

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Paulo Victor Hauck da Encarnação

**FSI: Uma Infraestrutura de Apoio ao Projeto  
FrameNet Utilizando Web Semântica.**

Juiz de Fora

2014

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Paulo Victor Hauck da Encarnação

**FSI: Uma Infraestrutura de Apoio ao Projeto  
FrameNet Utilizando Web Semântica.**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Juiz de Fora

2014

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Encarnação, Paulo Victor Hauck da.

FSI : Uma Infraestrutura de Apoio ao Projeto FrameNet Utilizando Web Semântica. / Paulo Victor Hauck da Encarnação. -- 2014.

147 p. : il.

Orientadora: Regina Maria Maciel Braga Villela

Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2014.

1. FrameNet. 2. Semântica de Frames. 3. Web Semântica. 4. Infraestrutura. I. Braga Villela, Regina Maria Maciel, orient. II. Título.

Paulo Victor Hauck da Encarnação

**FSI: Uma Infraestrutura de Apoio ao Projeto FrameNet  
Utilizando Web Semântica.**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Ciências Exatas da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Aprovada em 4 de Setembro de 2014.

BANCA EXAMINADORA

---

Profa. D.Sc. Regina Maria Maciel Braga Villela - Orientador  
Universidade Federal de Juiz de Fora

---

Profa. D.Sc. Fernanda Cláudia Alves Campos  
Universidade Federal de Juiz de Fora

---

Profa. D.Sc. Vera Maria Benjamim Werneck  
Universidade Estadual do Rio de Janeiro

*A todos aqueles que apoiaram e  
acreditaram.*

# AGRADECIMENTOS

Primeiramente agradeço a Deus pela oportunidade e capacidade dada a nós para que possamos realizar nossos sonhos. Aos meus pais Ieda e Carlos que sempre apoiaram os caminhos que escolhi para minha vida. Assim como devo agradecer a minha esposa, Rafaela, que sempre está ao meu lado independente da situação. A minha irmã, Marília, que mesmo indiretamente contribui para a realização de meus feitos.

Aos meus companheiros de turma e de laboratório que sempre nos animam e apoiam mesmo nos momentos em que tudo parece estar desmoronando. Aos professores do Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Juiz de Fora, que contribuíram enormemente para a evolução do meu conhecimento.

A minha orientadora, Profa. Dra. Regina Maria Maciel Braga Villela, soube ser paciente, e me encaminhar para a conclusão deste trabalho. Aos responsáveis e colaboradores do Projeto FrameNet Brasil que apoiaram o desenvolvimento deste trabalho, fornecendo material e também contribuindo para aquisição do conhecimento necessário para lidar com o projeto.

Aos membros da banca, que contribuem com a avaliação do trabalho, permitindo enriquecer este trabalho e os trabalhos que virão a ser realizados no futuro.

E por fim, a Universidade Federal de Juiz de Fora e a CAPES, pelo apoio financeiro.

# RESUMO

O projeto FrameNet é um projeto desenvolvido pelo International Computer Science Institute (ICSI) em Berkeley, com o objetivo de documentar frames da língua inglesa baseando-se no conceito de frames semânticos da Inteligência Computacional. Sendo também estendido para outras línguas, como, por exemplo, o Projeto FrameNet-Br, desenvolvido na Universidade Federal de Juiz de Fora, que tem como foco a documentação dos frames linguísticos em português do Brasil. O recurso lexical construído pelo FrameNet pode ser utilizado em diversas aplicações de Processamento de Linguagem Natural, como traduções, sumarização, dentre outras. A utilização de tecnologias de web semântica, como ontologias e dados ligados podem trazer diversos benefícios que colaboram com o compartilhamento das informações, principalmente considerando a possibilidade de uso de mecanismos de inferência para validação dos dados e obtenção de novas informações. Além disso, o uso de anotações semânticas em serviços Web também é considerado uma tecnologia promissora para facilitar a integração de recursos computacionais na web, servindo como mecanismo para facilitar a interação entre ferramentas de software. Estas anotações permitem que ferramentas possam compreender a estrutura dos serviços e como executá-los de maneira automática. Desta forma, considerando as vantagens apresentadas por estas tecnologias, é possível associá-las de maneira a criar uma infraestrutura que permita a utilização de recursos lexicais em conjunto com recursos da web semântica para facilitar a compreensão e busca por informações em um dado domínio. Neste trabalho foi especificada uma infraestrutura baseada em serviços, que busca aliar as tecnologias da web semântica aos dados da base do projeto FrameNet, a partir da hipótese de que a aplicação de tecnologias como ontologias, dados ligados, e anotações em serviços Web podem contribuir com a construção e reuso de recursos lexicais baseados na semântica de frames, sendo estas contribuições tanto voltadas para a confiabilidade dos dados, quanto para o enriquecimento das informações mantidas pela base lexical.

**Palavras-chave:** FrameNet. Semântica de Frames. Web Semântica. Ontologia. Dados Ligados.

# ABSTRACT

The FrameNet project is developed by the International Computer Sciences Institute (ICSI) in Berkeley, with the goal of documenting frames of the English language based on the concept of semantic frames of Computational Intelligence. The FrameNet has been well translated for other languages, such as the FrameNet-BR Project, developed at the Universidade Federal de Juiz de Fora (UFJF), which focus on the documentation of linguistic frames in Portuguese from Brazil. The lexical resource built by FrameNet can be used in many applications of natural language processing, such translations, summarization, among others. The use of semantic web technologies, such ontologies and linked data can bring many benefits that contribute to the sharing of information, specially considering the possibility of the use of inference mechanisms to validate data and retrieve new information. The use of semantic annotations for Web services is also considered a promising technology to facilitate the integration of computational resource on the web, serving as mechanism to facilitate interaction between software tools. These annotations allow tools understand the structure of services and how to execute them automatically. Thus, considering the advantages presented by these technologies, it is possible to associate them in order to create an infrastructure that enables the use of lexical resources in conjunction with semantic web resources to facilitate understanding and search for information in a given domain. In this work was specified a service-based infrastructure that seeks to combine the technologies of semantic web to the database of the FrameNet project, considering the hypothesis that the application of technologies such as ontologies, linked data, and annotations in Web services can contribute to the construction and reuse of lexical resources based on semantic frames, and these contributions are both related to the reliability of data, and for the enrichment of the his information kept by lexical base.

**Keywords:** FrameNet. Frame Semantics. Web Semantic. Ontology. Linked Data.



# LISTA DE FIGURAS

2.1	Frame de Turismo do Fn-BR(GAMONAL, 2011). . . . .	25
2.2	Exemplo de regra SWRL com suas partes identificadas(O’CONNOR et al., 2005). . . . .	29
2.3	Modelo de exemplo para definições de OWA e CWA. . . . .	30
2.4	Visão geral da rede do Linked Open Data (CYGANIAK; JENTZSCH, 2011). . . . .	31
2.5	Representação da Ontologia de serviços (MARTIN et al., 2004). . . . .	33
2.6	Organização hierárquica das classes na ontologia SWAN. . . . .	36
2.7	Organização hierárquica das classes na ontologia OnTour. . . . .	38
2.8	Visão geral dos módulos da PROTON (TERZIEV et al., 2005). . . . .	39
2.9	Ontologia do frame ”Atack” e alguns frames relacionados (SCHEFFCZYK et al., 2006). . . . .	41
2.10	Ontologia do frame ”Atack” e alguns frames relacionados (detalhado) (SCHEFFCZYK et al., 2006). . . . .	41
2.11	Ontologia de anotações para a setença: ”48 Kuwaiti jet fighters managed to escape the Iraqi invasion” (SCHEFFCZYK et al., 2006). . . . .	42
2.12	Conceitos das transformações feita pelo Semion (NUZZOLESE et al., 2011). . . . .	43
2.13	Regras de transformação de LOD para KP para base do FrameNet (NUZZOLESE et al., 2011). . . . .	44
2.14	Thai FrameNet Ontology (LEENOI et al., 2011). . . . .	45
2.15	Anotação exportada em XML pela ferramenta da TFN (LEENOI et al., 2011). . . . .	46
2.16	Ferramenta de apoio à visualização de Frames da TFN (LEENOI et al., 2011). . . . .	46
2.17	Analysis Patterns Reuse Infrastructure (APRI) (VEGI et al., 2012). . . . .	47
3.1	Classes da ONTO-FRAME-BR. . . . .	57
3.2	Classes e restrições declaradas na ontologia com uso do Protège. . . . .	58
3.3	Restrições para a relação Perspective_on no Protège. . . . .	60
3.4	Regra para verificação da herança de frame causativo. . . . .	60
3.5	Regra verificando a inexistência de herança em frame causativo. . . . .	61
3.6	Definição da classe <i>IncoativeBaseFrame</i> . . . . .	61
3.7	Regra verificação do frame incoativo. . . . .	62

3.8	Regras para verificar se EFs que participam de uma das relações entre EF pertencem ao mesmo Frame. . . . .	63
3.9	Regra para verificação da não exclusão e dependência mutua entre EFs. . . . .	63
3.10	Regras definidas em SWRL. . . . .	63
3.11	Resumo com as principais Classes da ontologia de ONTO-FRAME-BR. . . . .	64
3.12	Regra de validação de múltiplos lexemas anotados em uma mesma sentença. . . . .	65
3.13	Regra de validação de múltiplos frames em uma mesma anotação. . . . .	66
3.14	Principais Classes da Ontologia de ONTO-ANNOTATION-BR. . . . .	66
3.15	Visão geral da arquitetura da FSI. . . . .	67
3.16	Uso de Ontologias de domínio para restrição de tipo semântico para os Elementos de Frame. . . . .	68
3.17	Representação de fragmentos de anotações sob o padrão de dados ligados. . . . .	69
3.18	Visão Geral do Framework CADMOS (CATALDI et al., 2011). . . . .	75
3.19	Processo de desambiguação de termos na CADMOS (CATALDI et al., 2011). . . . .	75
3.20	Processo de desambiguação de termos na CADMOS com inclusão da obtenção dos dados de Frames via FSI. . . . .	76
3.21	Fluxo de execução de serviços para a obtenção dos dados do Frame e seus EFs pela FSI. . . . .	77
3.22	Tela Inicial do Portal da FSI. . . . .	77
3.23	Menu Principal da ferramenta de documentação de frames. . . . .	78
3.24	Tela para documentação de frame. . . . .	79
3.25	Detalhe do cadastro de Elementos de frame. . . . .	80
3.26	Tela de justificativa da inferência realizada. . . . .	80
3.27	Tela de associação de classes de ontologias de domínio com EFs. . . . .	81
3.28	Tela de cadastro de UL. . . . .	81
3.29	Tela de cadastro de Lexemas. . . . .	82
3.30	Tela para anotação de sentença. . . . .	83
3.31	Tela com sugestão de lexema a partir da palavra selecionada. . . . .	84
3.32	Tela apresentando a coloração de cada termo em relação ao EF. . . . .	84
3.33	Seleção de Recurso para o termo "A seleção brasileira". . . . .	85
3.34	Associação de recursos externos a classes das ontologias de domínio. . . . .	86
3.35	Tela de consulta de Frames do FnSViewer. . . . .	86

3.36	Tela de consulta a anotações com o detalhe de um recurso associado a um fragmento de sentença no popup. . . . .	87
4.1	Representação dos Frames Serviço_Turístico e Entretenimento e seus EFs. . .	93
4.2	Associação de herança entre EFs. . . . .	94
4.3	Representação hierárquica do Frame "Cenário de Turismo" e seus subframes. .	94
4.4	XML da resposta da consulta SPARQL com subframes de "Cenários de Turismo".	95
4.5	Frame "Copa do Mundo Jogar" com seus EFs nucleares e as relações entre eles.	95
4.6	Lógico de classificação do Frame como incorreto pela ferramenta. . . . .	96
4.7	Tela de anotação de Sentenças. . . . .	97
4.8	Sugestão para seleção de lexema. . . . .	98
4.9	Tela de anotação de sentenças após identificação do lexema. . . . .	98
4.10	Verificação de incoerência durante anotação. . . . .	99
4.11	Sentença completamente anotada. . . . .	100
4.12	Confirmação de EFs com instanciações nulas. . . . .	100
4.13	Opção de associar classes a EFs para restrição semântica. . . . .	102
4.14	Seleção de classes ontológicas para o EF "Seleções1". . . . .	102
4.15	Tela para cadastro de recursos externos em relação a ontologias. . . . .	104
4.16	Seleção de recursos para fragmentos de anotação. . . . .	105
4.17	Descrição do recurso " <i>BrazilNationalFootballTeam</i> " da DBPedia. . . . .	106
4.18	Visão da arquitetura da ferramenta do COPA 2014 (TORRENT et al., 2014).	107
4.19	Arquitetura da ferramenta do COPA 2014 modificado para consumo da ca- mada de serviços da FSI. . . . .	108
4.20	Fluxo de comunicação entre a ferramenta, o Serviço intermediador e a camada de serviços da FSI para consulta a palavras. . . . .	109
4.21	Fluxo de comunicação entre a ferramenta, o Serviço intermediador e a camada de serviços da FSI para consulta a frames. . . . .	110
A.1	Visão geral das principais classes e seus relacionamentos. . . . .	125
A.2	Representação da classe Frame no Protégè. . . . .	126
A.3	Representação da classe FrameElement no Protégè. . . . .	127
A.4	Representação da classe SemanticType no Protégè. . . . .	128
A.5	Representação da classe Lexeme no Protégè. . . . .	129

A.6	Representação da classe LexUnit no Protègè. . . . .	130
A.7	Representação da classe WordForm no Protègè . . . . .	131
A.8	Representação da classe PartOfSpeech no Protègè. . . . .	132

# LISTA DE TABELAS

2.1	Lista de classes definidas na ontologia RANWEZ. . . . .	34
2.2	Lista de propriedades definidas na ontologia RANWEZ. . . . .	35
2.3	Lista de classes definidas na ontologia Harmonise. . . . .	37
2.4	Lista de propriedades definidas na ontologia Harmonise. . . . .	37
2.5	Mapeamento entre os padrão Dublin Core e de Pantoquilha. . . . .	49
2.6	Regras de Aplicação do DC2AP. . . . .	50
2.7	Comparativo com trabalhos relacionados. . . . .	52
3.1	Lista de métodos presentes no serviço de acesso. . . . .	70
3.2	Lista de métodos presentes no serviço de acesso. . . . .	71
3.3	Lista de métodos presentes no serviço de acesso. . . . .	74
4.1	Associação de classes ontológicas para EFs. . . . .	103
4.2	Recursos De dados ligados cadastrados. . . . .	104

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>17</b>
1.1	MOTIVAÇÃO	17
1.2	JUSTIFICATIVA	17
1.3	HIPÓTESE	18
1.4	OBJETIVOS	19
1.5	ESTRUTURA DA DISSERTAÇÃO	19
<b>2</b>	<b>PRESSUPOSTOS TEÓRICOS</b>	<b>21</b>
2.1	ESTRUTURA DA DISSERTAÇÃO	21
2.2	FRAMENET E FRAMENET BRASIL	22
2.3	WEB SEMÂNTICA	26
2.3.1	Ontologias	27
2.3.1.1	Semantic Web Rule Language (SWRL)	28
2.3.1.2	Pressuposto De Mundo Aberto (Open World Assumption)	29
2.3.2	Dados Ligados	30
2.3.3	Serviços Web Semânticos	32
2.4	ONTOLOGIAS DE DOMÍNIO	33
2.4.1	Ranwez Soccer Ontology	34
2.4.2	Swan Soccer Ontology	35
2.4.3	Harmonise Ontology	36
2.4.4	Ontour Ontology	37
2.4.5	Proto Ontology	39
2.5	TRABALHOS RELACIONADOS	40
2.5.1	Ontology-Based Reasoning About Lexical Resources	40
2.5.2	Gathering Lexical Linked Data And Knowledge Patterns From Framenet ...	42
2.5.3	Thai Framenet Construction And Tools	44
2.5.4	APRI e DC2AP	46
2.5.4.1	Analysis Patterns Reuse Infrastructure (Apri)	46
2.5.4.2	A Dublin Core Application Profile To Analysis Patterns (DC2AP)	48

2.5.5	Discussões Sobre Os Trabalhos Relacionados .....	50
2.5.6	Considerações Finais do Capítulo .....	53
<b>3</b>	<b>FRAMENET SEMANTIC INFRASTRUCTURE.....</b>	<b>54</b>
3.1	INTRODUÇÃO .....	54
3.2	ONTOLOGIAS .....	55
3.2.1	Ontologia de Metadados Do FrameNet.....	55
3.2.1.1	Relações entre frames .....	59
3.2.2	Relações Entre Elementos de Frame .....	62
3.2.2.1	Ontologia de Anotações.....	64
3.3	ARQUITETURA .....	67
3.3.1	Camada de Dados.....	67
3.3.2	Camada de Serviços .....	69
3.3.2.1	Serviço de Acesso .....	70
3.3.2.2	Serviço de Visualização.....	70
3.3.2.3	Serviço de Ontologias e Dados Ligados .....	71
3.3.2.4	Serviço de Descoberta.....	73
3.3.2.5	Uso da camada de serviços .....	73
3.3.3	Portal .....	77
3.3.3.1	FrameNet Semantic Editor (FnSEditor) .....	78
3.3.3.2	FrameNet Semantic Viewer (FnSViewer) .....	86
3.4	TECNOLOGIAS .....	87
3.5	CONSIDERAÇÕES FINAIS DO CAPÍTULO .....	88
<b>4</b>	<b>AVALIAÇÃO DA PROPOSTA .....</b>	<b>90</b>
4.1	DEFINIÇÃO E PLANEJAMENTO DO MÉTODO DE PESQUISA .....	90
4.2	PROVAS DE CONCEITO .....	92
4.2.1	Cenários para as provas de conceito (PoC).....	92
4.2.2	Primeira PoC: Documentação de Frames.....	92
4.2.2.1	Evidências observadas na PoC 1 .....	96
4.2.3	Segunda PoC: Anotação de Sentenças.....	97
4.2.3.1	Evidências observadas na PoC 2.....	101
4.2.4	Terceira PoC: Relação de Dados Ligados e Ontologias de Domínio .....	101

4.2.4.1	Evidências Observadas na PoC 3.....	105
4.2.5	Quarta PoC: Uso da FSI por Aplicações de Software.....	107
4.2.5.1	Evidências Observadas na PoC 4.....	110
4.3	ANÁLISE DOS ESTUDOS .....	111
4.4	CONSIDERAÇÕES FINAIS DO CAPITULO .....	112
<b>5</b>	<b>CONCLUSÕES E TRABALHOS FUTUROS.....</b>	<b>114</b>
5.1	CONTRIBUIÇÕES .....	114
5.2	LIMITAÇÕES .....	115
5.3	TRABALHOS FUTUROS .....	116
	<b>REFERÊNCIAS .....</b>	<b>118</b>
	<b>APÊNDICES .....</b>	<b>124</b>



# 1 INTRODUÇÃO

## 1.1 MOTIVAÇÃO

O projeto FrameNet é um projeto desenvolvido pelo International Computer Science Institute (ICSI) em Berkeley, com o objetivo de documentar frames da língua inglesa baseando-se no conceito de frames semânticos da Inteligência Computacional. O FrameNet também foi estendido para outras línguas, como, por exemplo, o Projeto FrameNet-Br, desenvolvido na Universidade Federal de Juiz de Fora, que tem como foco a documentação dos frames linguísticos em português do Brasil (GAMONAL, 2011). Neste contexto, o recurso lexical construído pelo FrameNet pode ser utilizado em diversas aplicações de Processamento de Linguagem Natural, como traduções, sumariamento, dentre outras.

A utilização de tecnologias de web semântica, como ontologias aplicadas a um determinado domínio pode trazer diversos benefícios que colaboram com o compartilhamento das informações, principalmente considerando a possibilidade de uso de mecanismos de inferência para validação dos dados e obtenção de novas informações. O uso de anotações semânticas em serviços Web também é considerada uma tecnologia promissora para facilitar a integração de recursos computacionais na web, servindo como mecanismo para facilitar a interação entre ferramentas de software. Estas anotações permitem que ferramentas possam compreender a estrutura dos serviços e como executá-los de maneira automática. Por outro lado, a utilização do padrão de dados ligados permite uma maior integração entre recursos presentes na web, permitindo a obtenção de novas informações a partir da relação entre recursos.

Desta forma, considerando as vantagens apresentadas por estas tecnologias, é possível associa-las de maneira a criar uma infraestrutura que permita a utilização de recursos lexicais em conjunto com recursos da web semântica para facilitar a compreensão e busca por informações em um dado domínio.

## 1.2 JUSTIFICATIVA

Atualmente tem-se diversos trabalhos relacionados ao projeto FrameNet. Dentre estes trabalhos, alguns buscam aprimorar a reusabilidade dos dados deste projeto buscando

utilizar tecnologias que facilitem a reutilização das informações contidas em sua base. Dentre estas tecnologias, merece destaque o uso de tecnologias relacionadas à Web semântica. O uso de tecnologias da Web semântica enfatizam características como reuso e obtenção de novo conhecimento como possíveis ganhos no contexto do FrameNet.

Neste contexto, dentre as características da Web semântica que facilitam o uso dos dados do FramNet, podemos considerar: i) o formalismo provido pelas ontologias, que permitem um detalhamento formal e compartilhado de conceitos relacionados a um dado domínio, e também permitem o uso de máquinas de inferência para validação semântica dos dados e obtenção de informações implícitas; ii) os dados ligados, que promovem uma maior integração dos dados com recursos de outras bases de informações presentes na Web como a DBPedia, GeoNames, entre outros; e iii) serviços web, que permitem a integração entre ferramentas baseada na troca de mensagens, possibilitando a comunicação independente da linguagem de programação e ambiente em que estas ferramentas foram desenvolvidas.

Considerando este contexto, neste trabalho especificamos uma infraestrutura baseada em serviços, que busca aliar as tecnologias de web semântica aos dados da base do projeto FrameNet, a partir da hipótese de que a aplicação de tecnologias como ontologias, dados ligados, e anotações em serviços web podem contribuir com a construção e reuso de recursos lexicais baseados na semântica de frames. Sendo estas contribuições tanto voltadas para a confiabilidade dos dados, quanto para o enriquecimento das informações mantidas pela base lexical.

### 1.3 HIPÓTESE

A hipótese deste trabalho está relacionada aos benefícios que pode-se obter quanto a aplicação de tecnologias de web semântica no contexto da FrameNet, tanto no processo de documentação quanto na consulta dos frames. Sendo assim, a hipótese é: *Ao utilizarmos tecnologias de web semântica, como ontologias e dados ligados, em conjunto com recursos lexicais baseados na semântica de frames, podemos facilitar a utilização dos recursos lexicais por aplicações e ou usuários.*

## 1.4 OBJETIVOS

Tomando como base os benefícios associados ao uso de ontologias, dados ligados e serviços web com anotações semânticas para o compartilhamento e validade dos dados de um determinado domínio, essa dissertação tem como objetivo primário propor a construção de uma infraestrutura baseada nos conceitos de web semântica que possa contribuir com o projeto FrameNet promovendo a reusabilidade dos dados do projeto. Esta infraestrutura visa prover duas interfaces de interação, uma voltada para interação com outras ferramentas de software, através de uma camada de serviços, e outra interface para interação direta com o usuário, permitindo a manutenção dos dados, aproveitando também os benefícios do uso de ontologias para esta tarefa.

Sendo assim, podemos subdividir o objetivo principal deste trabalho da seguinte maneira:

1. Propor uma infraestrutura que apoie o projeto FrameNet a partir do uso de tecnologias de Web Semântica;
2. Prover maior formalismo aos dados da FrameNet a partir do uso de ontologias para descrever suas estruturas;
3. Promover a aplicação dos dados do FrameNet por ferramentas externas por meio do uso de serviços web;
4. Disponibilizar ferramentas que contribuam para a documentação de frames e para a realização de anotações de sentenças;
5. Reduzir a probabilidade de inserção de erro por falha humana nas anotações de sentenças, a partir da realização de validações por máquinas de inferências;
6. Proporcionar ao usuário uma nova experiência de consulta aos dados da FrameNet, a partir do uso de dados ligados, possibilitando a descoberta de novas informações.

## 1.5 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está organizada em quatro capítulos, além desta introdução.

O capítulo 2 apresenta os principais conceitos relacionados ao trabalho proposto nesta dissertação, incluindo a Web Semântica e suas tecnologias, além de conceitos relacionados

a semântica de frames e ao projeto Framenet. Também são apresentados neste capítulo os trabalhos relacionados.

O capítulo 3 apresenta a proposta da infraestrutura FSI (Framenet Semantic Infrastructure), onde são descritos cada um de seus componentes, como ontologias, serviços, anotações de serviços e o portal.

O capítulo 4 apresenta um conjunto de provas de conceitos realizadas na base de dados do Projeto Copa 2014 FrameNet Brasil, com o intuito de validar a proposta deste trabalho, contemplando desde o processo de documentação de um frame até a realização de anotações de sentenças e validações semânticas de seus dados.

Por fim, o capítulo 5 apresenta as considerações finais e trabalhos futuros.

## 2 PRESSUPOSTOS TEÓRICOS

Este capítulo tem como objetivo apresentar os principais conceitos relacionados à proposta desta dissertação, considerando a semântica de frames, os projetos FrameNet e COPA2014 e a utilização de tecnologias da Web semântica no contexto destes projetos. Para isso, o capítulo foi dividido nas seguintes seções: a seção 2.1 trata da semântica dos frames e sua importância no contexto da linguística e da Ciência da Computação. A seção 2.2 detalha o projeto FrameNet e os subprojetos relacionados. A seção 2.3 apresenta o contexto da web semântica e suas tecnologias, como ontologias, linguagens e padrões para definição de ontologias, dados ligados e serviços web semânticos. Na seção 2.4 são apresentadas algumas ontologias que são importantes para a validação do projeto. Na seção 2.5 são apresentados os trabalhos relacionados. E por fim, têm-se as considerações finais do capítulo.

### 2.1 ESTRUTURA DA DISSERTAÇÃO

A semântica de frames (FILLMORE, 1982) que, no contexto da Ciência da Computação, é derivada de pesquisas em Inteligência Artificial, propõe que o conhecimento humano não é construído a partir de conceitos isolados, e sim sobre um todo compartilhado. Esse conhecimento é especificado em estruturas complexas, denominados frames (FILLMORE, 1982) (FILLMORE, 1977).

Neste contexto, os frames constituem um sistema complexo de conceitos relacionados de maneira que para se compreender um deles é necessário entender toda a estrutura na qual o frame se encaixa (FILLMORE, 1982). Um exemplo destacado por Lage (2011) é a definição de andarilho pelo dicionário convencional: "aquele que anda". Entretanto não são todas as pessoas que caminham longas distâncias em períodos regulares que podem ser caracterizados como andarilhos, apesar do dicionário nos permitir essa interpretação. Nesse caso o frame de andarilho nos daria uma definição mais precisa, excluindo por exemplo, atletas ou praticantes de atividades físicas. Outro exemplo, também apresentado por Lage (2011) detalha um pouco melhor a necessidade de compreender como os frames podem ser relacionados, considerando a busca pela aceção do termo "mãe solteira", na visão da

semântica de frames, onde não é suficiente a enumeração dos definidores deste lexema<sup>1</sup> (mulher + progenitora + não casada) mas sim a definição de Frames como Casamento, Família, etc. E por fim, Fillmore (1982) demonstra como um frame denota uma abstração de cenas, como no caso do frame de Transação comercial e quatro cenas referentes a este frame:

1. Eu comprei uma dúzia de rosas;
2. Eu paguei a Harry cinco dólares;
3. Eu paguei uma dúzia de rosas para Harry por cinco dólares;
4. Eu paguei a Harry cinco dólares por uma dúzia de rosas.

As cenas estão relacionadas, respectivamente, a pessoa interessada em trocar dinheiro por alguma mercadoria (o comprador), a pessoa interessada em trocar mercadorias por dinheiro (o vendedor), aquilo que o comprador poderá adquirir (as mercadorias) e o dinheiro adquirido pelo vendedor (o dinheiro).

## 2.2 FRAMENET E FRAMENET BRASIL

O FrameNet (RUPPENHOFER et al., 2006) é um projeto desenvolvido no International Computer Science Institute, em Berkeley - Estados Unidos, sendo liderado pelo professor Charles Fillmore. Trata-se do desenvolvimento de um recurso lexical para língua inglesa baseada na teoria de semântica de frames discutida por Fillmore (1977) com o objetivo de:

1. Descrever as Unidades Lexicais (UL) em função dos frames que as evocam, os quais também seriam descritos a partir dos Elementos de Frames (EF) que os compõem;
2. Legitimar essas descrições através de pesquisa baseada em corpus;
3. Estabelecer as possibilidades combinatórias das Unidades Lexicais realizando, para tanto, uma anotação em camadas na qual se identifiquem os Elementos de Frames, seus Tipos Sintagmáticos e Funções Gramaticais;

---

<sup>1</sup>Palavra ou parte da palavra que serve de base ao sentido por ela expresso. Também se diz semantema.

4. Disponibilizar os resultados como entradas lexicais que representariam os padrões de valência de cada Unidade Lexical;
5. E determinar as relações entre os frames participantes da rede.

O projeto FrameNet constituiu uma base de informações, organizada em Frames, que demonstra sua importância pela sua utilização em aplicações como extração de informações, tradução automática e dicionário de valências, sendo também expandido para outras línguas, como alemão<sup>2</sup>, japonês<sup>3</sup>, francês<sup>4</sup> e espanhol<sup>5</sup>. Foi também desenvolvida uma versão para a língua portuguesa, denominada FrameNet-Brasil (SALOMAO, 2009) que é gerenciada pelo grupo de linguística da Universidade Federal de Juiz de Fora (UFJF). Atualmente o banco de dados do FrameNet conta com aproximadamente 1160 frames documentados, com 1675 relações entre eles, e 12600 unidades lexicais (ICSI, 2012). Estes dados podem ser utilizados em diversas aplicações, como por exemplo, no auxílio à rotulação semântica automática de textos, possibilitando desenvolvimento em vários setores de Processamento de Linguagem Natural (PLN), entre eles, a tradução automática, extração de informações, sumarização de textos, etc. Dentre as estruturas que compõem um frame, temos os Elementos do Frame (EF) que são fragmentos que compõem as cenas que podem ser representadas pelo frame. Sendo assim um frame não pode ser construído sem a presença de seus elementos, pois sem eles não é possível construir suas cenas. Os EF podem ser classificados em três tipos: os nucleares, os periféricos e os extratemáticos (GAMONAL, 2011).

Os EFs nucleares são aqueles que são indispensáveis para a criação de qualquer cena referente ao seu frame. No frame "Ataque"(Attack), por exemplo, temos "Assaltante" e "Vitima" como elementos centrais, que representam informações que não podem estar ausentes em cenas que sejam instanciadas a partir deste frame, pois prejudicariam o sentido da cena. Os periféricos ou não nucleares, são aqueles que acrescentam informação aos EF centrais, e diferentemente dos nucleares, sua ausência não prejudica a construção da cena. Por fim, têm-se os EF extratemáticos, que embora participem do frame, não pertencem ao frame em questão. Por exemplo, a cena descrita como: "Ele me assaltou duas vezes nessa rua". O número que determina quantas vezes a ação ocorreu é um EF

---

<sup>2</sup><http://www.laits.utexas.edu/gframenet/>

<sup>3</sup><http://jfn.st.hc.keio.ac.jp/>

<sup>4</sup><https://sites.google.com/site/anrasfalda/>

<sup>5</sup><http://sfn.uab.es:8080/SFN/>

extratemático chamado de Iteração (GAMONAL, 2011). Como exemplo temos a definição do frame de "Turismo" do FrameNet-BR apresentado na Figura 2.1.

Segundo (RUPPENHOFER et al., 2006) existem alguns fatores que podem levar frames já criados a serem divididos, tais como diferença de perspectiva, variação na estrutura argumental, alternância causativo-incoativa e distinção ontológica dos EF. A fim de atender este último fator também é adotado na FrameNet a atribuição de Tipos Semânticos para alguns EF, permitindo restringir as instancias de um EF de modo que estas também deverão ser consideradas instancias de seu Tipo Semântico.

Segundo ainda (RUPPENHOFER et al., 2006) e (LEENOI et al., 2011), os frames documentados no FrameNet podem ser conectados por sete possíveis relações, sendo elas:

- Inheritance (Herança): Com esta relação qualquer característica que for estritamente semântica do frame pai deverá ser igual ou mais específica no frame filho, como participação de EFs (com exceção dos extra-temáticos e nucleares não expressivos), relações com outros frames, relacionamentos entre seus EFs e os tipos semânticos de seus EFs.
- Perspective on (Perspectiva sobre): Relação que normalmente liga um frame neutro a no mínimo dois frames não neutros, com exceção de quando seus EFs podem adotar mais de um ponto de vista sem a necessidade desrespeitar a estrutura do Frame.
- SubFrame: Relação entre um ou mais Frames que compõe um Frame mais complexo.
- Precedes: Relação entre subframes de um frame que denota sequencia cronológica dos subframes de um frame complexo.
- Inchoative of: Relação de um frame com seu frame incoativo, sendo que o frame incoativo deverá herdar do frame Event, State ou Gradable\_attributes.
- Causative of: Relação de um frame com seu frame causativo, sendo que o frame causativo deve herdar do frame Transitive\_Action.
- Using: Utilizado quando um frame evoca outro frame, entretanto diferentemente da relação de subframe, o frame evocado não necessariamente faz parte do frame que o evocou



<b>Turismo (Touring)</b>	
<b>Definição:</b>	
Um <b>turista</b> visita ou experiencia uma <b>atração</b> um local com uma história peculiar ou um caráter individual reconhecido socialmente, com o objetivo de ver e aprender sobre ele. Normalmente, a <b>atração</b> tem uma fonte de informação tal como um guia, panfletos ou displays.	
<b>Elementos do Frame:</b>	
<b>Nucleares:</b>	
<b>Atração</b> Tipo Semântico: Lugar	O local peculiar socialmente reconhecido que o <b>turista</b> visita.
<b>Turista</b> Tipo Semântico: Consciente	O indivíduo que está buscando uma experiência em uma <b>atração</b> .
<b>Não-Nucleares:</b>	
<b>Coparticipante</b> Tipo Semântico: Consciente	O <b>coparticipante</b> é quem participa de modo coordenado, no caso em uma visita, com o <b>turista</b> .
<b>Estado</b>	O estado do <b>turista</b> ou da <b>atração</b> durante a visita.
<b>Duração</b> Tipo Semântico: Duração	A quantidade de tempo que dura a atividade de turismo.
<b>Base</b>	O background perceptivo pela qual a <b>atração</b> é experimentada pelo <b>turista</b> .
<b>Modo</b> Tipo semântico: Modo	Qualquer descrição dos detalhes do evento em termos de como eles se comparam com outros eventos (da mesma forma) ou em termos de como o estado do <b>turista</b> os afeta (felizmente, distraidamente).
<b>Meio</b> Tipo semântico: Estado_de_Coisas	O <b>Meio</b> indica o método que o <b>turista</b> está usando para participar da visita.
<b>Iteração_especifica</b>	Expressões marcadas com este EF extratemático modificam um uso não-iterativo do alvo e indica que ele é concebido como incorporado dentro de uma série iterada de eventos similares ou estados.
<b>Lugar</b> Tipo semântico: Relação_locativa	O local onde o turismo acontece.
<b>Propósito</b> Tipo semântico: Estado_de_Coisas	Alguma ação que o <b>turista</b> está tentando realizar com o turismo. Isso geralmente é (no mínimo) para fins de entretenimento e aprendizado.
<b>Tempo (Tim)</b> Tipo semântico: Tempo	O tempo que o turismo acontece.
<b>Relações entre frames:</b>	
Herda de: Perception_active	
É herdado de:	
Perspectivado em:	
É perspectivado por:	
Usa: Visiting	
É usado por:	
Subframe:	
Tem os subframes:	
Precede:	
É precedido por:	
É incoativo de:	
É causativo de:	
Veja também:	

Figura 2.1: Frame de Turismo do Fn-BR(GAMONAL, 2011).

- See Also: Relação normalmente utilizada em casos onde frames são semelhantes e deve-se ter atenção em distingui-los ou compará-los.

Além das adaptações do projeto para outras línguas, muitos outros projetos nasceram com base no FrameNet, por exemplo o *Kicktionary* (SCHMIDT, 2009) que busca construir um dicionário temático no domínio do futebol baseado nas línguas Inglesa, Alemã e Francesa, utilizando a ideia de frames e UL como unidades básicas de análise. E também o projeto "Copa 2014 FrameNet-BR" que se trata de um esforço para o desenvolvimento de um dicionário trilingue (português, inglês e espanhol) relacionado aos domínios de Futebol e Turismo que tinha como um de seus objetivos prover auxílio a pessoas que participaram da Copa do Mundo FIFA 2014, tanto para organização do evento, recepção de turistas e para os turistas (GAMONAL, 2011) (SALOMAO et al., 2011) (TORRENT et al., 2014).

## 2.3 WEB SEMÂNTICA

A World Wide Web da maneira como foi amplamente difundida pode ser considerada um artefato simples, definido como um conjunto de dados distribuídos, que podem ser ligados com simples referências entre eles por meio de hyperlinks. Os dados disponíveis podem ser acessados fazendo uso de buscas por palavras-chaves e navegação pelas ligações entre estes dados. Conforme o volume de dados da Web foi crescendo, se tornou cada vez mais difícil localizar informações específicas. Como exemplo, podemos citar a busca por informações de uma determinada pessoa utilizando apenas seu nome. Essa tarefa pode se mostrar como uma tarefa difícil, uma vez que serão retornados dados de vários homônimos existentes (HORROCKS, 2008).

Buscando resolver esta deficiência, foi proposto um novo modelo para representação dos dados disponibilizados na Web, denominado Web Semântica (BERNERS-LEE et al., 2001), onde as ligações entre os dados passam a ter expressividade semântica. O uso de ontologias é apontado como essencial para expressar a semântica dos termos presentes nas relações entre os dados, uma vez que as ontologias permitem formalizar os conceitos dos termos, facilitando o compartilhamento de informações e ajudando a evitar problemas inerentes a linguagem natural tais como homonímia, sinonímia, metonímia, etc. (ALVES, 2005).

A presença destas relações permite que as buscas, que na Web convencional são baseadas em palavras-chaves, possam ser realizadas aproveitando estas relações semânticas,

tornando-as mais específicas de maneira que parte dos resultados indesejados sejam eliminados. Além disto, a Web Semântica visa possibilitar, por meio da carga semântica de seus dados, que sistemas computacionais possam interpretar os dados e inferir informações, colaborando com a realização de tarefas que no modelo da Web convencional dependeriam completamente da interpretação humana (BERNERS-LEE et al., 2001).

Neste contexto, o uso de ontologias tem papel importante. Além de possibilitar o compartilhamento de conceitos que são utilizados nas relações entre dados, as ontologias também podem ser utilizadas para descrever a estrutura de um serviço web, o que além de facilitar a compreensão do serviço por humanos, também possibilita que ferramentas computacionais possam interpretar e consumir o serviço sem a necessidade de intervenção humana (MCILRAITH et al., 2001).

### 2.3.1 ONTOLOGIAS

O termo ontologia foi inicialmente relacionado à filosofia, sendo o nome de uma subárea da filosofia que considerava o estudo da natureza da existência. É ainda considerado um ramo da metafísica, que se preocupa com a identificação, em termos gerais, dos tipos de objetos existentes e como descrevê-los (ANTONIOU; HARMELEN, 2004).

Com o passar do tempo, este termo foi adotado pela Ciência da Computação com um sentido técnico específico, que é bastante diferente do original. Sendo uma das definições clássicas a definição de (GUARINO, 1998) onde uma ontologia é descrita como uma especificação formal e explícita de uma conceitualização compartilhada. Um detalhamento desta definição é apresentado por (DIETER, 2003): "conceitualização" é definido como um modelo abstrato de um determinado fenômeno onde são identificados os conceitos relevantes deste fenômeno; "explícito" significa que os tipos dos conceitos e as suas restrições de uso estão explicitamente descritos pelo modelo; "formal" significa que o modelo deve estar descrito de maneira que seja legível para máquinas; e "compartilhada" significa que a os conceitos e restrições capturados pelo modelo deverá ser aceito de maneira coletiva e não apenas por um indivíduo.

A utilização de ontologias traz como benefícios a possibilidade de compartilhamento de conhecimento sobre os conceitos de um determinado domínio, além da reutilização do conhecimento, fornecimento de meios para codificar o conhecimento e a semântica de forma que máquinas possam entender, além de fazer com que processamento de máquina

em larga escala possa ser realizado (YU, 2011).

Segundo (MORAIS e AMBROSIO, MORAIS; AMBROSIO) existem diversas linguagens para descrever uma ontologia, e estas podem ser agrupadas em três categorias, sendo elas: Linguagens de ontologias tradicionais, que conta com linguagens como a KIF (Knowledge Interchange Format) (GENESERETH et al., 1992) e a Ontolingua (GRUBER, 1992); Linguagens Web, com a RDF (KLYNE; CARROLL, 2006) e XML (BRAY et al., 2008); e Linguagens de ontologias baseadas em Web, que inclui as linguagens OIL (FENSEL et al., 2001), DAML+OIL (HORROCKS et al., 2002) e OWL (BECHHOOFER et al., 2004), por exemplo.

De acordo com o W3C, a linguagem padrão para definição de ontologias é a OWL (BECHHOOFER et al., 2004), que atualmente está na versão 2.0. Esta linguagem foi projetada para ser utilizada por aplicações que necessitam processar o conteúdo da informação e não somente para apresentar informações para os humanos, e também para resolver limitações das linguagens que a precederam. E como esta linguagem é baseada em lógica descritiva, é possível a utilização de mecanismos de inferência, os quais permitem explicitar conhecimentos que estão implícitos em uma base de conhecimento (COSTA, 2013). Para isso, podem-se utilizar tanto as restrições e propriedades especificadas na linguagem, incluindo o uso de property chains da OWL 2.0, quanto utilizar linguagens lógicas associadas, como é o caso da SWRL.

### 2.3.1.1 Semantic Web Rule Language (SWRL)

A linguagem SWRL foi submetida ao W3C em 2004, apresentando uma proposta baseada na associação de modelos OWL com uma sublinguagem da Rule Markup Language, a Datalog RuleML (BOLEY et al., 2010), com o objetivo de estender o modelo semântico da OWL. Essa extensão permite que regras descritas em SWRL possam determinar novas interpretações sobre os indivíduos e classes definidas no modelo OWL (BECHHOOFER et al., 2004).

As regras descritas em SWRL são baseadas nas cláusulas de Horn, sendo assim, cada regra é composta por duas partes, sendo a primeira chamada de antecedente, onde é definido um conjunto de átomos<sup>6</sup> que podem ser verificações de classe, propriedades de objetos, funções da própria linguagem, entre outros, e a segunda parte chamada de conse-

---

<sup>6</sup>Uma expressão do Tipo  $P(x, y)$  onde  $P$  é um predicado, ou  $C(x)$  onde  $C$  é uma definição de classe

quente, que é composta por um único átomo que pode ser uma classificação de um determinado indivíduo ou a declaração de propriedades para indivíduos, conforme o exemplo apresentado na Figura 2.2. Sendo assim se a primeira parte for verdadeira, a segunda parte também será (O’CONNOR et al., 2005).

$$\text{parent}(?x, ?y) \wedge \text{brother}(?y, ?z) \Rightarrow \text{uncle}(?x, ?z)$$

*antecedent*  *consequent*

Figura 2.2: Exemplo de regra SWRL com suas partes identificadas(O’CONNOR et al., 2005).

### 2.3.1.2 Pressuposto De Mundo Aberto (Open World Assumption)

Segundo (DRUMMOND; SHEARER, 2006) existem dois tipos de pressupostos quanto ao tratamento de informações não declaradas na representação de conhecimento em modelos computacionais, são eles: o *Open World Assumption* (OWA) e o *Closed World Assumption* (CWA).

O CWA, ou Pressuposto de Mundo Fechado, considera que se uma determinada informação não está declarada no modelo então ela é uma negação, por exemplo, caso existam as classes Pessoa e Pai, e um indivíduo João no modelo, e sobre o indivíduo João temos declarado no modelo que este faz parte da classe Pessoa (Figura 2.3). Ao solicitar duas listas para o modelo, uma com os indivíduos pertencentes à classe Pai, e uma segunda lista com os indivíduos que não pertencem a classe Pai, o sistema trará o indivíduo João apenas na segunda lista. Uma vez que a relação inexistente de João e Pai resulta na negação da mesma.

Enquanto o OWA, ou Pressuposto de Mundo Aberto, determina que as informações não declaradas em um banco não podem ser negadas, entretanto também não significa que sejam afirmativas, por exemplo, utilizando o mesmo cenário utilizado para exemplificar o CWA, ao solicitarmos ao modelo a lista de indivíduos que participam da classe Pai e uma segunda lista dos que não participam da classe Pai, o indivíduo João não estará presente em nenhum das listas, uma vez que a inexistência da relação entre João e Pessoa não significa que João efetivamente não seja da classe Pai e vice-versa.

Sendo assim, a informação inexistente no OWA é tratada como incerta, o que segundo (DRUMMOND; SHEARER, 2006) é o que mais se aproxima da representação do conhecimento na vida real, já que a ausência da informação significa falta de conhecimento e não

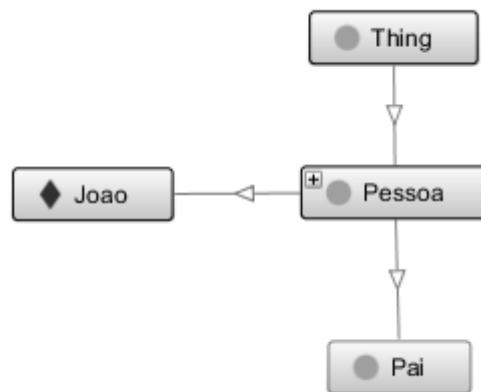


Figura 2.3: Modelo de exemplo para definições de OWA e CWA.

a negação sobre a informação ausente. Enquanto o CWA é mais voltado para aplicações onde se considera que tudo que é necessário se saber no momento está definido, o OWA é justamente ao contrário, o que o torna mais adequado ao uso no contexto de ontologias, uma vez que suas características promovem a expansibilidade, reuso e compartilhamento de informações.

O OWA merece destaque no contexto desta dissertação, considerando questões específicas identificadas no projeto FrameNet, que serão passíveis de entendimento, considerando os conceitos de OWA, conforme veremos no capítulo 4 desta dissertação.

### 2.3.2 DADOS LIGADOS

A natureza heterogênea dos dados presentes na Web convencional e a concentração destes dados por parte de diferentes serviços de busca, como Google, Yahoo, entre outros, causou a fragmentação dos dados presentes na Web (HEATH; BIZER, 2011). Considerando a Web semântica, é possível manter a natureza heterogênea dos dados e evitar que o surgimento de grandes datasets acabe levando a Web semântica para o mesmo problema, uma vez que o modelo para publicação de documentos na Web semântica, denominado dados ligados (BERNERS-LEE, 2009), tem como padrão para descrição da relação entre os documentos da rede o RDF (*Resource Description Framework*). O modelo de dados ligados também define um conjunto de regras que permite o crescimento da teia semântica formada, sem dificultar a exploração e utilização de dados obtidos através da navegação por esta teia.

A primeira regra do modelo de dados ligados diz respeito a identificação dos docu-

mentos ou recursos disponíveis na web, onde todos os documentos devem ser identificados por uma URI (*Uniform Resource Identifier*) de maneira a facilitar que recursos de fontes diferentes possam se referenciar entre si. A segunda regra diz respeito à acessibilidade do recurso, sugerindo que todas as URIs sigam o padrão HTTP (*HyperText Transfer Protocol*) a fim de permitir que pessoas possam acessar seu recurso. A terceira regra diz respeito ao provimento de informações, e diz que quando alguém acessar uma URI, informações úteis sobre este recurso deverão ser apresentadas com base nos padrões RDF e/ou SPARQL. Já a quarta regra diz respeito a formação da web em si, onde define que o recurso deverá possuir relações com outros recursos a fim de permitir que mais informações possam ser descobertas a partir destas ligações.

Baseando-se neste modelo, a W3C (*World Wide Web Consortium*) iniciou um projeto chamado *Linked Open Data* (LOD), onde o objetivo é capturar dados que estão sendo mantidos no modelo da web convencional e que possuem licenças públicas, e convertê-los para o padrão de dados ligados, contribuindo assim para o enriquecimento da teia semântica (Figura 2.4).

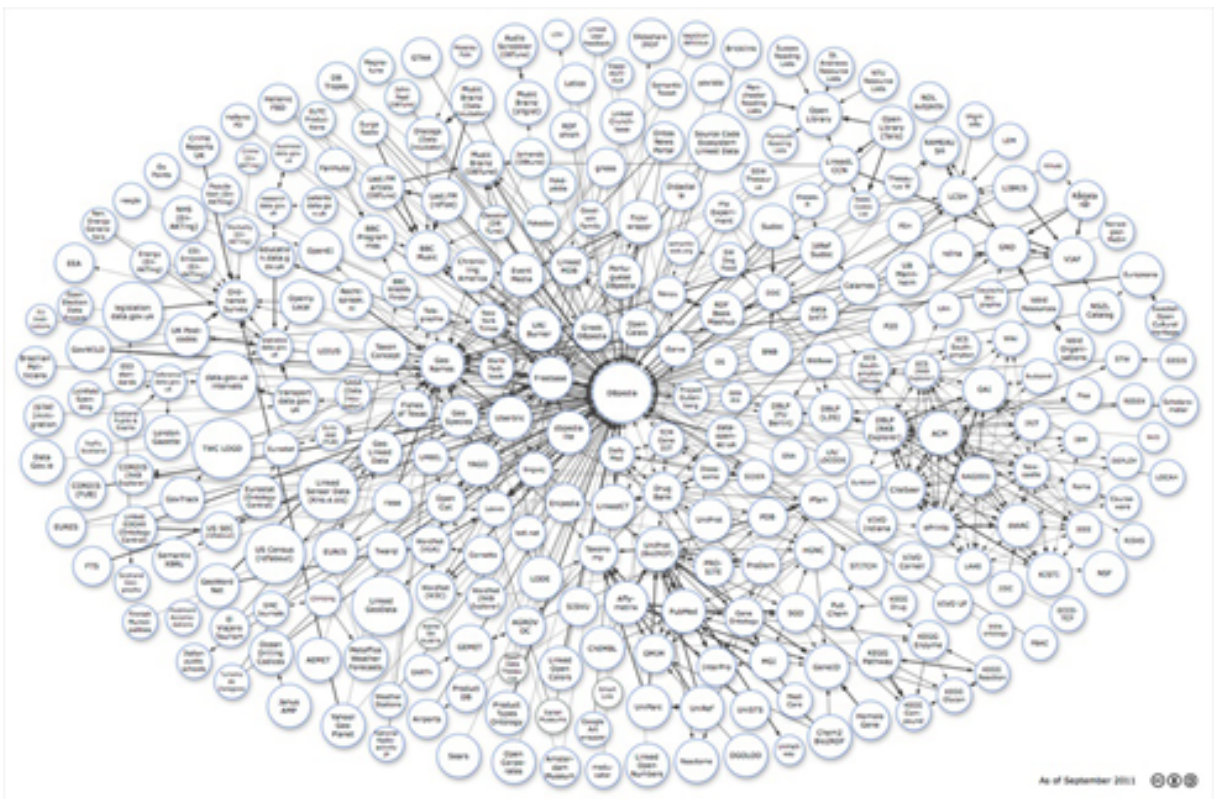


Figura 2.4: Visão geral da rede do Linked Open Data (CYGANIAK; JENTZSCH, 2011).

De maneira semelhante ao LOD, a *DBpedia* (AUER et al., 2007) surgiu com o objetivo

de criar uma rede de dados ligados a partir dos dados disponíveis na Wikipédia.

O uso de dados ligados também pode ser combinado com ontologias, como apresentado por (ZHAO; HARTIG, 2012) onde é proposta uma maneira de publicar dados de proveniência de maneira interoperável fazendo uso dos vocabulários *Open Provenance Model* (MOREAU et al., 2011) e o *Provenance Vocabulary* (HARTIG; ZHAO, 2010) definido por ontologias, com intuito de promover o compartilhamento destas informações.

### 2.3.3 SERVIÇOS WEB SEMÂNTICOS

Serviços Web ou Web Services são sistemas de software desenvolvidos com o objetivo de dar suporte a interação entre computadores por meio de uma rede. Esta interação é possível graças a descrição de sua interface em um formato processável por máquina, e pela adoção de padrões Web, como exemplo a adoção do padrão SOAP para troca de mensagens e o padrão HTTP para comunicação (BOOTH et al., 2004).

A descrição da interface destes serviços é geralmente especificada sob o padrão WSDL (*Web Services Description Language*), onde são descritas informações operacionais do serviço (CHINNICI et al., 2007).

Os serviços Web semânticos são serviços Web onde são associadas anotações semânticas a sua descrição. Um dos objetivos dessa associação é possibilitar a descoberta, execução e composição de forma automatizada, e facilitar a interoperabilidade da aplicação (MCIL-RAITH et al., 2001). Como a linguagem WSDL não é capaz de descrever semanticamente a estrutura do serviço, a solução para realizar esta anotação é fazer o mapeamento entre os elementos do serviço descritos pelo WSDL, como operações, mensagens, precondições e efeitos, para conceitos ontológicos (SILVA, 2010). Dentre as abordagens para a realização do processo de descrição semântica dos serviços web com uso de ontologias, podemos destacar a OWL-S (MARTIN et al., 2004), WSMO (BRUIJN et al., 2003) e SAWSDL (FARRELL; LAUSEN, 2007).

O OWL-S é um dos padrões que surgiu com o objetivo de tentar formalizar as anotações semânticas de serviços. A OWL-S é uma ontologia de serviços que baseia a representação das anotações em três subontologias: *ServiceProfile*, que descreve o que o serviço faz; *ServiceModel* que descreve como o serviço é usado; e *ServiceGrounding* que descreve como interagir com o serviço (Figura 2.5).

Dentre estes três tipos de ontologias, a *ServiceProfile* é responsável por informações



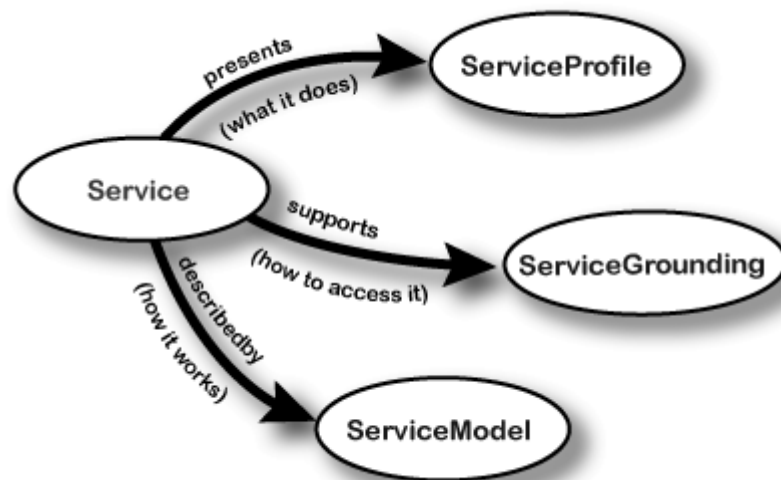


Figura 2.5: Representação da Ontologia de serviços (MARTIN et al., 2004).

abstratas, enquanto a *ServiceModel* e a *ServiceGrounding* são responsáveis por informações sobre a estrutura de interação e canal de comunicação do serviço, permitindo a descoberta e execução do serviço.

## 2.4 ONTOLOGIAS DE DOMÍNIO

Conforme dito no início deste capítulo, utilizamos no contexto desta dissertação, algumas ontologias de domínio para o processo de validação da proposta. Sendo assim, algumas ontologias foram analisadas e comparadas nos domínios de futebol e turismo, onde foram escolhidas duas delas.. No domínio de futebol apenas duas ontologias detalhadas na literatura, eram formalmente descritas e disponíveis para utilização. Sendo elas a *RANWEZ Soccer Ontology* (<http://www.daml.org/ontologies/273>) e a *SWAN Soccer Ontology* (<http://sw.deri.org/2005/05/swan/soccer/ontology/soccer.owl>).

Já para atender ao domínio de turismo foram estudadas três ontologias, considerando o número de citações destas em trabalhos relacionados à área de turismo. Sendo que, dentre estas, apenas a *Harmonise Ontology* (HÖPKEN; CLISSMANN, 2006) e a *On-Tour Ontology* (PRATNER, 2013) são focadas apenas no domínio de turismo, enquanto a *PROTo Ontology* (TERZIEV et al., 2005) é uma ontologia mais abrangente, não tendo foco específico para o domínio de turismo.

### 2.4.1 RANWEZ SOCCER ONTOLOGY

A *RANWEZ Soccer Ontology*<sup>7</sup> foi apresentada em 2002 e é uma ontologia que representa os termos do domínio do futebol para que possam ser utilizados em anotações de vídeos para criação de sumários personalizados de partidas de futebol.

Dentre as ontologias encontradas no domínio em questão, a RANWEZ foi uma das poucas disponíveis formalmente (em DAML). Além disso, esta ontologia também é utilizada como base para outras ontologias do mesmo domínio, por exemplo, a SWAN que será apresentada na próxima seção.

Na Tabela 2.1 temos, em ordem alfabética, a lista de classes, e na Tabela 2.2 as propriedades definidas nesta ontologia.

Tabela 2.1: Lista de classes definidas na ontologia RANWEZ.

Classes				
Action	Corner_Flag	Goal_kick	Organization	Shoot
Actor	Corner_kick	Goal_Line	OtherPlayer	Shoulder_charge
Advantage_rule	Counter_attack	Hacking	Other_player_action	Shout
Applause	Cross	Halftime	Outlet_passes	Side_tackle
Area	Cut_down_the_angle	Half_volley_shot	Out_of_bounds	Sliding_tackle
Attacker	Dangerous_play	Handball	Overlap	Soccer_rule_attribute
AttackingMidfielder	DefensiveMidfielder	Header	Overtime	Speaker
Attribute	Directed	Hit	Pass	Spectator
Audience	Direct_free_kick	Hold_opponent	Penalty	Spectator_action
Back	Direct_free_kicking	Hook	Penalty_Arc	Square_pass
Back_header	Dive	Indirect_free_kick	Penalty_Area	Start_time
Back_tackle	Diving_header	Indirect_free_kicking	Penalty_kick	Steal
Ball	Do_foul	Injury_time	Penalty_spot	Stop
Banana_kick	Do_handball	Inswinger	Penetrate	Stoppage
Beat	Do_obstruction	Integer-type	Penetrating_pass	Stopper
Begin	Do_Professional_foul	Juggling	Period	Substitute
Be_Offside	Dribble	Kick	Person	Sudden_death
Be_replaced	Drop_ball	Kickoff	Place	Supporter
Bicycle_kick	Drop_kick	Lead_pass	Play	Sweeper
Block	Duration	League	Player	Tackle
Boo	Encourage	Line	Player_Action	Tallied_shoot
Boolean-type	End	Linesman	Point	Team
Breakaway	End_Time	Linesman_action	Point_in_time	Thigh_trap
Cap	Entity	Linkman	President	Throw_in
Caution	False_value	Lob	Punch_out	Time
Center	Fault	Long_pass	Push_pass	Timekeeper
Center_Circle	Feint	Marking	Qualitative_attribute	Touch_Line
Center_kick	Field	Match	Real-type	Tournament
Center_line	Flick_header	Match_action	Red_Card	Trainer
Center_Spot	Foot_trap	Measure	Referee	Trap
CentralDefender	Forward	Midfield	Referee_action	True_value
Charge	Free_kick	MidfieldAnchor	Rise_the_flag	TV_spectator
Chest_trap	Free_kicking	Midfielder	Run	VIP
Chip_pass	Front_header	Net	Run_upfield	Volley
Chip_shoot	Front_tackle	Nutmeg	Sanction	Wall
Clear	Give_Sanction	Object	Save	Wall_pass
Club	Goal	Official	Scissor_kick	Whistle
Corner	Goalkeeper	Official_action	Score	Wing
Corner_Arc	Goalkeeper_action	Offside	Sending_off	Yellow_Card
Corner_Area	Goal_Area	Offside_trap	Shielding	

<sup>7</sup><http://www.daml.org/ontologies/273>

Tabela 2.2: Lista de propriedades definidas na ontologia RANWEZ.

Propriedades		
Agent	Height	national_team
aggregate	match_date	object
club_colors	match_first_team	supporter_colors
club_name	match_hour	team_club
club_president	match_name	team_goalkeeper
club_team	match_referee	team_substitute
club_town	match_second_team	team_supported
date_of_birth	match_tournament	team_trainer
ephemeral_instance	match_town	town_club
forname	name	user_specific
	nationality	weight

Entretanto, nesta ontologia os termos foram representados sem organização hierárquica e sem as restrições para existência de indivíduo, e as propriedades não estão relacionadas aos termos. Sendo assim, esta ontologia pode ser classificada apenas como um vocabulário de termos do domínio, determinando apenas quais são os termos possíveis, mas não a semântica relacionada a eles.

#### 2.4.2 SWAN SOCCER ONTOLOGY

A *SWAN Soccer Ontology*<sup>8</sup> foi publicada em 2005 e foi construída com base na Ontologia MUMIS (THIERRY; WITTENBURG, 2001) e como extensão da ontologia PROTON (TERZIEV et al., 2005). A MUMIS é uma ontologia com foco no futebol que foi construída com o objetivo de servir como recurso para anotação automatizada de arquivos multimídia no Projeto MUMIS. Já a PROTON foi desenvolvida como parte do projeto Semantically-Enabled Knowledge Technologies (SEKT), e busca representar conceitos em geral, não possuindo foco em um determinado domínio.

Na SWAN os termos são organizados em uma hierarquia, respeitando as ontologias que foram reaproveitadas para sua construção, e com propriedades e restrições bem definidas. Sendo assim, esta ontologia é mais expressiva em comparação a RANWEZ, uma vez que torna possível a verificação semântica e inferência sobre seus termos. Na Figura 2.6 temos uma representação das suas principais classes e como estão organizadas.

Devido a maior riqueza apresentada pela SWAN em comparação a RANWEZ, esta foi selecionada para ser utilizada no processo de validação da proposta desta dissertação no

<sup>8</sup><http://sw.deri.org/2005/05/swan/soccer/ontology/soccer.owl>

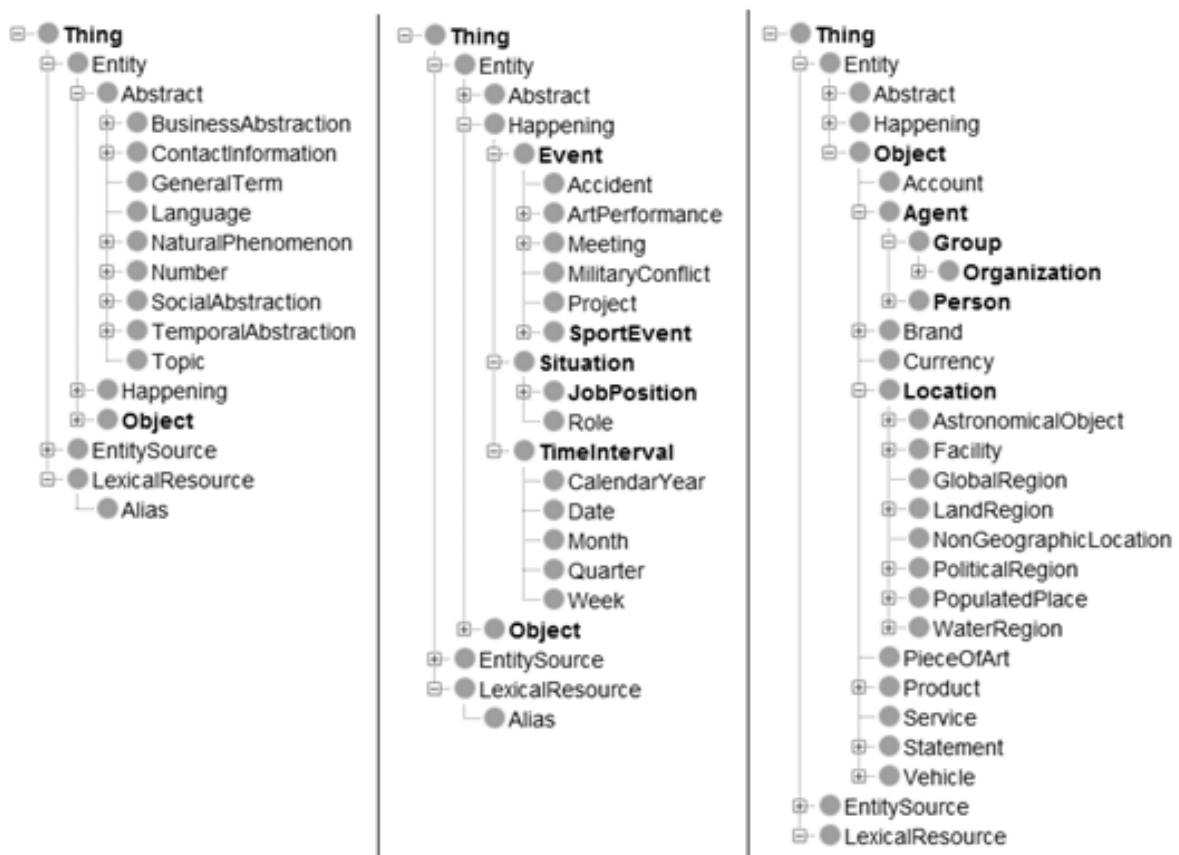


Figura 2.6: Organização hierárquica das classes na ontologia SWAN.

domínio de futebol.

### 2.4.3 HARMONISE ONTOLOGY

A Harmonise Ontology foi construída como parte do projeto Harmonise que tem como objetivo de expressar as informações no domínio de turismo e viagens de maneira que permita a desambiguação de termos que estão disponíveis em diferentes documentos ou diferentes formatos, facilitando a obtenção e troca de informações sobre este domínio para os membros participantes da rede do Harmonise (HÖPKEN; CLISSMANN, 2006).

Nesta ontologia, os termos não estão organizados em uma hierarquia, e as restrições de suas propriedades são parcialmente declaradas, possuindo na maioria delas apenas o campo "range" ou "domain" definido. Na Tabela 2.3 são exibidas, em ordem alfabética, todas as classes e na Tabela 2.4 as propriedades de objetos definidas na ontologia.

Esta ontologia nos apresenta um vocabulário formal, entretanto com as relações de seus termos com uma expressividade bem limitada, o que dificulta sua utilização em um projeto relacionado a Web semântica.

Tabela 2.3: Lista de classes definidas na ontologia Harmonise.

<b>Lista de classes definidas na ontologia Harmonise</b>			
AccessDirections	DaysOfWeek	MultiLanguageText	ReferencedValue
Accommodation	Description	MultiMediaItem	Service
Address	Event	Organisation	StreetAddress
Attraction	Facility	Participant	SupportAgency
Award	Gastro	Period	TelecomNumber
Category	IDComponent	PersonName	Telecoms
Contact	LanguageText	PointOfTime	TimeOfDay
Coordinates	LegalIdentifier	Position	TimeUnit
Cost	Link	Price	Timeline
Date	ListValue	PriceRange	Title
DateList	Location	Profile	Transport
DateRange	Measure	ProfileField	Unit
DayTimes	Messenger	ReferencePoint	XY

Tabela 2.4: Lista de propriedades definidas na ontologia Harmonise.

<b>Lista de classes definidas na ontologia Harmonise</b>			
AccessDirections	DaysOfWeek	MultiLanguageText	ReferencedValue
Accommodation	Description	MultiMediaItem	Service
Address	Event	Organisation	StreetAddress
Attraction	Facility	Participant	SupportAgency
Award	Gastro	Period	TelecomNumber
Category	IDComponent	PersonName	Telecoms
Contact	LanguageText	PointOfTime	TimeOfDay
Coordinates	LegalIdentifier	Position	TimeUnit
Cost	Link	Price	Timeline
Date	ListValue	PriceRange	Title
DateList	Location	Profile	Transport
DateRange	Measure	ProfileField	Unit
DayTimes	Messenger	ReferencePoint	XY

#### 2.4.4 ONTOUR ONTOLOGY

A OnTour Ontology (PRATNER, 2013) foi construída como parte integrante do projeto OnTour, onde a ontologia é utilizada com o objetivo de verificar a utilidade da aplicação de tecnologias de Web Semântica no campo do eTourism, demonstrando ganhos em relação à estruturação e recuperação das informações. Devido a natureza do eTourism, esta ontologia está mais relacionada a informações de pacotes de viagens, hospedagem, e outros serviços de turismo, ou seja, seus dados estão mais voltados para a área comercial do turismo. Na Figura 2.7 são apresentadas as classes e a hierarquia da ontologia.

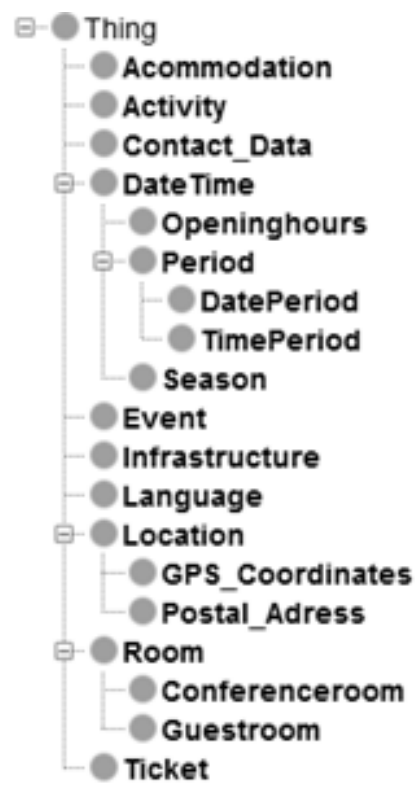


Figura 2.7: Organização hierárquica das classes na ontologia OnTour.

De maneira semelhante a Harmonise Ontology, esta ontologia define um vocabulário formal, entretanto sem definir as relações entre os seus termos e a semântica destas relações. Como esta ontologia está mais voltada para a área comercial do turismo, o conjunto de termos definidos foram considerados, durante a análise, insuficientes para atender ao domínio de Turismo utilizado na proposta.

#### 2.4.5 PROTO ONTOLOGY

A PROTON (*PROTo Ontology*) (TERZIEV et al., 2005) foi desenvolvida como parte do projeto *Semantically-Enabled Knowledge Technologies*(SEKT) que é um projeto que visa desenvolver e explorar as tecnologias de gestão do conhecimento. Esta ontologia é uma evolução da KIMO (*KIM Ontology*) (POPOV et al., 2003) e busca representar conceitos em diversos domínios, não possuindo foco em um domínio específico. Conforme apresentado na Figura 2.8, a PROTON foi dividida em quatro módulos: *System Module*, *Top Module*, *Upper Module* e o *Knowledge Management (KM) Module*.

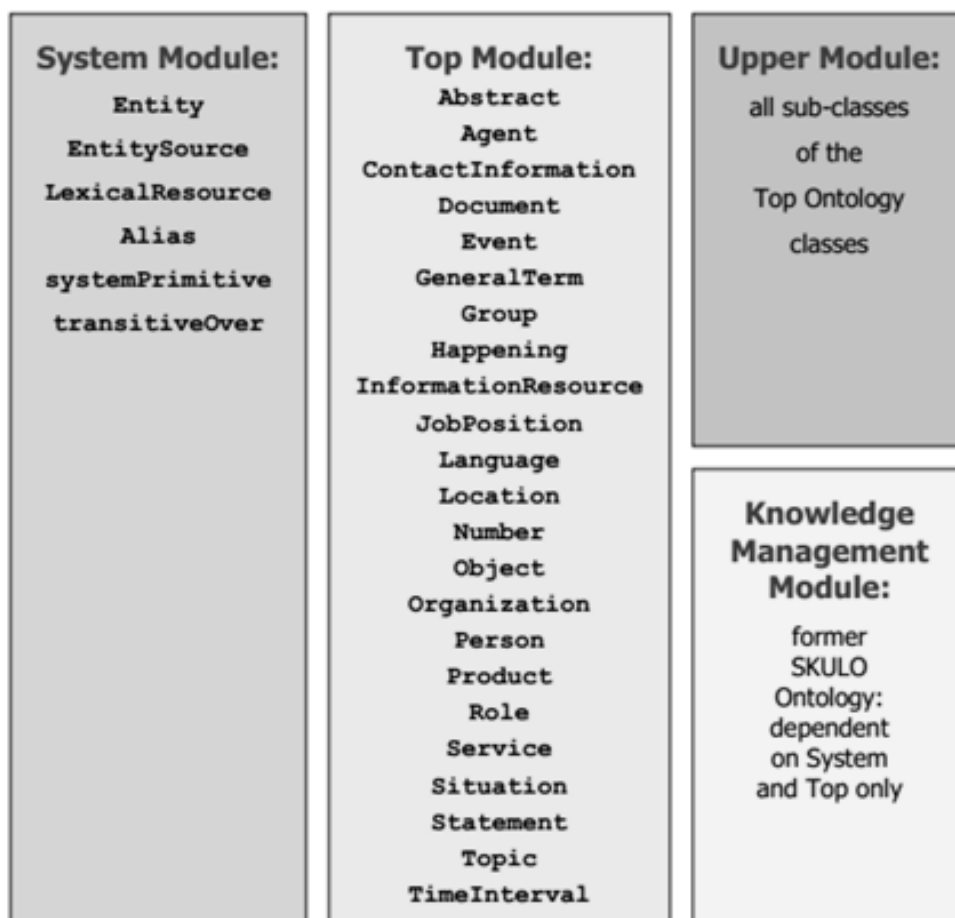


Figura 2.8: Visão geral dos módulos da PROTON (TERZIEV et al., 2005).

No *System Module* são definidos os conceitos de meta-level, ou seja, conceitos que normalmente são definidos via código em aplicações baseadas no uso de ontologias. O *Top Module* é considerado um nível mais alto na ontologia, onde o objetivo é assegurar um bom equilíbrio da utilidade, independência de domínio e facilidade no entendimento e uso. Sendo assim, esta normalmente é a melhor camada para realização de alinhamento com outras ontologias e esquemas. Já o *Upper Module* é onde se concentra o maior número de classes de domínios variados. E por fim, o *KM Module* compreende as classes que são específicas para uso em tarefas e aplicações de gestão de conhecimento.

Analisando-se a estrutura da ontologia PROTON, pode-se observar que esta, além de ser mais completa do que as demais ontologias analisadas, atende satisfatoriamente o domínio de turismo, e também complementa parte da ontologia SWAN selecionada para o domínio de futebol. Sendo assim, a PROTON foi selecionada para atender ao domínio de turismo para a validação da proposta deste trabalho.

## 2.5 TRABALHOS RELACIONADOS

Nesta seção são apresentadas algumas propostas encontradas na literatura e um comparativo com a proposta apresentada nesta dissertação.

### 2.5.1 ONTOLOGY-BASED REASONING ABOUT LEXICAL RESOURCES

(SCHEFFCZYK et al., 2006) propõe a construção de ontologias em OWL-DL a partir da transcrição das informações expressadas pelos frames do FrameNet. Além disso, (SCHEFFCZYK et al., 2006) propõe também a integração das ontologias criadas a partir desta transcrição, com outras ontologias OWL, fazendo assim com que possam ser interpretadas também as limitações para os EFs a partir de seus Tipos Semânticos. Como exemplo desta proposta temos o uso do frame "Attack" (Figura 2.9), que possui dois EFs nucleares denominados "Assailant" e "Victim", e seus frames relacionados.

A construção de uma ontologia baseada no frame "Attack" é apresentada na Figura 2.10, onde é possível fazer validações de sentenças de acordo com a estrutura definida na ontologia deste frame, como exemplo, a validação da presença de seus EFs nucleares.

A partir das ontologias de frames, é possível também a criação de uma ontologia



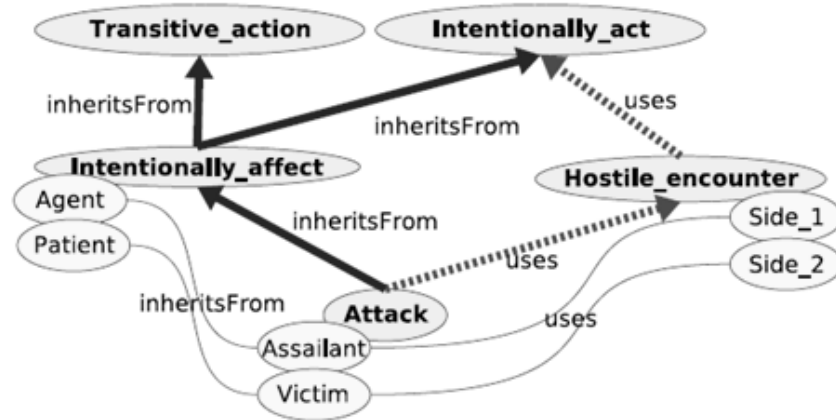


Figura 2.9: Ontologia do frame "Attack" e alguns frames relacionados (SCHEFFCZYK et al., 2006).

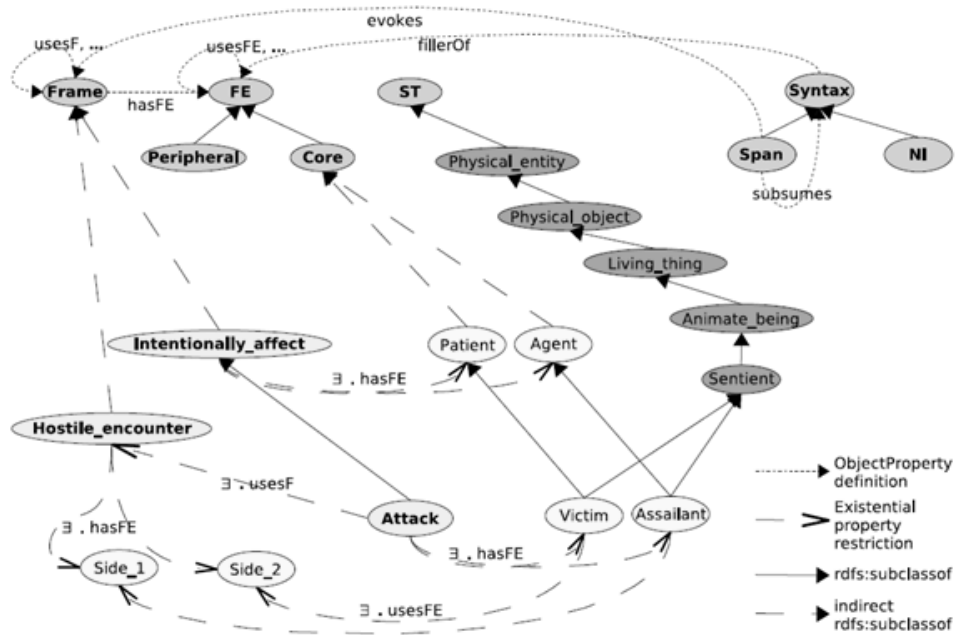


Figura 2.10: Ontologia do frame "Attack" e alguns frames relacionados (detalhado) (SCHEFFCZYK et al., 2006).

de anotações. Anotações é o processo de identificação dos fragmentos de uma sentença que representam os elementos que detalham a realização do contexto representado por um frame. Ou seja, a ontologia do frame serve como uma especificação formal para a construção da ontologia de anotações. Um exemplo desta ontologia pode ser visto na Figura 2.11.

A partir da ontologia de anotações, (SCHEFFCZYK et al., 2006) propõe que é possível realizar a busca por respostas a questões utilizando um reasoner, como o RacerPro

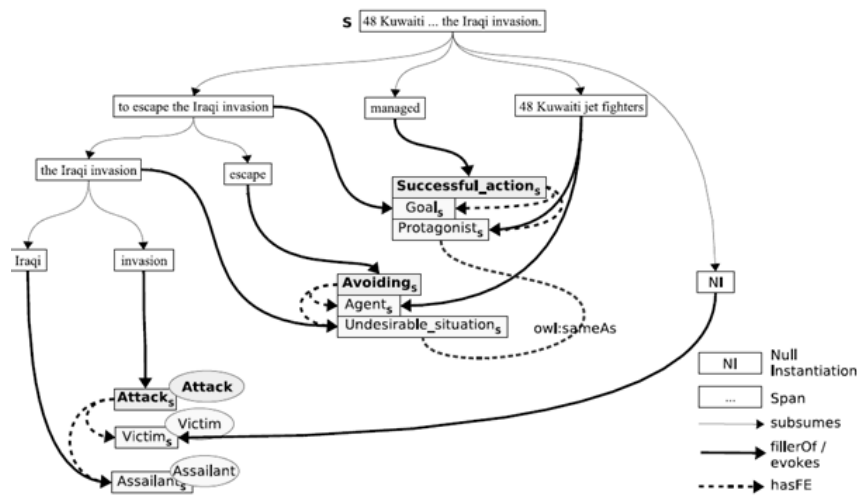


Figura 2.11: Ontologia de anotações para a sentença: "48 Kuwaiti jet fighters managed to escape the Iraqi invasion" (SCHEFFCZYK et al., 2006).

(WESSEL; MÖLLER, 2005). O processo de busca por respostas potenciais é definido em quatro etapas:

1. Identificação dos frames que serão utilizados pela pergunta, retornando também seus EFs e associando os fragmentos da pergunta a cada EF.
2. Para cada frame e EF, determinar qual a sua classe dentro da ontologia.
3. Para cada classe de frame, retornar suas instâncias dentro da base de anotações.
4. Tentar preencher as lacunas do frame com os EFs de cada instância, e verificar a compatibilidade com o EF da classe.

(SCHEFFCZYK et al., 2006) aponta que como o RacerPro é um reasoner e não uma ferramenta de Processamento de Linguagem Natural (PLN), a sua verificação de compatibilidade das lacunas de cada frame pode ser verificada apenas sintaticamente, e para o refinamento deste processo deve ser utilizada uma ferramenta de PLN.

## 2.5.2 GATHERING LEXICAL LINKED DATA AND KNOWLEDGE PATTERNS FROM FRAMENET

Observando a importancia da base de dados do FrameNet para aplicações de PLN, (NUZZOLESE et al., 2011) propõe uma transformação da base de dados do projeto FrameNet

para ontologias e dados ligados fazendo uso de uma ferramenta de triplificação RDF chamada Semion (NUZZOLESE et al., 2011).

Na abordagem proposta por Nuzzolese, cada frame é transformado em um conjunto simples de triplas RDF pela ferramenta Semion fazendo uso de uma ontologia que define a estrutura de um frame, e em seguida é possível que esta base RDF, gerada pela primeira etapa, seja alinhada a uma ontologia de domínio também pela ferramenta Semion, transformando estes dados em uma base RDF que expressa o conhecimento que está representado na base de origem na visão do domínio a qual foi aplicado (Figura 2.12).

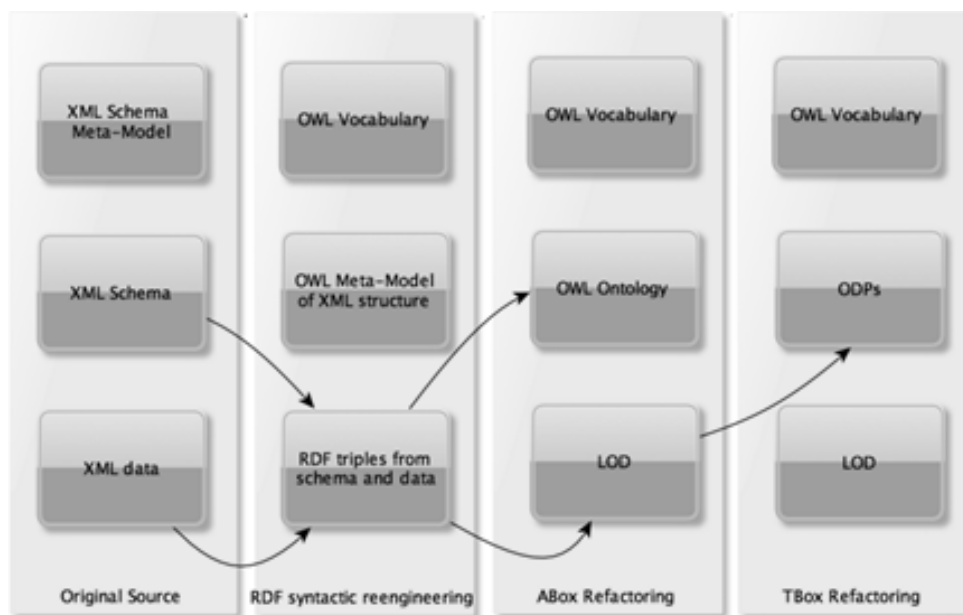


Figura 2.12: Conceitos das transformações feita pelo Semion (NUZZOLESE et al., 2011).

Conforme apontado pelo autor, o uso da ferramenta requer um conhecimento prévio sobre o domínio ao qual a base RDF será alinhada, pois o processo de alinhamento é semi-automático, sendo necessário que o usuário defina regras de transformação a qual a base será submetida.

Após o alinhamento, a base RDF poderá ter dois formatos resultantes, uma base de dados ligados, ou então uma base de *Knowledge Patterns*(KP). Como exemplo da transformação dos dados em RDF para o formato de KP, temos a transformação da base de RDF gerada a partir da base do FrameNet para Knowledge Patterns na Figura 2.13.

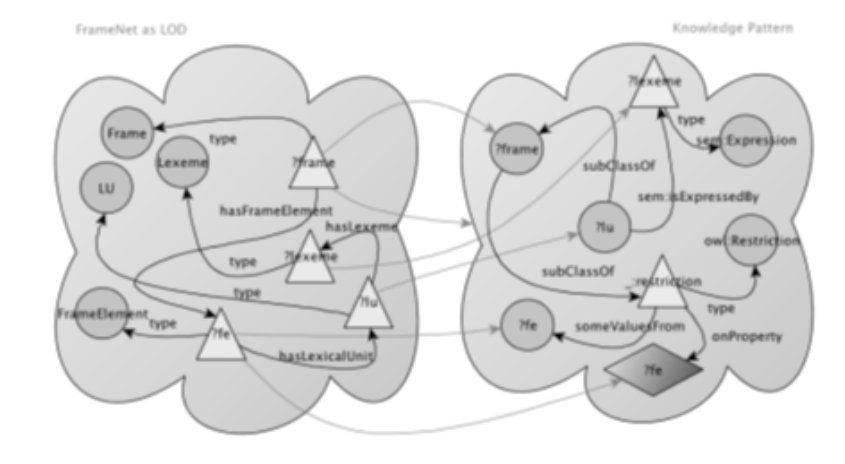


Figura 2.13: Regras de transformação de LOD para KP para base do FrameNet (NUZZOLESE et al., 2011).

### 2.5.3 THAI FRAMENET CONSTRUCTION AND TOOLS

(LEENOI et al., 2011) apresenta uma abordagem para a construção da FrameNet da língua Tailandesa baseada nas abordagem de merge e expansão. Isso permite que a construção do TFN (*Thai FrameNet*) siga uma metodologia que busca traduzir os frames da língua inglesa para o tailandesa e expandir sua carga realizando anotações em um corpora local, além do desenvolvimento de frames diretamente no TFN e definição da relação com outros frames em momento posterior. Além desta abordagem, também são definidos vocabulários a partir de ontologias que permitem a formalização dos elementos que participam de um frame (Figura 2.14).

Também são propostas duas ferramentas para apoio ao desenvolvimento do TFN. A primeira ferramenta visa apoiar a extração de sentenças de bases de sentenças para realização da anotação a partir da identificação da Unidade Lexical evocada na sentença. A partir desta identificação, a ferramenta lista os EF relacionados ao frame que possui essa UL, permitindo ao usuário identificá-los manualmente na sentença. Depois desse processo, a sentença pode ser exportada para um arquivo XML, conforme apresentado na figura 2.15.

Já a segunda ferramenta visa apoiar a visualização dos dados de cada frame e seus Elementos de Frame, e qual a sua participação na rede de frames, construída a partir do seu relacionamento com outros frames (Figura 2.16).



Figure 2.14: Thai FrameNet Ontology (LEENOI et al., 2011).

```

<extracted_sentence id=1 frame=การค้าขาย word=ขาย>
<sentence>เขา ขาย ก๊วย</sentence>
<annotation_result>
  <chunk>
    <word>เขา</word>
    <semantic_role>Seller [Sfr]</semantic_role>
    <phrase_type>NP </phrase_type>
    <grammatical_function>SUBT </grammatical_function>
  </chunk>
  <chunk>
    <word>ก๊วย</word>
    <semantic_role>Goods [Gds]</semantic_role>
    <phrase_type>NP </phrase_type>
    <grammatical_function>OBJT </grammatical_function>
  </chunk>
</annotation_result>
</extracted_sentence>

```

Figura 2.15: Anotação exportada em XML pela ferramenta da TFN (LEENOI et al., 2011).

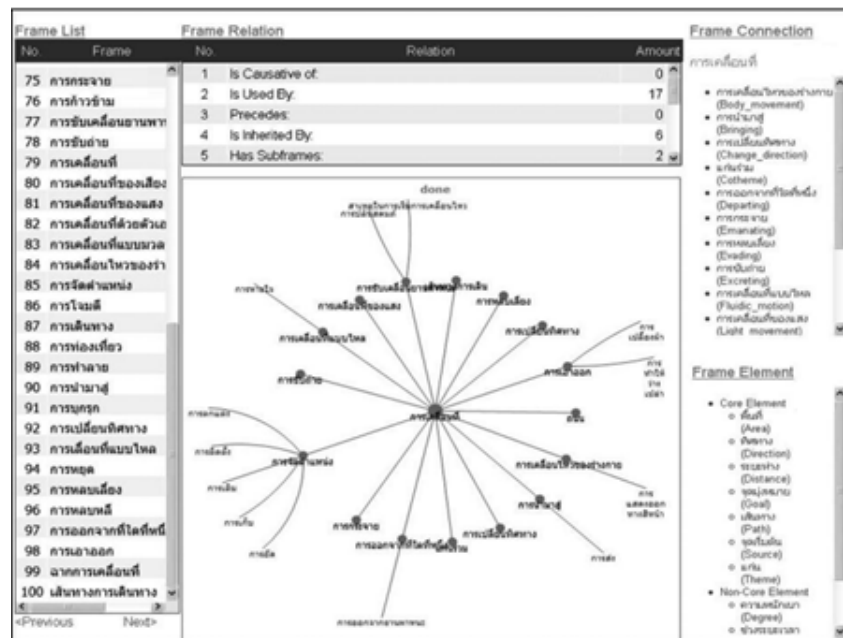


Figura 2.16: Ferramenta de apoio à visualização de Frames da TFN (LEENOI et al., 2011).

## 2.5.4 APRI E DC2AP

### 2.5.4.1 Analysis Patterns Reuse Infrastructure (Apri)

A APRI foi proposta por (VEGI et al., 2012) com a finalidade de prover uma infraestrutura que faz uso de serviços Web e uma representação de metadados para a especificação de padrões de análise a fim de apoiar a catalogação e reutilização destes. A construção da APRI foi baseada em Infraestruturas de Dados Espaciais (IDE), que segundo (INFRAS-

TRUCTURES, 2004) provêm uma base para descoberta de dados espaciais, avaliação e aplicação para usuários de vários tipos de instituições. Muitas das IDEs desenvolvidas atualmente são baseadas em *Service-Oriented Architecture*(SOA), o que torna possível a criação de ambientes compartilhados, distributos e interoperáveis (DAVIS; ALVES, 2005); Além disso o uso de IDE permite a disponibilidade de bases de dados espaciais de diferentes fontes, facilitando ao usuário a aquisição de novos datasets, sem a necessidade de reconstruí-los ou convertê-los, reduzindo assim a sua carga de trabalho (RAJABIFARD; WILLIAMSON, 2001).

A APRI é descrita por (VEGI et al., 2012) como uma infraestrutura para catalogação de padrões de análise baseada na arquitetura SOA que possibilita que seus dados sejam acessados tanto por clientes humanos ou softwares. Possui os seguintes componentes (Figura 2.17):

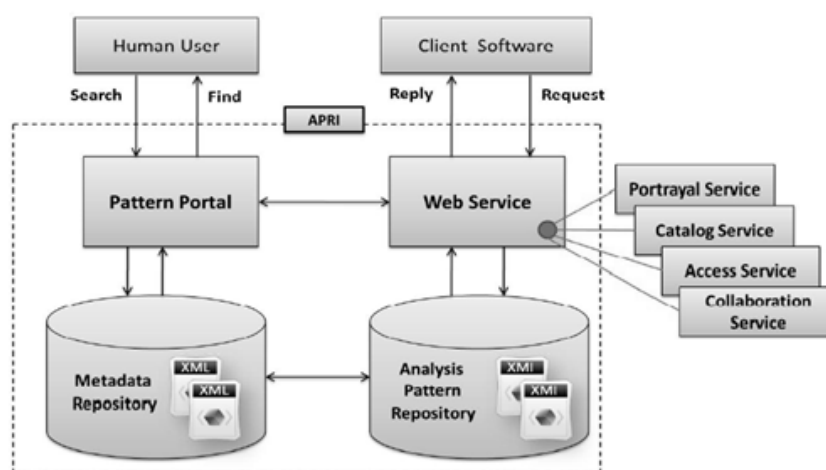


Figura 2.17: Analysis Patterns Reuse Infrastructure (APRI) (VEGI et al., 2012).

1. **Patten Portal:** Provê uma interface para utilização dos serviços disponíveis da APRI.
2. **Repositório de Metadados:** Repositório onde estão armazenados os metadados, em XML, que descrevem os padrões de análise ou serviços disponíveis do APRI
3. **Repositório de Padrões de Análise:** Repositório onde são armazenados os Padrões de Análise no formato XMI (XML Metadata Interchange), possibilitando o seu uso para visualização e serviços colaborativos

4. **Serviço de apresentação:** Serviço que provê apoio a visualização dos padrões de análise da APRI
5. **Serviço de catalogação:** Serviço que possibilita a descoberta e uso dos padrões de análise e serviços da APRI a partir de seus metadados.
6. **Serviço de acesso:** Serviço que controla o acesso e download dos padrões de análise.
7. **Serviço de colaboração:** Serviço que permite que o projetista experiente possa contribuir aprimorando os padrões de análise.

#### 2.5.4.2 A Dublin Core Application Profile To Analysis Patterns (DC2AP)

(VEGI, 2012) ressalta que um dos principais problemas no desenvolvimento de software está na fase de análise e projeto, e a cada etapa em que o problema gerado persiste, seu custo é aumentado consideravelmente (BOEHM; BASILI, 2005). Uma maneira de tornar o processo de análise mais ágil e preciso é fazendo uso de padrões de análise (FERNANDEZ; YUAN, 2000). Entretanto conforme apontado por (BLAIMER et al., 2010) os meios de acesso a padrões de análise são muito pobres. Buscando resolver o problema de acesso e reuso aos padrões de análise, Vegi et al. (2011) propõe a utilização de uma infraestrutura especializada denominada APRI, apresentada na seção anterior. No entanto, mesmo considerando o uso de uma infraestrutura como a APRI, não existe um consenso sobre a forma como os metadados que descrevem estes padrões de análise devem ser descritos, fazendo com que cada dataset de padrões de análises seja descrito conforme a preferência de seu autor (BLAIMER et al., 2010).

(VEGI, 2012) propõe então um padrão para documentação dos padrões de análise, o DC2AP, baseado no padrão Dublin Core e no padrão proposto em (PANTOQUILHO et al., 2003) e (RAMINHOS et al., 2006) para descrição de padrões de análise. Além disso o DC2AP é compatível com a infraestrutura APRI, tornando possível a criação de um dataset reutilizável e colaborativo entre outros usuários da APRI. O modelo padrão da DC2AP foi gerado a partir da junção de outros dois padrões para documentação conforme detalhado na Tabela 2.5.

Após a concepção deste novo padrão, alguns ajustes tiveram de ser efetuados. Alguns termos, como "*Applicability*" e "*Structural Adjustments*" foram considerados desnecessários e foram removidos. E então, com a finalidade de controlar a maneira como estes



Tabela 2.5: Mapeamento entre os padrão Dublin Core e de Pantoquilha.

<b>Simple Dublin Core</b>	<b>Pantoquilo's Template</b>
Title	1. Name 2. Also Know As
Creator	3. History (parc. equivalente)
Subject	4. Context
Description	5. Problem 6. Motivation 7. Context 8. Applicability 14. Examples (parc. equivalente) 18. Known Uses (parc. equivalente)
Publisher	(Nenhum equivalente)
Contributor	3. History (parc. equivalente)
Date	3. History (parc. equivalente)
Type	(Nenhum equivalente)
Format	(Nenhum equivalente)
Identifier	1. Name (parc. equivalente)
Source	15. Related Patterns (parc. equivalente)
Language	(Nenhum equivalente)
Relation	13. Anti-Patterns Trap (parc. equivalente) 15. Related Patterns (parc. equivalente) 16. Design Patterns (parc. equivalente)
Coverage	(Nenhum equivalente)
Rights	(Nenhum equivalente)
(Nenhum equivalente)	4. Structural adjustments 9. Requirements 9.1. Functional requirements 9.2. Non-functional requirements 9.3. Dependencies and contributions 9.4. Conflict identification guidance to resolution 9.5. Priorities 9.6. Participants 10. Modelling 10.1. Structure 10.1.1. Class diagram 10.1.2. Class description 10.2. Behaviour 10.2.1. Collaboration or sequence diagrams 10.2.2. Activity diagrams 10.2.3. State diagrams 10.3. Solution Variants 11. Resulting context 12. Consequences 17. Design guidelines

campos da documentação podem ser preenchidos foram propostas regras de obrigatoriedade, ocorrência e tipos conforme descrito na Tabela 2.6.

Tabela 2.6: Regras de Aplicação do DC2AP.

Rule Type	Rules and Acronyms	Description
Obligatoriness	Mandatory (M)	Element must always be filled in the documentation of an analysis patterns.
	Optional (O)	Element is optional, as it refers to additional information.
	Conditional (Cd)	Element must be filled if there is corresponding information.
Occurrence	Single (S)	Element can occur only once in the documentation of an analysis pattern.
	Multiple (Mu)	Element can occur multiple times in the documentation of an analysis pattern.
Value type	String (St)	Element should be filled with textual descriptions.
	Date (D)	Element must be filled with a date following a format previously specified.
	URI (U)	Element should be filled with a string used to identify a resource on the Internet.
	Null (N)	Element should not be filled with any value e because their data are concentrated in its element refinements.
	URI, Number or String (UNS)	Element can be filled with a resource identifier, with a number or a textual description.
	URI or String (US)	Element can be filled with a resource identifier or a textual description.

A fim de transformar os dados que serão descritos, respeitando o perfil apresentado acima, em um perfil que possa ser processado por maquina e atender a abordagem de Dados Ligados, cada elemento proposto pelo DC2AP deve ser associado a uma URI e descrito em RDF (*Resource Description Framework*).

A associação da APRI fazendo uso do DC2AP como padrão para descrição dos metadados de seus padrões de análise, transforma a APRI em um dataset de Dados Ligados, e então possibilitando a interoperabilidade com outros datasets de dados ligados.

### 2.5.5 DISCUSSÕES SOBRE OS TRABALHOS RELACIONADOS

A partir dos trabalhos apresentados na seção anterior, algumas ideias e características foram discutidas na especificação da proposta da FSI, incluindo, o uso de ontologias para formalização da estrutura dos frames e seus relacionamentos (SCHEFFCZYK et al., 2006) (NUZZOLESE et al., 2011); a construção de uma infraestrutura orientada a

serviços combinada com um modelo formal para descrição dos seus dados [(VEGI et al., 2012);(VEGI, 2012); e o desenvolvimento de uma ferramenta de apoio a documentação de frames e realização de anotações de sentenças (LEENOI et al., 2011).

Na proposta de (SCHEFFCZYK et al., 2006) são utilizadas ontologias para descrever formalmente a estrutura de um frame, o que traz benefícios como a possibilidade de verificação da validade da sua estrutura, e processamento dessas informações, a partir do uso de uma ferramenta relacionada a uma maquina de inferência, o que permite que sejam respondidas perguntas a partir da base de anotações de sentenças; além de permitir a restrição semântica da participação dos EFs em anotações de sentenças. Neste aspecto, aproveitamos a ideia de se criar uma ontologia para formalização da estrutura dos frames e de suas relações buscando obter um maior nível de confiabilidade para os dados armazenados, e permitindo que outras ferramentas possam tirar maior proveito dos dados graças a seu formalismo. Entretanto, buscamos ampliar a participação da ontologia, permitindo não apenas validar a estrutura de um frame, mas a relação com outros frames e também entre EFs, uma vez que a definição da semântica de frames ressalta que o frame também depende das suas relações com outros frames, e não apenas de seus componentes como os Elementos de Frame e Unidade Lexical.

Na proposta de (NUZZOLESE et al., 2011) os dados do projeto FrameNet foram transformados em recursos de dados ligados, utilizando ontologias, de maneira semi-automática por meio de uma ferramenta desenvolvida pelo autor. Segundo o autor, esta transformação possibilita uma maior integração destes dados com outras bases de dados ligados, provendo mais conhecimento, a partir da navegação nestes relacionamentos. De maneira semelhante a proposta de (NUZZOLESE et al., 2011), na FSI é utilizado também padrão de dados ligados, mas utilizando um vocabulário já disponível na FSI, a partir da integração dos dados de anotações de sentenças com outras bases de dados ligados. O diferencial da FSI neste caso, além da maior expressividade da ontologia que define o vocabulário formal dos dados, também está no uso de ontologias de domínio para permitir uma maior expressividade também a partir dos recursos externos ligados por meio do padrão de dados ligados, permitindo assim a formação de uma rede de conhecimento mais rica.

Considerando a proposta de (LEENOI et al., 2011), no contexto da FSI foram utilizadas ontologias para formalização de parte dos dados do FrameNet Tailandesa, e a construção de ferramentas com o objetivo de apoiar a documentação de Frames e realiza-

ção de anotação de sentenças. Na proposta deste trabalho também foram desenvolvidas ferramentas para apoiar o processo de documentação de frames, e anotações, sendo o principal diferencial de que neste passo também são utilizadas as informações semânticas para auxiliar ao usuário no processo de documentação de frames e anotações, garantindo uma maior confiabilidade dos dados, uma vez que desta maneira, através de inferência, as ontologias utilizadas no processo permitem fornecer ao usuário sinais de incoerências nos dados expressos.

(VEGI et al., 2012) propõe uma infraestrutura para controle e compartilhamento de padrões de análise com base nas suas descrições de metadados baseados em um vocabulário formal, e também aproveitando da orientação a serviços para criar uma interface de comunicação para uso de seus dados por ferramentas externas de maneira mais organizada. Assim como (VEGI et al., 2012), na FSI foram criados vocabulários formais para representação dos dados, mas com uma maior expressividade a partir do uso de OWL e SWRL por exemplo. Além disso, também foi utilizado o padrão de orientação a serviços para a construção de uma interface de comunicação com a infraestrutura de maneira organizada, promovendo assim maior disponibilidade para integração com outras ferramentas.

Na Tabela 2.7 é apresentada uma comparação entre os trabalhos relacionados e a proposta da FSI, considerando os seguintes aspectos: i) se estão abordando o domínio da FrameNet ou se estão tratando de outro domínio; ii) Se utilizam uma ontologia ou algum outro modelo para o vocabulário formal; iii) se adotam o padrão de dados ligados para integração de seus dados; iv) se apresentam uma interface de interação de ferramentas externas via serviços;

Tabela 2.7: Comparativo com trabalhos relacionados.

Trabalhos Relacionados	FrameNet	Ontologias	(Voc. Formal) Dados Ligados	Web Services
Scheffczyk et al. (2006)	Sim	Sim	Não	Não
Nuzzolese et al. (2011)	Sim	Não	Sim	Não
Leenoi et al. (2011)	Sim	Sim	Não	Não
Vegi et al. (2012)	Não	Não	Sim	Sim
FrameNet Semantic Infrastructure	Sim	Sim	Sim	Sim

## 2.5.6 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Neste capítulo foi apresentado o referencial teórico relacionado a proposta deste trabalho, bem como as principais pesquisas relacionadas.

Diversos conceitos utilizados nestes trabalhos também são utilizados na construção da FSI, como o uso de ontologias a fim de prover um formalismo aos dados; a aplicação do padrão de dados ligados para permitir uma maior integração dos dados locais com recursos da web dando mais riqueza a obtenção de informações; Serviços Web para permitir a interação da infraestrutura com outras ferramentas; entre outros.

No próximo capítulo serão abordadas as principais características arquiteturais deste trabalho, junto com seus conceitos e funcionalidades.

## 3 FRAMENET SEMANTIC INFRASTRUCTURE

Neste capítulo será descrita a infraestrutura proposta neste trabalho, sendo apresentada sua arquitetura e as tecnologias utilizadas para sua construção. Este capítulo se organiza da seguinte maneira: seção 3.1, introdução; seção 3.2, ontologias desenvolvidas; seção 3.3, arquitetura da FSI; e finalmente na seção 3.4 serão apresentadas as considerações finais para este capítulo.

### 3.1 INTRODUÇÃO

O FrameNet (RUPPENHOFER et al., 2006), conforme ressaltado no capítulo 2 desta dissertação, é um importante recurso lexical baseado na Semântica de Frames, que pode ser utilizado por diversas ferramentas de Processamento de Linguagem Natural, como traduções automáticas, extração de informações, sumarização de textos, entre outras. Além disso, o estabelecimento de projetos que objetivam adaptar o projeto FrameNet para outras linguagens confirma a importância da construção deste recurso.

Neste contexto, a aplicação de tecnologias advindas da web semântica, como ontologias e dados ligados, pode trazer benefícios como promover a legibilidade dos dados por parte de máquinas, promover o compartilhamento das informações contidas na base do projeto, e promover a confiabilidade dos dados por meio da validação semântica da composição dos frames e das relações entre eles. Estes benefícios podem ser observados nas propostas de (SCHEFFCZYK et al., 2006), (NUZZOLESE et al., 2011) e (LEENOI et al., 2011), apresentadas e discutidas no capítulo anterior.

Neste capítulo é apresentada a arquitetura da FrameNet Semantic Infrastructure, que é uma infraestrutura baseada no padrão SOA (ENDREI et al., 2004), e que busca aplicar conceitos de Web Semântica aos dados do projeto FrameNet de maneira a contribuir com a manutenção de seus dados e com a aplicabilidade destes dados em outras atividades, como, por exemplo no processamento de linguagem natural. Nesta infraestrutura, são estabelecidos dois tipos de ontologias para auxiliar no processo de manutenção e aplicabilidade dos dados: i) a ontologia de metadados do FrameNet, que busca descrever semanticamente a estrutura dos dados que compõe os frames e a semântica das relações entre eles, tornando estes dados legíveis por máquina devido ao formalismo imposto pela

ontologia e contribuindo para a confiabilidade dos dados, uma vez que a ontologia garante a validade semântica dos dados; e ii) ontologias de domínio, que servem como fonte para definição de tipo semântico dos elementos de frames, provendo uma expressividade semântica para os fragmentos da cena que referenciam cada um de seus frames.

Também é explorado por esta abordagem o padrão de dados ligados (BERNERS-LEE, 2009), permitindo ligar cada fragmento de uma cena que representa um EF a um recurso da Web, facilitando a obtenção de novas informações a partir da recuperação destes recursos.

No contexto da aplicação do projeto foi utilizada a base de dados do projeto COPA2014 (SALOMAO et al., 2011) considerando que seu padrão de manutenção é o do mesmo do FrameNet-BR.

## 3.2 ONTOLOGIAS

No contexto da infraestrutura FSI, são utilizadas dois tipos de ontologias, a ontologia de metadados do FrameNet, denominada Onto-Frame-BR, junto a ontologia de metadados de anotações, denominada Onto-Annotation-BR, e ontologias de domínio, de acordo com o domínio sendo modelado pela infraestrutura. A ontologia Onto-Frame-BR foi desenvolvida no contexto deste trabalho e será apresentada a seguir. Já as ontologias do domínio podem ser utilizadas as disponíveis na literatura, desde que as mesmas estejam modeladas em OWL. As ontologias de domínio utilizadas para a validação da infraestrutura serão apresentadas no capítulo 5 desta dissertação. Estas ontologias de domínio utilizadas são relacionadas aos domínios de futebol e turismo, que são relacionadas ao contexto do projeto COPA2014, projeto este ao qual esta dissertação está relacionada.

### 3.2.1 ONTOLOGIA DE METADADOS DO FRAMENET

A Ontologia ONTO-FRAME-BR é utilizada na FSI com o objetivo de prover uma base semântica para os dados. Como parte do desenvolvimento desta ontologia, algumas outras propostas de ontologias de metadados da FrameNet foram analisadas [(NUZZOLESE et al., 2011); (SCHEFFCZYK et al., 2006); (LEENOI et al., 2011)] a fim de verificar a viabilidade de uso na FSI. Das três ontologias analisadas apenas a apresentada por (NUZZOLESE et al., 2011) foi especificada integralmente. As ontologias de (SCHEFFCZYK

et al., 2006) e (LEENOI et al., 2011) não foram disponibilizadas em forma de arquivo, sendo apenas descritas superficialmente no artigo que as apresenta.

Apesar da ontologia de (NUZZOLESE et al., 2011) estar disponível, esta foi construída a partir da engenharia reversa dos dados da FrameNet, o que causou a perda de expressividade tanto nas relações e propriedades entre as estruturas presentes na semântica de frames, como nas relações entre frames, uma vez que por este processo não é possível identificar a semântica envolvida nas relações da base, como por exemplo não é possível verificar a restrição de transitividade ou reflexão de um propriedade de objeto, entre outras propriedades. Além disso, foram identificadas classes que não são compatíveis com a estrutura dos dados do FrameNet e que não são descritas pelo autor.

Considerando os problemas detectados na ontologia apresentada por (NUZZOLESE et al., 2011), optou-se por desenvolver uma nova ontologia. Para a construção desta ontologia foi adotada uma metodologia semelhante a utilizada por (NUZZOLESE et al., 2011), sendo realizada, inicialmente, uma engenharia reversa na base de dados do projeto FrameNet, identificando, no entanto, as inconsistências da base e assim tentando explicitar o conhecimento de maneira correta, com ajuda de especialistas. Entretanto, devido ao grande volume de dados da base do FrameNet, foi adotado a base do projeto COPA2014 FrameNet-BR (SALOMAO et al., 2011) que, por ser um projeto criado a partir do FrameNet respeita os mesmos padrões e possui um volume de dados menor, sendo, portanto, mais fácil de realizar a engenharia reserva.

Neste contexto, as entidades que compunham o modelo do banco de dados estritamente relacionadas a representação do frame foram mapeadas inicialmente como classes da ontologia, e cada uma das relações entre estas entidades são mapeadas como propriedades de objetos onde as classes equivalentes às entidades participantes da relação são declaradas como domínio e imagem desta propriedade. E, por consequência, os atributos das entidades são mapeados como propriedades de dados nesta ontologia.

Na Figura 3.1 temos a representação hierárquica das classes que compõe a ontologia ONTO-FRAME-BR, onde cada classe é representada por um círculo, sendo que os círculos cinza são classes equivalentes e os círculos brancos são classes definidas.

Após a realização da engenharia reversa nos dados do projeto Copa 2014, foi necessário realizar um refinamento da ontologia a partir de estudos da documentação do FrameNet apresentada em (RUPPENHOFER et al., 2006).



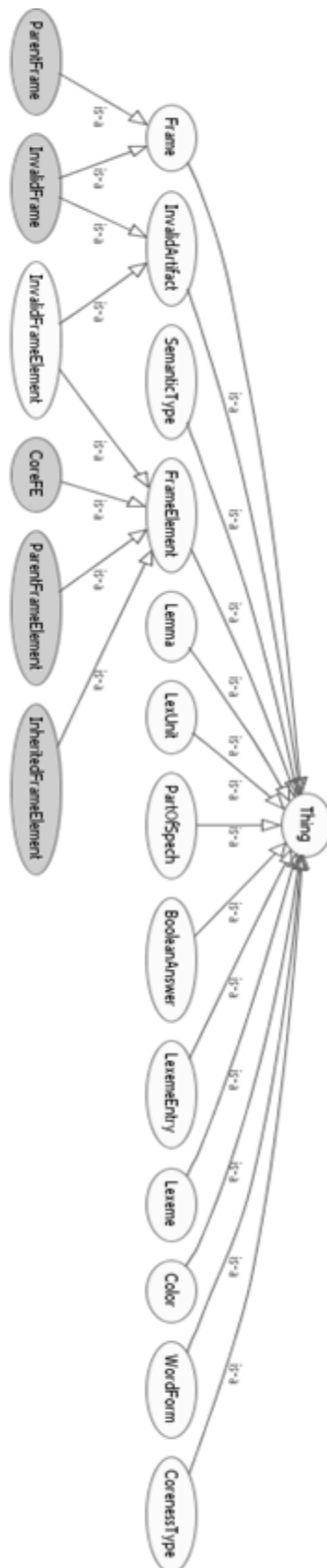


Figura 3.1: Classes da ONTO-FRAME-BR.

O primeiro passo no refinamento da ontologia foi na definição de restrições existenciais e universais das classes a fim de validar os indivíduos a partir dos requisitos mínimos para sua existência. Como exemplo, na Figura 3.2 são apresentadas as restrições para a classe ontológica *FrameElement*.

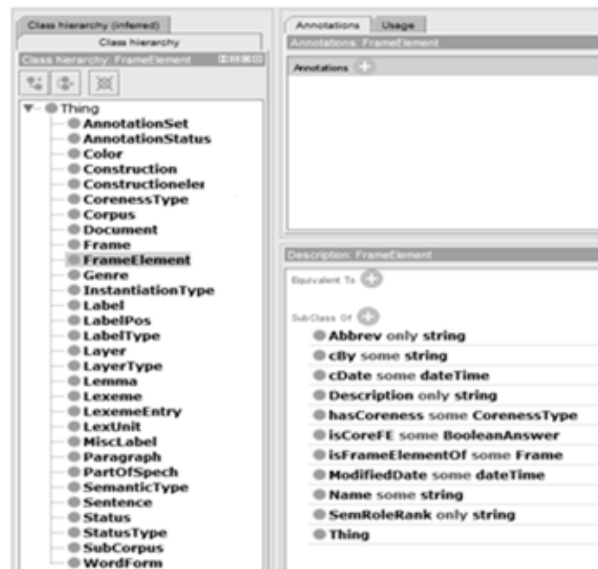


Figura 3.2: Classes e restrições declaradas na ontologia com uso do Protégé.

Em seguida, foi realizada a desambiguação das relações entre Frames e relações entre EFs, que na base obtida do projeto Copa 2014 eram mantidos de maneira inconsistente (não era possível distinguir essas relações), de maneira a evitar que relações sejam atribuídas incorretamente, e também a definição semântica destas relações com base na descrição de (RUPPENHOFER et al., 2006).

Sendo separadas assim as relações *Inheritance*, *Perspective\_on*, *SubFrame*, *Precedes*, *Inchoative\_of*, *Causative\_of*, *Using* e *See\_Also* como relações entre frames, e as relações *Inherits*, *Excludes*, *Requires* e *CoreSet* como relações entre EFs. Estas relações serão detalhadas nas seções 3.2.1.1 e 3.2.1.2.

Entretanto algumas definições semânticas não puderam ser totalmente especificadas apenas com o uso da linguagem OWL (MARTIN et al., 2004). Sendo assim, foi adotado o uso de regras definidas em SWRL<sup>9</sup> (HORROCKS et al., 2004) com o objetivo de classificar indivíduos ou identificar relações implícitas entre indivíduos com as quais não seriam possíveis apenas com o uso de OWL.

<sup>9</sup>Apesar de ter sido utilizada a OWL 2.0, apenas parte de seus novos recursos foram utilizados, como a cardinalidade qualificada por exemplo. O recurso *property chains* não foi utilizado, devido a não terem sido identificadas situações onde os mesmos poderiam ser aplicados durante a modelagem desta ontologia.

### 3.2.1.1 Relações entre frames

(RUPPENHOFER et al., 2006) definiu um documento para orientação da estrutura do projeto *FrameNet*, e (LEENOI et al., 2011) apresenta uma proposta de uma ontologia para manutenção da *FrameNet* Tailandesa. Ambos descrevem sete relações possíveis entre frames e as restrições destas relações. Considerando esta documentação e com o intuito de representar de maneira adequada a estrutura da *FrameNet*, foi definida na ontologia construída para a FSI, a semântica das relações entre frames com base nas definições apresentadas por estes autores conforme será detalhado abaixo. Com objetivo de permitir identificar os frames que desrespeitam estas ou qualquer outra das restrições definida por regras SWRL, criamos uma classe denominada *InvalidFrame*, onde serão classificados estes indivíduos.

A primeira relação tratada foi a de Herança, que é descrita como uma relação entre dois frames onde o frame que herda deverá possuir todos os EF, com exceção dos extra-temáticos, do frame que foi herdado. Sendo assim, foram criadas duas novas classes: a *InheritableFrameElement*, que é definida por qualquer EF que não seja extra-temático de um frame. Inicialmente foi criada uma regra SWRL para tratar essa restrição, onde seria explicitado que todos os EF do frame herdado que estão classificados como *InheritableFrameElement* possuem EF herdados no outro Frame, caso negativo o Frame deve ser classificado como *InvalidFrame*. Entretanto, na especificação desta regra, foi identificado um problema na expressividade da SWRL em relação à negação de relações, o que tornou a regra ineficiente, devendo essa, portanto ser tratada externamente pela ferramenta.

Já a relação *Perspective\_on* é descrita como uma relação entre um frame neutro e outro frame não neutro. Esta relação ocorre quando um frame neutro pode adotar mais de um ponto de vista. Desta forma, os EFs podem variar de acordo com o ponto de vista utilizado, não podendo coexistir no mesmo Frame. Para explicitar esta restrição, a propriedade equivalente a esta relação na ontologia foi descrita como não reflexiva, sem a necessidade de criação de regra SWRL (Figura 3.3).

Para as relações *Causative\_of* e *Inchoative\_of* foi necessária a criação de regras SWRL devido a sua semântica mais apurada, considerando que os frames causativos deverão herdar do frame *Transitive\_action*. Enquanto os frames incoativos devem herdar do frame *Event*, *State* ou *Gradable\_attributes*.

Conforme apresentado na figura 3.4, a regra para a relação "*Causative\_Of*" verifica se

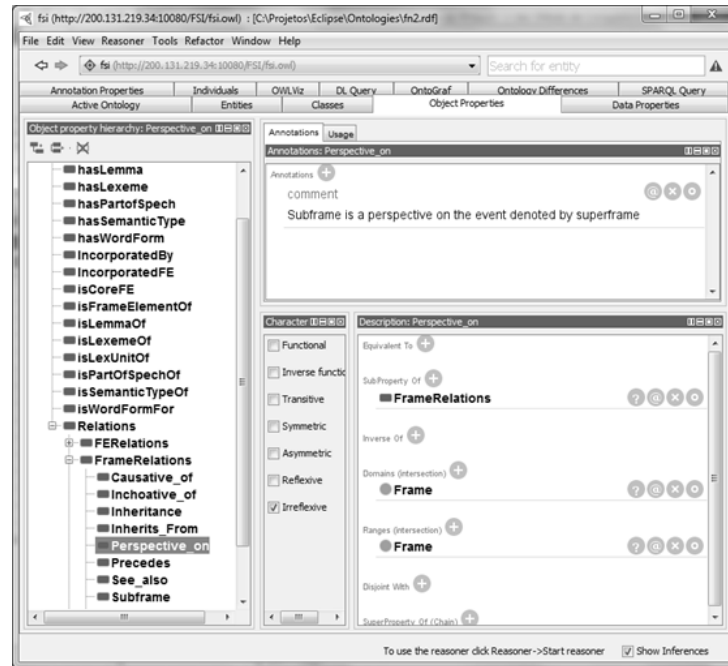


Figura 3.3: Restrições para a relação *Perspective\_on* no Protégè.

o frame que for definido como causativo de outro frame, herda de um frame que possua o nome diferente de "*Transitive\_action*", classificando o frame alvo da relação *causative\_of* na classe *Invalid\_Frame*. Entretanto, existe também a possibilidade do frame definido como causativo não possuir relação de herança com nenhum outro frame, que neste caso não é capturado por esta regra. Tentou-se criar uma regra para verificar esta situação (Figura 3.5). Entretanto, de acordo com o pressuposto de mundo aberto (apresentado na seção 2.1.1.2), a ausência da definição da herança de um frame não equivale a dizer que ela não exista, o que tornou impossível a verificação desta situação.

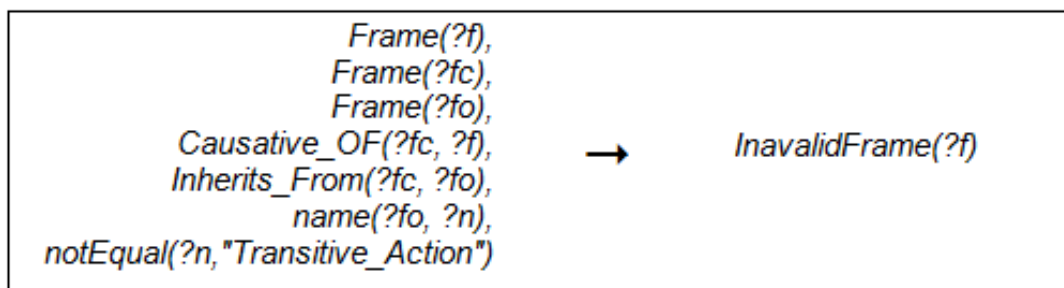


Figura 3.4: Regra para verificação da herança de frame causativo.

Já para os frames Incoativos foi necessária a criação de uma classe denominada *IncoativeBaseFrame* (Figura 3.6) que agrupa os *Frames Event*, *State* ou *Gradable\_attributes* que servem de base para os frames incoativos que devem herdar de um destes frames.

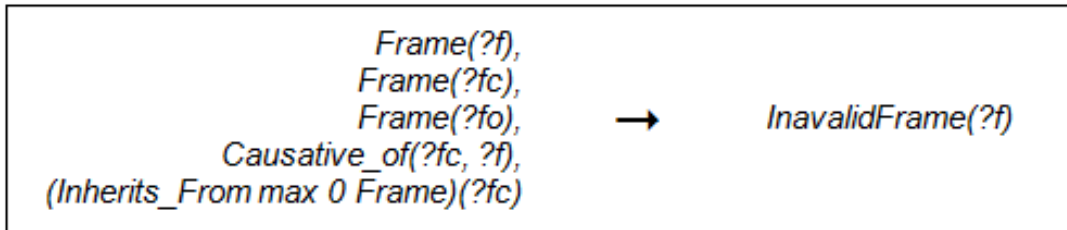


Figura 3.5: Regra verificando a inexistência de herança em frame causativo.

Após definida a classe foi também definida a regra SWRL que verifica se um frame declarado como incoativo de outro frame herda de um Frame que participe da classe definida acima (Figura 3.7).

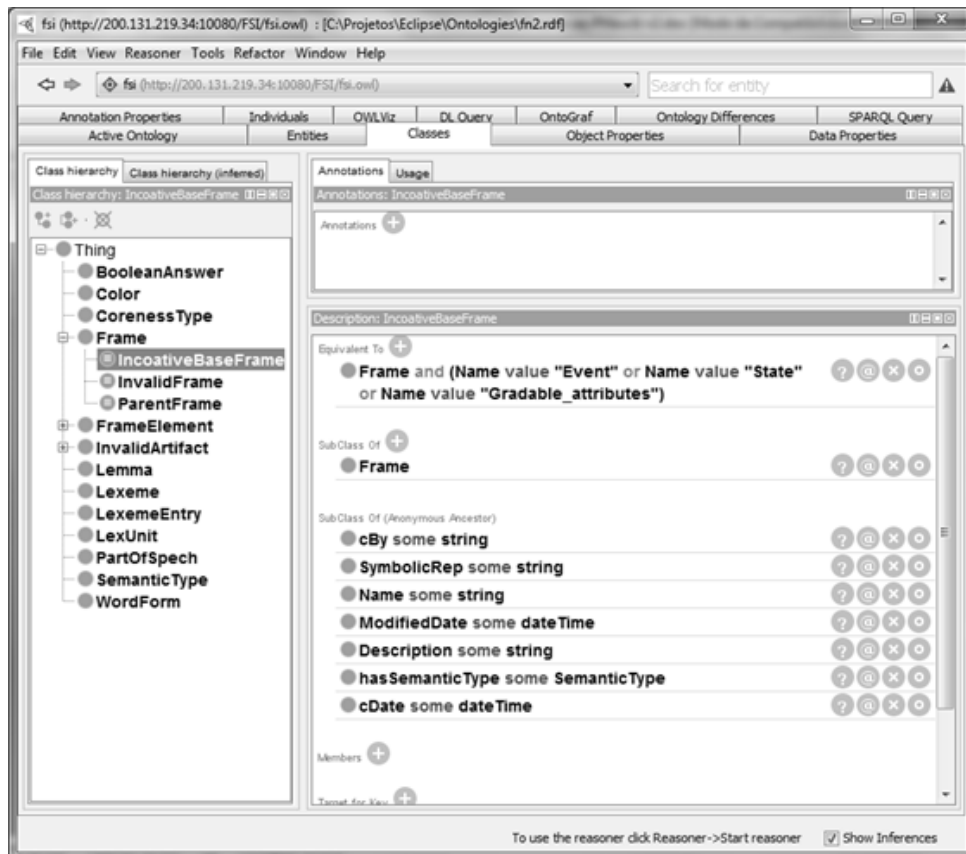


Figura 3.6: Definição da classe *IncoativeBaseFrame*.

A relação *SubFrame*, que ocorre quando um frame mais complexo é composto por outros frames, a relação *Precedes*, que denota sequencia cronológica entre os frames, e a relação *Using*, que ocorre quando um frame evoca outro, foram descritas na ontologia como transitivas, uma vez que o repetição delas em uma cadeia de frames equivale a uma relação entre as extremidades desta cadeia.

Já a relação *See\_Also* possui uma semântica simples, já que é apenas uma referencia

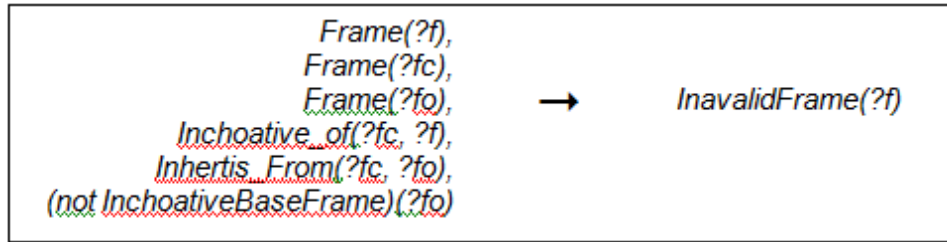


Figura 3.7: Regra verificação do frame incoativo.

para consulta que busca permitir um melhor entendimento do frame.

### 3.2.2 RELAÇÕES ENTRE ELEMENTOS DE FRAME

Assim como as relações entre frames, (RUPPENHOFER et al., 2006) e (LEENOI et al., 2011) também descrevem relações possíveis entre EFs de um mesmo frame. Para estas relações também foram realizadas descrições semânticas na ontologia, de maneira a respeitar a definição apresentada por estes autores. Com objetivo de permitir identificar os frames que desrespeitam estas ou outras restrições definidas por regras SWRL, foi criada uma classe denominada *InvalidFrameElement*.

A primeira relação entre EF declarada é a relação *Excludes* que define que a presença de um EF em uma cena não pode ocorrer na presença de outro EF. Assim, esta relação não pode ser definida entre dois EF nucleares, uma vez que para existência da cena todos os nucleares devem participar. Sendo assim, esta relação é definida como transitiva.

A relação *Requires* define uma relação de dependência entre dois EFs de um mesmo frame, onde a existência de um EF em uma cena obriga também a participação de outro EF. E a última relação é a *CoreSet* que é a relação que pode ocorrer entre os alguns EF nucleares de um mesmo frame.

Como todas estas relações são definidas entre EF em um mesmo frame, foi criada uma SWRL para cada uma das regras, primeiramente validando se os dois indivíduos participantes destas relações são participantes de um mesmo frame (Figura 3.8).

Para as relações *Requires* e *Excludes* foi criada uma regra para verificar se dois EF participam das duas relações ao mesmo tempo, uma vez que não é permitido que um EF dependa de outro que o exclui e vice-versa (Figura 3.9).

Na figura 3.10 podemos ver um resumo de todas as regras SWRL criadas para estas relações entre frames e entre EF.

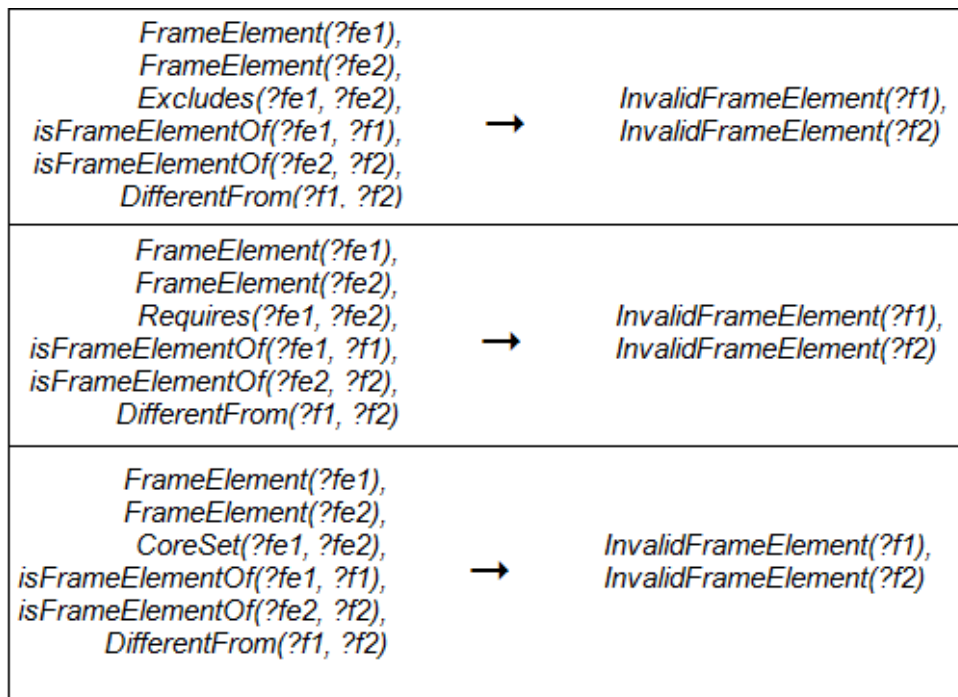


Figura 3.8: Regras para verificar se EFs que participam de uma das relações entre EF pertencem ao mesmo Frame.

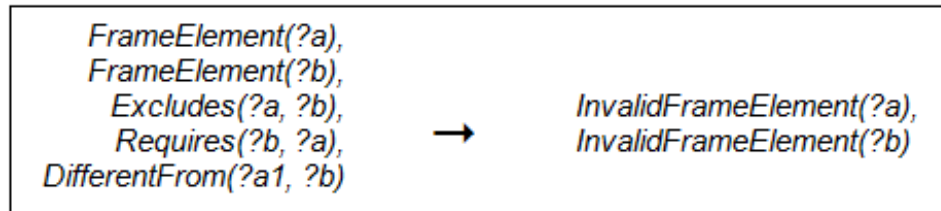


Figura 3.9: Regra para verificação da não exclusão e dependência mútua entre EFs.

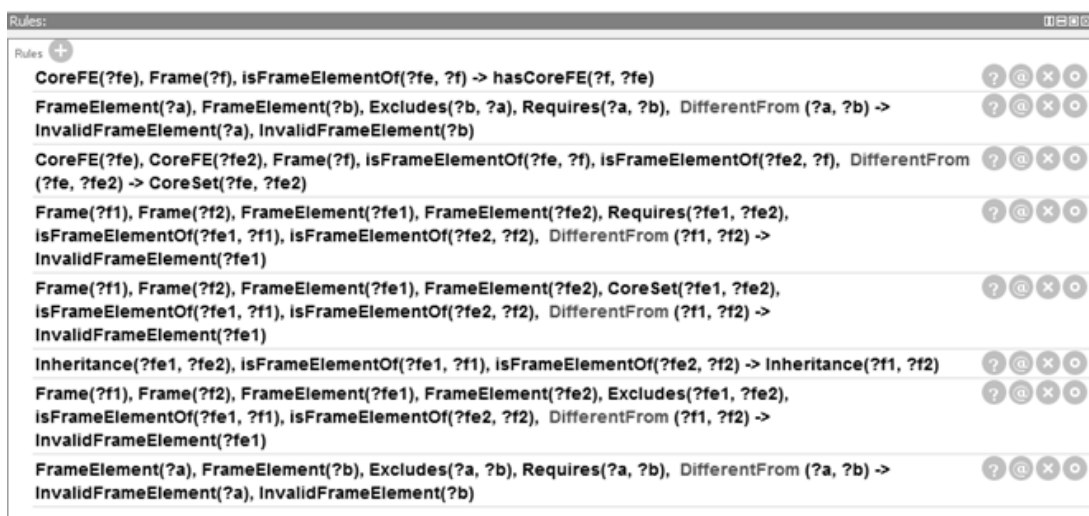


Figura 3.10: Regras definidas em SWRL.

Como resultado deste processo obtemos uma ontologia para representar os dados do projeto *FrameNet* de maneira concisa e semanticamente correta, que se difere das principais ontologias relacionadas ao *FrameNet* que foram encontradas na literatura [(LEENOI et al., 2011); (NUZZOLESE et al., 2011); (SCHEFFCZYK et al., 2006)], principalmente no que diz respeito ao detalhamento semântico das relações entre frames e entre elementos de frames. Neste contexto, considera-se que nas ontologias apresentadas por (NUZZOLESE et al., 2011) e (SCHEFFCZYK et al., 2006) estas relações não são expressas ou são expressas apenas como parte do vocabulário, sem restrições ou regras para validação das mesmas. Apenas em (LEENOI et al., 2011) as relações entre frames são discutidas. Entretanto o autor não deixa claro se elas foram tratadas na ontologia ou se foram apenas descritas para caráter informativo. Além disso, o autor não disponibiliza meio para obter ou reproduzir a ontologia.

Uma visão parcial desta ontologia com as principais classes e as relações entre elas é apresentada na Figura 3.1. A ontologia completa pode ser obtida em <http://goo.gl/qNZ8aq>.

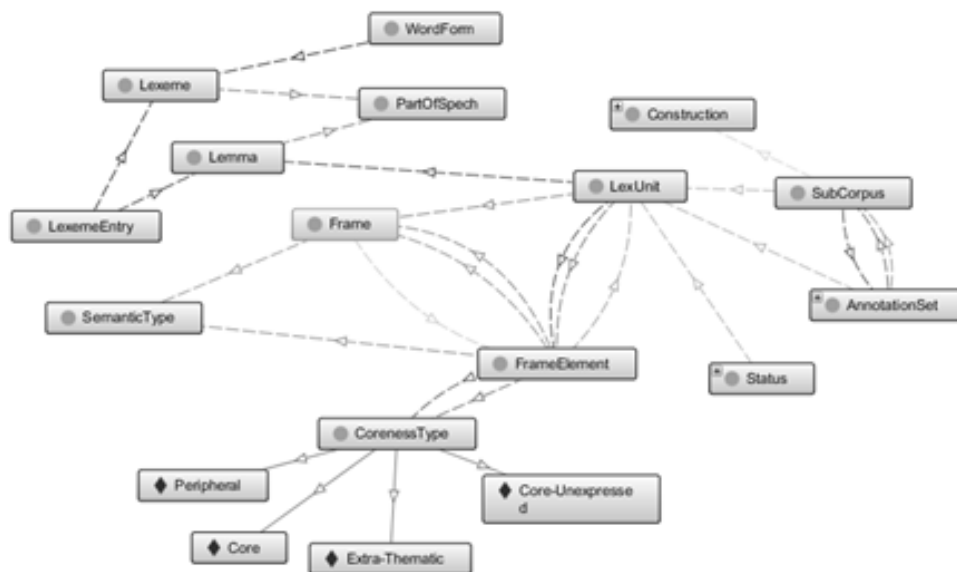


Figura 3.11: Resumo com as principais Classes da ontologia de ONTO-FRAME-BR.

### 3.2.2.1 Ontologia de Anotações

A ontologia ONTO-ANNOTATION-BR foi desenvolvida com o objetivo de complementar a ontologia ONTO-FRAME-BR de maneira a abranger as anotações de sentenças relativas a primeira camada, ou seja, definindo a participação de fragmentos como EFs e identifi-



cação do Frame, sendo a partir desta ontologia representadas as anotações das sentenças realizadas no projeto.

Esta ontologia é composta por apenas duas classes principais, uma que representa a sentença a ser anotada, e uma classe que representa cada fragmento anotado em uma sentença. A partir desta ontologia cada fragmento pode se relacionar a um EF ou Lexema presente na Ontologia de Metadados do FrameNet.

A fim de validar a semântica das anotações realizadas, foi necessário a criação de duas regras SWRL e uma classe, denominada *InvalidAnnotatedSentence*, para classificação das sentenças com anotações inválidas. Para esta ontologia foram criadas apenas duas regras SWRL para validação das anotações realizadas.

A primeira regra verifica se existe mais de um fragmento anotado relacionado a uma mesma sentença que referencia um lexema, tendo em vista que em uma sentença só é possível a identificação de um lexema, a situação definida nesta regra classificará a sentença na classe *InvalidAnnotatedSentence* (Figura 3.12).



Figura 3.12: Regra de validação de múltiplos lexemas anotados em uma mesma sentença.

A segunda regra verifica se em uma sentença anotada existe algum fragmento que referencia um EF que pertence a um Frame diferente do Frame associado ao o Lexema desta anotação, classificando também a sentença na classe *InvalidAnnotatedSentence* (Figura 3.13).

Com esta ontologia criou-se uma maneira de validar a semântica das anotações realizadas utilizando as duas ontologias construídas neste trabalho. Além disso, a partir dos fragmentos anotados em sentenças identificados nesta ontologia, torna-se possível a associação com recursos externos no padrão de Dados ligados (conforme será apresentado na seção 3.3.1).

Uma visão parcial desta ontologia com as principais classes e as relações entre elas é apresentada na Figura 3.14. A ontologia completa pode ser obtida em <http://goo.gl/qNZ8aq>.

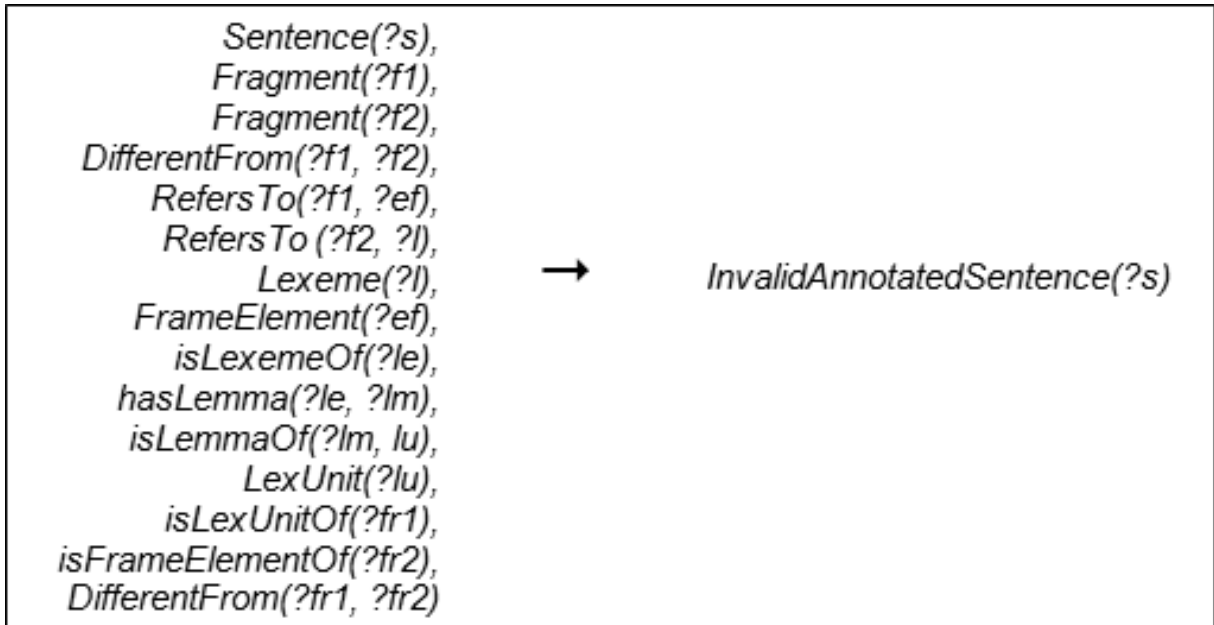


Figura 3.13: Regra de validação de múltiplos frames em uma mesma anotação.

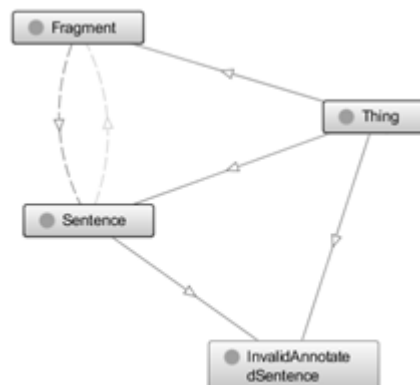


Figura 3.14: Principais Classes da Ontologia de ONTO-ANNOTATION-BR.

### 3.3 ARQUITETURA

Conforme apresentado anteriormente, o FSI é uma infraestrutura baseada no padrão *SOA*, que utiliza conceitos da Web Semântica aplicados aos dados do Projeto FrameNet. Na Figura 3.15 temos a visão geral da infraestrutura com seus principais componentes.

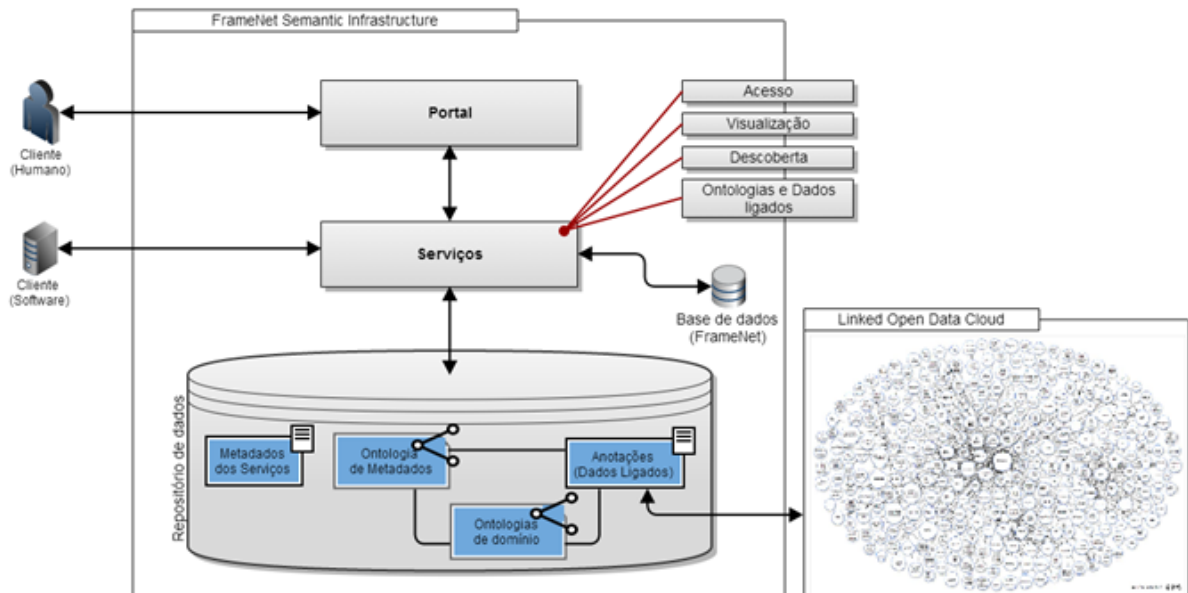


Figura 3.15: Visão geral da arquitetura da FSI.

A infraestrutura FSI pode ser dividida em três camadas: i) a camada de dados, onde ficam armazenados os dados tratados pela infraestrutura, como ontologias, recursos de dados ligados, anotações dos serviços e os dados relacionados ao controle de acesso; ii) a camada de serviços, cujo propósito é prover uma interface de interação com outras ferramentas de software independente da linguagem de programação utilizada; e por fim iii) o portal, onde é mantida uma interface amigável para utilização dos serviços da infraestrutura.

#### 3.3.1 CAMADA DE DADOS

A camada de dados da FSI é a camada onde são armazenados as ontologias de domínio, a ontologia de metadados da FrameNet e de anotações, os dados dos recursos de dados ligados, os metadados que descrevem os serviços disponíveis na infraestrutura, e também a base relacional que armazena as informações de permissões de acesso à infraestrutura.

Conforme já dito, na FSI é utilizado um conjunto de ontologias que visam prover uma estrutura formal e semântica para os dados que serão armazenados pela infraestrutura.

Uma das ontologias envolvidas neste conjunto, é a ontologia ONTO-FRAME-BR, que descreve a estrutura em que os dados do Projeto *FrameNet* são organizados, ou seja a forma no qual os Frames são definidos (Frames, Elementos de Frame, Unidades Lexicais, etc.) e a semântica das relações entre eles, além da ontologia associada a esta, que é a ontologia ONTO-ANNOTATION-BR, que define a estrutura das anotações.

As demais ontologias do conjunto são relacionadas aos domínios aos quais os frames podem se relacionar. Neste sentido, estas ontologias de domínio permitem que se defina, de acordo com o domínio, uma restrição semântica aos elementos do frame de maneira que torne possível avaliar se as anotações respeitam a semântica do frame que é referenciado. Por exemplo, na Figura 3.16, considerando o domínio de futebol, temos a representação do Frame "Jogar", em que seus EF "Seleções", "Seleção1" e "Seleção2" estão relacionados com o tipo ontológico "Seleção", que por sua vez está definido na ontologia de domínio de Futebol, definindo também uma restrição onde as instancias deste EF devem também ser instancias do termo "Seleção" descrito na ontologia. Da mesma maneira ocorre com o EF "Sede". Entretanto, neste caso, temos que ambos, "Cidade" e "País" são tipos ontológicos que podem ser aceitos para uma instancia deste EF<sup>10</sup>.

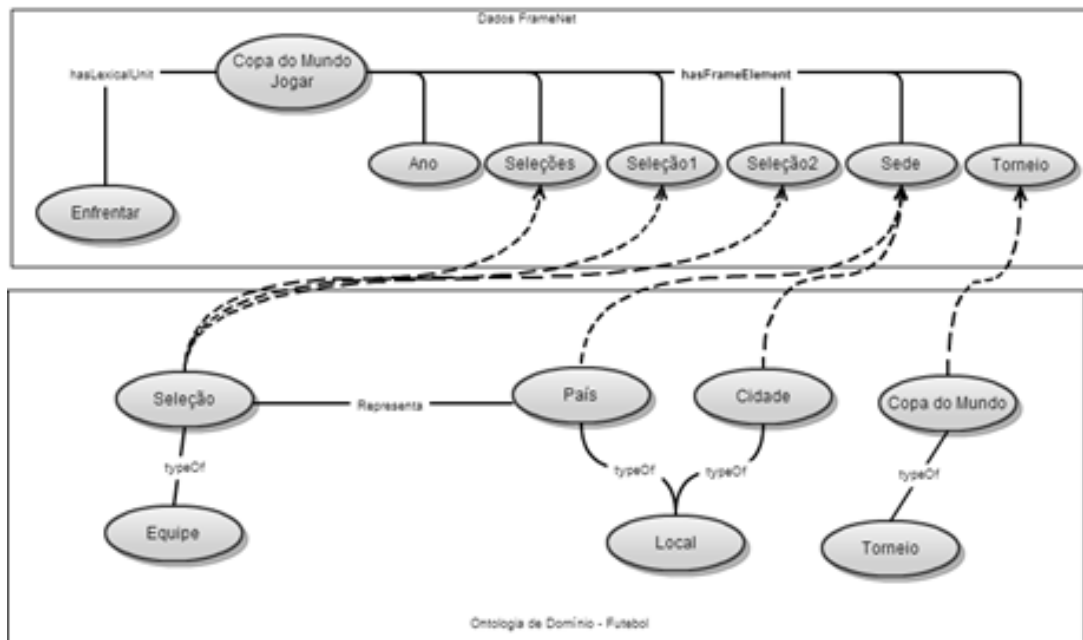


Figura 3.16: Uso de Ontologias de domínio para restrição de tipo semântico para os Elementos de Frame.

Para representação das anotações foi adotado o padrão de dados ligados (BERNERS-LEE, 2009) para os fragmentos que instanciam EFs. Sendo assim, cada fragmento deste

<sup>10</sup>O uso e importância das ontologias do domínio serão detalhados no capítulo 5 desta dissertação

estará ligado a, no mínimo, um outro termo, preferencialmente de uma base externa, que nos permita obter mais informações a partir da navegação entre estas ligações. Na Figura 3.17 é apresentado um exemplo desta abordagem para a anotação "A Seleção Brasileira enfrenta o EUA em Toronto.", onde partes desta anotação, como "Seleção Brasileira" e "O EUA" estão ligados por uma relação de equivalência com recursos de um dataset externo de dados ligados que representam estas seleções, e a partir destes recursos podemos obter novas informações considerando a rede semântica que é formada por este padrão. Como exemplo, podemos obter o nome do técnico ou até mesmo de jogadores de uma destas seleções aproveitando as ligações do recurso do dataset externo com outros recursos deste ou de outros datasets.

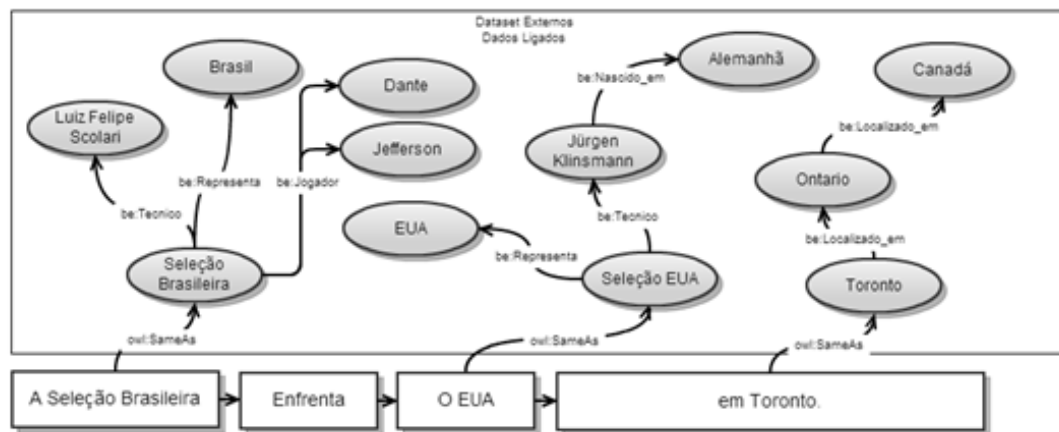


Figura 3.17: Representação de fragmentos de anotações sob o padrão de dados ligados.

### 3.3.2 CAMADA DE SERVIÇOS

Conforme já mencionado, as funcionalidades da FSI são disponibilizadas por meio de serviços, baseados no padrão SOA. Sendo assim, foram desenvolvidos quatro serviços com o objetivo de prover uma interface de comunicação para que ferramentas externas possam obter informações mantidas nos repositórios da infraestrutura.

Além da descrição WSDL de cada serviço, também foram associadas anotações semânticas aos principais métodos da infraestrutura, trazendo a possibilidade de descoberta, execução e composição de forma automatizada, e facilitando também a interoperabilidade da aplicação uma vez que sua interface está anotada semanticamente e, portanto preparada para os recursos da Web Semântica (MCILRAITH et al., 2001).

A seguir descreveremos cada um dos serviços com seus métodos e suas anotações

semânticas, quando existentes.

### 3.3.2.1 Serviço de Acesso

O serviço de acesso é o responsável por controlar o acesso das ferramentas às funcionalidades presentes na infraestrutura, evitando assim que modificações nos dados contidos nas ontologias sejam efetuadas por clientes não autorizados. Sendo assim, para interagir com alguns serviços disponibilizados pela FSI as ferramentas deverão solicitar permissão e validar a sua identidade durante o consumo dos serviços.

Este serviço disponibiliza apenas um método (Tabela 3.1), o *validateUser* que é responsável pela validação das credenciais de acesso aos serviços, retornando um código único, denominado *token*, que deverá ser utilizado nas próximas requisições a outros serviços que requeiram um *token* de usuário. Após 20 minutos sem a utilização de um token gerado para um usuário, ele é descartado pelo sistema sendo necessário que o usuário gere um novo token a partir deste mesmo método. Para utilização do método *validateUser* a aplicação deverá enviar ao serviço sua identificação e a senha de sua identificação<sup>11</sup> do mesmo em uma requisição, após validar o usuário os sistema retornará uma cadeia de caracteres com o valor do *token*.

Tabela 3.1: Lista de métodos presentes no serviço de acesso.

Método	Parâmetro	Tipo	Anotação	Descrição
<i>validateUser</i>	identifier	Input	xsd:string	Identificador do usuário
	password	Input	xsd:string	Senha
	token	Output	xsd:string	<i>Token</i> de acesso

### 3.3.2.2 Serviço de Visualização

O Serviço de visualização é responsável por prover formatos diversificados para os dados que podem ser obtidos da ontologia, incluindo a exibição de frames e sua estrutura no modelo padrão do portal do projeto FrameNet. Estes dados são obtidos a partir de uma resposta em XHTML, contando com dois métodos (Tabela 3.2).

O primeiro método, *formatAsRDF* é responsável por transpor uma entrada no formato XML padrão de respostas do serviço de Ontologias e Dados ligados (seção 3.2.3.3) para um formato RDF baseado em XML. Para que uma dada ferramenta possa utilizar este

<sup>11</sup>A definição de um identificador e senha é definida manualmente pelo administrador da infraestrutura.

serviço, a mesma deverá enviar a resposta obtida em XML. Caso seja possível transpor para RDF o método retornará estes dados no formato RDF.

Já o segundo método, *formatAsFN* é responsável por transpor a entrada proveniente do serviço de Ontologias e Dados Ligados num formato XHTML facilitando o entendimento dos dados. Esta transposição somente é possível quando a entrada representa alguns elementos específicos da FrameNet, como Frame, entrada lexical ou conjunto de anotações de uma Unidade Lexical. Para consumir este método a ferramenta deverá enviar o XML de um Frame ou Unidade Lexical, e o método retornará um XHTML que a ferramenta poderá utilizar para exibir de maneira amigável ao usuário.

Tabela 3.2: Lista de métodos presentes no serviço de acesso.

Método	Parâmetro	Tipo	Anotação	Descrição
<i>formatAsRDF</i>	XMLInput	Input	xsd:string	XML de entrada
	Result	Output	xsd:string	RDF/XML
<i>formatAsFN</i>	XMLInput	Input	xsd:string	XML de entrada
	Result	Output	xsd:string	XHTML

### 3.3.2.3 Serviço de Ontologias e Dados Ligados

O Serviço de Ontologias e Dados Ligados é o responsável por prover uma interface para obtenção dos dados contidos nas ontologias e também permitir que seja modificado seu conteúdo. Este serviço é o mais importante da infraestrutura contando com diversos métodos para obtenção de elementos da Framenet como Frames, UL, Sentenças e Anotações. A fim de facilitar a apresentação dos métodos deste serviço, agrupamos os mesmos em conjuntos.

O primeiro conjunto é composto pelos métodos que permitem modificar os dados da ontologia são eles: *createIndividual*, que cria ou atribui um indivíduo como instância de uma classe da Ontologia, sendo necessário que a ferramenta invoque este método informando qual a classe e a URI do indivíduo que será instanciado na classe; *setDataProperty*, que permite associar um valor a uma propriedade de dados de um recurso, sendo necessário informar qual a URI do indivíduo que possui a propriedade, qual a URI da propriedade de dados e, por fim, o valor que será atribuído a esta propriedade; *setObjectProperty*, que permite associar uma relação entre dois recursos, devendo ser informado qual a URI do indivíduo que possui a propriedade, qual a URI da propriedade de dados e, por fim, a URI do indivíduo que é o alvo desta relação; *addGraph*, que permite incluir na ontologia

um conjunto de declarações em RDF, sendo necessário informar no formato XML as tríades (sujeito-predicado-objeto) que deverão ser incluídas da ontologia, permitindo assim a inclusão de alterações em lote na ontologia ; *removeGraph*, que permite remover um conjunto de declarações em RDF, sendo necessário informar no formato XML as tríades (sujeito-predicado-objeto) que deverão ser removidas da ontologia; e *setLDRelation*, utilizada para atribuir a propriedade *RefersTo* entre um recurso presente na ontologia com um recurso qualquer da nuvem de dados ligados. Para uso deste último método é necessário que a ferramenta informe qual a URI do indivíduo sujeito e a URI do indivíduo objeto da relação.

O segundo conjunto engloba métodos voltados para obtenção dos dados presentes nas ontologias, são eles: *getIndividual*, que permite obter todos os dados e relações de um determinado indivíduo a partir de sua URI. Para a invocação deste método, deverá ser informado apenas a URI do indivíduo que se deseja obter as informações, sendo retornado à ferramenta um XML que descreve o indivíduo e as relações deste com outros indivíduos ; *getFrame*, que permite obter todos os dados e relacionamentos de um determinado frame a partir de sua URI, sendo estes retornados em formas de tríades no formato XML; *getFramebyName*, permite obter todos os dados e relacionamentos de um determinado frame a partir de seu nome, sendo estes dados retornados em formas de tríades no formato XML; *getLexUnit*, permite a obtenção de dados de uma Unidade Lexical, onde a invocação do método é feita fornecendo a URI da UL desejada e sendo retornado as informações na formas de tríades no formato XML; *getAnnotations*, que permite obter a anotação de uma determinada sentença, sendo fornecida a URI da sentença e sendo retornado os termos anotados; *getAnnotationsByFrame*, permite obter todas as anotações realizadas em sentenças relacionadas a um determinado frame a partir do URI deste frame; *getLDRelation*, retorna ao usuário todas as URIs de recursos associados a um determinado recurso das ontologias a partir de sua URI ; e *executeSPARQL*, que permite ao usuário executar uma consulta em formato SPARQL nas ontologias presentes na infraestrutura, sendo necessário informar durante a invocação do método a consulta em SPARQL, o nome da ontologia a ser consultada e se vai ser executada a inferência na consulta ou não, os dados retornados pelo método serão em formato RDF.

E por fim temos os métodos *verifyIntegrity* que verifica se um determinado recurso está associado a uma das classes que foram criadas como inválidas ou se existe alguma



inconsistência relacionada a este recurso durante a inferência, e o *explainInference* que verifica se o axioma informado pela ferramenta que consome este método foi inferido pela ontologia e caso positivo fornece a explicação lógica desta inferência em XML.

Na Tabela 3.3 podemos verificar um resumo dos parâmetros de entrada e saída de cada método deste serviço, e suas anotações semânticas.

### 3.3.2.4 Serviço de Descoberta

O serviço de descoberta é responsável por prover informações sobre os serviços da infraestrutura e seus métodos, incluindo suas anotações semânticas. Este serviço conta com dois métodos, sendo o primeiro *listServiceMethods* responsável por prover uma lista dos serviços e métodos disponíveis, e o segundo *showMethodData* responsável por prover os dados sobre a estrutura do método e suas anotações semânticas.

### 3.3.2.5 Uso da camada de serviços

Com o intuito de facilitar o entendimento de uso da infraestrutura FSI, nesta seção apresentamos um exemplo de como a interface provida pela camada de serviços (seção 3.3.2) pode ser utilizada em atividades de PLN por ferramentas externas.

Para esta finalidade será considerada a ferramenta CADMOS (*Character?centred Annotation of Dramatic Media ObjectS*) (CATALDI et al., 2011), que é um framework que tem como objetivo apoiar a anotação de recursos multimídia com base no uso de ontologias e identificação de cenas e desambiguação a partir de bases de recursos lexicais, como a *FrameNet*, *Wordnet*, entre outros. Uma visão geral da arquitetura do *framework* é apresentada na Figura 3.18.

Muitas das vezes durante a descrição da anotação da cena podem ser observados termos ou expressões que possuem sentidos ambíguos e podem ser interpretados de diferentes maneiras de acordo com o contexto ao qual se aplicam. Para este problema a CADMOS prevê um processo de desambiguação de termos que faz uso de diversos recursos lexicais, entre eles a *FrameNet* e a *WordNet* (Figura 3.19).

Conforme podemos observar nas Figuras 3.18 e 3.19, as anotações geradas durante o processo de anotação dos recursos de mídia são armazenados em tríades RDF e associadas a ontologias para delimitação de domínio. E, além disso, a identificação da cena é apoiada pelo *WordNet* e o *FrameNet*. Sendo assim, a FSI pode ser utilizada, como apresentado na

Tabela 3.3: Lista de métodos presentes no serviço de acesso.

Método	Parâmetro	Tipo	Anotação	Descrição
<i>createIndividual</i>	URI	Input	owl:Thing	URI do indivíduo
	classURI	Input	owl:Class	URI da classe
	token	Input	xsd:string	<i>Token</i> de acesso
	ontologyName	Input	xsd:string	Nome da Ontologia
<i>setDataProperty</i>	URLDomain	Input	owl:Thing	URI do Indivíduo
	propertyURI	Input	owl:DatatypeProperty	URI da propriedade
	data	Input	xsd:string	Valor da propriedade
	<i>token</i>	Input	xsd:string	<i>Token</i> de acesso
	ontologyName	Input	xsd:string	Nome da Ontologia
<i>setObjectProperty</i>	URLDomain	Input	owl:Thing	URI do Ind. domínio
	PropertyURI	Input	owl:ObjectProperty	URI da propriedade
	URLrange	Input	owl:Thing	URI do Ind. Imagem
	Token	Input	xsd:string	<i>Token</i> de acesso
	ontologyName	Input	xsd:string	Nome da Ontologia
<i>addGraph</i>	XML	Input	xsd:string	XML do Grafo
	ontologyName	Input	xsd:string	Nome da Ontologia
	token	Input	xsd:string	<i>Token</i> de acesso
<i>removeGraph</i>	XML	Input	xsd:string	XML do Grafo
	ontologyName	Input	xsd:string	Nome da Ontologia
	Token	Input	xsd:string	<i>Token</i> de acesso
<i>setLDRelation</i>	URLResource1	Input	owl:Thing	URI do Recurso local
	URLResource2	Input	owl:Thing	URI do Recurso externo
	token	Input	xsd:string	<i>Token</i> de acesso
<i>getIndividual</i>	URI	Input	owl:Thing	URI do indivíduo
	ontologyName	Input	xsd:string	Nome da Ontologia
	withInference	Input	xsd:boolean	Usar reasoner?
	Result	Output	owl:Thing	Indivíduo obtido
<i>getFrame</i>	URI	Input	owl:Thing	URI do Frame
	withInference	Input	xsd:boolean	Usar reasoner?
	Result	Output	fsi:Frame	Frame obtido
<i>getFramebyName</i>	Name	Input	fsi:Name	Nome do Frame
	withInference	Input	xsd:boolean	Usar reasoner?
	Result	Output	fsi:Frame	Frame obtido
<i>getLexUnit</i>	Name	Input	owl:Thing	URI da UL
	withInference	Input	xsd:boolean	Usar reasoner?
	Result	Output	fsi:LexUnit	UL obtida
<i>getAnnotation</i>	URI	Input	stc:Sentence	URI da Sentença
	Result	Output	owl:Thing	Anotação da sentença
<i>getAnnotations byFrame</i>	URI	Input	fsi:Frame	URI do Frame
	Result	Output	owl:Thing	Anotações e sentenças obtidas
<i>getLDRelation</i>	URI	Input	owl:Thing	URI do Recurso
	Result	Output	owl:Thing	Recursos obtidos
<i>executeSPARQL</i>	Query	Input	xsd:string	Consulta SPARQL
	ontologyName	Input	xsd:string	Nome da ontologia
	withInferences	Input	xsd:boolean	Usar reasoner?
	Result	Output	xsd:string	Resultado da consulta
<i>verifyIntegrity</i>	URI	Input	owl:Thing	URI do recurso
	Result	Output	xsd:boolean	Recurso valido?
<i>explainInference</i>	URI	Input	owl:Thing	URI do sujeito
	Property	Input	owl:Property	URI do predicado
	URLObject	Input	owl:Thing	URI do objeto
	Result	Output	xsd:string	Explicação da inferência

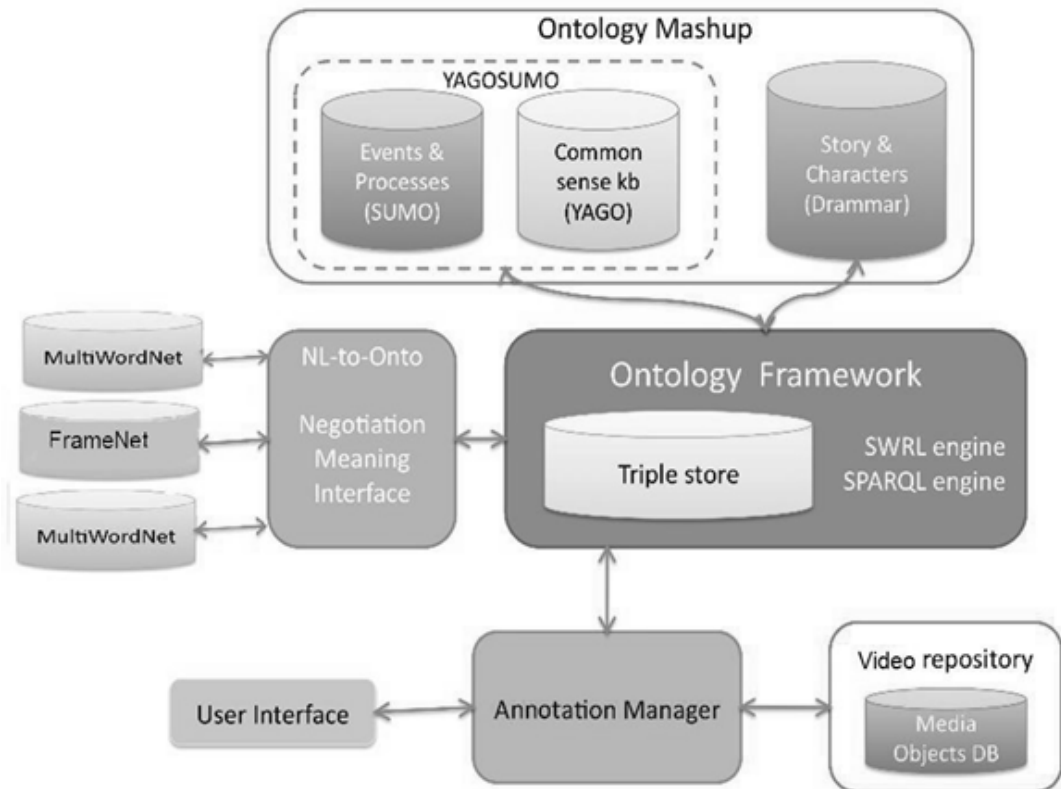


Figura 3.18: Visão Geral do Framework CADMOS (CATALDI et al., 2011).

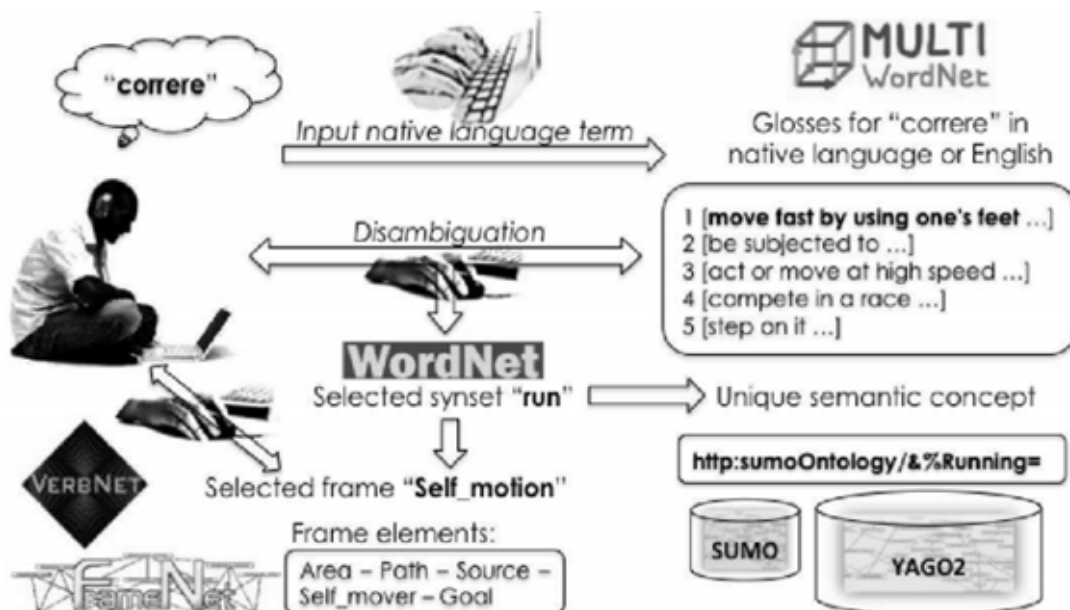


Figura 3.19: Processo de desambiguação de termos na CADMOS (CATALDI et al., 2011).

figura 3.20, no processo de desambiguação de frames na CADMOS. Uma das vantagens de se utilizar a FSI neste caso, é o aproveitamento das informações semânticas que podem ser obtidas ao se obter os EFs de um frame, uma vez que estes podem estar atribuídos a elementos das ontologias de domínio, tornando possível identificar melhor o contexto ao qual o frame se aplica.

Além disso, também seria possível aproveitar os dados de anotações da *FrameNet* nos frames armazenados da FSI, que estejam associados a recursos externos de dados ligados, de maneira a enriquecer também a anotações de mídia. Por exemplo, atribuindo a um elemento anotado da sentença da *FrameNet* a um elemento de anotação da CADMOS. O que torna possível a obtenção de informações acerca de um recurso.

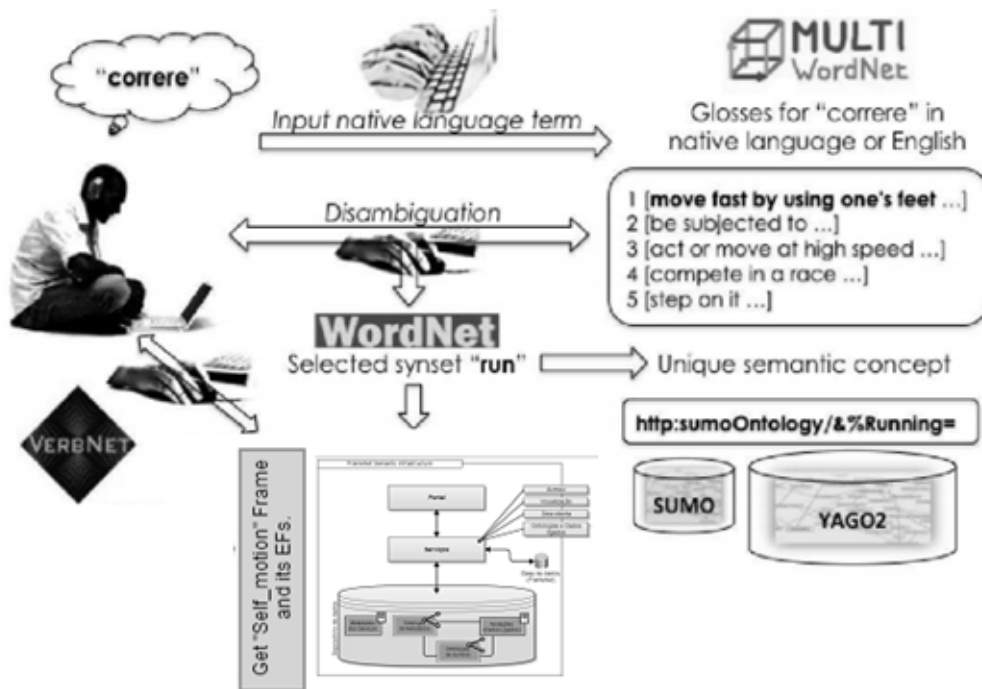


Figura 3.20: Processo de desambiguação de termos na CADMOS com inclusão da obtenção dos dados de Frames via FSI.

A Figura 3.21 detalha o fluxo de interação da ferramenta CADMOS com os métodos do serviço de ontologias e dados ligados da FSI para obtenção dos dados de Frames e dos seus EFs, no contexto do processo de desambiguação.

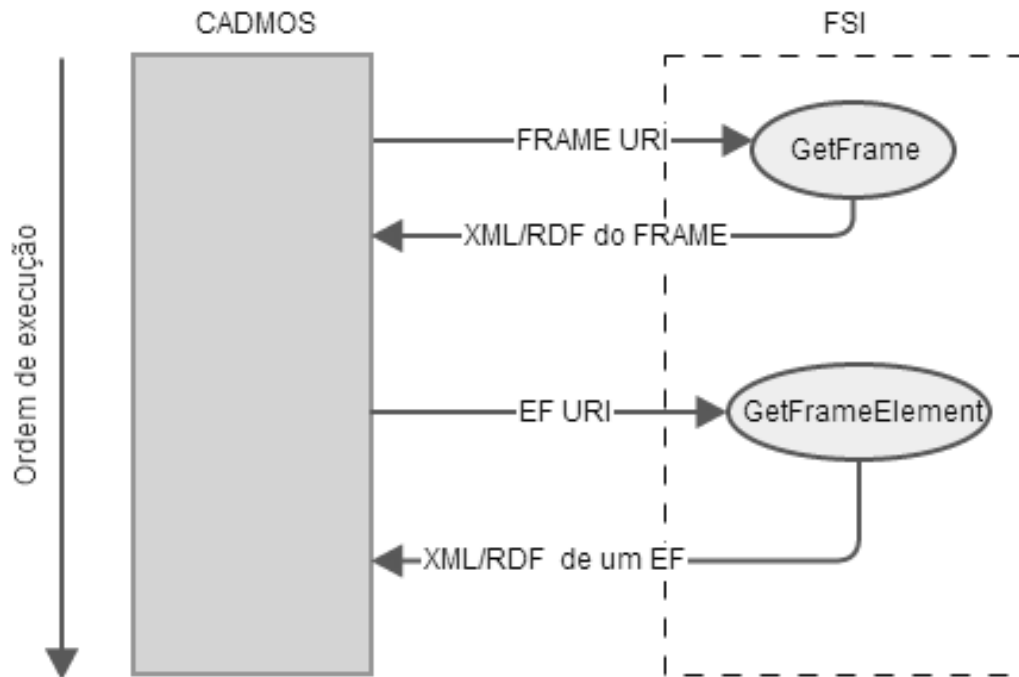


Figura 3.21: Fluxo de execução de serviços para a obtenção dos dados do Frame e seus EFs pela FSI.

### 3.3.3 PORTAL

O Portal da infraestrutura foi desenvolvido com o objetivo de prover uma interface que facilite a utilização das informações diretamente pelos usuários, sem a necessidade de utilizar ferramentas que façam papel de intermediário entre os serviços e o usuário (Figura 3.22).

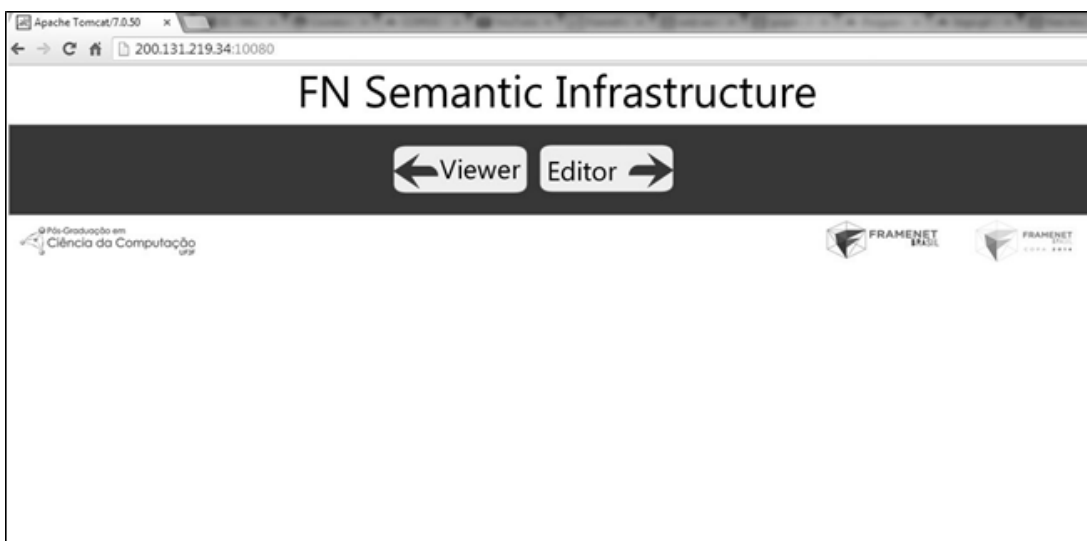


Figura 3.22: Tela Inicial do Portal da FSI.

Foram disponibilizados no portal duas ferramentas, a primeira delas denominada FnSEditor que permite a edição dos dados dos frames do projeto, e a realização de anotação de sentenças. A segunda ferramenta é uma ferramenta que simula a interface atual do *FrameNet* com o objetivo de permitir a consulta aos dados dos Frames e suas anotações.

### 3.3.3.1 FrameNet Semantic Editor (FnSEditor)

O FnSEditor (FrameNet Semantic Editor) é uma ferramenta que foi especificada com o intuito de prover ao usuário uma interface amigável para manutenção dos dados relativos aos frames, permitindo a criação de novos frames, definição de relacionamento entre eles, realização de anotações de sentenças, cadastro de recursos externos para relação com fragmentos de sentenças, entre outros.

As funcionalidades desta ferramenta são exibidas no menu principal, conforme apresentado na figura 3.23. São elas: "Frames" onde é realizada a documentação de novos frames e/ou a modificações dos atuais; "Sentences" onde é realizada a inclusão de novas sentenças e também realizada a anotação das sentenças existentes; e o "*Linked Data*" onde são relacionados recursos de dados ligados externos a classes das ontologias de domínio presentes na infraestrutura, a fim de permitir a associação aos fragmentos de sentença que representam EF durante a anotação. Estas funcionalidades são detalhadas abaixo. 3.23).

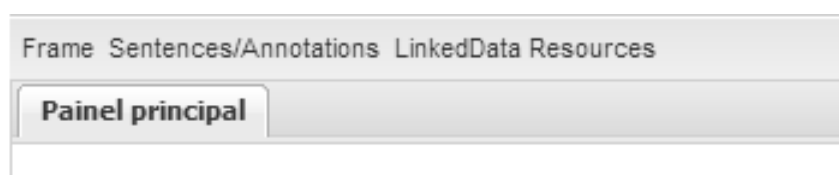


Figura 3.23: Menu Principal da ferramenta de documentação de frames.

A primeira funcionalidade presente na ferramenta é um conjunto de telas que permite a documentação de novos frames e a alteração dos frames existentes, de maneira a incluí-los na ontologia de metadados da *FrameNet*. Para a documentação dos frames, o portal disponibiliza telas de cadastro onde são preenchidos os dados básicos do frame, como nome do frame, descrição e tipo semântico, assim como os elementos que participam em sua estrutura, como seus elementos de frame, unidades lexicais e as relações com outros Frames 3.24.

Na Figura 3.25 é apresentada a tela de cadastro de um EF em um Frame, onde são

Figura 3.24: Tela para documentação de frame.

necessárias as informações básicas de um EF como título, tipo de participação no frame (*Core*, *Core-Unexpressed*, *Extra-Thematic*, *Peripheal*) e a cor selecionada para representação no fragmento. Além disso, temos a aba de relacionamentos, onde é possível selecionar o tipo de relacionamento entre este EF e os demais EFs deste frame, como exclusão e dependência.

Ao confirmar o cadastro de um EF, a ontologia que descreve o formato dos dados e regras das relações entre os elementos da FrameNet, cadastra e verifica se o EF foi inferido como participante da classe que o define como inválido (*InvalidFrameElement*), ou seja, se o EF desrespeita alguma das regras em SWRL definidas na ontologia. Caso o EF em questão seja classificado nesta classe, a ferramenta apresentará uma tela com a sequência de sentenças em formato RDF que descreve a lógica para chegar a conclusão desta classificação (Figura 3.26).

Outra funcionalidade da ferramenta é a possibilidade de associar uma ou mais classes das ontologias de domínio presentes na infraestrutura a um EF, (Figura 3.27) com o objetivo de prover durante a anotação, uma melhor identificação semântica do EF e permitindo associar a ele um recurso da web que esteja também identificado como um recurso daquele tipo semântico. Esta funcionalidade será detalhada mais a frente.

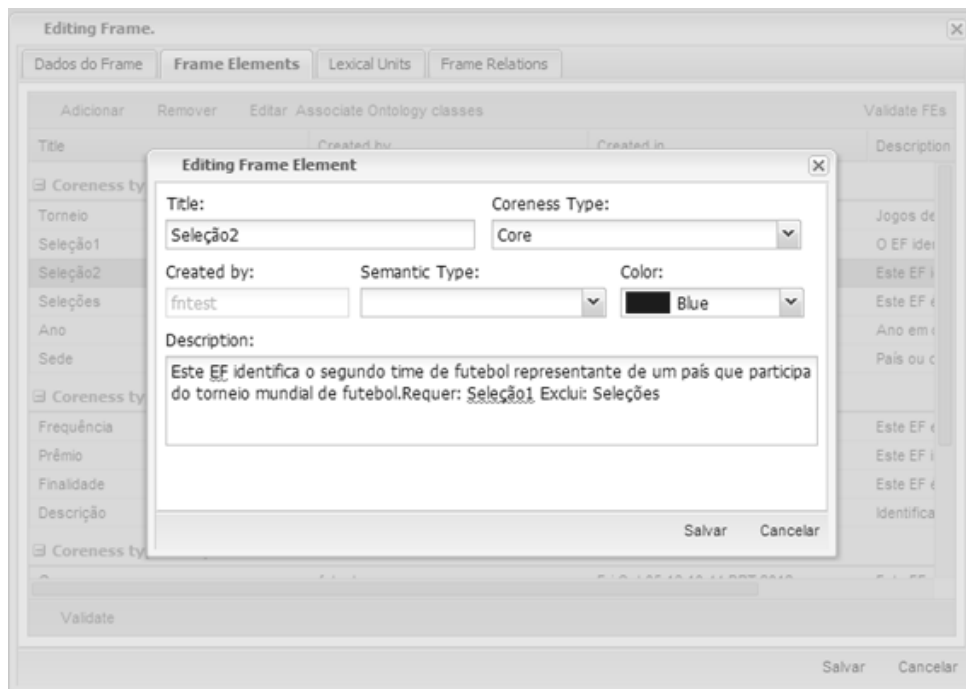


Figura 3.25: Detalhe do cadastro de Elementos de frame.

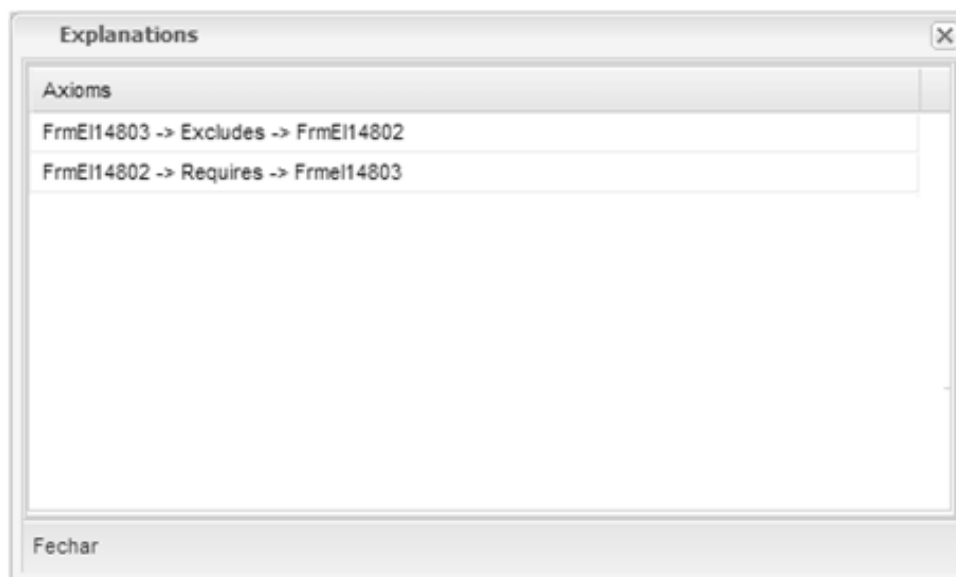


Figura 3.26: Tela de justificativa da inferência realizada.



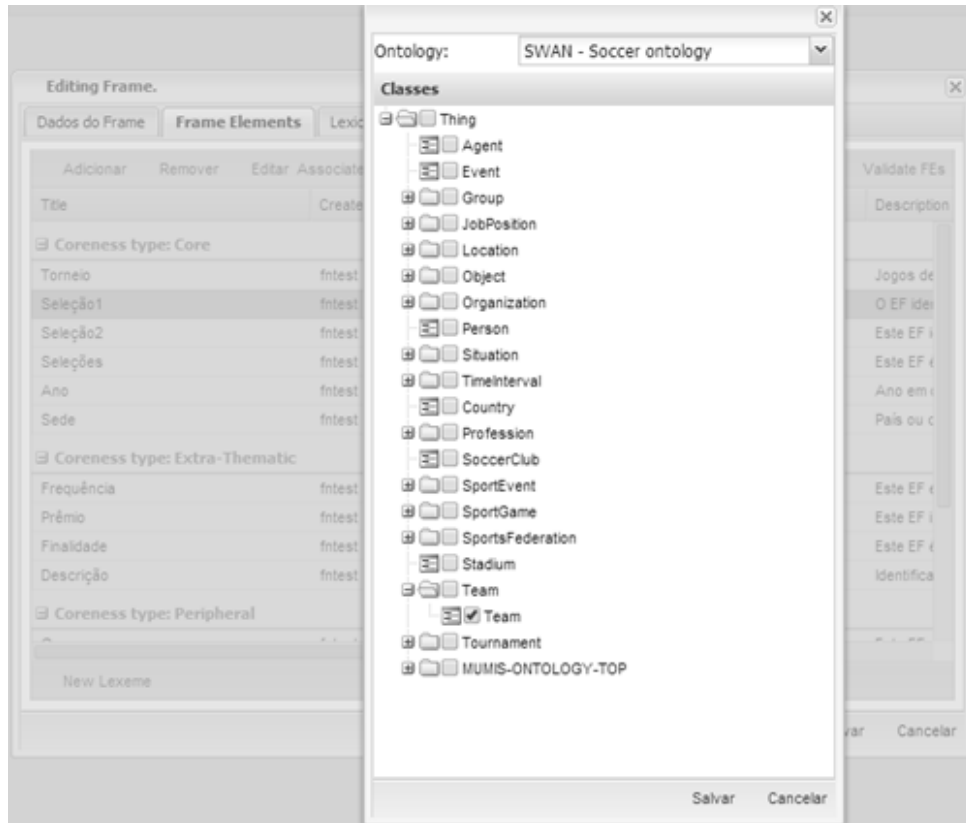


Figura 3.27: Tela de associação de classes de ontologias de domínio com EFs.

Além dos Elementos de Frame, também é possível cadastrar Unidades Lexicais (UL) que participem do Frame em questão. Durante o cadastro de uma UL, é necessário informar o título da UL, sua descrição, o EF que esta UL incorpora (quando for o caso), e os lexemas ao qual essa UL se relaciona neste Frame (Figura 3.28).

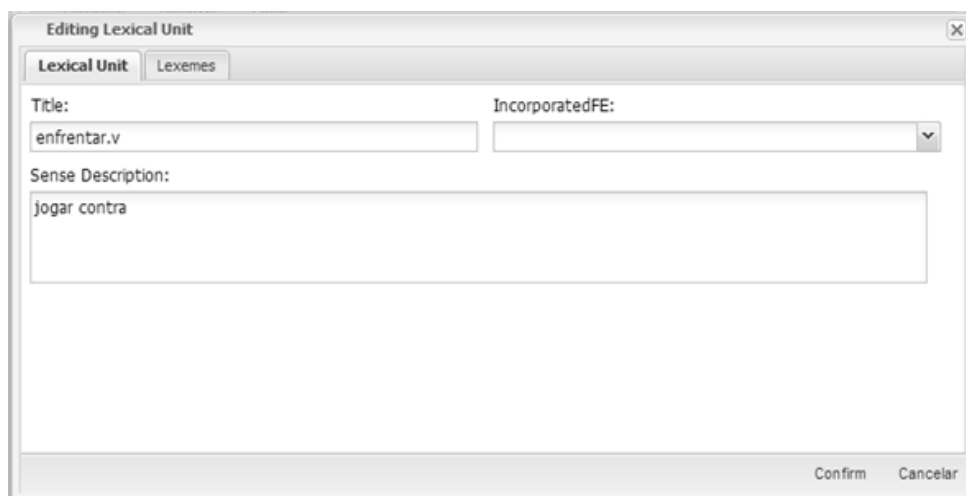


Figura 3.28: Tela de cadastro de UL.

Caso o lexema não se encontre na lista disponibilizada para associação ao UL, é possível

cadastrar um novo Lexema utilizando o botão "New Lexeme" localizado acima da lista. Durante o cadastro de Lexema é possível preencher seu título, sua classe lexical, e as formas de palavras que este lexema pode adotar quando participando de uma sentença (Figura 3.29).

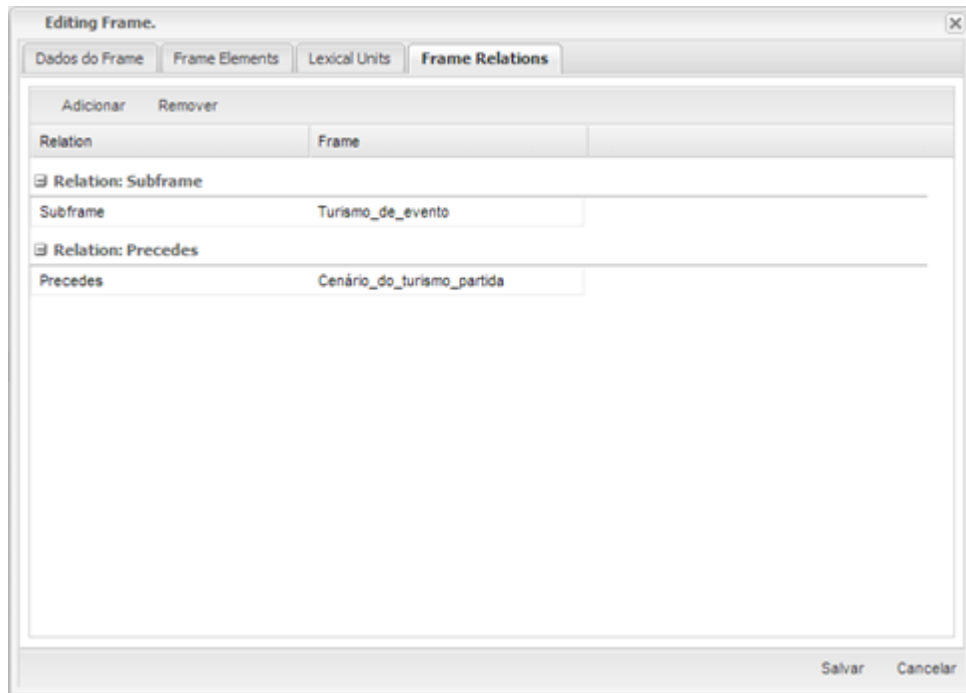


Figura 3.29: Tela de cadastro de Lexemas.

A última aba disponível do cadastro de frames é a aba onde é possível definir relações entre o frame atual e os demais frames da infraestrutura. Após a confirmação do cadastro do frame, a ferramenta realiza uma verificação semelhante a feita para os EFs, buscando identificar possíveis incoerências nos dados recém cadastrados ou a ocorrência de classificação por inferência deste frame na classe `InvalidFrame`. Caso esta verificação encontre algum destes casos, será exibida uma tela com a sequência lógica que levou a esta conclusão.

O foco desta funcionalidade é apoiar o processo de documentar frames, mantendo a estrutura lógica da FrameNet mas utilizando uma ontologia para formalizar estes dados, permitindo utilizar máquinas de inferências para verificação dos dados e garantir um maior nível de confiabilidade uma vez que a verificação de sua validade é feita de maneira automatizada, apresentando ao usuário qualquer incorreção, permitindo ainda sua correção.

Outra importante funcionalidade que esta ferramenta disponibiliza é um apoio para

realização de anotações de sentenças, adotando um sistema de sugestões. A tela da Figura 3.30 apresenta a sentença decomposta por palavras, onde é possível a seleção de uma ou mais palavras para a atribuição de um elemento do Frame.

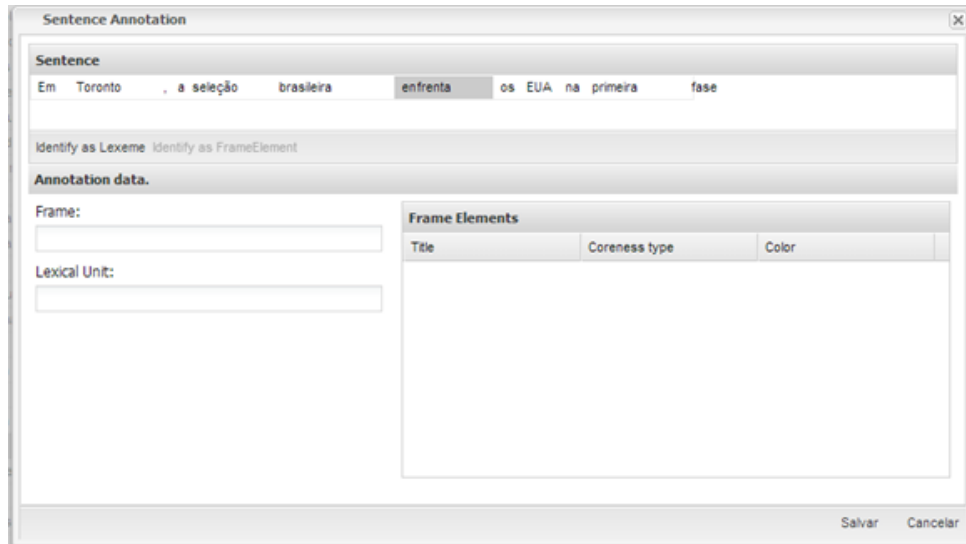


Figura 3.30: Tela para anotação de sentença.

O primeiro passo para a realização da anotação de uma sentença, é a seleção do termo que representa o lexema, fazendo assim a instanciação da UL, e por consequência definindo qual o frame que está sendo representado pela cena descrita na sentença em questão.

Ao selecionar um termo da sentença e sugeri-lo como Lexema, a ferramenta apresenta ao usuário um conjunto de Lexemas e o frame ao qual estes se relacionam. Os Lexemas apresentados são sugeridos a partir da comparação entre as formas das palavras associados aos lexemas na ontologia e a palavra selecionada na sentença (Figura 3.31). Caso o usuário não encontre o lexema desejado entre os que foram sugeridos pela ferramenta, ele poderá selecionar um a partir da lista de todos os lexemas.

Ao selecionar o lexema e conseqüentemente a UL, a ferramenta identifica o frame que se relaciona a ele, e retorna a lista de elementos de Frame para que possam ser identificados na sentença.

Para cada novo conjunto de palavras selecionadas na sentença, o usuário poderá marcar como um dos EFs do Frame. Ao realizar esta marcação, o conjunto selecionado apresentará a cor relacionada ao elemento de frame que este conjunto representa (Figura 3.32).

A cada novo conjunto de palavras identificado como instancia EF, a ferramenta exclui este EF da lista de pendentes, impedindo que o usuário marque mais de um conjunto

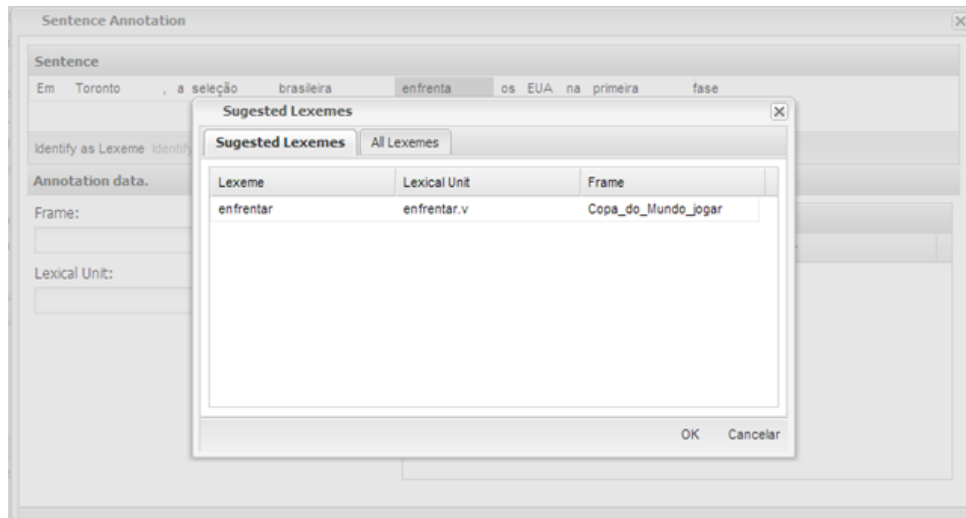


Figura 3.31: Tela com sugestão de lexema a partir da palavra selecionada.

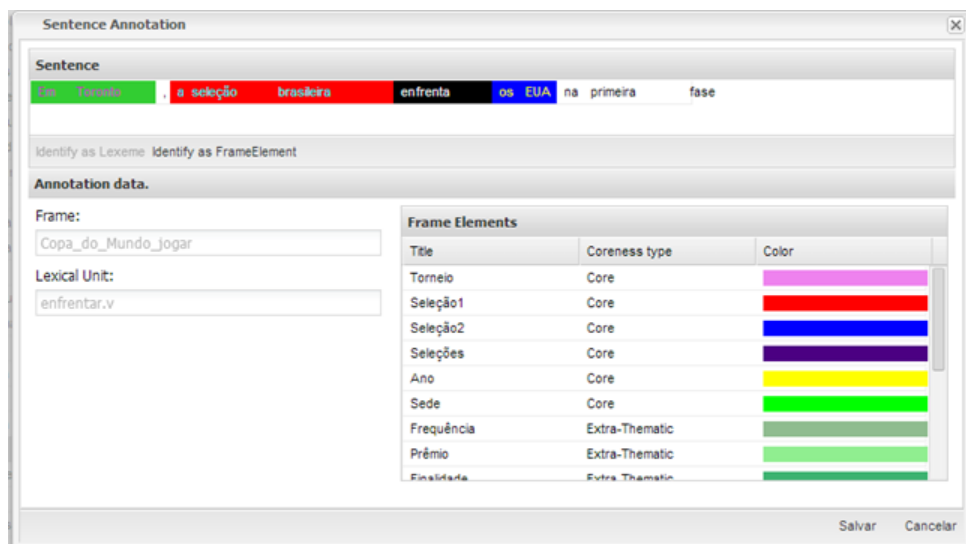


Figura 3.32: Tela apresentando a coloração de cada termo em relação ao EF.

como sendo parte do mesmo EF. Outra opção é a remoção de um conjunto já identificado, permitindo que um novo conjunto seja selecionado para aquele EF.

Outra funcionalidade da ferramenta é a atribuição de um recurso disponível na web que esteja cadastrado como sendo do tipo semântico equivalente ao daquele EF, permitindo que, durante uma consulta à anotação, o usuário possa também obter novas informações sobre o que a sentença representa. Um exemplo desta referencia pode ser observado na figura 3.33.

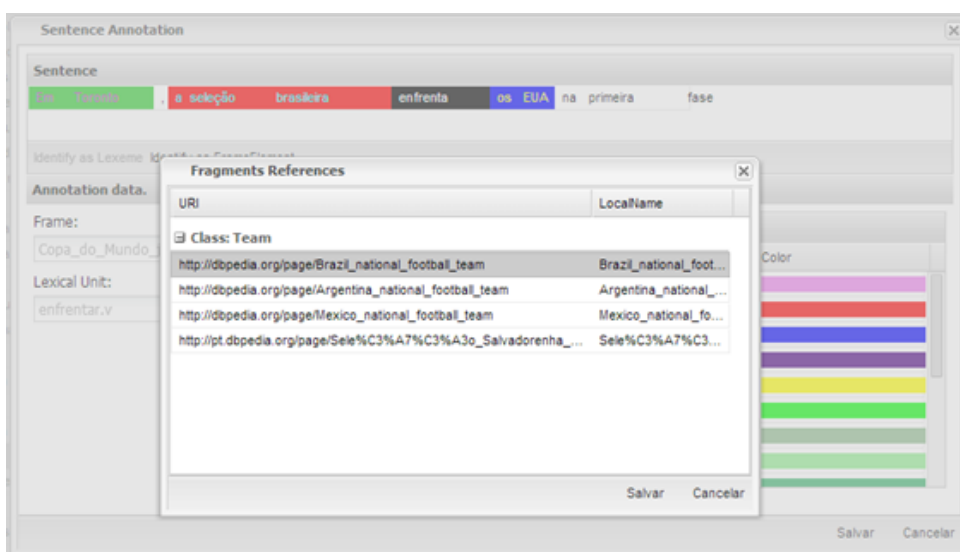


Figura 3.33: Seleção de Recurso para o termo "A seleção brasileira".

Esta referencia só é possível a partir do uso das ontologias de domínio que são utilizadas para a definição de tipos semânticos dos EFs conforme apresentado na seção 3.2.1.

Ao confirmar a conclusão da anotação da sentença em questão, o sistema realiza uma validação, verificando se pelo menos todos os EFs nucleares estão identificados na sentença, ou estar identificado como instância nula na sentença, ou seja, não estar explicitamente definido na sentença.

Neste contexto, a tela de LD *Resources* permite ao usuário selecionar uma ontologia de domínio que esteja sendo utilizada na infraestrutura e a partir dos termos desta, cadastrar URIs de recursos da Web relacionados aos seus termos. Assim, o usuário deverá selecionar uma das ontologias e suas classes serão apresentadas. Ao selecionar uma classe, serão exibidos os recursos cadastrados como relacionados a este termo ontológico, sendo possível também adicionar novos recursos àquele tipo semântico a partir de sua URI(Figura 3.34).

O objetivo desta associação é permitir que, durante o processo de anotação de sen-

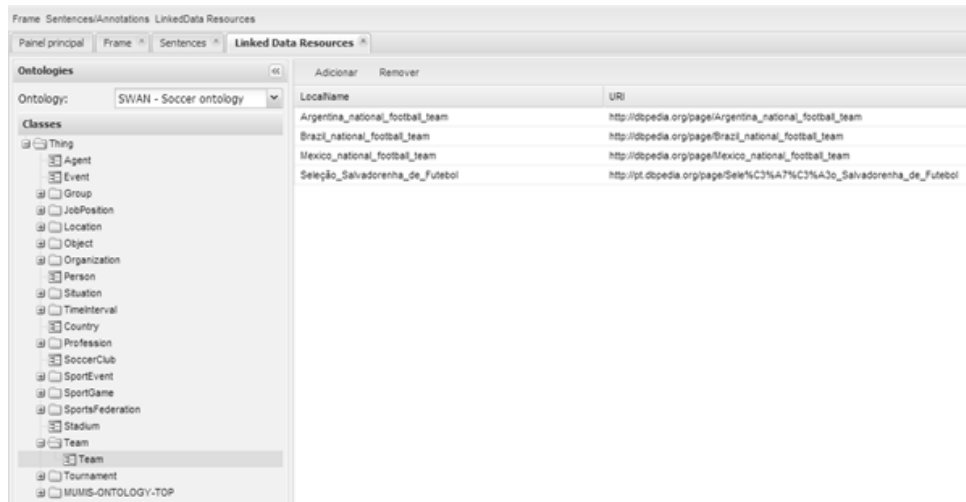


Figura 3.34: Associação de recursos externos a classes das ontologias de domínio.

tenças, estes recursos possam ser associados a conjuntos de palavras que referenciem um determinado EF, mantendo a consistência semântica daquele elemento, e permitindo ao usuário obter maiores informações sobre os elementos participantes daquela sentença.

### 3.3.3.2 FrameNet Semantic Viewer (FnSViewer)

Essa ferramenta apresenta uma interface semelhante à apresentada pela página de visualização de dados da FrameNet, exibindo a lista dos frames disponíveis para consulta na área esquerda da tela, e o detalhe do frame selecionado a direita, conforme apresentado na Figura 3.35 .

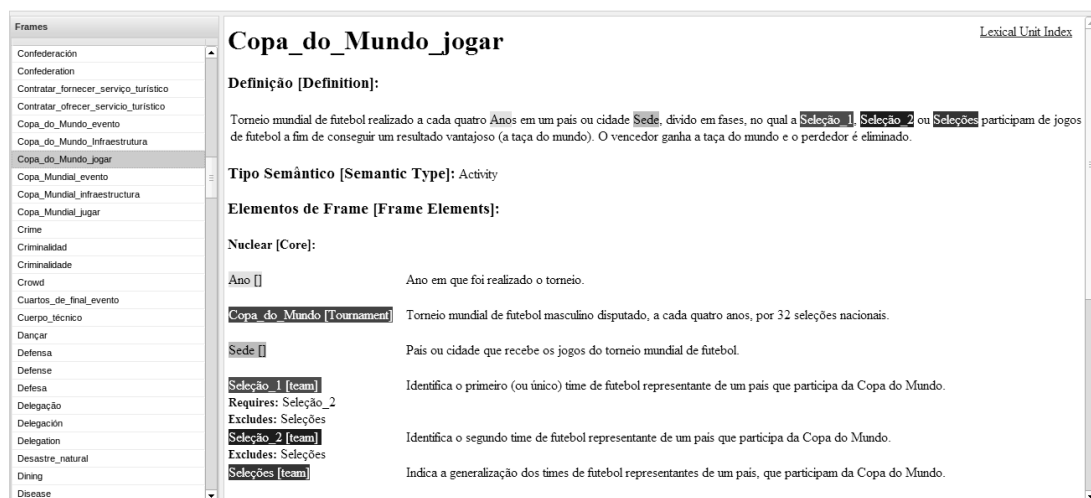


Figura 3.35: Tela de consulta de Frames do FnSViewer.

Assim como na interface do FrameNet, também é possível visualizar as anotações

de sentenças relacionadas ao frame selecionado. Entretanto na FnSViewer é possível consultar o recurso ligado ao fragmento da anotação graças à relação entre recursos e fragmentos de anotações que foi adotado na infraestrutura. Para isso basta clicar no fragmento e caso existam recursos associado a ele, estes serão exibidos em uma nova janela (Figura 3.36).



Figura 3.36: Tela de consulta a anotações com o detalhe de um recurso associado a um fragmento de sentença no popup.

### 3.4 TECNOLOGIAS

No processo de desenvolvimento da FSI diversas tecnologias foram utilizadas. A infraestrutura foi desenvolvida em JAVA (JAVA, 2014), sendo utilizados para tal fim duas ferramentas de desenvolvimento integrado NetBeans (NETBEANS, 2014), utilizado para desenvolvimento dos serviços e Eclipse (ECLIPSE, 2014), utilizado no processo de desenvolvimento do portal. E a utilização do TOMCAT como ambiente de execução de ambos.

Os serviços foram desenvolvidos com base na arquitetura JAXB (*Java Architecture for XML Binding*) e com uso da API JAX-WS (*JAVA API for XML Web Services*). Foram utilizadas também outras APIs tais como o JDOM2 (JDOM, JDOM) para manipulação de XML, *StarDog API* (STARDOG, 2013) para comunicação SNARL com o *StarDog Server*, e o JENA API (JENA, 2013) para leitura das ontologias que possuem a anotação dos serviços; o *MySQL Server Community Edition* (MYSQL, 2014), para manutenção de um banco de dados relacional onde estão armazenados os usuários e alguns logs de

utilização, e o *StarDog Server* (STARDOG, 2013), que foi utilizado com o objetivo de armazenar e realizar inferências sobre as ontologias que possuem um grande número de Indivíduos (ZHOU et al., 2006).

Para construção e definição das ontologias foram utilizadas as linguagens RDF e OWL, com uso da ferramenta Protege (PROTÉGÉ, 2012).

### 3.5 CONSIDERAÇÕES FINAIS DO CAPÍTULO

O FSI é uma infraestrutura, desenvolvida no contexto do projeto COPA2014, com o objetivo de apoiar o projeto FrameNet, e que foi construída a partir da utilização dos conceitos de web semântica. A FSI engloba três camadas distintas: i) a camada de dados, onde estão armazenadas as ontologias de domínio e as ontologias ONTO-FRAME-BR e ONTO-ANNOTATION-BR, e também os dados relativos a participação de recursos em fragmentos de anotações e os metadados dos serviços; ii) camada de serviços, onde estão disponíveis os serviços para comunicação com a infraestrutura; e iii) portal, que possui duas ferramentas: a FnSEditor para edição dos dados dos frames e realização das anotações, e o FnSViewer para visualização dos frames e suas anotações de maneira semelhante ao FrameNet. Apesar de (RUPPENHOFER et al., 2006) apresentar a anotação de sentenças na FrameNet baseadas em três camadas, onde são identificados em cada uma delas, respectivamente, os competentes do frame, a função gramatical e o tipo do sintagma, na ONTO-ANNOTATION-BR é focada apenas a primeira camada da anotação, não sendo abrangida neste momento a possibilidade de anotação nas outras duas camadas.

As ontologias ONTO-FRAME-BR e a ONTO-ANNOTATION-BR, foram detalhadas, bem como foi apresentada uma visão geral da arquitetura da infraestrutura, considerando cada um de seus componentes. Dentre as principais características da infraestrutura podemos ressaltar a utilização de ontologias para descrição da estrutura dos frames e anotações de sentenças; a utilização de serviços web semânticos, facilitando a descoberta semântica destes serviços e possível utilização destes por aplicações no contexto da Web semântica; e a criação de ferramentas que compõem o portal da infraestrutura que permitem a manutenção dos dados de frames e a realização das anotações.

Sendo também um diferencial quanto a realização de anotação de sentenças demonstrado pela ferramenta FnSEditor, quanto a outras ferramentas como a SALTO (ERK et al., 2003) e a *Framenet Desktop* (RUPPENHOFER et al., 2006), uma vez que o uso de



dados ligados e associação a ontologias de domínio nos permite enriquecer a obtenção de informações relacionadas à cena representada pela sentença, e ainda a formalização dos dados dos frames por meio do uso da ONTO-FRAME-BR e ONTO-ANNOTATION-BR permitem uma maior confiabilidade dos dados anotados.

## 4 AVALIAÇÃO DA PROPOSTA

Neste capítulo é apresentada uma avaliação preliminar do framework FSI, sendo definido um método de pesquisa com o objetivo de avaliar a hipótese apresentada.

No contexto desta avaliação preliminar, o framework FSI foi utilizado em conjunto com o projeto Copa 2014 FrameNet (SALOMAO et al., 2011), por ser um projeto desenvolvido com base na FrameNet-Brasil e por possuir foco no contexto da organização e realização da Copa do Mundo 2014 FIFA, o que permitiu delimitar os domínios a serem trabalhados, a saber, Futebol e Turismo. O projeto COPA2014 permite trabalhar com um menor volume de dados, permitindo, no entanto a formação de uma teia semântica concisa.

O objetivo deste estudo é demonstrar a validade da hipótese deste trabalho, ou seja, verificar se a aplicação de tecnologias de web semântica como ontologias, dados ligados, e anotações em serviços web podem contribuir com a construção e uso de recursos lexicais baseados na semântica de frames. Sendo estas contribuições voltadas para a confiabilidade dos dados, e para o enriquecimento das informações mantidas pela base lexical. Como nesta avaliação os domínios a serem tratados estão voltados para futebol e turismo, foram selecionadas duas ontologias, considerando as avaliadas na seção 2.4, que melhor representassem os conceitos destes domínios, sendo, portanto selecionadas as ontologias PROTON (TERZIEV et al., 2005), que apesar de ser uma ontologia que abrange diversos domínios, quando comparada as outras ontologias estudadas, atende melhor ao domínio de turismo, e para o domínio de futebol foi selecionada a ontologia *SWAN Soccer Ontology* (MÖLLER, 2005) que dentre as ontologias estudadas é a mais completa, e preenche a deficiência de termos do domínio de futebol na ontologia PROTON.

### 4.1 DEFINIÇÃO E PLANEJAMENTO DO MÉTODO DE PESQUISA

Conforme dito, o objetivo principal deste capítulo é avaliar a FSI, que utiliza elementos de Web Semântica, como ontologias, dados ligados e anotações semânticas de serviços, aplicados ao contexto da construção e aplicabilidade de recursos lexicais baseados na Semântica de Frames, como o FrameNet. A avaliação apresentada neste capítulo tem como propósito a verificação dos benefícios promovidos pela FSI tanto na manutenção e evolução do recurso lexical quanto na utilização de seus dados por terceiros.

Apresentamos a seguir uma estrutura com base no método GQM (BASILI, 1994) onde são definidos cinco elementos que compõe o estudo, sendo eles: o objeto do estudo; o objetivo do estudo; o foco do estudo; o ponto de vista do estudo; e por fim o contexto em que se aplica. Sendo assim, o estudo aplicado nesta avaliação foi estruturado da seguinte maneira:

*”Analisar a utilização de tecnologias de web semântica, principalmente ontologias e dados ligados, aplicadas a recursos lexicais baseados na Semântica de Frames para o propósito de facilitar a utilização de recursos lexicais por aplicações ou seres humanos com relação à utilização dos serviços oferecidos pela FSI (FrameNet Semantic Infrastructure) associada a ONTO-FRAME-BR do ponto de vista do usuário e aplicações que necessitam consultar recursos lexicais no contexto de manutenção dos frames, anotação de sentenças e consulta aplicados a uma base lexical”*

Para este estudo é apresentada uma hipótese com o intuito de oferecer respostas a problemas apresentados neste trabalho, sendo ela: Ao utilizarmos tecnologias de web semântica, como ontologias e dados ligados, em conjunto com recursos lexicais baseados na semântica de frames podemos facilitar a utilização dos recursos lexicais por aplicações e ou usuários.

O método de pesquisa adotado para verificar a hipótese deste trabalho é a realização de provas de conceitos (*Proof of Concept- PoC*). Para este fim, foram utilizados dados obtidos a partir da base do projeto COPA 2014 FrameNet-BR, que trata-se de um projeto derivado do projeto FrameNet-BR, relacionado especificamente aos domínios de futebol e turismo. Também foram utilizadas ontologias de domínio, além da ontologia ONTOFRAME-BR. Essas ontologias foram utilizadas com o intuito de prover um modelo semântico formal relacionado ao domínio abrangido pelos frames do COPA2014. Assim, foram selecionadas as ontologias PROTON (TERZIEV et al., 2005) para representar o domínio de turismo, e a ontologia *SWAN Soccer Ontology* (MÖLLER, 2005) para cobrir o domínio do futebol.

Apesar de não apresentar o formalismo de um estudo experimental, a realização de provas de conceitos neste domínio, contribuiu para a avaliação preliminar da hipótese formulada. Além disso, possibilitou também a verificação da viabilidade da solução proposta nesta dissertação.

Estas avaliações foram aplicadas em um problema real de documentação de frames, anotações de sentenças e consultas a bases lexicais. Quanto à generalidade, a avaliação foi realizada em um projeto específico, o COPA 2014 FRAMENET-BR. A dimensão observada é relacionada a eficácia no uso de recursos lexicais por aplicações e/ou usuários finais.

## 4.2 PROVAS DE CONCEITO

### 4.2.1 CENÁRIOS PARA AS PROVAS DE CONCEITO (POC)

O cenário das provas de conceito está inserido no contexto do projeto COPA2014, considerando a necessidade de anotação, evolução e consulta a recursos lexicais relacionados ao projeto. As provas foram realizadas pelo próprio desenvolvedor da proposta, que já possuía, portanto conhecimento prévio do funcionamento do framework e seu propósito específico. Assim, para a realização do estudo, uma instância do framework FSI foi disponibilizada para o a realização da prova juntamente com uma cópia da base de dados do projeto COPA2014. A PoC4 é um tipo especial de prova de conceito, onde não temos um usuário específico, mas sim uma aplicação acessando a interface programática do FSI. O objetivo desta última prova de conceito foi verificar a adequabilidade e facilidade de uso da FSI por aplicações na web.

### 4.2.2 PRIMEIRA POC: DOCUMENTAÇÃO DE FRAMES

A primeira prova de conceito foi relacionada a documentação de frames e suas relações e teve como objetivo avaliar a viabilidade de uso do FSI para a documentação de frames. Para isso, apresentamos a seguir as atividades desenvolvidas para a documentação de alguns frames. Esta PoC teve duração de aproximadamente 40 minutos e foi realizada com o propósito de verificar a contribuição da ferramenta quanto a verificação da validade dos dados de frames de acordo com o uso da ontologia ONTO-FRAME-BR para formalização da estrutura do frame. A primeira atividade desenvolvida foi a documentação do frame Entretenimento, que possui relação de herança com o frame Serviço\_Turistico (Figura 4.1). onde foi realizado o processo de cadastramento do frame e pode também utilizar as funcionalidades semânticas providas pela ONTO-FRAME-BR para avaliar a relação de herança. O primeiro passo para representação desta relação foi o cadastro dos frames

”Entretenimento” e ”Serviço\_Turístico”, e seus elementos de frames conforme apresentado pela figura 4.1. Após o cadastro destes frames, pode-se criar a relação de herança a partir da edição do frame que herda, que no caso é o ”Entretenimento”.

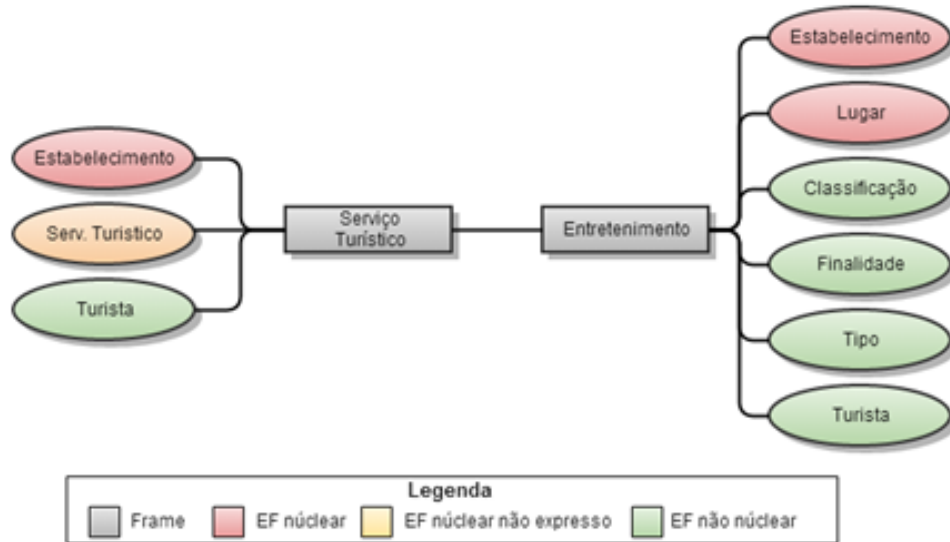


Figura 4.1: Representação dos Frames Serviço\_Turístico e Entretenimento e seus EFs.

Ao selecionar a relação de herança e o frame de ”Serviço\_Turístico” como frame herdado, a ferramenta requisitou a especificação da relação entre os elementos de frames nucleares presentes no Frame pai e o frame que herda. Assim, somente após a definição da relação da herança entre todos os frames nucleares do frame pai foi possível salvar a relação de herança entre os frames (Figura 4.2). Ao selecionar um EF de cada frame e clicar no botão ”Associate”, pode-se verificar que todos os EFS foram removidos das listas superiores e adicionados em conjunto na lista inferior.

Após a documentação do Frame de ”Entretenimento”, também foram documentados os frames contidos na Figura 4.3. Estes frames estão apresentados com a relação de SubFrame de acordo com a hierarquia apresentada.

A fim de averiguar se alguma informação implícita foi identificada por meio do reasoning da ONTO-FRAME-BR, após a definição dos frames e seus elementos de frame, foi realizada uma consulta em SPARQL utilizando o método *executeSPARQL* do serviço de Ontologias com a opção de inferência, e na resposta obtida (Figura 4.4 ) pode-se identificar a presença dos Frames da extremidade da hierarquia de subframes se relacionando com o Frame pai de hierarquia, demonstrando assim a transitividade desta relação. A mesma coisa pôde ser observada no caso das relações de *Using* e *Precedes*, uma vez que estas também são transitivas.

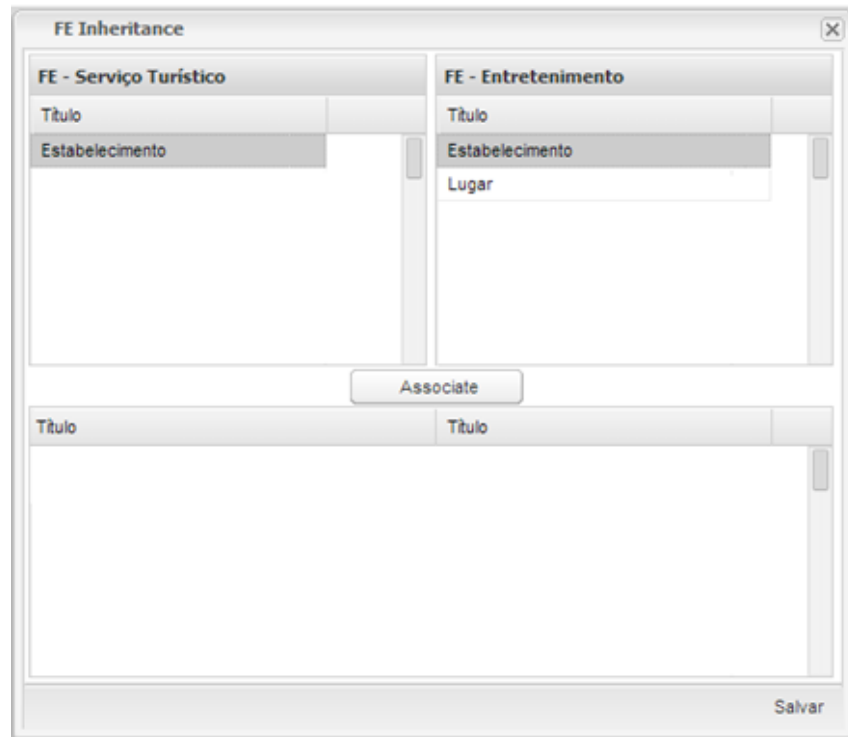


Figura 4.2: Associação de herança entre EFs.

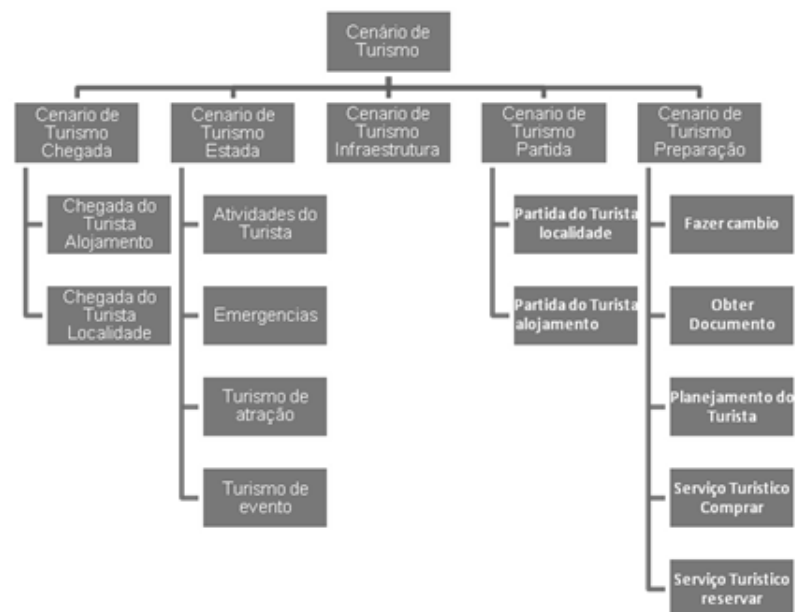


Figura 4.3: Representação hierárquica do Frame "Cenário de Turismo" e seus subframes.

```

<response>
<ResultSet>
  <set_1 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Cenário_do_turismo_partida" />
  <set_2 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Cenário_do_turismo_estada" />
  <set_3 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Cenário_do_turismo_chegada" />
  <set_4 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Cenário_do_turismo_planejamento" />
  <set_5 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Partida_do_turista_alojamento" />
  <set_6 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Partida_do_turista_localidade" />
  <set_7 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Turismo_de_atração" />
  <set_8 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Chegada_do_turista_localidade" />
  <set_9 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Chegada_do_turista_alojamento" />
  <set_10 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Turismo_de_evento" />
  <set_11 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Planejamento_do_turista" />
  <set_12 fName="Cenário_do_turismo" relation="http://200.131.219.34:10080/FSI/fai.owl/Subframe" subFName="Serviço_turistico_comprar" />
</ResultSet>
</response>

```

Figura 4.4: XML da resposta da consulta SPARQL com subframes de "Cenários de Turismo".

E por fim, foi feita a anotação do frame "Copa\_do\_Mundo\_Jogar" onde foi ressaltada a definição das relação "requires" e "excludes" presentes em alguns de seus EFs nucleares, conforme apresentado na Figura 4.5.

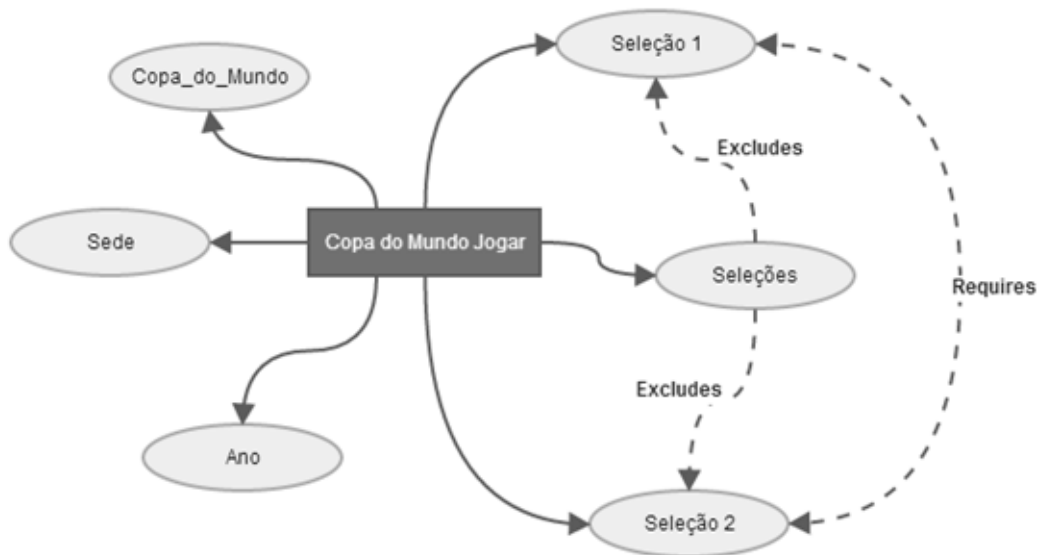


Figura 4.5: Frame "Copa do Mundo Jogar" com seus EFs nucleares e as relações entre eles.

Durante a documentação deste frame, além das estruturas originalmente obtidas a partir da base do COPA 2014, foi adicionada também uma relação de exclusão entre os EFs "Seleção 1" e "Seleção 2". Ao confirmar alteração do Frame com esta relação, a ferramenta considerou o Frame como "incorreto" e apresentou a sequência lógica apresentada na Figura 4.6 como justificativa para esta classificação.

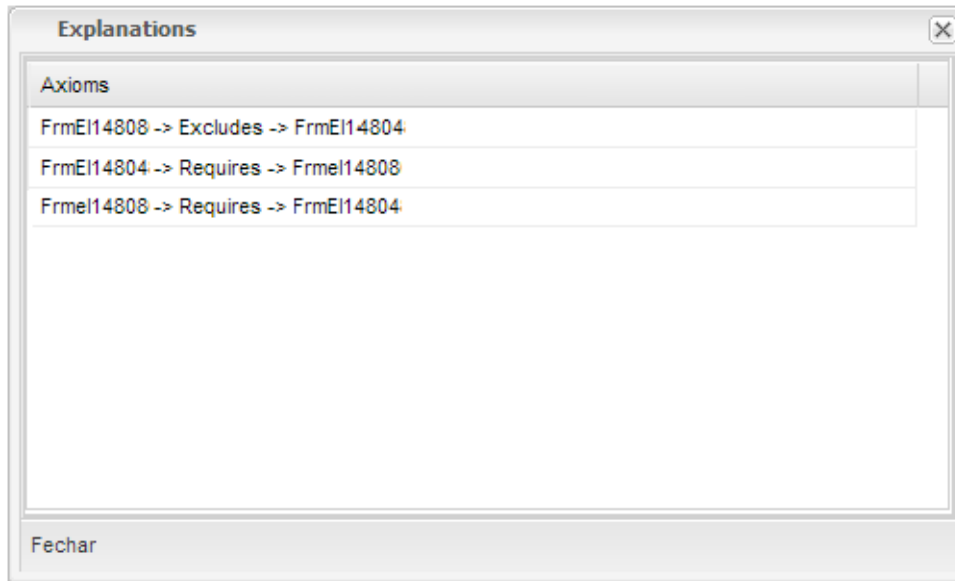


Figura 4.6: Lógico de classificação do Frame como incorreto pela ferramenta.

#### 4.2.2.1 Evidências observadas na PoC 1

Nesta primeira prova de conceito foram realizados os cadastros de frames capturados da base do COPA 2014 FRAMENET-BR, sendo respeitada a sua relação com os demais frames da base e suas estruturas. As ferramentas disponibilizadas pelo Portal da FSI foram utilizadas com o objetivo de apoiar a avaliação dos frames cadastrados.

Pode-se observar neste cenário que as definições semânticas das relações entre frames formalizadas na ONTOFRAME-BR contribuíram para o enriquecimento dos Frames, como foi observado na inferência da relação "*SubFrames*", fornecendo ao usuário informações implícitas que outrora dependia somente da percepção do usuário que realizava a documentação dos frames.

Outra importante observação feita durante a execução da PoC1 foi quanto a confiabilidade dos frames, uma vez que a partir da semântica das relações e estruturas dos frames definidas na ONTO-FRAME-BR é possível detectar possíveis incorreções existentes na rede de frames. Assim, considerando-se a anotação de frames, pode-se obter evidências quanto a viabilidade de uso da FSI para esse propósito. No entanto, existem outras atividades que envolvem a manutenção e utilização dos dados do recurso lexical. Sendo assim, foram efetuadas duas novas provas de conceito, detalhadas a seguir.



### 4.2.3 SEGUNDA POC: ANOTAÇÃO DE SENTENÇAS

Outra importante atividade no processo de manutenção dos dados de recurso lexical em questão é a anotação de sentenças, que trata da identificação do frame e de seus elementos que são evocados em uma sentença extraída de um corpora.

Neste cenário foi utilizada como sentença a ser anotada: "A seleção brasileira enfrentou a de El Salvador apenas uma vez." que está presente no corpus disponibilizado pela base do projeto Copa 2014. Para realizar essa anotação, foi utilizada a ferramenta FnSEditor presente no portal da FSI. Esta atividade teve duração de aproximadamente 15 minutos. Após a seleção da sentença dentre as sentenças disponíveis no corpus, foi definida cada parte da sentença, sendo estas compostas por no mínimo uma palavra, que representam algum artefato de um frame. Esta seleção se deu a partir da tela de anotação onde é possível selecionar cada palavra individualmente, ou então um grupo de palavras para ser identificado como um Lexema ou EF (Figura 4.7).

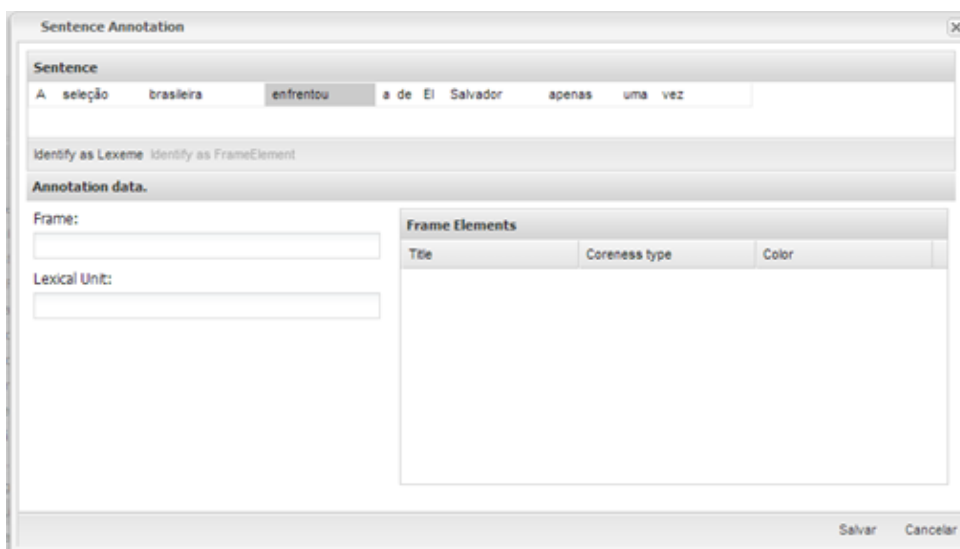


Figura 4.7: Tela de anotação de Sentenças.

No processo de anotação pela Ferramenta FnSEditor, o primeiro elemento a ser identificado na sentença é o lexema, assim, foi necessária a seleção de uma palavra e em seguida o clique no botão "Identify as Lexeme", o que no caso desta sentença foi indicado pela palavra "enfrentar", que é considerada a palavra que dá sentido à cena representada pela sentença. Quando foi selecionada a palavra, a ferramenta apresentou uma tela com uma sugestão de Lexema baseada na análise da forma da palavra selecionada e as formas de palavras associadas aos lexemas na Onto-Frame-BR (Figura 4.8).

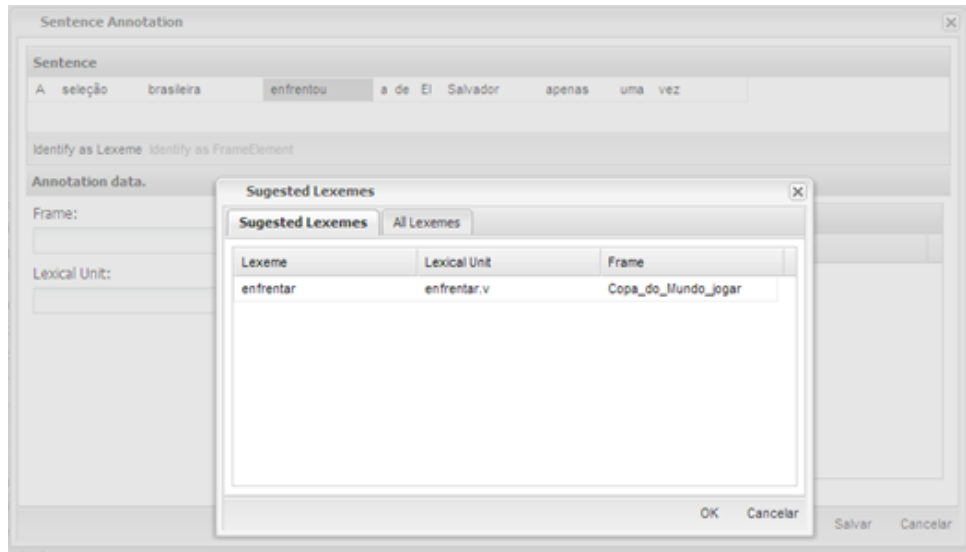


Figura 4.8: Sugestão para seleção de lexema.

O lexema selecionado nesta etapa foi o "enfrentar", que é relacionado a UL "enfrentar.v", e a parte da sentença que foi marcada como representante do lexema é representada pelo fundo de cor preta. Além disso, os dados do frame ao qual essa UL representa, "Copa\_do\_Mundo\_Jogar" neste caso, foram exibidos na parte inferior da tela (Figura 4.9).

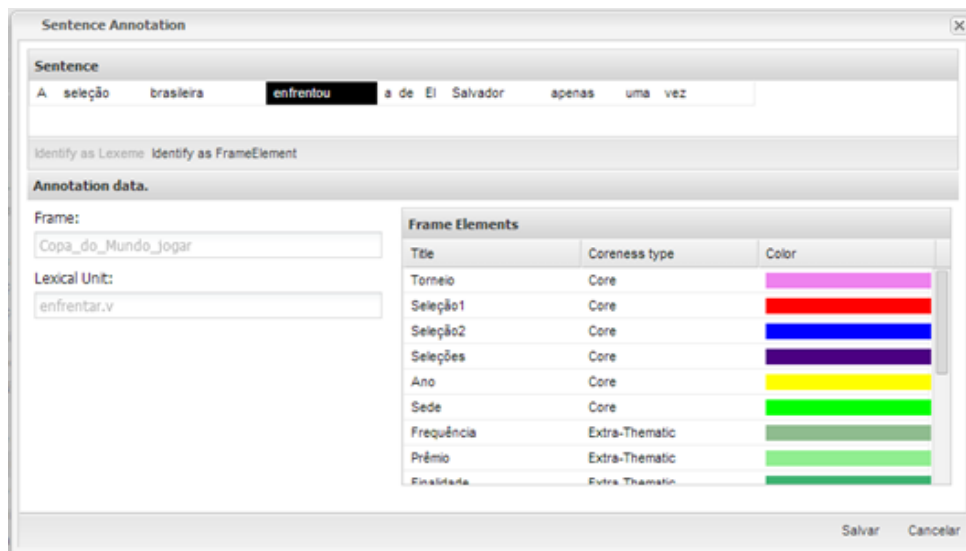


Figura 4.9: Tela de anotação de sentenças após identificação do lexema.

Após a identificação do lexema, foi necessário identificar também os conjuntos de palavras da sentença que representavam cada um dos EFs do frame selecionado. No caso deste frame, foi observada uma situação bem interessante, uma vez que existem as relações de dependência entre os EFs "seleções1" e "seleções2", e também a relação de exclusão do

”Seleções1” e ”Seleções2” com o ”Seleções”, mesmo sendo todos nucleares. Isso se dá considerando que na verdade o EF ”Seleções” é uma abstração dos dois EFS ”Seleções1” e ”Seleções2”, que é possível ser utilizado em caso de sentenças onde um conjunto de palavras contínuo pode representar as duas seleções envolvidas no confronto.

No caso desta sentença, foram utilizados o ”Seleções1” e ”Seleções2”, já que o conjunto de palavras que representam as seleções não é contínuo. Conforme podemos visualizar na Figura 4.10, ao confirmar a anotação incoerente, a ferramenta apresentou uma tela mostrando o fato que a torna incorreta.

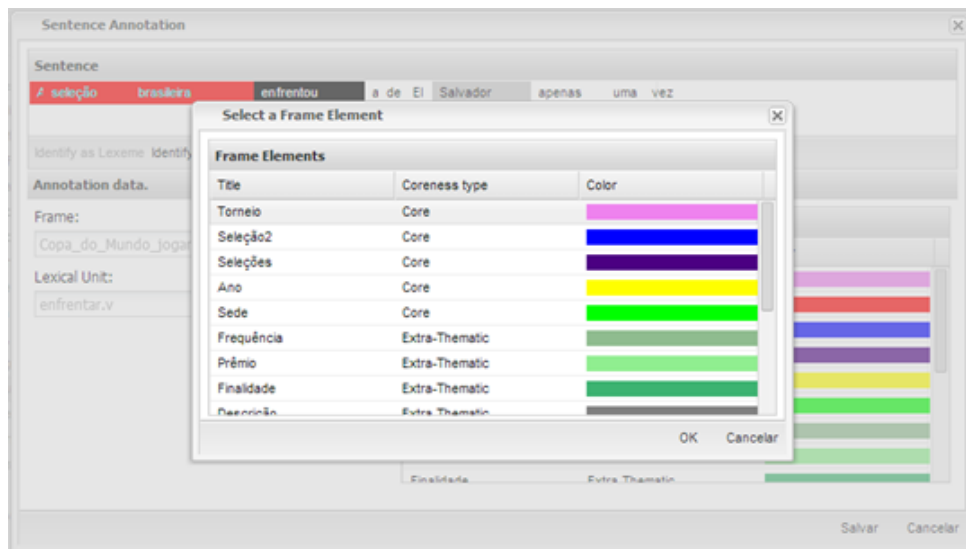


Figura 4.10: Verificação de incoerência durante anotação.

A figura 4.11 apresenta como ficou a sentença com todos os EFs participantes anotados. Assim, foi possível concluir a anotação da sentença, e considerando os EFs que não foram representados na cena e são nucleares, a ferramenta solicitou a confirmação se estes deveriam ser atribuídos como Instanciação Nula (Figura 4.12).

Dentre as características desta ferramenta, quando comparada as demais ferramentas de anotação baseadas em semânticas de frames, como a SALTO e a *Framenet Desktop* (RUPPENHOFER et al., 2006), podemos destacar a utilização das informações na ontologia ONTO-FRAME-BR e ONTO-ANNOTATION-BR, para sugestão de Unidades Lexicais a partir da comparação da forma de palavra da sentença selecionada como lexema, com as formas de palavras associadas aos Lexemas associados a ULs na ontologia. Além disso, outro diferencial é a possibilidade de averiguar se os fragmentos anotados da sentença estão satisfazendo a todos os EFs nucleares do frame associado a anotação, fornecendo ao usuário a informação sobre a incompletude desta anotação.

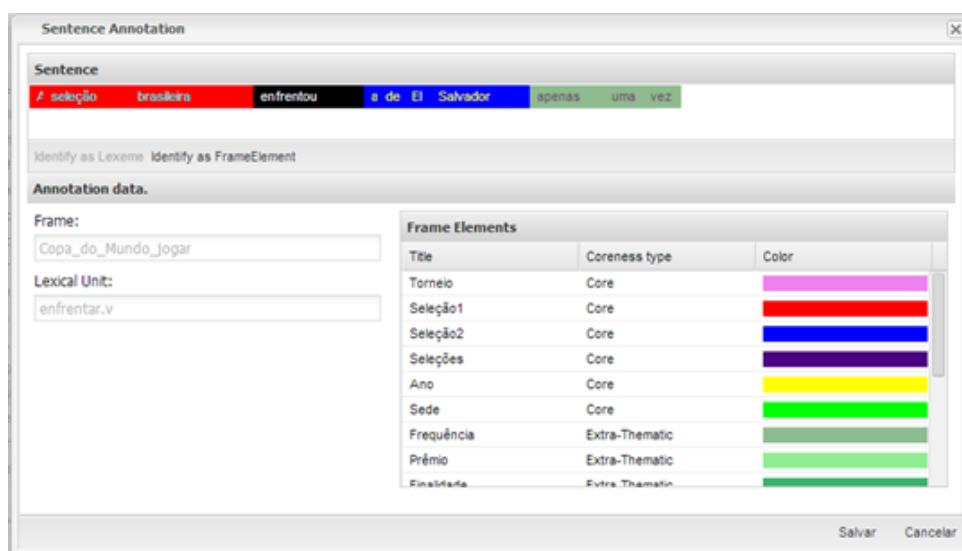


Figura 4.11: Sentença completamente anotada.

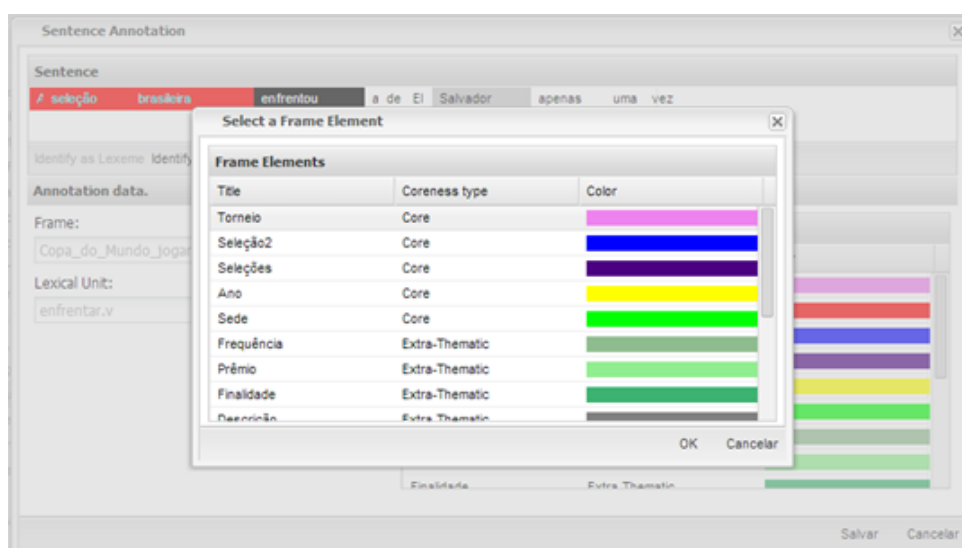


Figura 4.12: Confirmação de EFs com instanciações nulas.

### 4.2.3.1 Evidências observadas na PoC 2

Além das evidências observadas na primeira PoC, destaca-se na segunda POC, onde foi realizada a atividade de anotação de sentenças, a verificação do apoio da ferramenta aos anotadores, permitindo que sejam evitados possíveis erros por falha humana, uma vez que a ferramenta provê sugestões ao usuário de acordo com as informações armazenadas na ontologia que descreve os frames, e também valida e acusa possíveis incoerências de acordo com a semântica das relações entre os EFs.

Além disto, pode-se verificar a completude da anotação, a partir do uso das informações do frame contidas na ontologia de metadados de Frames, onde os EFs que são classificados como nucleares devem aparecer na sentença mesmo que seja como "Instanci-ção Nula", evitando assim a ocorrência de anotações incompletas, provendo assim uma maior confiabilidade quanto a validade da anotação realizada.

Considerando a PoC 2 e sua abrangência, consideramos que não foi avaliado o uso do padrão de dados ligados e a aplicação das ontologias de domínio em anotações. Assim, uma terceira POC foi executada e está detalhada a seguir.

## 4.2.4 TERCEIRA POC: RELAÇÃO DE DADOS LIGADOS E ONTOLOGIAS DE DOMÍNIO

Uma terceira prova foi realizada para avaliar o enriquecimento do vocabulário com as ontologias de domínio (SWAN e PROTON) associado ao uso do padrão de dados ligados. Para esta nova PoC foi utilizado o frame já documentado no cenário da primeira PoC, o frame "Copa do Mundo Jogar". Essa POC teve duração de cerca de 30 minutos.

Para isso, foi utilizada a opção "*Associate Ontology Classes*" (Figura 4.13), para associar um ou mais tipos ontológicos ao EF selecionado.

Os primeiros EFs a terem seus tipos ontológicos associados neste cenário foram "Seleções", "Seleção1" e "Seleção2", que foram associados aos tipos "*SoccerClub*" e "*Team*" da SWAN, e a "*SportClub*" da PROTON conforme apresentado na Figura 4.14 .

Em seguida o processo foi repetido para cada um dos EFs presentes no Frame selecionado, e seus tipos ontológicos foram definidos conforme descrito na Tabela 4.1.

Em seguida, para poder fazer a associação com fragmentos de anotações, foi necessário o cadastro dos recursos disponíveis em outras bases de dados ligados, como a DBPedia,

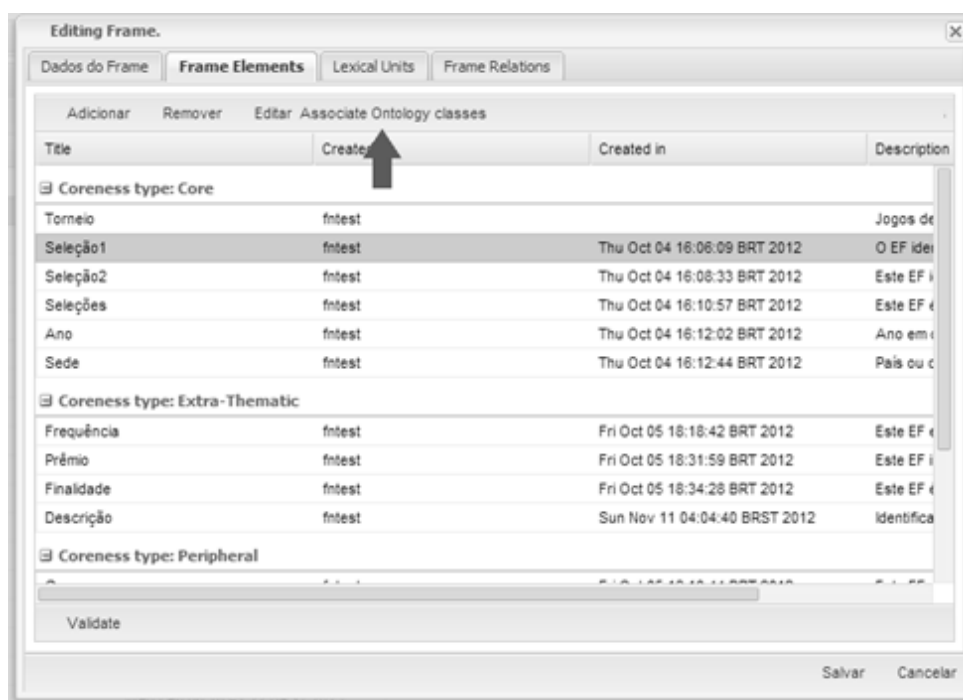


Figura 4.13: Opção de associar classes a EFs para restrição semântica.

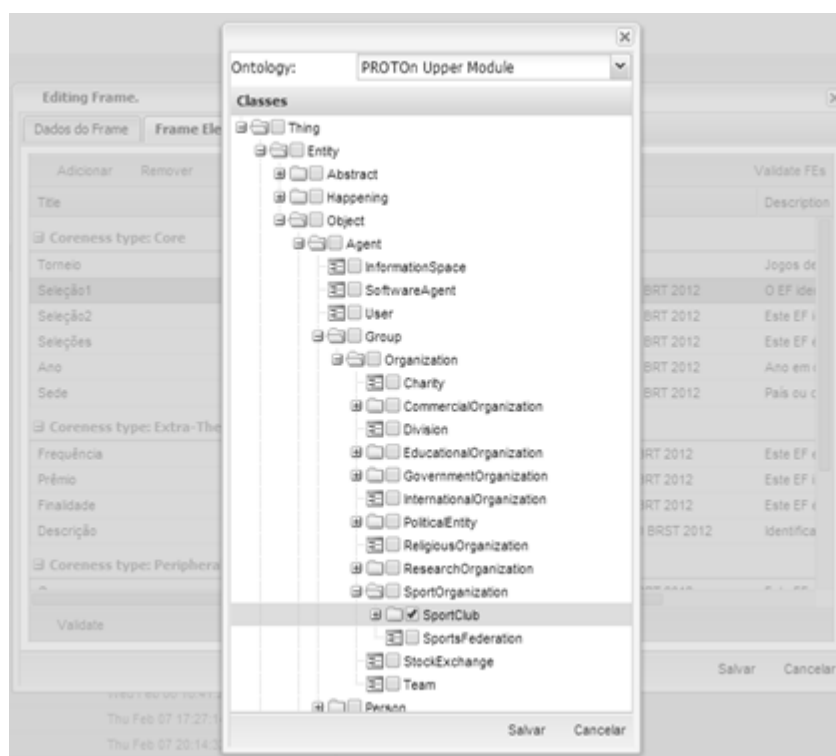


Figura 4.14: Seleção de classes ontológicas para o EF "Seleções1".

Tabela 4.1: Associação de classes ontológicas para EFs.

<b>Elemento de Frame</b>	<b>Ontologia</b>	<b>Classes</b>
Torneio	SWAN e PROTON	Tournament
Ano	PROTON	CalendarYear
Seleção1	SWAN	SoccerClub
		Team
Seleção2	PROTON	SportClub
		SoccerClub
Seleções	SWAN	SoccerClub
		Team
Sede	PROTON	SportClub
		SoccerClub
Frequência	—	—
Premio	SWAN	Trophy
Finalidade	—	—
Descrição	—	—
Grau	—	—
Duração	PROTON	TimeInterval
Maneira	—	—
Meio	—	—
Lugar	PROTON	PoliticalRegion
		PopulatedPlace
Tempo	PROTON	TimeInterval
Local	SWAN	Stadium

para que pudessem ter seu tipo semântico discriminado. Para isso, foi utilizada a opção "LinkedData Resources" (Figura 4.15). Foram cadastrados neste cenário os recursos apresentados na Tabela 4.2, todos sobre a Classe "Team" da ontologia SWAN.

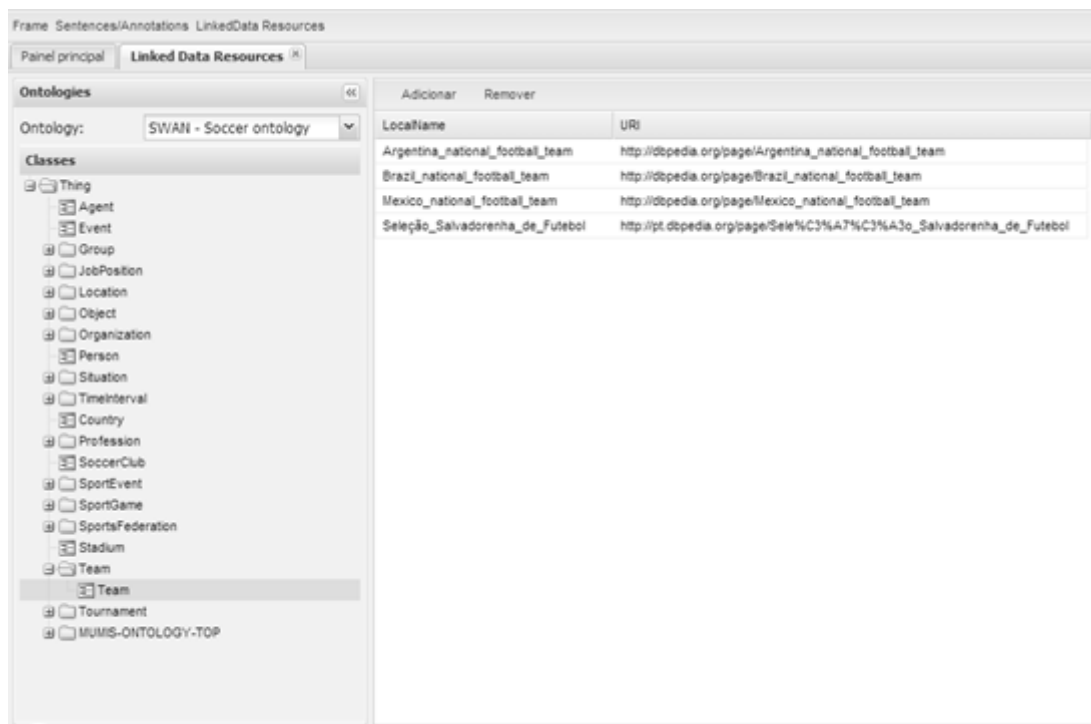


Figura 4.15: Tela para cadastro de recursos externos em relação a ontologias.

Tabela 4.2: Recursos De dados ligados cadastrados.

Title	URI
<i>Brazil_national_football_team</i>	http://dbpedia.org/page/Brazil_national_football_team
<i>Argentina_national_football_team</i>	http://dbpedia.org/page/Argentina_national_football_team
<i>Mexico_national_football_team</i>	http://dbpedia.org/page/Mexico_national_football_team
<i>Seleção_Salvadorenha_de_Futebol</i>	http://dbpedia.org/page/ Seleção_Salvadorenha_de_Futebol

Feito isto, na tela de anotações, a sentença anotada foi aberta na segunda PoC (seção 4.2.3), e foi definida a associação dos recursos cadastrados com a sentença. Ao efetuar um duplo clique em cada trecho anotado da sentença, foi exibida uma tela permitindo selecionar recursos que foram cadastrados na base e que possuísem o mesmo tipo ontológico que o EF selecionado, como podemos observar no exemplo da Figura 4.16, onde os recursos exibidos para a o fragmento "a seleção brasileira" estão referenciando o EF *Seleção1*.

Depois dessa associação, a ferramenta possibilitou que fosse possível consultar essa anotação e acessar a URI do recurso associado ao trecho da sentença anotada, permitindo



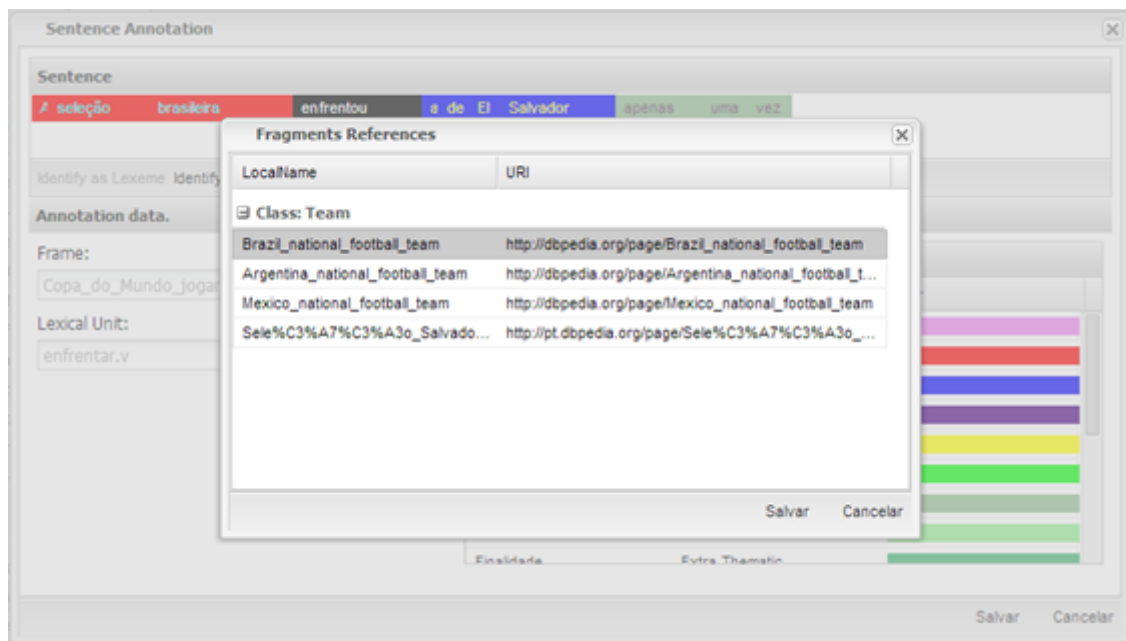


Figura 4.16: Seleção de recursos para fragmentos de anotação.

assim a obtenção de mais informações sobre um determinado termo de seu interesse, a partir da navegação pelos recursos associados ao termo. Nesta POC, foi selecionado o fragmento que o levou a descrição do recurso "Brazil\_national\_football\_team" provido pela DBPedia (Figura 4.17).

#### 4.2.4.1 Evidências Observadas na PoC 3

Com as duas provas de conceitos anteriores, buscamos validar a contribuição para o usuário principalmente quanto à confiabilidade dos dados documentados, entretanto nesta PoC o objetivo da avaliação é verificar a contribuição para integração dos dados com outras bases de conhecimento através das tecnologias aplicadas na proposta.

Podemos verificar nesta PoC que a associação dos padrões de dados ligados a ontologias de domínio no apoio a anotações promove um enriquecimento na experiência de consulta aos dados anotados na base lexical, uma vez que a partir a definição do tipo ontológico, a EF permite controlar melhor o tipo de recurso que pode participar nos frames, mas sem obrigar a participação de recursos nestes casos.

Além disso, como um dos princípios do padrão de dados ligados é permitir ao usuário encontrar cada vez mais informações relacionadas ao recurso atual, a partir das ligações entre os recursos, pode-se observar que a adoção deste padrão permite ao usuário uma experiência mais rica a partir do contexto de uma anotação de sentença.

The image shows a browser window displaying the DBpedia page for 'Seleção Brasileira de Futebol'. The browser's address bar shows the URL 'http://dbpedia.org/page/Brazil\_national\_football\_team'. The page title is 'About: Seleção Brasileira de Futebol'. Below the title, there is a description: 'An Entity of Type: agent from Named Graph: http://dbpedia.org, within Data Space: dbpedia.org'. The main content is organized into two columns: 'Property' and 'Value'. The 'Property' column lists 'dbpedia-owl:abstract'. The 'Value' column contains a detailed abstract in Portuguese, followed by a list of multilingual descriptions in various languages including Polish, Czech, Japanese, Chinese, and German. The abstract describes the Brazilian national football team, its history, and its status as the most successful national football team in the world.

**About: Seleção Brasileira de Futebol**  
An Entity of Type: [agent](#) from Named Graph: <http://dbpedia.org>, within Data Space: [dbpedia.org](#)

A Seleção Brasileira de Futebol é o time nacional do Brasil de futebol masculino, gerido pela Confederação Brasileira de Futebol (CBF), que representa o país nas competições de futebol organizadas pela CONMEBOL e pela FIFA. É o time mais bem-sucedido de futebol na história das Copas do Mundo, sendo a seleção nacional que mais vezes conquistou o Mundial com cinco títulos até então.

**Property**

**dbpedia-owl:abstract**

**Value**

- Het Braziliëans voetbalelftal is een team van voetballers dat Brazilië vertegenwoordigt bij internationale wedstrijden. Het is het succesvolste voetbalelftal van Latijns-Amerika.
- Reprezentacja Brazylii w piłce nożnej jest najbardziej uytuowana narodowa drużyna piłkarska na świecie. Jako jedyna w historii pięć razy zdobyła Puchar Świata, ponadto osiemkrotnie była najlepsza w rozgrywkach o Copa América. W roku 2009 Reprezentacja Brazylii zdobyła w RPA Puchar Konfederacji. Są jedyną reprezentacją, która uczestniczyła we wszystkich mistrzostwach świata w piłce nożnej.
- サッカー代表 (葡: Seleção Brasileira de Futebol) は、男子サッカー一連の大会を編成する男子サッカーの代表チームである。
- 巴西國家足球隊（葡語字源：Seleção Brasileira de Futebol）是巴西官方的男子國家足球代表隊，由巴西足球協會負責管理，並代表巴西參加大型國際性足球賽事。巴西隊被譽為世界上最有名及最成功的國家足球隊之一，它曾是自南美解放者盃世界盃冠軍的球隊，曾獲於1958年、1962年、1970年、1994年及2002年奪得冠軍，巴西球員是唯一一奪麥加迪所有國際世界盃決賽的球隊。巴西國家足球隊由巴西足球協會建立於1914年，並於1924年加入國際足聯。
- L'équipe du Brésil de football est constituée par une sélection des meilleurs footballeurs brésiliens sous l'égide de la Confédération Brasileira de Futebol. Favorité à chaque édition de la Coupe du monde, elle est l'unique sélection à avoir disputé toutes les phases finales (19) et a en avoir remporté cinq éditions en 1958, 1962, 1970, 1994 et 2002. La Selección est la seule détentrice du trophée Jules-Rimet, mis en jeu à partir de la Coupe du monde 1930, qu'elle a définitivement conservé en 1970 après ses trois premières victoires dans la compétition avec Pelé, le joueur le plus emblématique de son histoire et de celle du football mondial. L'équipe du Brésil a gagné la Coupe du monde sur quatre continents (Europe en 1958, Amérique du Sud en 1970 et 1994, Asie en 2002), et on dit communément à son propos « Les Anglais ont inventé le football, les Brésiliens l'ont perfectionné ». Souvent posté au 1 rang du classement mondial de la FIFA, le Brésil, également vainqueur à huit reprises de la Copa América, est considéré depuis plus de 50 ans comme la plus forte nation du football. Les surnoms Selección (Selection), Atimeiras (Vents et or) ou Scratch sont employés pour désigner cette équipe.
- ainsi que Canarino (« petit canari ») en référence à leur maillot jaune porté lorsqu'elle évolue à domicile.
- Die brasilianische Fußballnationalmannschaft, genannt Selección, ist ein vom brasilianischen Fußballtrainer zusammengestellter Kader von Spitzenspielern. Sie repräsentiert den brasilianischen Fußballverband, die Confederação Brasileira de Futebol (CBF), auf internationaler Ebene bei Freundschaftsspielen und internationalen Turnieren. Mit fünf genannten Titeln ist sie die weltweit erfolgreichste Fußball-Nationalmannschaft bei den Fußball-Weltmeisterschaften. Der letzte WM-Sieg wurde 2002 erreicht. Seit dem Start der FIFA-Weltmeisterschaft im August 1993 war die Mannschaft Brasieliens insgesamt in 141 von 220 Monaten, in denen eine Weltmeisterschaft stattfand, auf Rang 1. Dies entspricht 64 % der Zeit. Dabei waren sie ab Juli 1994 knapp über sechs Jahre (73 Monate) und ab Juli 2002 noch einmal viermonatig Jahre (54 Monate) ununterbrochen Erster. Zuletzt stand Brasilien im Mai 2010 auf Platz 1. Im Dezember 2012 wurde mit Platz 18 der bisher schlechteste Platz belegt, auch bedingt durch die Tatsache, dass Brasilien als Gastgeber der WM 2014 derzeit keine Pflichtspiele bestreiten muss.
- The Brazil national football team represents Brazil in international men's football. Brazil is made by the Brazilian Football Confederation (CBF), the governing body for football in Brazil. They are a member of the International Federation of Association Football (FIFA) since 1923 and also a member of the South American Football Confederation (CONMEBOL) since 1916. Brazil is the most successful national football team in the history of the FIFA World Cup, with five championships. They are also the most successful team in the FIFA Confederations Cup with three titles. Brazil are the current holders of the FIFA Confederations Cup after winning the 2005 and the 2009 edition of the tournament. Brazil is currently ranked eightheenth by FIFA. Brazil are the only national team to have played in every World Cup. Brazil is the only team to have won the championship in four different continents: once in Europe, once in North America and once in Asia. A common quip about football is: "Os ingleses o inventaram, os brasileiros o aperfeiçoaram" ("The English invented it, the Brazilians perfected it"). Brazil is scheduled to host the 2013 Confederations Cup and the 2014 World Cup and therefore, they are automatically qualified for the tournaments.
- La Selección de fútbol de Brasil (Seleção Brasileira de Futebol en portugués) es el equipo representativo del país en las competiciones oficiales. Su organización está a cargo de la Confederación Brasileira de Fútbol, perteneciente a la Confederación Sudamericana de Fútbol (Conmebol). Conocida como «La Verde-amarela» o «La Canarinhas» se encuentra

**DBpedia**

Figura 4.17: Descrição do recurso "BrazilNationalFootballTeam" da DBPedia.

Considerando estas três primeiras PoCs, a avaliação foi relacionada ao uso das funcionalidades da FSI por usuários humanos. Como o objetivo da FSI é também permitir a interação com clientes de software através da interface de serviços, percebeu-se a necessidade da realização de uma quarta PoC, onde é avaliada a interação entre a infraestrutura com uma ferramenta externa à FSI.

#### 4.2.5 QUARTA POC: USO DA FSI POR APLICAÇÕES DE SOFTWARE

Uma quarta prova de conceito foi realizada com o objetivo de avaliar a interação da FSI com outras ferramentas a partir de sua interface de comunicação baseada em serviços. Sendo assim, foi selecionada a ferramenta consultiva do dicionário Copa 2014 (TORRENT et al., 2014), onde foram necessárias modificações na ferramenta para consumo dos serviços da FSI como fonte de dados.

A ferramenta do Copa 2014 é composta de duas camadas, sendo a primeira camada onde estão armazenados os dados (frames e seus componentes) e serviços que controlam o acesso a estes dados, e a segunda camada sendo uma interface de interação com o usuário, que consome os dados da primeira camada (Figura 4.18).

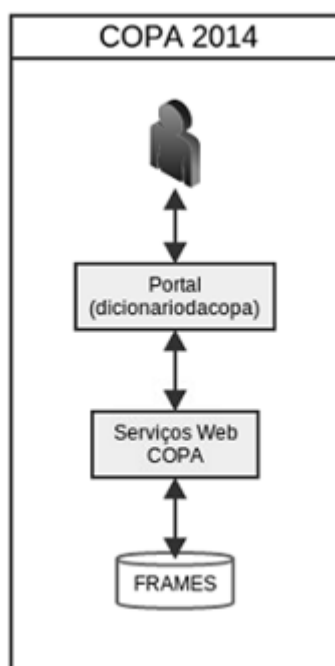


Figura 4.18: Visão da arquitetura da ferramenta do COPA 2014 (TORRENT et al., 2014).

Considerando o modelo arquitetural do Dicionário COPA 2014 e os dados disponi-

bilizados pela FSI (XML/RDF), optou-se por especificar um conector entre o FSI e o dicionário COPA 2014, uma vez que o dado recebido pelo COPA2014 deve ser disponibilizado em formato JSON e os serviços da FSI disponibilizam dados no formato XML/RDF (Figura 4.19). Além disso, como o objetivo da POC é verificar a viabilidade de utilização dos serviços da FSI por aplicações de software, foram utilizadas somente as funções da FSI que permitem a consulta de UL e Frames.

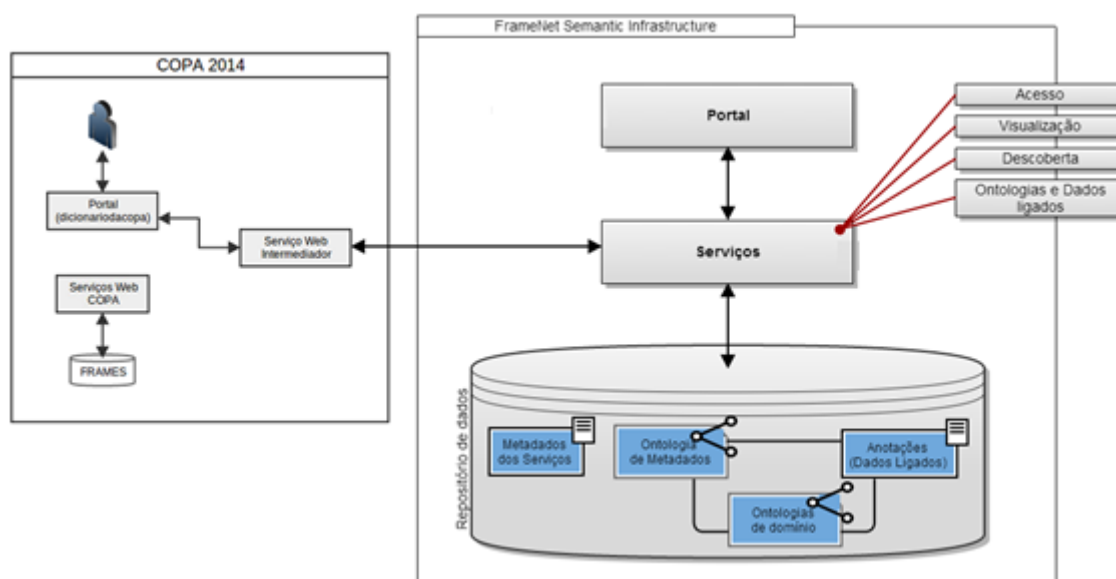


Figura 4.19: Arquitetura da ferramenta do COPA 2014 modificada para consumo da camada de serviços da FSI.

Após a realização destas modificações foi feito o uso do dicionário Copa 2014, considerando o uso deste conector entre os serviços. A duração dessa POC foi de aproximadamente 15 minutos, onde foram avaliadas as trocas de mensagens entre a FSI com o conector, e do conector com a interface do Copa 2014, com o objetivo de observar os benefícios do uso da FSI como fonte de dados.

Inicialmente o foi utilizada a função de consulta de palavras, onde foram consultadas algumas das palavras disponíveis na base do Copa 2014 utilizada para a FSI, e pode ser observado o fluxo de interação entre a ferramenta, o conector e a FSI conforme apresentado na figura 4.20.

Conforme é possível verificar no fluxo, inicialmente a ferramenta solicita ao conector a lista de palavras disponíveis na base. Sendo este o primeiro acesso do conector a FSI, é necessária a autenticação para em seguida solicitar via consulta SPARQL as unidades lexicais disponíveis na FSI, sendo esta resposta obtida no formato XML. Após converter

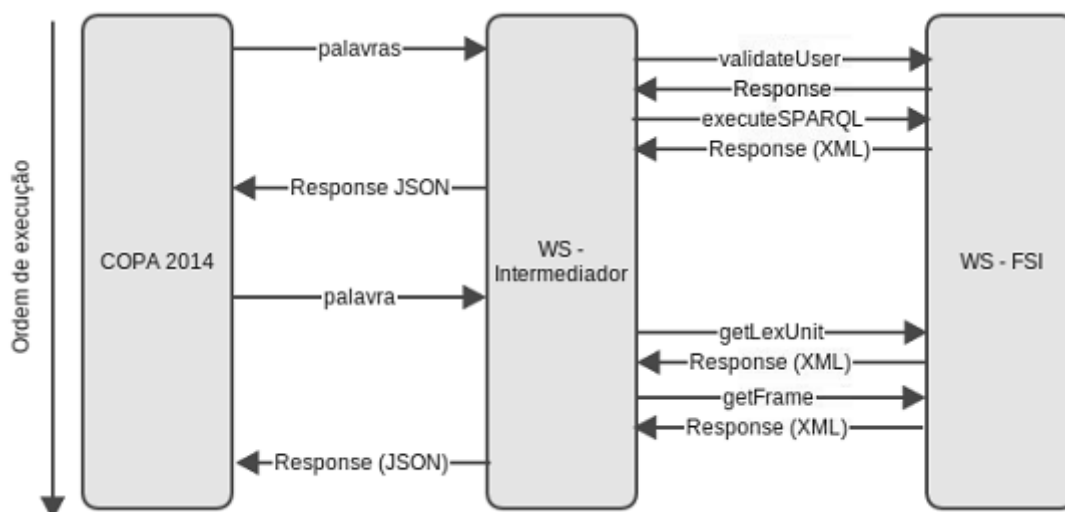


Figura 4.20: Fluxo de comunicação entre a ferramenta, o Serviço intermediador e a camada de serviços da FSI para consulta a palavras.

a resposta para o formato JSON o conector enviou as palavras ao COPA 2014. Em seguida, foi solicitado ao COPA o detalhamento de uma palavra, sendo então, disparado ao conector a solicitação do COPA2014 pelo detalhe da palavra, que no caso do COPA2014 são exibidas as informações da Unidade Lexical e o nome do frame relacionado a ela. Sendo assim o conector solicitou a FSI o detalhe da UL, via o método *getLexicalUnit()* e também informações sobre o frame relacionado a UL selecionada via o método *getFrame()*, sendo também estes dados recebidos em XML, condensados em um JSON e enviados de volta a ferramenta COPA 2014.

Em um segundo momento, foi realizada uma consulta aos Frames disponíveis via a opção "ver significado" da ferramenta, e eventualmente consultou o detalhe de alguns frames. Na figura 4.21 podemos ver o fluxo de comunicação verificado nesta ocasião.

Conforme é possível verificar no fluxo da figura 4.21, inicialmente a ferramenta solicita ao serviço intermediador a lista de frames divididos por domínio. Entretanto, como na versão da base disponibilizada para realização destas provas os frames ainda não estavam divididos por domínio, foi feita uma simplificação, considerando uma requisição com todos os frames em um mesmo domínio. A obtenção dos frames é feita via o método *executeSPARQL()*, onde foi enviada uma consulta SPARQL solicitando os indivíduos da classe "Frame", obtidos pelo conector no formato XML e transcrito pelo mesmo para o formato JSON. Após isso, foi solicitado a ferramenta o detalhe de alguns frames, onde foi solicitado pela ferramenta ao conector, e por sua vez o conector solicitou a FSI por meio

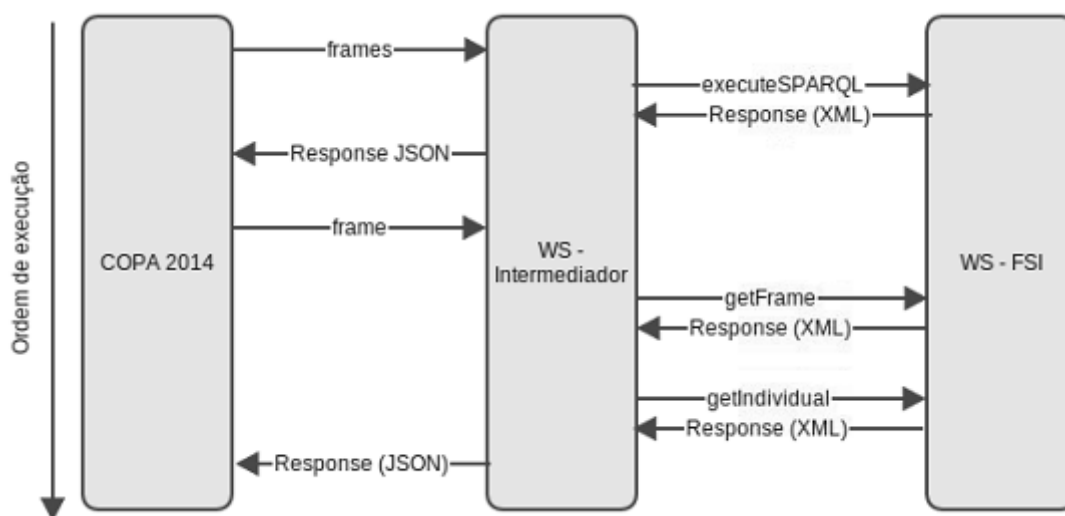


Figura 4.21: Fluxo de comunicação entre a ferramenta, o Serviço intermediador e a camada de serviços da FSI para consulta a frames.

dos métodos *getFrame()* para obter os dados relativos ao frame, e para cada EF solicitou um *getIndividual()* para recolher algumas informações relativas aos EFs (descrição, cor, etc). Estas informações foram obtidas em XML, e em seguida transcritas e condensadas em um JSON para a ferramenta.

#### 4.2.5.1 Evidências Observadas na PoC 4

Nas provas de conceitos anteriores, pode-se observar a contribuição para o usuário quanto a confiabilidade dos dados documentados e quanto a integração dos dados com outras bases de conhecimento, permitindo um enriquecimento no processo de consulta a estes dados. Nesta PoC buscamos verificar o quanto a criação de uma interface de interação por meio de serviços pode contribuir para ferramentas que utilizam como base recursos lexicais baseados na semântica de frames, como o caso da FrameNet e do COPA 2014.

A partir da realização desta PoC pudemos observar que a criação da interface baseada em serviços, permite que ferramentas externas utilizem os dados mantidos na FSI, aproveitando assim das contribuições verificadas nas provas anteriores, como a confiabilidade dos dados promovida pelo uso de ontologias e o enriquecimento de informações promovidos pela utilização do padrão de dados ligados e ontologias, sem a necessidade que estes projetos sejam desenvolvidos com o objetivo de manipular ontologias, por exemplo. No entanto, algumas considerações devem ser feitas, considerando a interação do Dicionário COPA2014 e o framework FSI:

- Foi necessário o desenvolvimento de um conector para que a interação entre as ferramentas fosse possível. Consideramos o uso de um conector uma desvantagem, uma vez que a FSI foi desenvolvida com o objetivo de disponibilizar dados diretamente para aplicações/serviços consumidores. No entanto, devemos ressaltar que, no contexto da FSI, os dados são disponibilizados em formato padrão da Web semântica, ou seja, XML/RDF. Desta forma, cabe aos serviços consumidores estarem preparados para receber os dados neste padrão.
- Outro ponto importante, é que não foi explorado nesta POC, é o uso dos recursos de dados ligados e ontologias disponibilizados pela FSI. Considerando as dificuldades encontradas pelo uso de conectores, não foi possível avaliar estas funcionalidades, que são as principais contribuições da FSI. Desta forma, é necessário o desenvolvimento de novos experimentos para avaliação destas funcionalidades a partir da interface programática da FSI.

### 4.3 ANÁLISE DOS ESTUDOS

A partir do estudo realizado por meio destas PoCs foi possível obter evidências a respeito da viabilidade da proposta deste trabalho, sendo possível observar os benefícios obtidos quanto a utilização de ontologias e dados ligados no processo de documentação de frames em bases lexicais baseadas na semântica de frames.

Uma das evidências observadas neste aspecto, foi o fornecimento de informações quanto a coerência das relações entre Frames, e/ou entre Elementos de Frames, através da definição semântica das relações na ontologia Onto-Frame-BR. Estas informações evidenciam para o usuário a existência de inconsistências na documentação de um frame, reduzindo a possibilidade de inclusão de dados incorretos na base lexical.

Além disso, a especificação da ontologia Onto-Annotation-BR permite também o fornecimento de informações ao usuário a respeito da completude das anotações das sentenças em relação a estrutura do frame, permitindo ao usuário perceber se a anotação a ser realizada se adéqua ao frame selecionado durante a anotação.

Durante a terceira PoC também foi possível observar que o uso do padrão de dados ligados aplicados aos fragmentos anotados, formalizados na Onto-Annotation-BR, permite ao usuário o enriquecimento das informações contidas em uma sentença, tornando a

experiência de utilização dos dados de anotações contidos na base mais ricos, a partir do fornecimento de ligações com recursos de outras bases de dados ligados. O que no contexto da FrameNet é interessante, uma vez que a base lexical tem, entre seus objetivos, servir como uma fonte de informações úteis para diversas aplicações. Assim, o enriquecimento destas informações pode tornar estes dados mais úteis ao utilizados.

E por fim, pode-se observar também que a interface de comunicação por serviços permitiu a obtenção das informações mantidas pela FSI por outras ferramentas.

Apesar de verificar a validade da proposta com a realização das provas de conceito, ainda foram identificados alguns pontos que podem ser melhorados na abordagem proposta, sendo eles:

- Ampliar a Onto-Annotation-BR de maneira a contemplar também a segunda e terceira camada de anotações de sentenças, permitindo assim um maior enriquecimento da experiência do usuário durante a anotação de uma sentença.
- Aprimorar também a ferramenta de anotação do FnSEditor de maneira a contemplar também a segunda e a terceira camada de anotação de sentenças.
- Estudar outras plataformas alternativas ao StarDog, para processamento de ontologias extensas, uma vez que o StarDog apresentou algumas limitações quanto ao número de ontologias e no processo de explicação sobre sentenças obtidas a partir do reasoning de regras em SWRL.

#### 4.4 CONSIDERAÇÕES FINAIS DO CAPÍTULO

Este capítulo detalhou 4 provas de conceito realizadas, considerando a aplicação da FSI e suas ferramentas a um recurso lexical baseado em semântica de frames, o COPA 2014 FRAMENET-BR, envolvendo os processos de documentação de frames, anotações de sentenças, e utilização da interface de comunicação com outras ferramentas. Foram utilizadas as ontologias de domínio SWAN e PROTON para apoiar a identificação dos elementos dos domínios de futebol e turismo durante as anotações. A análise sobre os estudos realizados discutiu as principais contribuições que puderam ser observadas, bem como alguns pontos que podem ser melhorados.

Através destas avaliações e, de acordo com a hipótese levantada nesta dissertação, conseguiu-se obter indícios de que ao aplicarmos tecnologias de web semântica, como on-



tologias e dados ligados, a recursos lexicais baseados na semântica de frames buscando representar a semântica de sua estrutura, podemos promover um maior nível de confiabilidade de seus dados e também apoiar a experiência do usuários destas bases a partir da possibilidade de integração destes dados com outras bases de dados ligados. No entanto, é importante destacar que a proposta apresentada nesta dissertação é susceptível a avaliações experimentais adicionais, para que a hipótese seja plenamente avaliada.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

O uso de recursos lexicais a partir do projeto FrameNet tem sido cada vez mais destacado. Neste contexto, diversos autores vêm contribuindo com o processo de melhoria de acesso aos recursos lexicais, seu uso em diferentes aplicações e compartilhamento de informações relacionadas.

Conforme pôde ser observado durante o levantamento dos trabalhos relacionados, percebe-se os benefícios que tecnologias da web semântica, como ontologias e dados ligados, aplicados ao FrameNet pode trazer para a formalização de parte da estrutura de frames, em alguns casos como vocabulários formais e em alguns outros como formalização semântica de sua estrutura.

A proposta deste trabalho utiliza tecnologias relacionadas a Web semântica, combinando: i) o uso de ontologias que descrevem a estrutura dos frames e a semântica das relações entre estes frames, associado ao uso de ontologias de domínio para restrição semântica dos EFs; ii) o uso de padrão de dados ligados para enriquecimento das anotações de sentenças; e iii) disponibilizando o acesso aos dados por meio de uma camada de serviços, que permite uma maior integração com outros serviços e aplicações. Além disso, o trabalho propõe o uso de duas ferramentas, disponibilizadas em um portal, com o objetivo de apoiar o processo de documentação de frames e anotação de sentenças.

Considerando a hipótese do trabalho, a saber, "Ao utilizarmos tecnologias de web semântica, como ontologias e dados ligados, em conjunto com recursos lexicais baseados na semântica de frames podemos facilitar a utilização dos recursos lexicais por aplicações e ou usuários", e as POCS discutidas no capítulo 4 desta dissertação, pode-se dizer que foram obtidos indícios da validade da hipótese. No entanto, estudos experimentais mais formais devem ser executados, com o intuito de validar de maneira definitiva a abordagem.

### 5.1 CONTRIBUIÇÕES

Como contribuições da proposta deste trabalho, pode-se destacar a construção de uma infraestrutura baseada em tecnologias de Web Semântica e arquitetura baseada em serviços para promover um maior acesso a informações do recurso lexical e uma maior confiabilidade nos processos de documentação de frames e de anotação de sentenças.

Outra importante contribuição a ser ressaltada é a construção de uma ontologia, a ONTO-FRAME-BR, que além de representar formalmente a estrutura de frames, também se preocupa com a semântica das relações entre frames, e entre seus elementos, permitindo que durante o processo de documentação de frames possa ser fornecido ao usuário indicativos de possíveis erros durante a realização deste processo. É também contribuição desta abordagem, a construção da ONTO-ANNOTATION-BR que permite estruturar as anotações de sentenças de maneira que seus fragmentos possam ser relacionados aos elementos de frames documentados na ONTO-FRAME-BR e também que sejam utilizados como recursos de dados ligados.

Outro destaque da proposta é a possibilidade de uso de ontologias de domínio para relacionar recursos de dados ligados externos, presentes em fontes de recursos na Web, a fragmentos de sentenças anotadas a partir da classificação semântica do Elemento de Frame que este representa na cena, enriquecendo assim o volume de informações que pode ser obtido a partir da consulta a anotações. A construção de ferramentas para documentação de frames e realização de anotações tirando proveito das ontologias construídas pela proposta também é uma importante contribuição, onde é possível obter os benefícios observados pelo uso das ontologias (ONTO-FRAME-BR e ONTO-ANNOTATION-BR).

E por fim, também foi observado que a construção de uma interface baseada em serviços Web, que busca permitir a comunicação com ferramentas externas à infraestrutura contribui para a reutilização das informações obtidas a partir do recurso lexical, tornando possível por exemplo, a utilização por ferramentas de PLN ou ferramentas consultivas.

## 5.2 LIMITAÇÕES

No desenvolvimento da infraestrutura e de suas ferramentas foram observadas algumas limitações tanto relativas a tecnologias utilizadas quanto ao escopo adotado para a proposta. Dentre elas podemos destacar:

- O não tratamento da relação de herança entre frames por parte da ontologia devido a limitações da OWL e SWRL quanto a negação da existência de relações entre indivíduos no modelo OWA.
- As anotações de sentenças na FSI foram abordadas apenas em uma camada, enquanto na FrameNet as anotações de sentenças são realizadas em três camadas.

- Existência de problemas para identificação de algumas inferências por parte da ferramenta StarDog, que em algumas situações não consegue apresentar uma explicação para um inferência realizada por ele, normalmente quando envolve regras em SWRL mais complexas .
- Não foram disponibilizados indicadores padrão para os usuários nas ferramentas disponibilizadas no portal da FSI, sendo necessário obter informações normalmente por meio de consultas SPARQL (numero de frames anotados e etc).

Apesar destes pontos a serem trabalhados, acreditasse que o trabalho atingiu os seus objetivos, proporcionando uma infraestrutura que contribui para o projeto FrameNet tanto na questão de manutenção do projeto quanto para a utilização das informações por parte de usuários externos ao projeto.

### 5.3 TRABALHOS FUTUROS

Como trabalhos futuros podemos citar:

- Buscar uma alternativa ao StarDog para manipulação de ontologias com um grande volume de dados na sua ABOX, uma vez que podemos identificar algumas limitações referentes ao uso do StarDog no processo de explicação de inferências com o uso de regras SWRL complexas, e também quanto a limitação da licença gratuita do StarDog que foi utilizado para desenvolvimento da proposta.
- Estudar a possibilidade de ampliar o escopo abordado sobre as anotações de sentenças, uma vez que na proposta foi abordada somente a primeira camada. Esta ampliação demandaria, inicialmente, uma evolução da ontologia de anotações para contemplar os artefatos envolvidos nas demais camadas, entretanto deve-se observar o impacto de tentar descrever semanticamente estes artefatos.
- Outra questão a ser observada na infraestrutura é que as ontologias são mantidas atualmente por uma ferramenta de terceiros, a StarDog, sendo possível alterar as ontologias mantidas por ela apenas por meio do acesso a uma interface da própria ferramenta, ou por meio de um dos serviços disponibilizados pela FSI. Sendo assim, sente-se a falta nas ferramentas disponibilizadas na FSI de uma interface intuitiva ao usuário para realizar modificações nas ontologias mantidas pela infraestrutura,

uma vez que a interface do StarDog é voltada para usuários com conhecimento na área de ontologias e na linguagem SPARQL.

- Além disto, existe a possibilidade de expansão da proposta quanto a melhorias na sugestão de anotações de sentenças por parte da ferramenta do portal da FSI, com a possibilidade de integrar com ferramentas de rotulação automáticas de sentenças e desambiguação semântica, ou até mesmo a realização de tarefas deste tipo com base nas informações semânticas providas pelas ontologia ONTO-FRAME-BR.

## REFERÊNCIAS

- ALVES, R. C. V. **WEB SEMÂNTICA: uma análise focada no uso de metadados**. Tese (Doutorado) — Dissertação (Mestrado em Ciência da Informação)—Faculdade de Filosofia e Ciências—Universidade Estadual Paulista-UNESP, Marília/SP, 2005.
- ANTONIOU, G.; HARMELEN, F. V. **A semantic web primer**, 2004.
- AUER, S.; BIZER, C.; KOBILAROV, G.; LEHMANN, J.; CYGANIAK, R.; IVES, Z. **Dbpedia: A nucleus for a web of open data**, 2007.
- BASILI, V. **GQM Approach Has Evolved To Include Models**, 1994. Disponível em: <<http://www.cs.umd.edu/basili/publications/technical/T86.pdf>>.
- BECHHOOFER, S.; HARMELEN, F.; HENDLER, J.; HORROCKS, I.; MCGUINNESS, D. L.; PATEL-SCHNEIDER, P. F.; STEAIN, L. A. **OWL Web Ontology Language Reference**. 2004. Disponível em: <<http://www.w3.org/TR/owl-ref/>>.
- BERNERS-LEE, T. **Linked Data - Design Issues**. jun. 2009. Disponível em: <<http://www.w3.org/DesignIssues/LinkedData.html>>.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 1–5, 2001.
- BLAIMER, N.; BORTFELDT, A.; PANKRATZ, G. **Patterns in object-oriented analysis**, 2010. Disponível em: <<https://www.fernuni-hagen.de/wirtschaftswissenschaft/download/beitraege/db451.pdf>>.
- BOEHM, B.; BASILI, V. R. Software defect reduction top 10 list. **Foundations of empirical software engineering: the legacy of Victor R. Basili**, v. 426, 2005.
- BOLEY, H.; PASCHKE, A.; SHAFIQ, O. Ruleml 1.0: the overarching specification of web rules. **Lecture Notes in Computer Science**, v. 6403, n. 4, p. 162–178, 2010.

- BOOTH, D.; HAAS, H.; MCCABE, F.; NEWCOMER, E.; CHAMPION, M.; FERRIS, C.; ORCHARD, D. **Web Service Architecture**. 2004. Disponível em: <<http://www.w3.org/TR/ws-arch/>>.
- BRAY, T.; PAOLI, J.; SPERBERG-MCQUEEN, C. M.; MALER, E.; YERGEAU, F. **Extensible Markup Language (XML) 1.0 (Fifth Edition) - W3C Recommendation**. nov. 2008. Disponível em: <<http://www.w3.org/TR/XML/>>.
- BRUIJN, J.; BUSSLER, C.; DOMINGUE, J.; FENSEL, D.; HEPP, M.; KELLER, U.; KIFER, M.; KÖNIG-RIES, B.; KOPECKY, J.; LARA, R.; LAUSEN, H.; OREN, E.; POLLERES, A.; ROMAN, D.; SCICLUNA, J.; STOLLBERG, M. **Web Service Modeling Ontology (WSMO)**. jun. 2003. Disponível em: <<http://www.w3.org/submission/wsmo>>.
- CATALDI, M.; DAMIANO, R.; LOMBARDO, V.; PIZZO, A.; SERGI, D. Integrating commonsense knowledge into the semantic annotation of narrative media objects. In: **AI\* IA 2011: Artificial Intelligence Around Man and Beyond**, 2011. p. 312–323.
- CHINNICI, R.; MOREAU, J. J.; RYMAN, A.; WEERAWARANA, S. **Web Services Description Language (WSDL) Version 2.0 Part 1: Core Language**. jun. 2007. Disponível em: <<http://www.w3.org/TR/wsdl20/>>.
- COSTA, G. C. B. **Uma Abordagem para Linha de Produtos de Software Científico Baseada em Ontologia e Workflow**. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, 2013.
- CYGANIAK, R.; JENTZSCH, A. **Linking Open Data cloud diagram**. 2011. Disponível em: <<http://lod-cloud.net/>>.
- DAVIS, C. A.; ALVES, L. L. Local spatial data infrastructures based on a service-oriented architecture. In: **GeoInfo**, 2005. p. 30–48.
- DIETER, F. **Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce**. 2. ed., 2003. ISBN 3540003029.

- DRUMMOND, N.; SHEARER, R. The open world assumption. In: **eSI Workshop: The Closed World of Databases meets the Open World of the Semantic Web**, 2006.
- ECLIPSE. **Eclipse - The Eclipse Foundation open source community website**. 2014. Disponível em: <<https://www.eclipse.org>>.
- ENDREI, M.; ANG, J.; ARSANJANI, A.; CHUA, S.; COMTE, P.; KROGDAHL, P.; LUO, M.; NEWLING, T. **Patterns: service-oriented architecture and web services**, 2004.
- ERK, K.; KOWALSKI, A.; PADÓ, S.; PINKAL, M. Towards a resource for lexical semantics: A large german corpus with extensive semantic annotation. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1**, 2003. p. 537–544.
- FARRELL, J.; LAUSEN, H. **Semantic Annotations for WSDL and XML Schema. W3c Recommendation**. ago. 2007. Disponível em: <<http://www.w3.org/TR/sawSDL/>>.
- FENSEL, D.; HARMELEN, F. V.; HORROCKS, I.; MCGUINNESS, D. L.; PATEL-SCHNEIDER, P. F. Oil: An ontology infrastructure for the semantic web. **IEEE intelligent systems**, IEEE Computer Society, v. 16, n. 2, p. 38–45, 2001.
- FERNANDEZ, E. B.; YUAN, X. Semantic analysis patterns. In: **Conceptual Modeling - ER 2000**, 2000. p. 183–195.
- FILLMORE, C. Topics in lexical semantics. 1977.
- FILLMORE, C. J. Frame semantics. In: HANSHIN PUBLISHING CO. **Linguistics in the Morning Calm**, 1982. p. 111–137.
- GAMONAL, M. A. Análise da unidade lexical visitar do frame de turismo. **Anais do SILEL**, n. 2, 2011.
- GENESERETH, M. R.; FIKES, R. E. et al. Knowledge interchange format-version 3.0: Reference manual. Computer Science Department, Stanford University San Francisco, CA, 1992.



- GRUBER, T. R. **Ontolingua: A mechanism to support portable ontologies**, 1992.
- GUARINO, N. **Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy**, 1998.
- HARTIG, O.; ZHAO, J. Publishing and consuming provenance metadata on the web of linked data. In: **Provenance and annotation of data and processes**, 2010. p. 78–90.
- HEATH, T.; BIZER, C. Linked data: Evolving the web into a global data space. **Synthesis lectures on the semantic web: theory and technology**, Morgan & Claypool Publishers, v. 1, n. 1, p. 1–136, 2011.
- HÖPKEN, W.; CLISSMANN, C. **Tourism Harmonisation Trans-European Network - Final Ontology Report**. jan. 2006. Disponível em: <[http://euromuse.harmonet.org/c/document\\_library/get\\_file?uuid=00d5d76d-f150-4632-ab1e-b29ffac89818&groupId=14](http://euromuse.harmonet.org/c/document_library/get_file?uuid=00d5d76d-f150-4632-ab1e-b29ffac89818&groupId=14)>.
- HORROCKS, I. Ontologies and the semantic web. **Communications of the ACM**, ACM, v. 51, n. 12, p. 58–67, 2008.
- HORROCKS, I. et al. Daml+oil: a description logic for the semantic web. **IEEE Data Eng. Bull.**, v. 25, n. 1, p. 4–9, 2002.
- HORROCKS, I.; PATEL-SCHNEIDER, P. F.; BOLEY, H.; TABET, S.; GROSOFF, B.; DEAN, M. et al. Swrl: A semantic web rule language combining owl and ruleml. **W3C Member submission**, v. 21, p. 79, 2004.
- ICSI. **The framenet project**. 2012. Disponível em: <<https://framenet.icsi.berkeley.edu/fndrupal/>>.
- INFRASTRUCTURES, D. S. D. the sdi cookbook. **GSDI/Nebert**, 2004.
- JAVA. **java.com: Java + Você**. 2014. Disponível em: <[http://java.com/pt\\_BR/](http://java.com/pt_BR/)>.
- JDOM. **JDOM**. Disponível em: <<http://www.jdom.org/> Year = 2012>.
- JENA. **Apache Jena - Home**. 2013. Disponível em: <<http://www.jena.apache.org/>>.
- KLYNE, G.; CARROLL, J. J. Resource description framework (rdf): Concepts and abstract syntax. 2006.

- LAGE, L. M. Frames e construções: Um estudo de caso da construção [sn vdeixar para/por vinfinitivo] e do frame deixado\_por\_fazer. **Anais do SILEL**, n. 2, 2011.
- LEENOI, D.; JUMPATHONG, S.; PORKAEW, P.; SUPNITHI, T. Thai framenet construction and tools. **Int. J. of Asian Lang. Proc.**, v. 21, n. 2, p. 71–82, 2011.
- MARTIN, D.; BURSTEIN, M.; HOBBS, J.; LASSILA, O.; MCDERMOTT, D.; MCILRAITH, S.; NARAYANAN, S.; PAOLUCCI, M.; PARSIA, B.; PAUNE, T.; SIRIN, E.; SRINIVASAN, N.; SYCARA, K. **OWL-S: Semantic Markup for Web Services**. nov. 2004. Disponível em: <<http://www.w3.org/submission/owl-s/>>.
- MCILRAITH, S. A.; SON, T. C.; ZENG, H. Semantic web services. **IEEE intelligent systems**, IEEE Computer Society, v. 16, n. 2, p. 46–53, 2001.
- MÖLLER, K. **SWAN Soccer Ontology**. maio 2005. Disponível em: <<http://sw.deri.org/2005/05/swan/soccer/ontology/soccer.owl/>>.
- MORAIS, E. A. M.; AMBROSIO, A. P. L. **Ontologias: conceitos, tipos, metodologias, ferramentas e linguagens**.
- MOREAU, L.; CLIFFORD, B.; FREIRE, J.; FUTRELLE, J.; GIL, Y.; GROTH, P.; KWASNIKOWSKA, N.; MILES, S.; MISSIER, P.; MYERS, J. et al. The open provenance model core specification (v1. 1). **Future Generation Computer Systems**, Elsevier, v. 27, n. 6, p. 743–756, 2011.
- MYSQL. **MySQL:: The world's most popular open source database**. 2014. Disponível em: <<http://www.mysql.com/>>.
- NETBEANS. **Welcome to NetBeans**. 2014. Disponível em: <<http://www.netbeans.org/>>.
- NUZZOLESE, A. G.; GANGEMI, A.; PRESUTTI, V. Gathering lexical linked data and knowledge patterns from framenet. In: ACM. **Proceedings of the sixth international conference on Knowledge capture**, 2011. p. 41–48.
- O'CONNOR, M.; KNUBLAUCH, H.; TU, S.; GROSOFF, B.; DEAN, M.; GROSSO, W.; MUSEN, M. Supporting rule system interoperability on the semantic web with swrl. In: **The Semantic Web–ISWC 2005**, 2005. p. 974–986.

- PANTOQUILHO, M.; RAMINHOS, R.; ARAUJO, J. Analysis patterns specifications: filling the gaps. In: **VIKING PLOP CONFERENCE**, 2003. v. 2, p. 169–180.
- POPOV, B.; KIRYAKOV, A.; KIRILOV, A.; MANOV, D.; OGNYANOFF, D.; GORANOV, M. Kim–semantic annotation platform. In: **The Semantic Web-ISWC 2003**, 2003. p. 834–849.
- PRATNER, K. **OnTour: The Ontology**. 2013. Disponível em: <<http://e-tourism.derit.at/ont/docu2004/OnTour - The Ontology.pdf>>.
- PROTÉGÉ. **Protégé**. 2012. Disponível em: <<http://protege.stanford.edu/>>.
- RAJABIFARD, A.; WILLIAMSON, I. P. Spatial data infrastructures: concept, sdi hierarchy and future directions. 2001.
- RAMINHOS, R.; PANTOQUILHO, M.; ARAÚJO, J.; MOREIRA, A. A systematic analysis patterns specification. In: **ICEIS (3)**, 2006. p. 453–456.
- RUPPENHOFER, J.; ELLSWORTH, M.; PETRUCK, M. R.; JOHNSON, C. R.; SCHEFFCZYK, J. **FrameNet II: Extended Theory and Practice**, 2006. Distributed with the FrameNet data.
- SALOMAO, M. M. M. Framenet brasil: um trabalho em progresso. **Calidoscópico**, v. 7, n. 3, p. 171–182, 2009.
- SALOMAO, M. M. M.; TORRENT, T.; CAMPOS, F.; BRAGA, R.; VIEIRA, M. Copa 2014 framenet brasil. **Anais do SILEL**, n. 2, 2011.
- SCHEFFCZYK, J.; BAKER, C. F.; NARAYANAN, S. Ontology-based reasoning about lexical resources. In: CITESEER. **Proc. of OntoLex**, 2006. p. 1–8.
- SCHMIDT, T. The kicktionary—a multilingual lexical resource of football language. de Gruyter, 2009.
- SILVA, G. C. B. D. **Composer-science: um framework para a composição de workflows científicos**. Dissertação (Mestrado) — Universidade Federal de Juiz de Fora, 2010.

- STARDOG. **StarDog: The RDF Database**. 2013. Disponível em: <<http://stardog.com/>>.
- TERZIEV, I.; KIRYAKOV, A.; MANOV, D. Base upper-level ontology (bulo) guidance. **SEKT deliverable**, v. 1, n. 1, 2005.
- THIERRY, D.; WITTENBURG, P. Mumis—a multimedia indexing and searching environment. In: **Proc. of The First International Workshop on Multimedia Annotation, Japon**, 2001.
- TORRENT, T. T.; SALOMÃO, M. M. M.; CAMPOS, F. C.; BRAGA, R. M.; MATOS, E. E.; GAMONAL, M. A.; GONÇALVES, J. A.; SOUZA, B. C.; GOMES, D. S.; PERON, S. R. Copa 2014 framenet brasil: a frame-based trilingual electronic dictionary for the football world cup. **COLING 2014**, p. 10, 2014.
- VEGI, L. **Technical description of Dublin Core application profile to analysis patterns(DC2AP)**. 2012.
- VEGI, L. F.; PEIXOTO, D. A.; SOARES, L. S.; LISBOA-FILHO, J.; OLIVEIRA, A. de P. An infrastructure oriented for cataloging services and reuse of analysis patterns. In: SPRINGER. **Business Process Management Workshops**, 2012. p. 338–343.
- WESSEL, M.; MÖLLER, R. A high performance semantic web query answering engine. **Description Logics**, v. 147, 2005.
- YU, L. **A developer’s guide to the semantic web**, 2011.
- ZHAO, J.; HARTIG, O. Towards interoperable provenance publication on the linked data web. In: **LDOW**, 2012.
- ZHOU, J.; MA, L.; LIU, Q.; ZHANG, L.; YU, Y.; PAN, Y. Minerva: A scalable owl ontology storage and inference system. In: **The Semantic Web—ASWC 2006**, 2006. p. 429–443.

## Apêndice A - ONTOLOGIA ONTO-FRAME-BR

A ontologia ONTO-FRAME-BR foi construída com o objetivo de representar semanticamente os termos utilizados na documentação de frames no projeto FrameNet, assim como definir a semântica das relações entre frames e relações entre Elementos de Frame.

Neste documento é representada a documentação das principais classes da ontologia por meio da ferramenta Protégè.

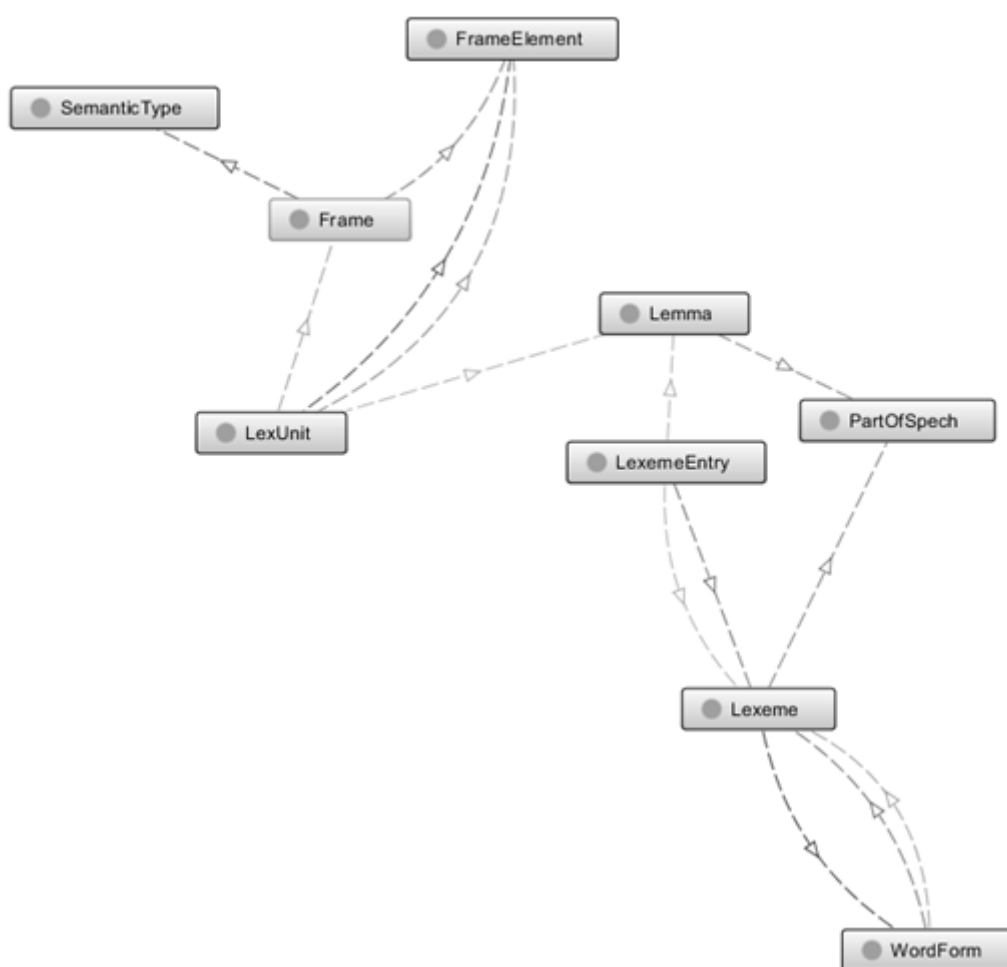


Figura A.1: Visão geral das principais classes e seus relacionamentos.

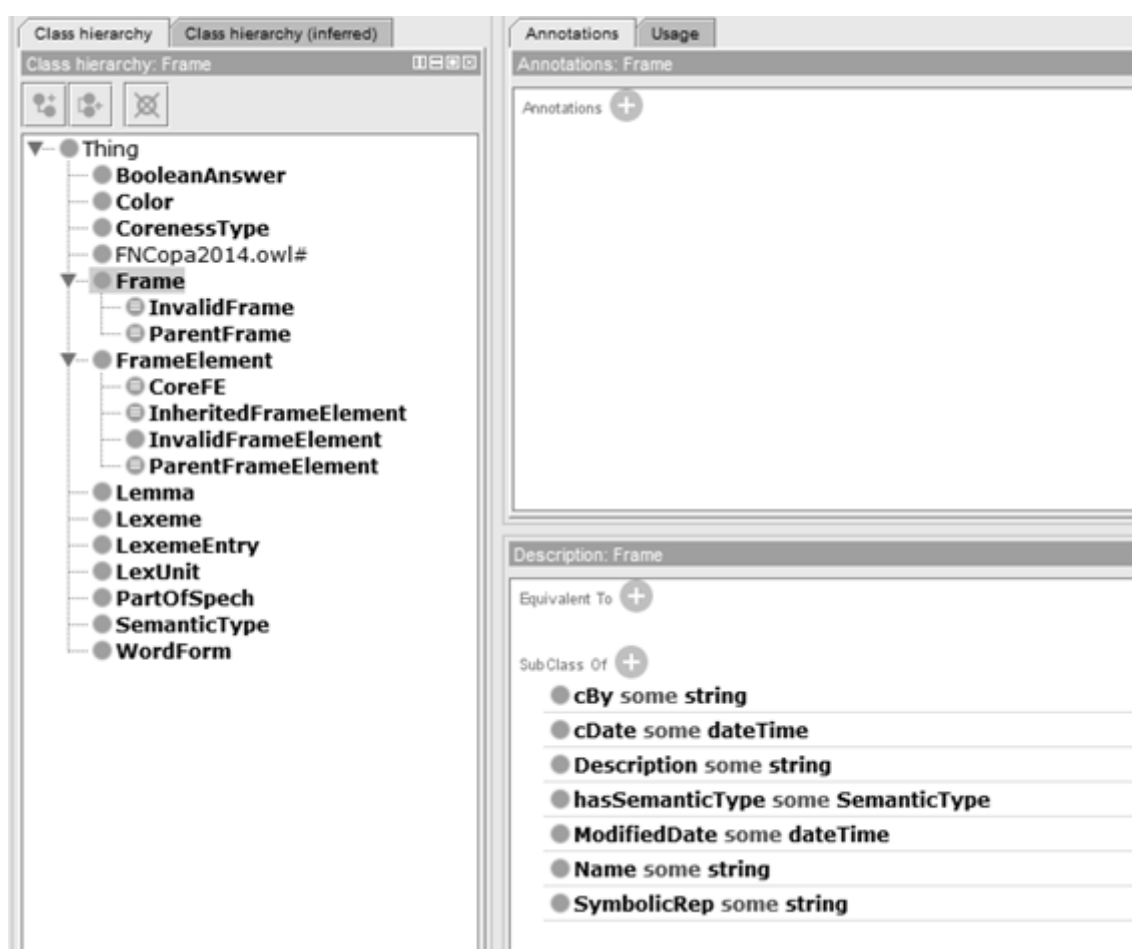


Figura A.2: Representação da classe Frame no Protégè.

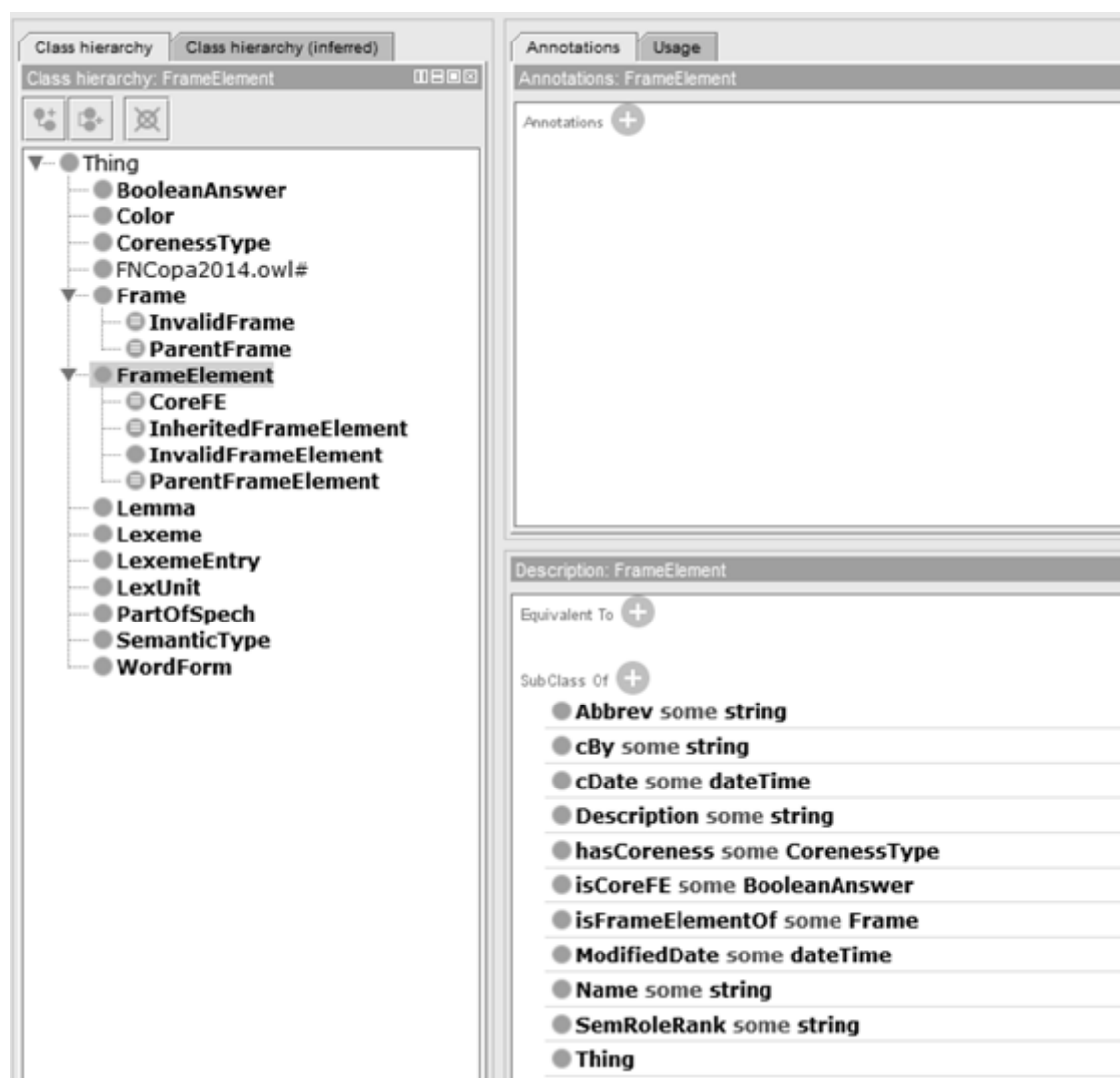


Figura A.3: Representação da classe `FrameElement` no Protégè.

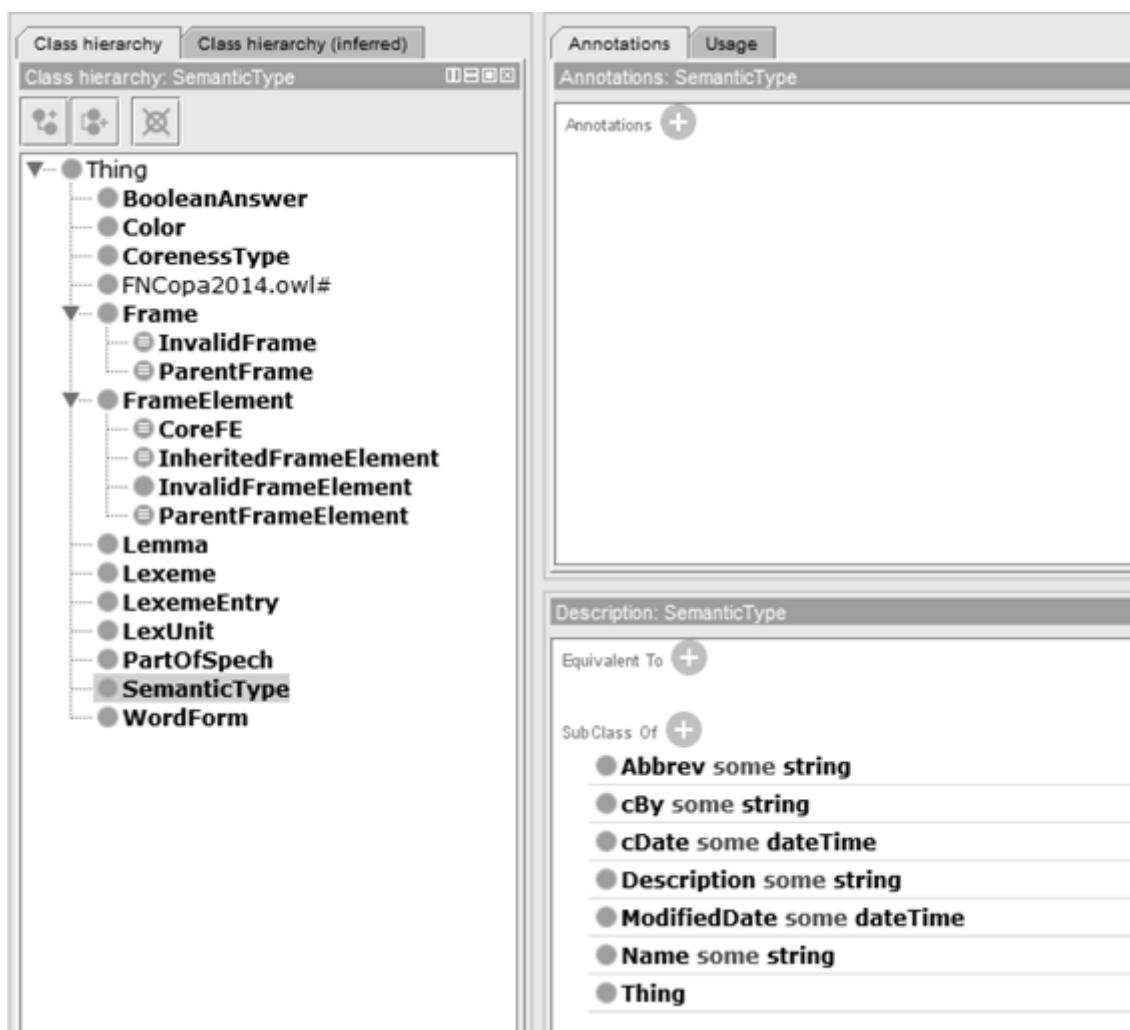


Figura A.4: Representação da classe `SemanticType` no Protégè.



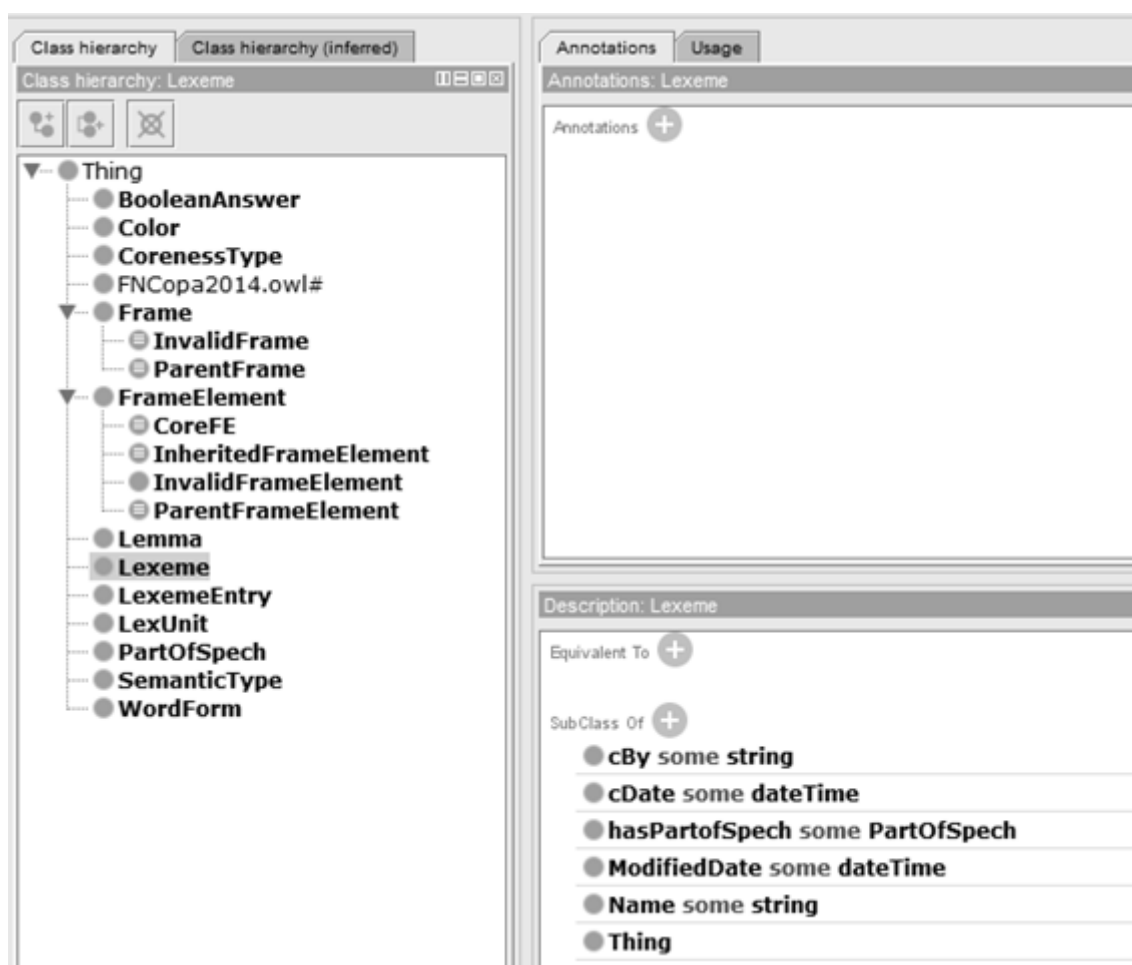


Figura A.5: Representação da classe Lexeme no Protégè.

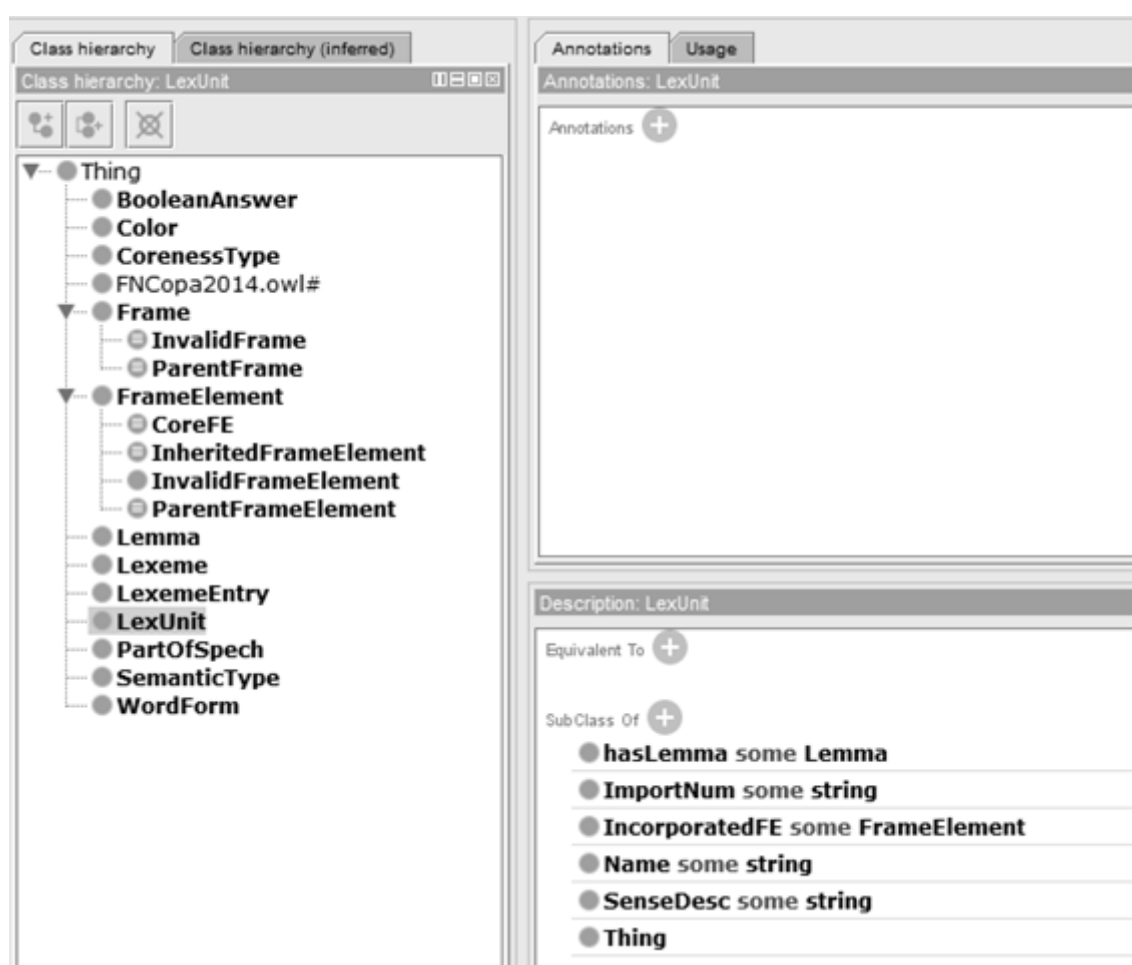


Figura A.6: Representação da classe LexUnit no Protégè.

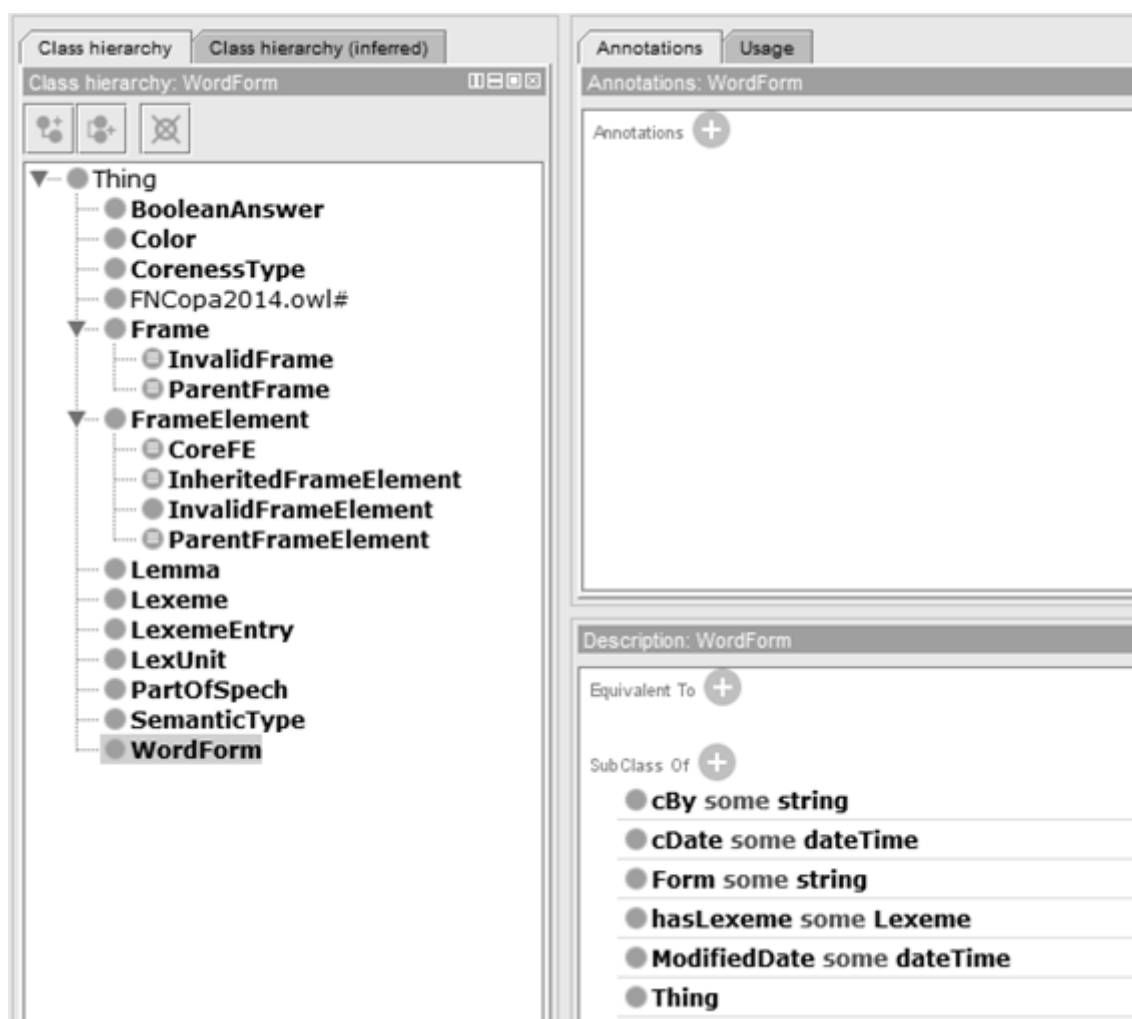


Figura A.7: Representação da classe WordForm no Protégè

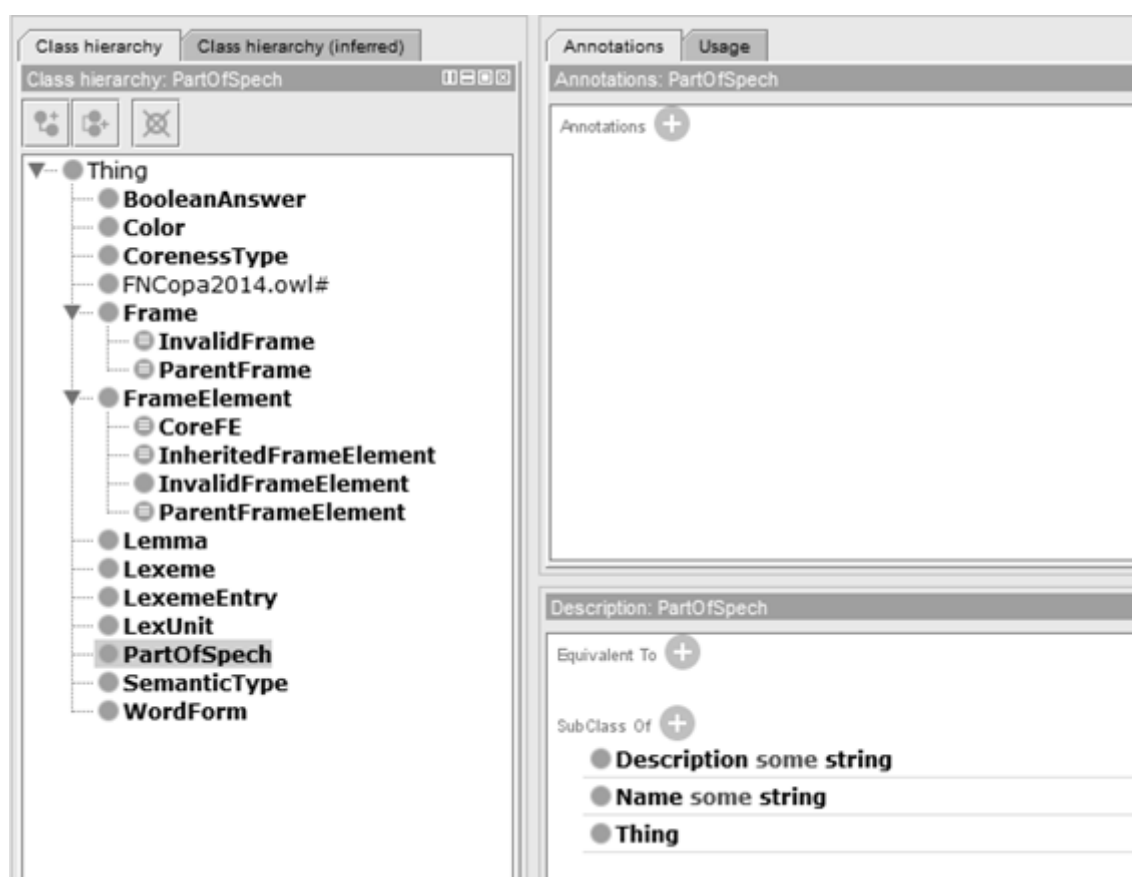


Figura A.8: Representação da classe PartOfSpeech no Protégè.