
@prefix tssam: <http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#>.
@prefix ontoex: <http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#>.

[regra1: (?recurso1 ontoex:p1 ?recurso2)
  (?anotacao tssam:hasSemanticBody ?recurso1)-> (?anotacao tssam:hasSemanticBody ?recurso2)]

[regra2: (?recurso1 ontoex:p2 ?recurso2)
  (?recurso2 ontoex:p2 ?recurso3)
  (?anotacao tssam:hasSemanticBody ?recurso1)-> (?anotacao tssam:hasSemanticBody ?recurso3)]
```

Figura 3.11: Exemplo de arquivo de regras.

O exemplo mostra que, usando o Algoritmo 1, o resultado de uma busca por séries anotadas pode ser expandido.

3.4 Arquitetura do TS^3 Annotation

A Figura 3.13 ilustra a arquitetura do *framework* proposto. Nossa arquitetura estende a arquitetura de Silva [36], adicionando os módulos de processamento de ontologias, *reasoner* e alterando o módulo de anotações para que trate de anotações semânticas (as anotações de Silva [36] são textuais). Além disso, armazena anotações e ontologias em um banco de dados RDF, separado das séries temporais. Estes componentes estão destacados

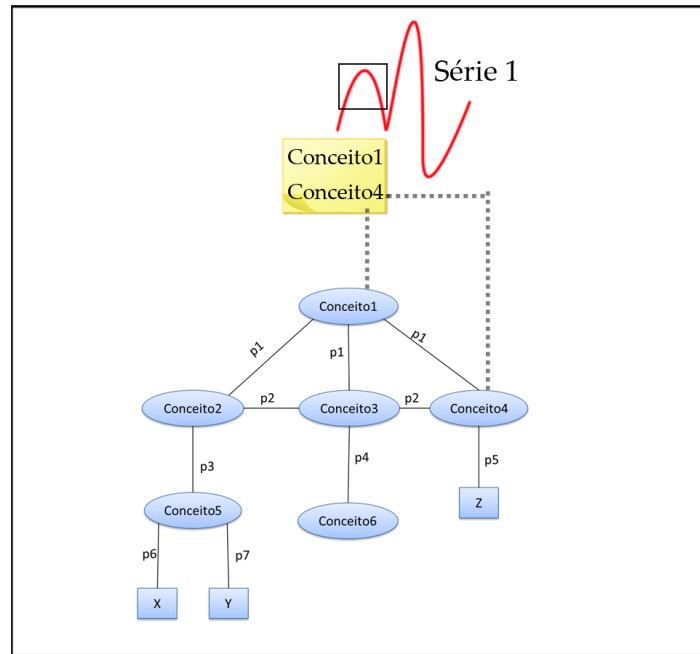


Figura 3.12: Representação de uma anotação semântica expandida.

em vermelho. As setas largas correspondem ao fluxo de dados e as setas pontilhadas correspondem à invocação de serviços e respostas.

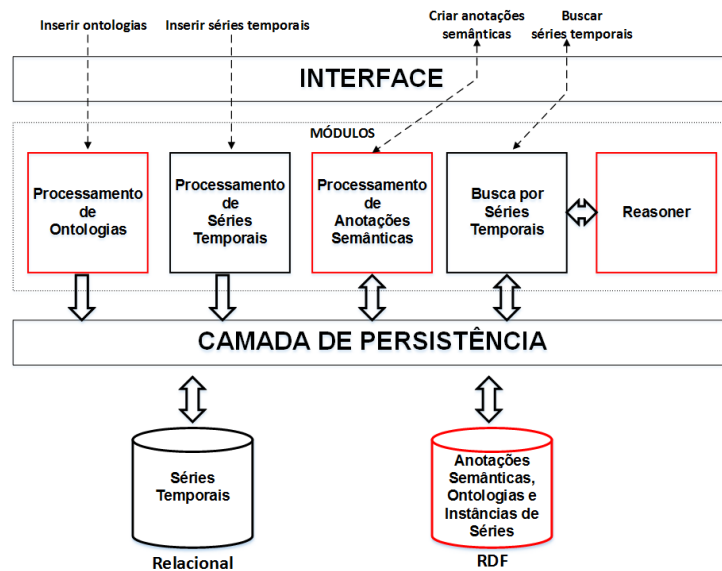


Figura 3.13: Arquitetura do TS^3 Annotation.

A arquitetura é composta por três camadas:

- **Interface** - Provê aos usuários acesso aos serviços do *TS³Annotation*;
- **Módulos** - Provê os serviços necessários para o gerenciamento de anotações semânticas e busca por séries temporais:
 - **Processamento de Ontologias** - Módulo que processa e armazena as ontologias que serão utilizadas para anotar séries temporais;
 - **Processamento de Séries Temporais** - Módulo que processa e armazena as séries temporais e cria as instâncias de séries;
 - **Processamento de Anotações Semânticas** - Módulo que cria, armazena e versiona anotações semânticas associando-as às séries;
 - **Busca por séries temporais** - Módulo que utiliza anotações semânticas na busca por séries temporais;
 - **Reasoner** - Módulo que aplica as regras de inferência em ontologias, gerando novas conexões na ontologia.
- **Camada de Persistência** - Provê aos Módulos acesso aos bancos de dados nos quais as anotações semânticas, as instâncias de séries, as ontologias e as séries temporais estão armazenadas.

Para melhor entendimento da arquitetura, retornaremos ao exemplo da seção 3.3. Inicialmente, através da Interface, os especialistas acionam o Módulo de Processamento de Séries Temporais para armazenar as séries temporais. Este módulo recebe como parâmetro um arquivo contendo uma série temporal, processa este arquivo e armazena a série temporal no banco de dados relacional (como feito por Silva [36]). Além disso, cria e armazena uma instância desta série no banco de dados RDF.

Os especialistas acionam o Módulo de Processamento de Ontologias para armazenar ontologias (neste caso, a OntoEx) que serão utilizadas para anotar as séries temporais. Este módulo recebe como parâmetro uma URL ou o arquivo que contém a ontologia, realiza o *download* da ontologia ou processa o arquivo e armazena esta ontologia no banco de dados RDF.

Os especialistas, a partir do Módulo de Processamento de Anotações Semânticas, realizam anotações semânticas das séries temporais utilizando a OntoEx. Este módulo recebe as informações do usuário, cria a anotação semântica e a armazena no banco de dados RDF. A Figura 3.4 (primeiro protótipo) ilustra este processo de anotação. No caso de versionamento de anotações, este módulo cria a nova anotação, altera a data de modificação e os demais componentes adequados e conecta a anotação à sua nova versão e vice-versa.

A partir deste momento, um especialista que deseja buscar séries temporais utilizando o *TS³Annotation* utiliza a interface para acionar o Módulo de Busca por Séries Temporais informando o termo que deseja buscar, o arquivo de regras e a ontologia envolvida. Este módulo é responsável por retornar as séries temporais que satisfazem as condições de busca. Além disso, se comunica com o Módulo *Reasoner* para obter termos semanticamente relacionados com os termos contidos nas anotações. Por fim, o Módulo *Reasoner* é responsável por aplicar as regras nas ontologias expandindo as anotações semânticas. Detalhes de implementação são apresentados no Capítulo 4.

3.5 Conclusões

Este capítulo apresentou detalhes do *framework* proposto para auxiliar especialistas a realizar anotações semânticas e busca por séries temporais. O capítulo apresentou o modelo de anotação semântica criado, o algoritmo utilizado na busca por séries temporais e a arquitetura do *framework*. Além disso, mostrou exemplos para melhor compreensão de seu funcionamento. O próximo capítulo apresenta os detalhes de implementação e um estudo de caso criado para validar o *framework*.

Capítulo 4

Detalhes de Implementação e Estudo de Caso

Este capítulo apresenta os detalhes de implementação para a construção do *TS³Annotation*. A seção 4.1 apresenta as tecnologias utilizadas na implementação do *framework*. A seção 4.2 mostra o estudo de caso criado para validar esta dissertação, com algumas imagens de tela do sistema. Por fim, a seção 4.3 apresenta as conclusões do capítulo.

4.1 Detalhes de Implementação

O *framework* foi implementado como uma aplicação *Web*. O sistema foi implementado em uma arquitetura de três camadas, conhecida como MVC - *Model-View-Controller*. A camada *model* é responsável pelo acesso aos dados; *view* pela apresentação do conteúdo ao usuário; e *controller* é responsável pelo fluxo da aplicação, por responder aos eventos do usuário e integrar *model* e *view* [26]. Para a implementação do sistema utilizamos o Java Server Faces - JSF. De acordo com Jacobi e Fallows[15], JSF é um *framework* MVC que permite a construção de interfaces de usuário baseada em componentes.

A Figura 4.1 ilustra a arquitetura do *TS³Annotation* com as tecnologias adotadas para a implementação das três camadas do *framework*: Interface, Módulos e Camada de Persistência. Para a construção da interface utilizamos: o Twitter Bootstrap¹, um *framework* que possui HTML (*HyperText Markup Language*), CSS (*Cascading Style Sheets*) e JavaScript embutido facilitando o desenvolvimento de interfaces responsivas (que se adaptam a diferentes tamanhos de tela); e o PrimeFaces², um *framework* que estende o JSF com componentes mais sofisticados. A camada de Módulos e a Camada de Per-

¹<http://getbootstrap.com/2.3.2/index.html>

²<http://primefaces.org/>

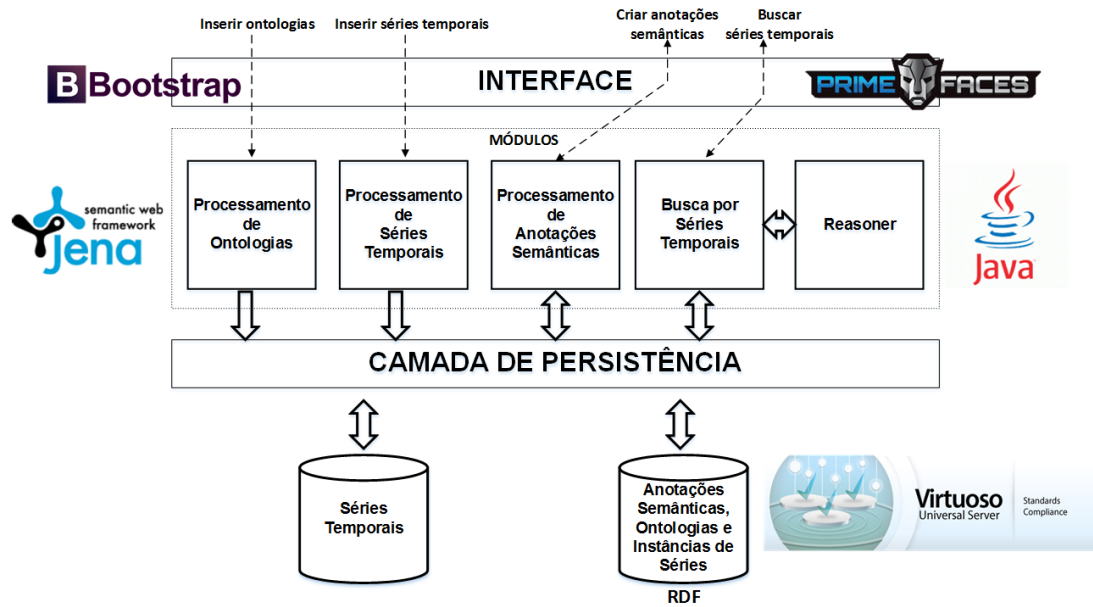


Figura 4.1: Tecnologias utilizadas na implementação do *framework*.

sistência foram implementadas utilizando o Jena³, um *framework* baseado em Java para construir aplicações da *Web Semântica* [2]. De acordo com Yu [44], o Jena possui os seguintes componentes:

1. uma API (Application Programming Interface) RDF;
2. uma API OWL;
3. leitura e escrita de RDF nos formatos RDF/XML, N3 e N-triplas;
4. criação de modelos RDF em memória ou armazenados em um banco de dados;
5. apoio a consultas em SPARQL⁴ (*SPARQL Protocol and RDF Query Language*);
6. mecanismo de inferência baseado em regras.

Na arquitetura do *TS³Annotation*, o Módulo de Processamento de Ontologias utiliza a API OWL do Jena para obter uma ontologia via arquivo ou URI e armazená-la no banco de dados. Os Módulos de Processamento de Séries Temporais e Anotações Semânticas utilizam os componentes 1 e 4 (acima) do Jena para, respectivamente, criar e armazenar

³<https://jena.apache.org/>

⁴<http://www.w3.org/TR/rdf-sparql-query/>

o conteúdo RDF da instância de uma série temporal e criar e armazenar o conteúdo RDF das anotações semânticas.

O Módulo de Busca por Séries Temporais utiliza o componente de consultas SPARQL do Jena (componente 5) para procurar anotações semânticas. SPARQL é uma linguagem de consulta RDF e protocolo de acesso a dados para a *Web Semântica* [44]. Utilizamos SPARQL para acessar o banco de dados RDF e procurar séries temporais anotadas ou semanticamente relacionadas com o termo informado pelo usuário.

O Módulo de Reasoner utiliza o componente 6 do Jena para realizar inferências em um modelo RDF. A Figura 4.2 ilustra a arquitetura básica de um *reasoner*. De acordo com Yu [44], um *reasoner* é utilizado para derivar triplas adicionais inerentes ao grafo RDF junto com as ontologias e usa suas próprias regras, axiomas e métodos de encadeamento (por exemplo, encadeamento para frente ou para trás). O *framework* Jena possui quatro *reasoners* pré-definidos:

1. Reasoner Transitivo: implementa apenas a transitividade e reflexividade das propriedades rdfs: subPropertyOf e rdfs: subClassOf;
2. Reasoner RDFS: implementa um conjunto configurável de implicações RDFS;
3. Reasoner OWL, OWL Mini, OWL Micro: contém uma implementação incompleta do OWL/Lite;
4. Reasoner de regras genéricas: suporta regras definidas pelo usuário, com encadeamento para frente, para trás e híbrido.

O *TS³ Annotation* utiliza o *reasoner* genérico. Dessa maneira, os usuários, além de utilizar as regras RDFS/OWL, podem construir suas próprias regras ou inserir regras RDFS/OWL que não estão implementadas no Jena. Regras são construídas seguindo a sintaxe descrita na Figura 4.3. Detalhes sobre a arquitetura do *reasoner* do Jena, outras primitivas e mais informações sobre o mecanismo de inferência do Jena podem ser encontrados em <https://jena.apache.org/documentation/inference/>.

O *TS³ Annotation* utiliza o Virtuoso Universal Server⁵ para armazenar seus dados. O Virtuoso possui um banco de dados de triplas RDF, suporta consultas SPARQL e pode ser integrado com o Jena. A opção alternativa, de usar um banco de dados relacional para armazenar o modelo RDF, foi descartada por razões de desempenho (ver, por exemplo, Yu [44] ou Souza [37] a respeito).

⁵<http://virtuoso.openlinksw.com/>

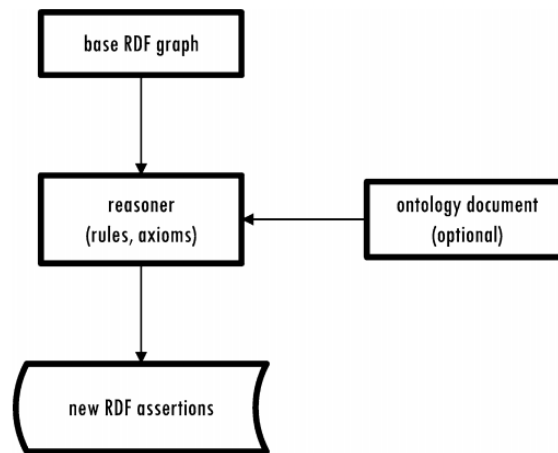


Figura 4.2: Arquitetura básica de um reasoner. Extraída de [44].

```

Rule      := bare-rule .
           or [ bare-rule ]
           or [ ruleName : bare-rule ]

bare-rule := term, ... term -> hterm, ... hterm // forward rule
           or bhterm <- term, ... term // backward rule

hterm     := term
           or [ bare-rule ]

term      := (node, node, node) // triple pattern
           or (node, node, functor) // extended triple pattern
           or builtin(node, ... node) // invoke procedural primitive

bhterm    := (node, node, node) // triple pattern

functor   := functorName(node, ... node) // structured literal

node      := uri-ref // e.g. http://foo.com/eg
           or prefix:localname // e.g. rdf:type
           or <uri-ref> // e.g. <myscheme:myuri>
           or ?varname // variable
           or 'a literal' // a plain string literal
           or 'lex'^^typeURI // a typed literal, xsd:* type names supported
           or number // e.g. 42 or 25.5
  
```

Figura 4.3: Descrição da sintaxe de regras. Extraída de <https://jena.apache.org/documentation/inference/>.

4.2 Estudo de Caso

Especialistas, muitas vezes, analisam ou criam associações entre séries e anotações. Além disso, buscam e correlacionam séries temporais de diferentes categorias (por exemplo, séries de temperatura, umidade, produção de grãos, culturas cultivadas etc) para estudar algum problema. Diante deste cenário, utilizamos o *TS³Annotation* para apoiar os especialistas na criação de anotações semânticas e busca por séries temporais. A subseção 4.2.1 descreve os dados utilizados para este estudo de caso e a subseção 4.2.2 mostra os testes realizados para validar a proposta desta dissertação, com algumas imagens do protótipo desenvolvido.

4.2.1 Dados utilizados

Os dados utilizados para validar esta dissertação foram fornecidos pela EMBRAPA. Trata-se de um conjunto de 417 arquivos, cada um contendo uma série temporal NDVI - *Normalized Difference Vegetation Index*, extraídas de imagens de satélite. De acordo com a NASA [28], o NDVI determina a densidade de verde em uma região e é calculado pela diferença da luz visível e infravermelho refletida pela vegetação. Quando a luz solar atinge uma vegetação, determinados comprimentos de onda são absorvidos e outros são refletidos. O valor do NDVI varia de -1 a +1 e, quanto mais verde estiver uma região, mais próximo de +1 será o seu valor.

As 417 séries temporais são da região do Mato Grosso, no período de 2000 a 2012 e contém ao total 2009 anotações com 29 conceitos anotados. A Tabela 4.1 mostra os conceitos anotados e a quantidade de anotações com aquele conceito. Os conceitos podem ser combinados em uma única anotação, por exemplo, existem anotações do tipo “Milho” e anotações do tipo “Soja / Milho”. Neste caso, foram contabilizadas duas anotações para o conceito de “Milho”.

Para possibilitar a recuperação de outras séries temporais, além das séries NDVI, simulamos duas séries temporais de temperatura e duas séries de produção de grãos. Cada série de temperatura possui uma anotação com o conceito de “Graus Dia” e cada série de produção de grãos possui uma anotação com o conceito de “Cercospora Zea Mays”. “Graus Dia” é um cálculo realizado, a partir de uma temperatura base, para determinar a quantidade de energia necessária para uma planta se desenvolver [27]. “Cercospora Zea Mays” é uma doença que afeta a plantação de Milho [10]. Portanto, temos um total de 421 séries temporais de diferentes categorias (NDVI, temperatura e produção de grãos) com 2013 anotações.

Para validar a proposta desta dissertação, as anotações textuais foram substituídas por anotações semânticas. As anotações com o conteúdo “Milho”, “Soja”, “Graus Dia” e “Cercospora Zea Mays” tiveram seu conteúdo semanticamente representado, enquanto

Tabela 4.1: Conceitos anotados e a quantidade de anotações.

	Conceitos	Quantidade de anotações
1	Milho	646
2	Pasto Degradado	17
3	Arroz	62
4	Soja	1611
5	Milheto	502
6	Soja Precoce	6
7	Boi	48
8	Algodão	168
9	Feijão	38
10	Juquira	13
11	Pasto	217
12	Abertura	29
13	Capim	1
14	Pasto Brizantão	2
15	Sorgo	44
16	Mata	24
17	Pousio	143
18	Enleiramento	2
19	Corte Seletivo	1
20	Crotalaria	20
21	Girassol	46
22	Nabo	20
23	Pé de Galinha	1
24	Sorgo Branco	1
25	Eucalipto	3
26	Milho Branco	1
27	Pipoca	35
28	Amendoim	7
29	Reflor	4

que as demais anotações tiveram seu conteúdo substituído por “OutroConteúdo”. A Figura 4.4 ilustra a instância da ontologia criada para semanticamente anotar as séries. Criamos esta instância a partir de informações extraídas do site da EMBRAPA⁶ e do Zoneamento Agrícola do Ministério de Agricultura⁷. Nesta Figura, é possível observar

⁶http://www.cnpms.embrapa.br/publicacoes/milho_6_ed/manejomilho.htm http://www.cnpms.embrapa.br/publicacoes/milho_6_ed/doencas.htm

⁷<http://www.agricultura.gov.br/politica-agricola/zoneamento-agricola/>

a representação dos conceitos utilizados nas anotações semânticas e os relacionamentos entre esses conceitos. Por exemplo, “Cercospora Zea Mays” é uma praga que afeta “Milho”; e “Milho” possui diferentes nomes, como “Corn” e “Maize”, sendo no Brasil muitas vezes utilizado na rotação de cultura junto com “Soja”. Além disso, possui um ciclo de plantio, que é o tempo desde a semeadura até a colheita, que pode ser classificado como “Precoces”, “Semi-precoces”, “Super precoces” e “Normais”. Cada ciclo possui diferentes exigências térmicas (intervalo de “Graus Dia”) para o desenvolvimento da planta.

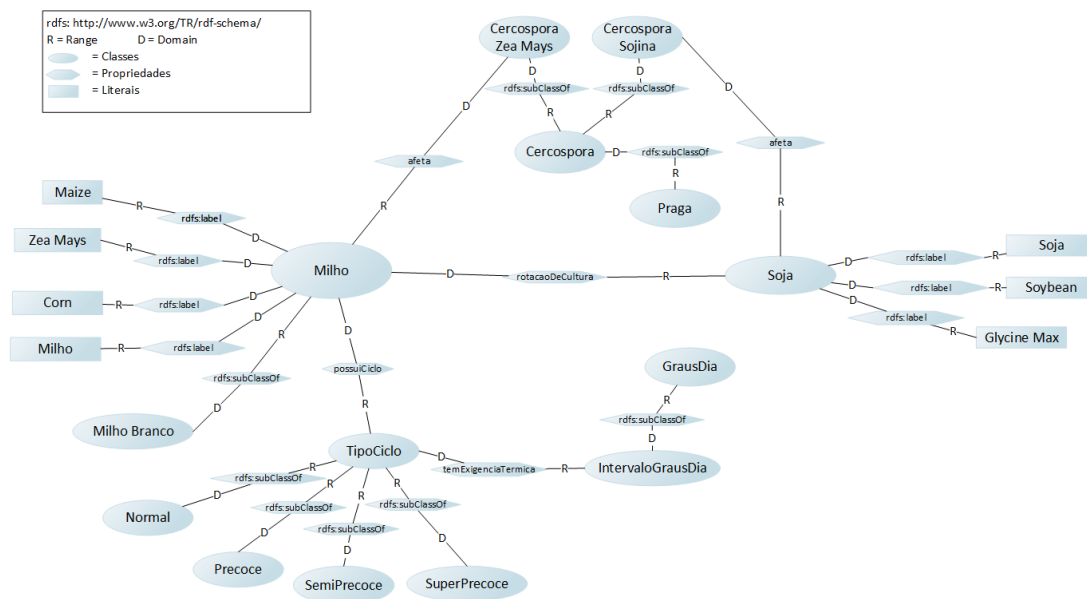


Figura 4.4: Instância de ontologia utilizada na criação das anotações semânticas.

Em resumo, os testes descritos na subseção 4.2.2 foram realizados em uma base de dados contendo 421 séries temporais do Estado do Mato Grosso, com período de 2000 a 2012. As séries possuem três categorias: NDVI, temperatura e produção de grãos. Existem 2013 anotações semânticas associadas às séries temporais: 646 anotações possuem como parte de seu conteúdo o conceito de “Milho”; 1611 anotações possuem o conceito “Soja”; 2 anotações possuem o conceito “Intervalo Graus Dia”; 2 anotações possuem o conceito “Cercospora Zea Mays”; 1 anotação possui o conceito “Milho Branco”; por fim, o termo “OutroConteudo” foi simbolicamente inserido no conteúdo das demais anotações.

4.2.2 Testes e Validação

O *TS³Annotation* possui duas funcionalidades básicas: anotação semântica e busca por séries temporais. Para anotar semanticamente uma série temporal, os especialistas inserem no sistema as séries temporais e a ontologias que serão utilizadas como conteúdo da

anotação. Para esta etapa, ainda não foi desenvolvido um protótipo de tela. Portanto, as séries temporais e a ontologia utilizadas para os testes foram inseridas via código Jena.

A Figura 4.5 ilustra o protótipo de tela desenvolvido para a criação de anotações semânticas. Na esquerda, o usuário seleciona a série temporal que deseja anotar. Esta série é graficamente exibida no componente à direita do formulário, com representações gráficas das anotações associadas a ela. Na parte inferior, é possível observar com mais detalhes as anotações da série escolhida. O usuário fornece os dados para a criação de uma anotação semântica pelo formulário localizado à esquerda. Vários elementos de ontologias devem ser separados por “;” e os demais componentes da anotação, que não aparecem no formulário, são automaticamente criados. Na atual versão do *framework*, a única característica da série temporal que está sendo automaticamente extraída é o ponto máximo. Além disso, o componente gráfico exibe apenas uma série temporal e apenas um ponto das anotações.

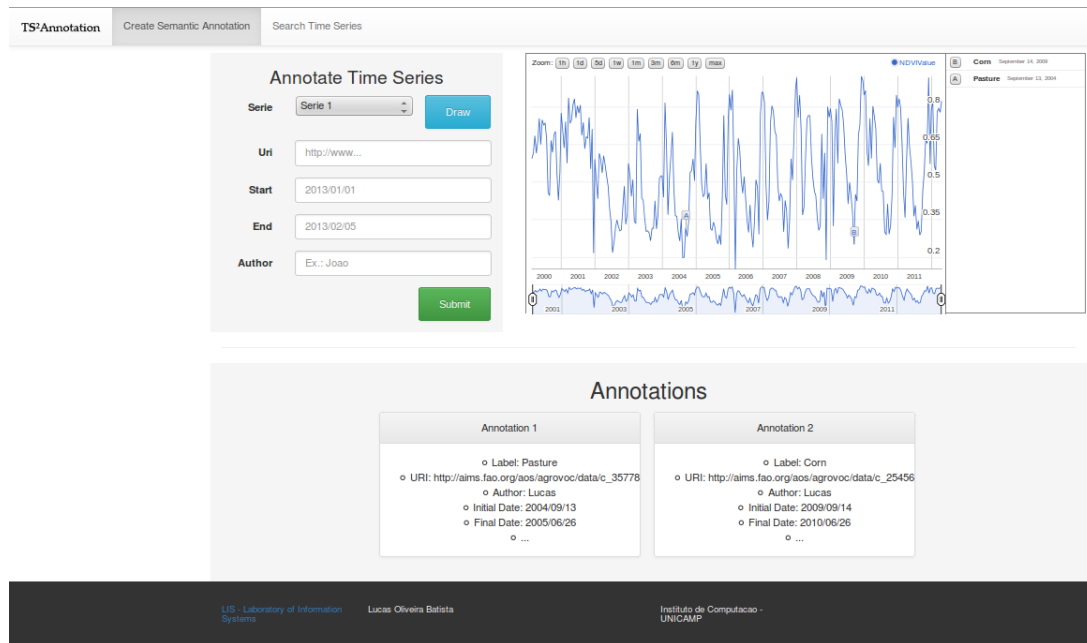


Figura 4.5: Tela do protótipo para criação de anotações semânticas.

A Figura 4.6 mostra uma anotação semântica, no formato RDF/XML, criada pelo *TS³Annotation*, com alguns comentários para melhor entendimento e sem as definições dos tempos da instância de série, pois são similares aos da anotação. Esta anotação é identificada pela URI <http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3>. O seu conteúdo possui o conceito de “Soja”, o autor é “Lucas Batista”, foi criada e modificada em “2015-05-07”, está associada à série “TSerie_area_1” no intervalo de “20070914 - 20080625”, o número de sua versão é “1” e, por fim, tem como

ponto máximo “0.9456”.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:annotate="http://www.w3.org/2000/10/annotation-ns#"
  xmlns:tssam="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#"
  xmlns:time="http://www.w3.org/2006/time#"
  xmlns:estudocaso="http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" >

  ### ANOTAÇÃO SEMÂNTICA ###
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3">
    <tssam:semanticBody rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#Soja"/>
    <tssam:hasNumberVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</tssam:hasNumberVersion>
    <annotate:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-07T12:00:00-0300</annotate:modified>
    <annotate:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-07T12:00:00-0300</annotate:created>
    <annotate:author rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Lucas Batista</annotate:author>
    <annotate:annotates rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSerie_area_1"/>
    <tssam:hasTSProperties rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSProperties_0.9456"/>
    <tssam:hasInterval rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#IntervaloAnotacao_2007091420080625"/>
    <rdf:type rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation"/>
  </rdf:Description>

  ### INSTÂNCIA DA SÉRIE TEMPORAL ###
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSerie_area_1">
    <tssam:tsEnd rdf:resource="http://www.w3.org/2006/time#20120422"/>
    <tssam:tsBegin rdf:resource="http://www.w3.org/2006/time#20000218"/>
    <geo:long>-55.380001</geo:long>
    <geo:lat>-11.773094</geo:lat>
    <rdf:type rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TimeSeries"/>
  </rdf:Description>

  ### PROPRIEDADES DA SUBSÉRIE ASSOCIADA À ANOTAÇÃO ###
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSProperties_0.9456">
    <tssam:hasMaxPoint>0.9456</tssam:hasMaxPoint>
    <rdf:type rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TimeSeriesProperties"/>
  </rdf:Description>

  ### INTERVALO DA ANOTAÇÃO ###
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#IntervaloAnotacao_2007091420080625">
    <time:hasEnd rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#InstanteFinalAnotacao_20080625"/>
    <time:hasBeginning rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#InstanteIncialAnotacao_20070914"/>
    <rdf:type rdf:resource="http://www.w3.org/2006/time#Interval"/>
  </rdf:Description>

  ### INSTANTES INICIAL E FINAL DA ANOTAÇÃO ###
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#InstanteFinalAnotacao_20080625">
    <time:inXSDDateTime rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2008-06-25T12:00:00-0300</time:inXSDDateTime>
    <rdf:type rdf:resource="http://www.w3.org/2006/time#Instant"/>
  </rdf:Description>

  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#InstanteIncialAnotacao_20070914">
    <time:inXSDDateTime rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2007-09-14T12:00:00-0300</time:inXSDDateTime>
    <rdf:type rdf:resource="http://www.w3.org/2006/time#Instant"/>
  </rdf:Description>

</rdf:RDF>
```

Figura 4.6: Anotação semântica em RDF/XML.

Além de anotar séries temporais, os especialistas podem colaborar a partir do versionamento de anotações. Um especialista pode alterar o conteúdo de uma anotação semântica

a partir de seu versionamento. Relembrando, o versionamento de uma anotação ocorre criando uma nova anotação a qual está relacionada à anotação de versão anterior pelas propriedades “hasNewVersion” e “hasOldVersion”. A Figura 4.7 ilustra o versionamento, feito via código Jena, da anotação mostrada na Figura 4.6. O *TS³Annotation* cria a nova anotação (...*SemanticAnnotation3_2*) com o *SemanticAnnotation* igual ao da anotação com versão anterior (...*SemanticAnnotation3*) acrescido de um *underline* e o novo número de versão. Na Figura 4.7, a nova versão da anotação possui como novo conteúdo o conceito de “Milho”, o autor “Mateus Batista”, o número de sua versão é “2” e a data de modificação é “2015-05-14”. Os demais conteúdos permaneceram os mesmos. Além disso, é possível observar que a propriedade “hasNewVersion” foi adicionada à anotação de versão número “1”.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:annotate="http://www.w3.org/2000/10/annotation-ns#"
  xmlns:tssam="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#"
  xmlns:time="http://www.w3.org/2006/time#"
  xmlns:estudocaso="http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#" >

  ## ANOTAÇÃO NA VERSÃO 1##
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3">
    <annotate:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-07T12:00:00-0300</annotate:created>
    <annotate:annotates rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#Serie_area_1"/>
    <tssam:hasInterval rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#IntervaloAnotacao_2007091420080625"/>
    <annotate:author rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Lucas Batista</annotate:author>
    <tssam:hasTSProperties rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSProperties_0.9456"/>
    <tssam:hasSemanticBody rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#Soja"/>
    <annotate:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-07T12:00:00-0300</annotate:modified>
    <rdf:type rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotationIdentifier"/>
    <tssam:hasNewVersion rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3_2"/>
    <tssam:hasNumberVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</tssam:hasNumberVersion>
  </rdf:Description>

  ## ANOTAÇÃO NA VERSÃO 2##
  <rdf:Description rdf:about="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3_2">
    <annotate:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-07T12:00:00-0300</annotate:created>
    <tssam:hasInterval rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#IntervaloAnotacao_2007091420080625"/>
    <annotate:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-05-14T12:00:00-0300</annotate:modified>
    <annotate:annotates rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#Serie_area_1"/>
    <annotate:author rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Mateus Batista</annotate:author>
    <tssam:hasSemanticBody rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#Milho"/>
    <rdf:type rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation"/>
    <tssam:hasTSProperties rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#TSProperties_0.9456"/>
    <tssam:hasOldVersion rdf:resource="http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#SemanticAnnotation3"/>
    <tssam:hasNumberVersion rdf:datatype="http://www.w3.org/2001/XMLSchema#int">2</tssam:hasNumberVersion>
  </rdf:Description>

  [...]
```

Figura 4.7: Anotação semântica versionada.

Para validar a principal contribuição desta dissertação, comparamos a busca por séries temporais utilizando anotações de texto livre e utilizando o *TS³Annotation*. O termo de busca utilizado foi “Milho”, sobre a base anotada de Silva [36], séries e a ontologia

mostrada na Figura 4.4.

Utilizando a comparação de *string* entre “Milho” e o conteúdo das anotações em texto livre foi possível recuperar 264 séries temporais NDVI. Estas séries têm 647 anotações que possuem como parte de seu conteúdo o termo buscado. Utilizando o *TS³Annotation*, sem aplicar as regras de expansão, conseguimos recuperar as mesmas séries temporais. Em um segundo teste, executou-se a busca usando como entrada sinônimos de “Milho”, como “Corn” ou “Maize”. Neste caso, a busca via texto livre não encontra séries e o *TS³Annotation* encontra as mesmas 264 séries. As tabelas 4.2 e 4.3 resumem, respectivamente, o primeiro e o segundo teste realizados.

Tabela 4.2: Resultados obtidos utilizando o parâmetro “Milho”, sem utilizar as regras de expansão do *TS³Annotation*.

	Séries Temporais			Anotações
	NDVI	Temperatura	Produção de Grãos	
Anotações Texto Livre	264	0	0	647
<i>TS³Annotation</i> (sem expansão)	264	0	0	647

Tabela 4.3: Resultados obtidos utilizando os parâmetros “Corn” e “Maize”, sem utilizar as regras de expansão do *TS³Annotation*.

	Séries Temporais			Anotações
	NDVI	Temperatura	Produção de Grãos	
Anotações Texto Livre	0	0	0	647
<i>TS³Annotation</i> (sem expansão)	264	0	0	647

O terceiro teste verifica a busca com expansão das anotações semânticas. A Figura 4.8 ilustra as regras utilizadas para expandir as anotações semânticas. A regra 0 não expande anotações, mas representa uma transitividade que relaciona um elemento (?ontologyElement) com as classes superiores (?ontologyElement3) utilizando a mesma propriedade (?propriedade) dos elementos (?ontologyElement2) que diretamente se relacionam com ele. A regra 1 é uma transitividade que expande anotações de conteúdo “Graus Dia” adicionando ao seu conteúdo o conceito “Milho”. A regra 2 expande anotações que possuem elementos que são utilizados na rotação de cultura. Por fim, a regra 3 permite derivar que anotações de pragas que afetam “Milho” devem também ter em seu conteúdo o conceito de “Milho”.

```

@prefix tssam: <http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMOnto#>.
@prefix esc: <http://www.lis.ic.unicamp.br/~lucasbatista/EstudoCasoOnto#>.
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>

[regra0: (?ontologyElement ?propriedade ?ontologyElement2)
      (?ontologyElement2 rdfs:subClassOf ?ontologyElement3)-> (?ontologyElement ?propriedade ?ontologyElement3)]

[regra1: (esc:Milho ?propriedade ?elemento) (?elemento ?propriedade2 esc:GrausDia)
      (?anotacao tssam:hasSemanticBody esc:GrausDia)->
      (?anotacao tssam:hasSemanticBody esc:Milho)]

[regra2: (esc:Milho esc:rotacaoDeCultura ?ontologyElement)
      (?anotacao tssam:hasSemanticBody ?ontologyElement)->
      (?anotacao tssam:hasSemanticBody esc:Milho)]

[regra3: (?ontologyElement rdfs:subClassOf esc:Praga)
      (?ontologyElement esc:afeta esc:Milho)
      (?anotacao tssam:hasSemanticBody ?ontologyElement)->
      (?anotacao tssam:hasSemanticBody esc:Milho)]

```

Figura 4.8: Regras utilizadas na busca por séries temporais.

Tabela 4.4: Resultados obtidos utilizando o parâmetro “Milho”, utilizando as regras de expansão do *TS³Annotation*.

	Séries Temporais			Anotações
	NDVI	Temperatura	Produção de Grãos	
Anotações Texto Livre	264	0	0	647
<i>TS³Annotation</i> (com expansão)	382	2	2	1633

A tabela 4.4 mostra os resultados obtidos para o terceiro teste. Efetuando novamente a consulta utilizando o parâmetro “Milho” no *TS³Annotation*, com a aplicação das regras de expansão, recuperamos 386 séries. Do conjunto de séries recuperadas, 1633 anotações contêm “Milho”, “Soja”, “Cercospora Zea Mays” ou “Intervalo Graus Dia” como parte de seu conteúdo.

A Figura 4.9 ilustra quatro séries temporais anotadas semanticamente utilizando o TSSAM. As séries 1 e 3 são séries NDVI anotadas com “Milho” e “Soja”, respectivamente; a série 2 é uma série de temperatura anotada com “Intervalo Graus Dia”; e, por fim, a série 4 é uma série de produção de grãos anotada com “Cercospora Zea Mays”. Do conteúdo das anotações semânticas, apenas o elemento da ontologia foi graficamente representado para uma melhor visualização.

O *TS³Annotation* recupera a série 1 usando a comparação direta do conteúdo da anotação semântica com o termo buscado (caso 1 do Algoritmo 1). As demais séries são recuperadas pelo relacionamento semântico entre os termos da ontologia e o termo de busca (caso 2 do Algoritmo 1). Assim, além das séries NDVI de “Milho” (recuperadas a partir da comparação direta), será retornado ao usuário séries NDVI de “Soja” (recuperadas a partir da regra 2), séries de temperatura (recuperadas a partir das regras 0 e 1)

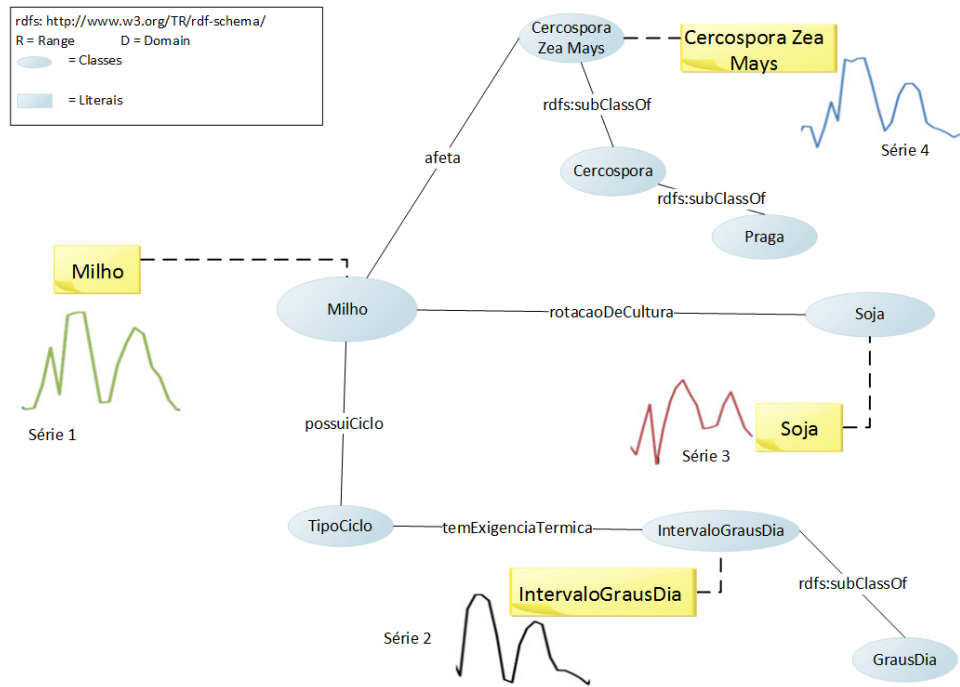


Figura 4.9: Exemplos de séries temporais e suas anotações semânticas utilizadas no estudo de caso. Todas são recuperadas pela busca por “Milho”.

e séries da produção de grãos (recuperadas a partir da regra 3) totalizando as 386 séries temporais recuperadas.

A Figura 4.10 ilustra a aplicação das regras 0 e 1 para expandir a anotação semântica da série 2, com o conceito de “Milho”. Após a expansão, a série é adicionada ao conjunto resultado devido ao relacionamento semântico do conteúdo de sua anotação com o termo buscado. Notar que, comparando com a Figura 4.9, na Figura 4.10 a expansão da anotação da série 2 associa o conceito de “Milho” ao seu conteúdo. Esta expansão está simbolicamente representada pela linha verde pontilhada.

Apenas as últimas versões das anotações são consideradas na busca por séries temporais. De maneira similar, navegando nas ontologias e expandindo anotações, séries de outras categorias (por exemplo, séries de umidade, financeiras, etc.) poderiam ser recuperadas pelo *framework*.

A Figura 4.11 ilustra o protótipo de tela criado para buscar séries temporais. Na barra superior o especialista informa o termo que deseja buscar e abaixo o resultado é exibido com as séries temporais e suas respectivas anotações semânticas.

A Figura 4.12 ilustra 10 instâncias de séries temporais recuperadas pelo *TS³ Annotation*. Em www.lis.ic.unicamp.br/~lucasbatista/estudocaso/ encontram-se os dados utilizados neste estudo de caso: um arquivo contendo as 386 instâncias de séries recuperadas;

```
[regra0: (?ontologyElement ?propriedade ?ontologyElement2)
        (?ontologyElement2 rdfs:subClassOf ?ontologyElement3)->
        (?ontologyElement ?propriedade ?ontologyElement3)]

[regra1: (esc:Milho ?propriedade ?elemento) (?elemento ?propriedade2
esc:GausDia)
        (?anotacao tssam:hasSemanticBody esc:GausDia)->
        (?anotacao tssam:hasSemanticBody esc:Milho)]
```

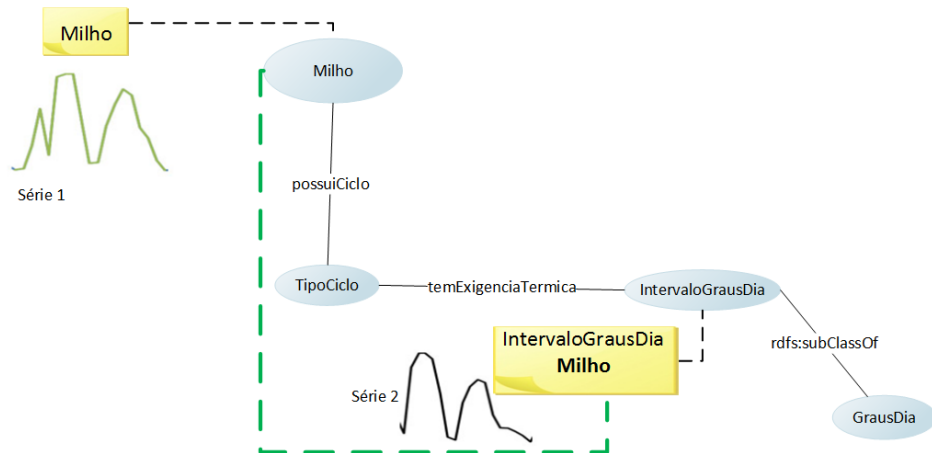


Figura 4.10: Representação gráfica da expansão de uma anotação semântica.

um arquivo com as regras; um arquivo com o modelo em RDF/XML que contém as anotações semânticas, a ontologia e as instâncias de séries temporais.

Figura 4.11: Tela utilizada na busca por séries temporais.

```
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_298 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_270 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_184 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_185 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_259 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_404 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_48 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_218 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_152 . }  
{ http://www.lis.ic.unicamp.br/~lucasbatista/TSSAMonto#TSerie_area_65 . }
```

Figura 4.12: Conjunto de 10 séries temporais presentes no resultado de busca do *TS³Annotation*.

4.3 Conclusões

Este capítulo apresentou os aspectos de implementação da arquitetura do *TS³Annotation*, incluindo as tecnologias utilizadas na sua implementação e os repositórios utilizados pelo *framework*. O capítulo também apresentou a arquitetura básica de um *reasoner* e a sintaxe para a criação de regras. Além disso, discutiu os testes realizados para a validação da proposta desta dissertação. O próximo capítulo apresenta as conclusões e trabalhos futuros desta dissertação.

Capítulo 5

Conclusões e Trabalhos Futuros

Durante a análise de séries temporais, muitas vezes, especialistas associam anotações em texto livre às séries. Além disso, precisam buscar e correlacionar diversas séries temporais para estudar algum problema. Anotações em texto livre dificultam a busca, o compartilhamento e a integração de dados e os sistemas de busca por séries não permitem buscar séries temporais utilizando informações complementares nem sempre diretamente associadas às séries. Esta dissertação é um passo para resolver estes problemas permitindo que séries temporais sejam anotadas semanticamente e apresentando uma nova abordagem na busca por séries temporais a partir das anotações semânticas.

5.1 Conclusões

O *TS³ Annotation* é um primeiro passo para permitir que especialistas estudem correlações entre séries temporais. O *framework* proposto nesta dissertação permite que o usuário recupere séries temporais que possuam alguma relação com um termo de busca utilizado. Além disso, especialistas podem criar anotações semânticas de séries temporais e colaborar a partir do versionamento de anotações. As contribuições desta dissertação são:

- Para a Computação:
 - Um modelo em RDF para anotações semânticas de séries temporais. O modelo apresentado permite associar anotações semânticas a diferentes trechos de uma série temporal. Além disso, permite o versionamento de anotações e é um modelo flexível, possibilitando que novos elementos possam ser adicionados de acordo com a necessidade;
 - Um algoritmo que utiliza anotações semânticas na busca por séries temporais, tirando proveito das informações complementares que estão associadas às séries. O algoritmo proposto utiliza regras de inferência para permitir a busca

por séries utilizando informações que nem sempre estão diretamente associadas às séries.

- Para as outras áreas:
 - Um *framework* que utiliza anotações semânticas como uma nova abordagem na busca por séries temporais. Além disso, torna possível a criação de anotações semânticas em séries temporais e a colaboração entre especialistas por meio do versionamento de anotações. Este *framework* possibilita aos especialistas economizar tempo na busca por séries e recuperar outras séries temporais que podem ser relevantes para uma análise diferente da sua pesquisa.

5.2 Trabalhos Futuros

Esta pesquisa tem várias possibilidades de extensão:

- **Anotações semânticas auxiliando o reconhecimento automático de culturas em séries NDVI** - Creus *et al.* [7], utilizam aprendizado de máquina para reconhecer as culturas presentes em uma série temporal NDVI, a partir das características das séries. Anotações semânticas poderiam ser incluídas como um parâmetro no aprendizado de culturas, possibilitando aumentar as chances do reconhecimento de uma determinada cultura;
- **Visualização de Séries Temporais** - Não é possível visualizar graficamente as séries temporais que são encontradas em uma determinada busca feita pelo *TS³ Annotation*. Portanto, uma possível extensão deste trabalho é na visualização do resultado de uma busca que consiste de séries temporais e suas anotações. Vários desafios estão envolvidos, por exemplo, como visualizar séries temporais com diferentes períodos de tempo, como visualizar uma grande quantidade de séries temporais e como visualizar, de maneira eficiente, as anotações que estão associadas às séries;
- **Extração automática de regras** - Atualmente as regras utilizadas pelo *TS³ Annotation* são informadas pelo usuário. Uma possível extensão deste trabalho é a extração automática de regras das ontologias, como realizado por Carlson *et al.* [5], potencializando a busca realizada pelo *framework*;
- **Desenvolvimento de ferramenta para facilitar a especificação de regras** - As regras específicas, na versão atual, são inseridas manualmente pelo usuário. Uma extensão seria desenvolver uma ferramenta para facilitar esta atividade;

- **Linked Data** - Outra possível extensão é conectar as ontologias informadas pelo usuário com outras ontologias, por exemplo a Dbpedia, aumentando a possibilidade de encontrar novos termos que podem ser utilizados para buscar séries temporais. Ainda, outra extensão neste sentido seria associar autores de séries às ontologias;
- **Apoio à busca por séries temporais utilizando um termo e um período de tempo ou localização** - Na versão atual o *TS³Annotation* considera apenas o termo de busca informado, ignorando informações espaço-temporais associadas às anotações. Assim, outra extensão é adicionar o período de tempo e/ou localização aos parâmetros de busca;
- **Comentário das anotações** - Atualmente o *framework* apenas armazena a mudança que ocorreu na anotação. Entretanto, poderiam ser armazenados comentários sobre uma anotação. Estes comentários poderiam ser a discussão, entre especialistas, que gerou a mudança no conteúdo da anotação.
- **Experimentos de usabilidade e acessibilidade** - Não foram realizados testes com usuários. Este seria mais uma extensão do trabalho;
- **Melhoras na interface do *framework*** - Existem algumas melhorias que necessitam ser feitas na interface do *TS³Annotation*. Por exemplo, o componente utilizado para exibir graficamente uma série temporal na tela, para criar uma anotação semântica, não permite visualizar todo o intervalo ao qual a anotação está associada. Além disso, seria importante o usuário ter alguma possibilidade de visualização gráfica do resultado de uma busca, incluindo a visualização gráfica das séries retornadas em uma busca e suas anotações. Um trabalho de iniciação científica foi iniciado visando melhorias na interface do *TS³Annotation* e a implementação de alguma técnica de visualização de séries temporais. Ainda, outra extensão nesse sentido é integrar ao *TS³Annotation* as interfaces (mostradas nas Figuras 5.2, 5.4, 5.1 e 5.3) desenvolvidas na iniciação científica por Larissa Toome.

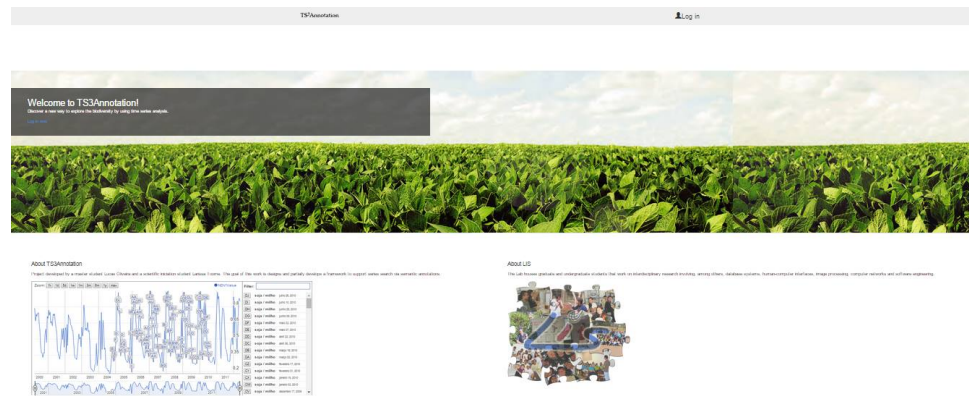


Figura 5.1: Tela inicial - proposta de melhoria de interface.

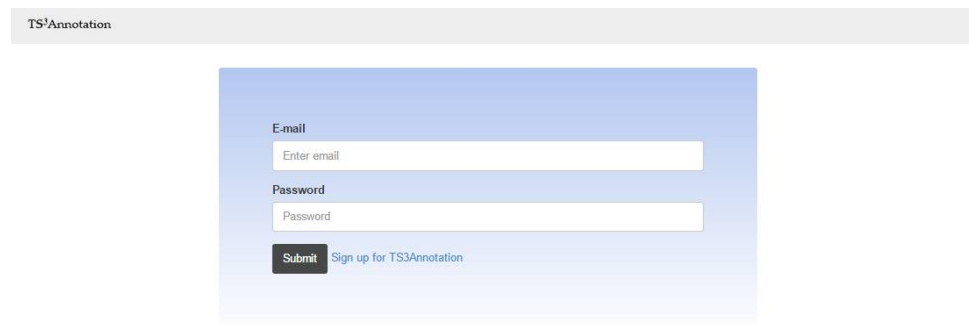


Figura 5.2: Tela de Login - considera a inclusão de usuários para controle de autorização.

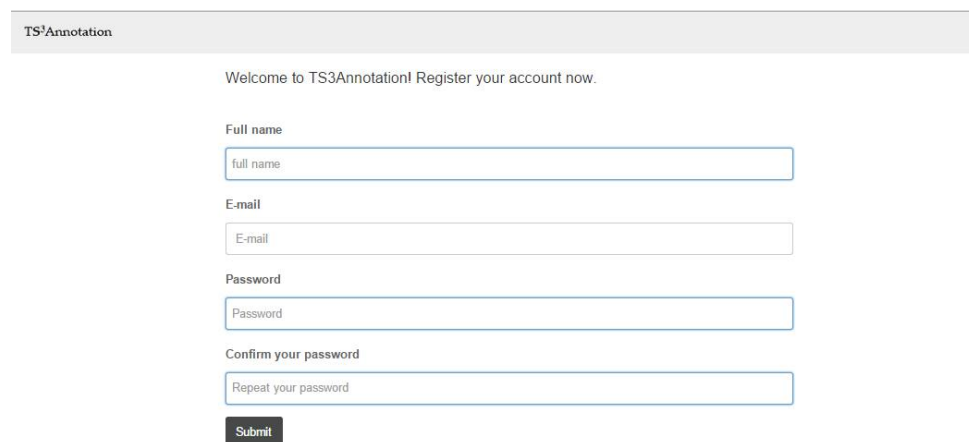


Figura 5.3: Tela para a criação de uma conta.



Figura 5.4: Tela para a criação de anotações em séries temporais e visualização gráfica das anotações, associadas a uma série, com seu tempo inicial e final. A parte esquerda mostra uma inserção de anotação. À direita, todas as anotações da série ilustrada no centro da tela.

Referências Bibliográficas

- [1] Ryo Aoto, Toshiyuki Shimizu, and Masatoshi Yoshikawa. Propagation of multi-granularity annotations. In Abdelkader Hameurlain, StephenW. Liddle, Klaus-Dieter Schewe, and Xiaofang Zhou, editors, *Database and Expert Systems Applications*, volume 6861 of *Lecture Notes in Computer Science*, pages 589–603. Springer Berlin Heidelberg, 2011.
- [2] Apache. Apache jena. <https://jena.apache.org/>. Acessado: 10-05-2015.
- [3] Johannes Aßfalg, Hans-Peter Kriegel, Peer Kröger, Peter Kunath, Alexey Pryakhin, and Matthias Renz. Tquest: threshold query execution for large sets of time series. In *Advances in Database Technology-EDBT 2006*, pages 1147–1150. Springer, 2006.
- [4] Alice Berard and Georges Hebrail. Searching time series with hadoop in an electric power company. In *Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, pages 15–22. ACM, 2013.
- [5] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka Jr, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, volume 5, page 3, 2010.
- [6] Chris Chatfield. *The Analysis of Time Series: An Introduction, Sixth Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2003.
- [7] Jordi Creus-Tomàs, Fabio Augusto Faria, Júlio César Dalla Mora Esquerdo, Alexandre Camargo Coutinho, and Claudia Bauzer Medeiros. Sircub – brazilian agricultural crop recognition system. In *XVII Simpósio Brasileiro de Sensoriamento Remoto*, 2015.
- [8] Secretaria do Tesouro Nacional. Séries temporais. http://www3.tesouro.fazenda.gov.br/series_temporais/principal.aspx#ancora_consulta. Acessado: 03-09-2013.

- [9] Alexandros Doulamis, Nikos Pelekis, and Yannis Theodoridis. Easytracker: An android application for capturing mobility behavior. In *Informatics (PCI), 2012 16th Panhellenic Conference on*, pages 357–362. IEEE, 2012.
- [10] EMBRAPA. Cultivo do milho. http://www.cnpms.embrapa.br/publicacoes/milho_6_ed/doencas.htm. Acessado: 10-05-2015.
- [11] Jérôme Euzenat. Eight questions about semantic web annotations. *Intelligent Systems, IEEE*, 17(2):55–62, 2002.
- [12] Like Gao and Xiaoyang Sean Wang. Time series query. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 3114–3119. Springer US, 2009.
- [13] Baris Guc, Michael May, Yucel Saygin, and Christine Körner. Semantic annotation of gps trajectories. In *11th AGILE international conference on geographic information science*, 2008.
- [14] Harry Hochheiser and Ben Shneiderman. Dynamic query tools for time series data sets: timebox widgets for interactive exploration. *Information Visualization*, 3(1):1–18, 2004.
- [15] Jonas Jacobi and John R Fallows. *Pro JSF and Ajax: building rich Internet components*. Apress, 2006.
- [16] Christian S. Jensen and Richard T. Snodgrass. Semantics of time-varying information. *INFORMATION SYSTEMS*, 21:311–352, 1996.
- [17] José Kahan, Marja-Riita Koivunen, Eric Prud’Hommeaux, and Ralph R. Swick. Annotea: an open {RDF} infrastructure for shared web annotations. *Computer Networks*, 39(5):589 – 608, 2002.
- [18] Atanas Kiryakov, Borislav Popov, Ivan Terziev, Dimitar Manov, and Damyan Ognyanoff. Semantic annotation, indexing and retrieval. *Journal of Web Semantics*, 2(1):49–79, 2004.
- [19] Gavin CKW Koh, Gemma Hawthorne, Alice M Turner, Heinke Kunst, and Martin Dedicoat. Tuberculosis incidence correlates with sunshine: an ecological 28-year time series study. *PloS one*, 8(3):e57752, 2013.
- [20] Yongxin Liao, Mario Lezoche, Hervé Panetto, and Nacer Boudjlida. Semantic annotation model definition for systems interoperability. In Robert Meersman, Tharam Dillon, and Pilar Herrero, editors, *On the Move to Meaningful Internet Systems:*

- OTM 2011 Workshops*, volume 7046 of *Lecture Notes in Computer Science*, pages 61–70. Springer Berlin Heidelberg, 2011.
- [21] Yun Lin. *Semantic Annotation for Process Models : Facilitating Process Knowledge Management via Semantic Interoperability*. PhD thesis, Norwegian University of Science and Technology, Department of Computer and Information Science, 2008.
- [22] André Lourenço, Hugo Plácido da Silva, Carlos Carreiras, Ana Priscila Alves, and Ana LN Fred. A web-based platform for biosignal visualization and annotation. *Multimedia Tools and Applications*, 70(1):433–460, 2014.
- [23] Andreas Lykke-Olesen and Jesper Nielsen. bibphone: adding sound to the children’s library. In *Proceedings of the 6th international conference on Interaction design and children*, IDC ’07, pages 145–148, New York, NY, USA, 2007. ACM.
- [24] Xiaoyi Ma, Haejoong Lee, Steven Bird, and Kazuaki Maeda. Models and tools for collaborative annotation. *CoRR*, cs.CL/0204004, 2002.
- [25] Carla Geovana Nascimento Macario. *Anotação Semântica de Dados Geoespaciais*. PhD thesis, Universidade Estadual de Campinas - UNICAMP, 2009.
- [26] Kito D Mann. *Java Server Faces in Action*. Dreamtech Press, 2005.
- [27] Marcelo Notti Miranda and José Holanda Campelo Júnior. Soma térmica para o sub-período semeadura-maturação de feijão cv. carioca em colorado do oeste, rondônia. *Pesquisa Agropecuária Tropical (Agricultural Research in the Tropics)*, 40(2):DOI–10, 2010.
- [28] NASA. Ndvi - normalized difference vegetation index. http://earthobservatory.nasa.gov/Features/MeasuringVegetation/measuring_vegetation_2.php. Acessado: 10-05-2015.
- [29] Tripti Negi and Veena Bansal. Time series: Similarity search and its applications. In *Proceedings of the International Conference on Systemics, Cybernetics and Informatics: ICSCI-04, Hyderabad, India*, pages 528–533, 2005.
- [30] Bank of Japan. Boj time-series data search. http://www.stat-search.boj.or.jp/index_en.html. Acessado: 03-09-2013.
- [31] Eyal Oren, Knud Möller, Simon Scerri, Siegfried Handschuh, and Michael Sintek. What are semantic annotations. *Relatório técnico. DERI Galway*, 2006.

- [32] Tomás Pariente, José María Fuentes, María Angeles Sanguino, Sinan Yurtsever, Giuseppe Avellino, Andrea E Rizzoli, and Saša Nešić. A model for semantic annotation of environmental resources: the tatoo semantic framework. In *Environmental Software Systems. Frameworks of eEnvironment*, pages 419–427. Springer, 2011.
- [33] William B. S. Pressly, Jr. Tspad: a tablet-pc based application for annotation and collaboration on time series data. In *Proceedings of the 46th Annual Southeast Regional Conference on XX*, ACM-SE 46, pages 166–171, New York, NY, USA, 2008. ACM.
- [34] Davood Rafiei and Alberto O. Mendelzon. Querying time series data based on similarity. *Knowledge and Data Engineering, IEEE Transactions on*, 12(5):675–693, 2000.
- [35] Lawrence Reeve and Hyoil Han. Survey of semantic annotation platforms. In *Proceedings of the 2005 ACM symposium on Applied computing*, pages 1634–1638. ACM, 2005.
- [36] Felipe Henriques Silva. Serial annotator: Managing annotations of time series. Master’s thesis, Universidade Estadual de Campinas - UNICAMP, 2013.
- [37] Sidney Roberto Sousa. Gerenciamento de anotações semânticas de dados na web para aplicações agrícolas. Master’s thesis, Universidade Estadual de Campinas - UNICAMP, 2010.
- [38] Hassina Nacer Talantikite, Djamil Aissani, and Nacer Boudjlida. Semantic annotations for web services discovery and composition. *Computer Standards & Interfaces*, 31(6):1108 – 1117, 2009.
- [39] Michael Tsang, George W. Fitzmaurice, Gordon Kurtenbach, Azam Khan, and Bill Buxton. Boom chameleon: simultaneous capture of 3d viewpoint, voice and gesture annotations on a spatially-aware display. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, UIST ’02, pages 111–120, New York, NY, USA, 2002. ACM.
- [40] Victoria Uren, Philipp Cimiano, Jose Iria, Siegfried Handschuh, Maria Vargas-Vera, Enrico Motta, and Fabio Ciravegna. 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*, 4(1):14–28, 2006.
- [41] Tatiana von Landesberger, Viktor Voss, and Jörn Kohlhammer. Semantic search and visualization of time-series data. In *Networked Knowledge-Networked Media*, pages 205–216. Springer, 2009.

- [42] Martin Wattenberg. Sketching a graph to query a time-series database. In *CHI'01 Extended Abstracts on Human factors in Computing Systems*, pages 381–382. ACM, 2001.
- [43] Zhixian Yan, Dipanjan Chakraborty, Christine Parent, Stefano Spaccapietra, and Karl Aberer. Semitri: a framework for semantic annotation of heterogeneous trajectories. In *Proceedings of the 14th international conference on extending database technology*, pages 259–270. ACM, 2011.
- [44] Liyang Yu. *A developer's guide to the semantic Web*. Springer Science & Business Media, 2011.