



UNIVERSIDADE FEDERAL DO CEARÁ
CAMPUS DE RUSSAS
CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FRANCISCA TÁGILA LIMA DA SILVA

**CONSTRUÇÃO DE UMA ONTOLOGIA PARA RESPONDER PERGUNTAS
FACTUAIS SOBRE A COPA DA RÚSSIA 2018 EM LINGUAGEM NATURAL**

RUSSAS

2019

FRANCISCA TÁGILA LIMA DA SILVA

CONSTRUÇÃO DE UMA ONTOLOGIA PARA RESPONDER PERGUNTAS FACTUAIS
SOBRE A COPA DA RÚSSIA 2018 EM LINGUAGEM NATURAL

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em ciência da computação
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em ciência da computação.

Orientador: Dr. Alexandre Matos Arruda

RUSSAS

2019

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- S58c Silva, Francisca Tágila Lima da.
Construção de uma ontologia para responder perguntas factuais sobre a copa da Rússia 2018 em linguagem natural / Francisca Tágila Lima da Silva. – 2019.
69 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas, Curso de Ciência da Computação, Russas, 2019.
Orientação: Prof. Dr. Alexandre Matos Arruda.
1. Processamento de Linguagem Natural. 2. Ontologia. 3. Linguagem Natural Controlada. 4. SPARQL.
I. Título.

CDD 005

FRANCISCA TÁGILA LIMA DA SILVA

CONSTRUÇÃO DE UMA ONTOLOGIA PARA RESPONDER PERGUNTAS FACTUAIS
SOBRE A COPA DA RÚSSIA 2018 EM LINGUAGEM NATURAL

Trabalho de Conclusão de Curso apresentado ao
Curso de Graduação em ciência da computação
do Campus de Russas da Universidade Federal
do Ceará, como requisito parcial à obtenção do
grau de bacharel em ciência da computação.

Aprovada em: 19 de junho de 2019

BANCA EXAMINADORA

Dr. Alexandre Matos Arruda (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Marcio Costa Santos
Universidade Federal do Ceará (UFC)

Prof. Ms. Tatiane Fernandes Figueiredo
Universidade Federal do Ceará (UFC)

Aos meus pais, irmãs, amigos e ao meu namorado que mesmo diante das dificuldades que enfrentei, acreditaram em mim quando eu não mais acreditava e que estenderam suas mãos nos momentos mais difíceis.

AGRADECIMENTOS

Primeiramente quero agradecer a Deus pela oportunidade, por me guiar em cada passo dessa jornada e me mostrar que para ele nada é impossível.

A Universidade Federal do Ceará pelo curso de Ciência da Computação que me proporcionou vários momentos de aprendizado.

Ao Prof. Dr. Alexandre Matos Arruda pela sua orientação neste trabalho.

A todo o corpo docente que mantiveram firme o dever de ensinar e passar o melhor para seus alunos.

A minha família que acompanharam de perto toda a minha jornada. Ao meu pai João Batista da Silva e minha mãe Maria de Fátima Lima pelo apoio em cada decisão e paciência constante. As minhas irmãs e amigas de nascença, Tatiana, Thais, Rosiane e Elenia que fizeram parte de cada momento e me deram força constante, tanto financeira quanto familiar.

Ao meu namorado Nathanael, pelo seu amor e cuidado, que esteve comigo em cada situação, seja de tristeza, alegria ou até mesmo nos momentos de desespero.

Aos amigos conquistados em sala de aula, Marília Cristina, Paloma Bispo, Erik Almeida, Carlos Victor, Guilherme Sombra, Hugo Venâncio, Igor Mendes, Isaias Ferreira, Marcos Paulo, Thomas Dillan, Vinicius Almeida, Isaac Rahel, Alex Frederico, Marcos de Alencar, Mateus Oliveira, Lavinia Matoso, Alex Felipe, Thiago Hellen e Bárbara Feijão, que foram fonte de inspiração e de luta durante todo o curso, superando juntos as dificuldades. Aos amigos conquistados ao final da graduação Elis Ionara, Nathalia Gonçalves, Alex Santos, Sabrina Oliver e Paula Gama, que acreditaram em mim e no meu potencial me proporcionando muitas conversas, conselhos e risadas. E a minha amiga de longa data Williana Paulino pelos momentos de lazer e bagunça.

A todos, o meu imenso obrigado!

“Mas graças a Deus, que sempre nos conduz victoriosamente em Cristo e por nosso intermédio exala em todo lugar a fragrância do seu conhecimento.”

(2 Coríntios 2:14)

RESUMO

A Web é algo hoje bastante cotidiano na vida das pessoas, possibilitando a compra de produtos online sem sair de casa ou até mesmo as trocas de mensagens através das redes sociais. Para o olho humano é comum o entendimento de uma página da web, mas para uma máquina tudo não passa de conteúdos com representações formais ou estruturais. Com isso o surgimento da Web Semântica vem para trazer meios de introduzir o lado semântico e o significado das informações contidas nas páginas web, a fim de que seja feito um mapeamento correto do que está sendo transferido e o que está sendo entendido. As máquinas devem estar preparadas para essa representação do conhecimento. Com isso, formalismos foram desenvolvidos para a inclusão da semântica na representação do conhecimento através das páginas web, como por exemplo o RDF - Resource Description Framework e as ontologias. O primeiro representa triplas incluindo sujeito, predicado e objeto, podendo formar uma grande quantidade de informações que possuem relação uma com as outras. Já o segundo representa uma coleção de informações para um determinado assunto, podendo ser usada para melhorar as buscas na web através da linguagem natural. Portanto, o trabalho aqui descrito tem como objetivo desenvolver uma ontologia para um domínio específico referente à Copa Russa 2018, onde as consultas podem ser feitas em linguagem natural e transformadas em uma linguagem de consulta formal, optamos pelo SPARQL devido ao seu grande uso na Web Semântica. acesso a ontologias.

Palavras-chave: Processamento de Linguagem Natural. Ontologia. Linguagem Natural Controlada. SPARQL.

ABSTRACT

The Web is something quite common in people's lives, making it possible to buy products online without leaving home or even exchanging messages through social networks. For the human eye is common the understanding of a web page, but for a machine everything is merely content with formal or structural representations. With this, the emergence of the Semantic Web comes to bring the means to introduce the semantic side and the meaning of the information contained in the web pages, so that a correct mapping of what is being transferred and what is being understood is done. The machines must be prepared for this representation of knowledge. Thus, formalisms have been developed for the inclusion of semantics in the representation of knowledge through web pages, such as the RDF - Resource Description Framework and the ontologies. The first represents triples including subject, predicate and object, being able to form a great amount of information that has relation with each other. The second represents a collection of information for a particular subject, and can be used to improve web searches through natural language. Therefore, the work described here goal to develop an ontology for a specific domain referring to the 2018 Russian Cup where queries can be made in natural language and transformed into a formal query language, we chose SPARQL due to its great use in the Semantic Web for access to ontologies.

Keywords: Natural language processing. Ontology. Controlled natural language. SPARQL.

LISTA DE FIGURAS

Figura 1 – Relação entre propriedades e indivíduos	21
Figura 2 – Ferramentas mais utilizadas nas organizações segundo entrevista.	26
Figura 3 – Tela para definição de classes no Protégé.	27
Figura 4 – Respostas relacionadas as partidas da Copa.	29
Figura 5 – Respostas relacionadas aos Jogadores da Copa.	29
Figura 6 – Respostas relacionadas a cada Seleção da Copa.	30
Figura 7 – Respostas relacionadas a Copa em Geral.	30
Figura 8 – Classes <i>disjoint</i> da classe time na ontologia.	33
Figura 9 – Relação binária entre Jogador e Time, através da <i>Object Property</i> tem_jogador.	33
Figura 10 – Exemplo de uma propriedade inversa da ontologia.	34
Figura 11 – Exemplo de uma propriedade funcional da ontologia.	34
Figura 12 – Hierarquia de classes da ontologia.	37
Figura 13 – Relações da instância Tite da classe Tecnico.	40
Figura 14 – Relações da instância Brasil da classe Time.	40
Figura 15 – Relações da instância Neymar da classe Jogador.	40
Figura 16 – Relações da instância Final da classe Final subclasse de Partida.	40
Figura 17 – Árvore de derivação para a sentença, <i>qual o técnico do Brasil?</i>	42
Figura 18 – Árvore de derivação para a sentença <i>qual é a posição do jogador Neymar?</i>	43
Figura 19 – Árvore de derivação para a sentença, <i>qual os times que jogaram na partida final?</i>	45
Figura 20 – Árvore de derivação para a sentença, <i>quais os times das oitavas de final 1?</i>	47
Figura 21 – Árvore de derivação para a sentença <i>qual o time vencedor na partida entre França e Croácia?</i>	48
Figura 22 – Árvore de derivação para o SPARQL da sentença S1.	51
Figura 23 – Árvore de derivação para o SPARQL da sentença S2.	51
Figura 24 – Árvore de derivação para a sentença, <i>quem é o técnico do Brasil?</i>	60
Figura 25 – Árvore de derivação para a sentença, <i>quem são os jogadores do Brasil?</i>	61
Figura 26 – Árvore de derivação para a sentença, <i>que horas aconteceu a partida final?</i>	62
Figura 27 – Árvore de derivação para a sentença, <i>em que dia aconteceu a partida da semifinal 1?</i>	64
Figura 28 – Árvore de derivação para a sentença, <i>onde aconteceu a partida final?</i>	65

Figura 29 – Árvore de derivação para a sentença, <i>quando aconteceu a partida de oitavas de final 2?</i>	66
Figura 30 – Árvore de derivação para a sentença, <i>quantos gols a França fez na partida final?</i>	67
Figura 31 – Árvore de derivação para a sentença, <i>quantos títulos tem o Brasil?</i>	69

LISTA DE TABELAS

Tabela 1 – Classes da ontologia e suas características	38
Tabela 2 – Propriedades de objeto (<i>Object Property</i>) da ontologia.	39
Tabela 3 – Propriedades de dados (<i>Data property</i>) da ontologia.	39
Tabela 4 – Elementos sintáticos da sentença, <i>qual o técnico do Brasil?</i>	41
Tabela 5 – Agrupamento dos sintagmas da sentença, <i>qual o técnico do Brasil?</i>	41
Tabela 6 – Elementos sintáticos da sentença, <i>qual é a posição do jogador Neymar?</i>	42
Tabela 7 – Agrupamento dos sintagmas da sentença, <i>qual é a posição do jogador Neymar?</i>	42
Tabela 8 – Elementos sintáticos da sentença, <i>quais os times que jogaram na partida final?</i>	44
Tabela 9 – Agrupamento dos sintagmas da sentença, <i>quais os times que jogaram na partida final?</i>	44
Tabela 10 – Elementos sintáticos da sentença, <i>quais os times das oitavas de final 1?</i>	46
Tabela 11 – Agrupamento dos sintagmas da sentença, <i>quais os times das oitavas de final 1?</i>	46
Tabela 12 – Elementos sintáticos da sentença, <i>qual o time vencedor na partida entre França e Croácia?</i>	47
Tabela 13 – Agrupamento dos sintagmas da sentença <i>qual o time vencedor na partida entre França e Croácia?</i>	47
Tabela 14 – Exemplo de dicionário para a ontologia.	50
Tabela 15 – Elementos sintáticos da sentença, <i>quem é o técnico do Brasil?</i>	59
Tabela 16 – Agrupamento dos sintagmas da sentença, <i>quem é o técnico do Brasil?</i>	59
Tabela 17 – Elementos sintáticos da sentença, <i>que horas aconteceu a partida final?</i>	61
Tabela 18 – Agrupamento dos sintagmas da sentença, <i>que horas aconteceu a partida final?</i>	61
Tabela 19 – Elementos sintáticos da sentença, <i>em que dia aconteceu a partida da semifinal 1?</i>	62
Tabela 20 – Agrupamento dos sintagmas da sentença, <i>em que dia aconteceu a partida da semifinal 1?</i>	63
Tabela 21 – Elementos sintáticos da sentença, <i>onde aconteceu a partida final?</i>	63
Tabela 22 – Agrupamento dos sintagmas da sentença, <i>onde aconteceu a partida final?</i>	64
Tabela 23 – Elementos sintáticos da sentença, <i>quando aconteceu a partida de oitavas de final 2?</i>	65
Tabela 24 – Elementos sintáticos da sentença, <i>quantos gols a França fez na partida final?</i>	66

Tabela 25 – Agrupamento dos sintagmas da sentença, <i>quantos gols a França fez na partida final?</i>	67
Tabela 26 – Elementos sintáticos da sentença, <i>quantos títulos tem o Brasil?</i>	68
Tabela 27 – Agrupamento dos sintagmas da sentença, <i>quantos títulos tem o Brasil?</i>	68

LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Sintaxe da linguagem SPARQL	23
Código-fonte 2 – Hierarquia de classes	27
Código-fonte 3 – Consulta SPARQL após as devidas alterações.	52

LISTA DE ABREVIATURAS E SIGLAS

LN	Linguagem Natural
LNC	Linguagem Natural Controlada
OWL	Web Ontology Language
PLN	Processamento de Linguagem Natural
RDF	Resource Description Framework
SQL	Structured Query Language
URI	Uniform Resource Identifier
W3C	World Wide Web
WS	Web Semântica

SUMÁRIO

1	INTRODUÇÃO	17
2	OBJETIVOS	19
2.1	Objetivo geral	19
2.2	Objetivos específicos	19
3	FUNDAMENTAÇÃO TEÓRICA	20
3.1	Processamento de Linguagem Natural	20
3.2	Linguagem Natural Controlada	20
3.3	Ontologias	20
3.4	Gramáticas	21
3.5	Consultas SPARQL	22
3.5.1	<i>Sintaxe</i>	22
4	TRABALHOS RELACIONADOS	24
5	METODOLOGIA	26
5.1	Estudo de Ferramentas	26
5.1.1	<i>Protégé</i>	26
5.1.2	<i>Reasoner</i>	28
5.2	Levantamento de dados para o domínio	28
5.3	Construção da Ontologia	31
5.4	Definição e construção da Linguagem Controlada	34
5.5	Construção das Consultas SPARQL	35
5.6	Validação da ontologia	36
6	RESULTADOS	37
6.1	Base ontológica	37
6.2	Conversão da LNC para o SPARQL	41
6.2.1	<i>Tratando os tipos de consultas da LNC</i>	41
6.2.1.1	<i>Limitações da LNC</i>	49
6.2.2	<i>Gerando consultas SPARQL</i>	49
6.2.3	<i>Considerações sobre a conversão</i>	52
7	CONCLUSÕES E TRABALHOS FUTUROS	53
	REFERÊNCIAS	54

APÊNDICES	56
APÊNDICE A – Questionário utilizado para coletar dados sobre a copa da Rússia 2018	56
APÊNDICE B – Continuação das análises da Subseção 6.2.1	59
B.0.1 <i>Tratando consultas do tipo Quem?</i>	59
B.0.2 <i>Tratando consultas do tipo Que?</i>	61
B.0.3 <i>Tratando consultas do tipo Onde?</i>	63
B.0.4 <i>Tratando consultas do tipo Quando?</i>	65
B.0.5 <i>Tratando consultas do tipo Quanto?</i>	66

1 INTRODUÇÃO

As informações na Web têm crescido bastante à medida em que novas tecnologias são desenvolvidas, esse enorme volume de informações dificulta bastante quando se deseja buscar o que quer (VANTI, 2002).

Novas técnicas e ferramentas são desenvolvidas nas áreas de Processamento de Linguagem Natural (PLN) e Recuperação da Informação com o objetivo de facilitar o compartilhamento e a reutilização do conhecimento através de consultas em linguagem natural. Cada conjunto de técnicas possui a sua própria gama de problemas a solucionar. Na representação do conhecimento através da web também não é diferente e com isso a Web Semântica (WS) vêm para trazer meios de introduzir o lado semântico e o significado das informações contidas nas páginas web, a fim de que seja feito um mapeamento correto do que está sendo transferido e o que está sendo entendido.

Segundo Berners-Lee *et al.* (2001) o desafio da Web Semântica é fornecer uma linguagem que expresse dados e regras para raciocinar sobre os dados e que permita que regras de qualquer sistema de representação de conhecimento existente sejam exportadas para a Web.

Alguns formalismos são usados para compartilhamento e representação de conhecimento pela Web, e, seguindo por esse caminho a World Wide Web (W3C) propõe os seguintes padrões: a linguagem Resource Description Framework (RDF) e a Web Ontology Language (OWL). O primeiro é um modelo simples criado para a troca de dados na Web, já o segundo é projetado para representar o conhecimento sobre as coisas e suas devidas relações. Atualmente ambos são popularmente usados para descrever e representar as Ontologias.

Uma Ontologia é uma descrição formal de uma conceitualização (GRUBER, 1993), conceito esse referente a representação de um domínio, definindo suas relações existentes conforme algum modelo abstrato do mundo. Com o uso de ontologias é possível especificar conceitos e relações em uma base de conhecimento através de inferências. As ontologias podem melhorar de forma significativa as buscas por informações nessa imensidão que é a web, na qual um mecanismo de busca poderá apenas se atentar a conteúdo específicos e não varrer toda a web para encontrar o que deseja.

As consultas sobre os dados representados na Web Semântica podem ser efetuadas pela linguagem SPARQL (do acrônimo recursivo (em inglês) SPARQL Protocol and RDF Query Language). O SPARQL atua em dados representados em RDF e podem ser utilizados em consultas a ontologias.

O trabalho aqui descrito propõe uma ontologia para o domínio da Copa da Rússia 2018 voltada para a língua portuguesa, contribuindo assim para a comunidade, devido a sua carência na área. É desenvolvida uma Linguagem Natural Controlada (LNC), um subconjunto da linguagem natural, a fim de mitigar a complexidade do português para a realização de consultas na ontologia através da linguagem de consulta SPARQL.

A ontologia abrangerá um domínio ainda não explorado para uma língua pouco utilizada e de grande complexidade. Transformar uma linguagem natural para uma linguagem de consulta não é algo trivial, mas será mostrado que dado uma modelagem correta da ontologia é possível sim sua conversão.

Este trabalho é organizado da seguinte estrutura: a seção 3 descreve todo o fundamento e definições que abrangerá este trabalho; a seção 4 apresenta alguns trabalhos encontrados na literatura que fazem uso de técnicas para construção de ontologias e transformação da linguagem natural para consultas; a seção 5 faz uma breve descrição de como será feita a pesquisa; a seção 6 apresenta com detalhe todo o desenvolvimento da pesquisa; e por fim, a seção 7 relata as conclusões e trabalhos futuros.

2 OBJETIVOS

2.1 Objetivo geral

Construir uma ontologia para efetuar perguntas do tipo factoid em uma Linguagem Natural Controlada, através de consultas SPARQL.

2.2 Objetivos específicos

- Levantar dados sobre assuntos relacionado a futebol.
- Elaborar uma base ontológica para a Copa da Rússia 2018.
- Elaborar a gramática da linguagem controlada.
- Elaborar as consultas em SPARQL.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Processamento de Linguagem Natural

Segundo JOSHI (1991) a PLN – *Processamento de Linguagem Natural* - é o estudo da modelagem matemática e computacional de vários aspectos da linguagem e uma vasta gama de sistemas. O ser humano tem a capacidade de desenvolver linguagem naturalmente seja ela através de línguas faladas, ou até mesmo escritas. Assim, a PLN tem como objetivo analisar e verificar as formas de transformar a linguagem natural humana em algo compreensível pelo computador.

PLN abrange várias e complexas áreas de conhecimento (SILVA *et al.*, 2007), mas possui um papel em especial na ciência da computação, pois muitos aspectos do campo procuram modelar características linguísticas da computação (JOSHI, 1991).

3.2 Linguagem Natural Controlada

De acordo com Schwitter (2010) a LNC são subconjuntos da linguagem natural, onde as gramáticas e dicionários são restringidos de forma a reduzir ou eliminar a ambiguidade e complexidade. Esse tipo de linguagem foi tradicionalmente agrupadas em dois grupos: as LNCs orientadas para o ser humano com objetivo de melhorar a legibilidade e compreensibilidade da documentação técnica e simplificar e padronizar a comunicação humano-humana para fins específicos e as LNCs orientadas para máquinas a fim de melhorar a capacidade de tradução de documentos técnicos e aquisição, representação e processamento de conhecimento (SCHWITTER, 2010) para sistemas de conhecimento (FUCHS *et al.*, 2008) como também para a Web Semântica.

Devido a esse controle da linguagem pode-se facilitar a modelagem de sistemas usando PLN. Assim, este trabalho propõe uma linguagem natural controlada para o português, um subconjunto da língua portuguesa, a fim de que possa diminuir sua complexidade e ambiguidade, explorando suas restrições gramaticais e lexicais.

3.3 Ontologias

O termo surgiu na filosofia como o estudo das coisas existentes no mundo, hoje o termo foi adaptado e usado em áreas de Inteligência Artificial. Usadas na computação para

capturar conhecimento de um determinado domínio, Paralic e Kostial (2003) afirmam que seu uso permite definir conceitos e relações representando conhecimento a respeito de um documento ou domínio específico. Segundo Gruber (1993) ontologia é uma descrição formal de uma conceitualização.

Uma ontologia possui três elementos principais, sendo eles:

- **Indivíduos**, que são instâncias das classes, representam os objetos no domínio.
- **Propriedades**, também chamados de papéis, são relações binárias entre indivíduos. Pode ser dividida em propriedade primitiva (Datatype Property) ou propriedade de objeto, em que a relação é com atributos de dados ou relação com outro indivíduo respectivamente (HORRIDGE *et al.*, 2011). Por exemplo, na Figura 1, a propriedade *hasSibling* (temIrmão) pode relacionar o indivíduo Matthew com o indivíduo Gemma; ou a propriedade *livesIn* (viveEm) relaciona o indivíduo Matthew com o indivíduo England.
- **Classes**, são conjuntos que contêm os indivíduos. Podem ser organizadas de forma hierárquica em superclasse-subclasse, conhecidas como taxonomias. As subclasses são especialização de suas superclasses (HORRIDGE *et al.*, 2011).

Figura 1 – Relação entre propriedades e indivíduos



Fonte: Horridge *et al.* (2011).

3.4 Gramáticas

As gramáticas são um conjunto de regras de produção para cadeias de caracteres em uma linguagem formal. Chomsky (1956) define 4 tipos de hierarquia para as gramáticas: Gramática do Tipo-1, do Tipo-2, do Tipo-3 e do Tipo-4, mas o foco do nosso trabalho é nas gramáticas do tipo-2 conhecidas como gramáticas livre de contexto (GLC). Essas gramáticas podem ser utilizadas para descrever fragmentos de uma língua, como por exemplo a língua portuguesa. Segundo Sipser (2007, p. 106) uma gramática livre de contexto é definida formalmente como:

Uma gramática livre de contexto é uma 4-upla (V, Σ, R, S) , onde V é um conjunto finito de símbolos não terminais denominados variáveis, Σ é um conjunto finito, disjunto de V , denominados terminais, R é o conjunto finito de regras, com cada regra sendo uma variável e uma cadeia de variáveis e terminais, e S que pertence a V e é a variável inicial.

3.5 Consultas SPARQL

É uma linguagem para consultas capaz de recuperar informações no formato RDF (Resource Description Framework). Segundo Prud'hommeaux e Seaborne (2008) RDF é um formato de dados rotulado e direcionado para representar informações na Web.

3.5.1 *Sintaxe*

A sintaxe do SPARQL é similar ao Structured Query Language (SQL) e possui diversos elementos que a compõe. Será descrita nesta seção uma noção básica de como as consultas SPARQL são construídas e como é o seu funcionamento.

As consultas SPARQL podem ser baseadas em triplas RDF, como uma lista composta por sujeito, predicado e objeto separadas por espaços em branco (KLYNE; CARROLL, 2004), onde o sujeito pode ser um indivíduo ou classe da ontologia, os predicados são relações que define propriedades, e o objeto são valores dessas propriedades. Exemplo:

Código-fonte 1 – Sintaxe da linguagem SPARQL

```
1 PREFIX   dc: <http://purl.org/dc/elements/1.1/>
2 SELECT   ?title
3 WHERE    { <http://example.org/book/book1> dc:title ?title }
```

Fonte: Disponível em: <<https://www.w3.org/2001/sw/DataAccess/rq23/examples.html>>.

O exemplo acima mostra uma consulta simples em SPARQL, onde é feita uma busca por todos os livros que tenha a propriedade em dc declarada, dc é a Uniform Resource Identifier (URI), o identificador de recurso da ontologia.

Prud'hommeaux e Seaborne (2008) descreve o PREFIX como uma palavra-chave que associa um rótulo de prefixo da URI, a cláusula SELECT, identifica as variáveis que serão exibidas ao final da consulta, e a cláusula WHERE onde será fornecido o padrão gráfico básico que corresponderá ao gráfico de dados, representado nas triplas RDF's. Além da cláusula SELECT, existem vários outros tipos de consultas, como, o CONSTRUCT, que retorna um único gráfico construído, substituindo variáveis em um conjunto de modelos de triplas, o ASK que retorna um booleano (Verdadeiro ou Falso) indicando se um padrão de consulta é correspondente ou não, e o DESCRIBE que retorna um gráfico de RDF que descreve os recursos encontrados.

4 TRABALHOS RELACIONADOS

Sistemas que possuem base de conhecimento descritos usando ontologias, têm recebido bastante atenção nos últimos anos, sendo aplicados tanto para domínios abertos como também para domínios específicos.

Barbalho *et al.* (2018) elabora uma investigação do uso de ontologias para a classificação de padrões. Neste trabalho Barbalho *et al.* (2018) desenvolve uma ontologia capaz de analisar e classificar os movimentos e sinais sonoros do processo de mastigação e deglutição em sólidos ou líquidos. O desenvolvimento utilizou de uma metodologia baseada em um processo iterativo, realizando várias ações até que a ontologia alcançasse seu objetivo. Além disso, também foram realizados testes em ambientes reais e com pessoas saudáveis com o intuito de verificar se a ontologia está classificando corretamente como o esperado.

Semelhante ao nosso trabalho utilizaremos da mesma metodologia de Noy *et al.* (2001) que traz uma forma bem dividida e intuitiva no desenvolvimento da ontologia, contribuindo assim para uma maior facilidade e melhor entendimento da ontologia.

No trabalho do Torres (2017) foi proposto um sistema de pergunta-resposta focado no processamento da consulta. Tal sistema utilizou-se de gramáticas elaboradas de forma manual e do padrão *interpreter* que segundo Gamma (2009) esse padrão torna as gramáticas mais fáceis de entender e mudar, para transformar uma consulta em linguagem natural para uma linguagem formal usando o SPARQL, e assim fazer a consulta de ontologias. O trabalho se baseia em 5 temas para construir a base de ontologias, mas, foca-se em apenas um, que é o de Comida Peruana, em língua espanhola.

Diferentemente do proposto nesta pesquisa, não se utilizará do padrão *interpreter* para a construção das gramáticas, como também o domínio abordado para a criação da ontologia será voltado para a Copa da Rússia 2018 e devido sua carência na comunidade se utilizará a língua portuguesa.

Luz (2013) propôs também um sistema de domínio ontológico onde desenvolve um modelo de busca para a realização de consultas, utilizando uma Linguagem Natural Controlada (LNC) e aproveita-se da expressividade das ontologias para inferir consultas através do SPARQL. Com esse intuito, foi desenvolvido uma gramática controlada (GLNC) baseada em gramáticas livre de contexto (GLC) especificando vocabulários e regras de produções. O foco principal foi atender consultas do tipo Qual?, Quem? e Quantos?. O uso da LNC simplifica a linguagem reduzindo ou até eliminando a quantidade de ambiguidades existentes no português, linguagem

utilizada em sua pesquisa. Assim, o uso de LNC e GLNC no modelo possibilitou uma consulta controlada, que só faz sentido se soubermos como os dados estão estruturados na base.

Semelhante ao nosso trabalho construiremos uma LNC do português para diminuir sua complexidade, principalmente quando se diz respeito a ambiguidades e assim utilizar-se desta LNC para efetuar buscas usando o SPARQL.

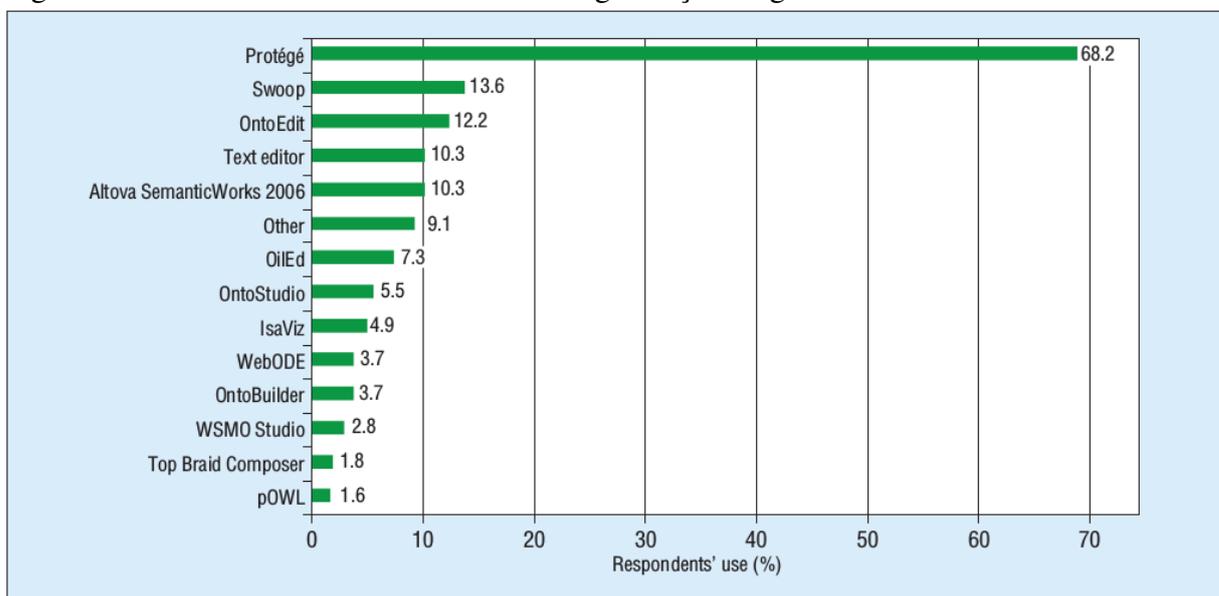
5 METODOLOGIA

5.1 Estudo de Ferramentas

Um estudo das ferramentas existentes foi realizado a fim de conhecer quais as mais utilizadas na construção de ontologias e de consultas através do SPARQL. Nesse estudo foi verificado como é o funcionamento dessas ferramentas e como aplicá-las em nosso problema.

Atualmente há uma vasta quantidade de ferramentas para construção e elaboração de uma ontologia como, Protégé, OilEd, OntoEdit, WebODE, entre outras (ver Figura 2). Pode-se usar diferentes critérios para comparar determinadas ferramentas como, generalidade, expressividade, complexidade, entre outras, Alatrish (2013) em seu trabalho compara 5 ferramentas considerando a sua facilidade de uso e a sua disseminação. Nesta sessão será mostrado e descrito apenas a ferramenta utilizada neste trabalho, uma breve descrição entre as ferramentas existentes pode ser encontrado no trabalho de Almeida e Bax (2003).

Figura 2 – Ferramentas mais utilizadas nas organizações segundo entrevista.



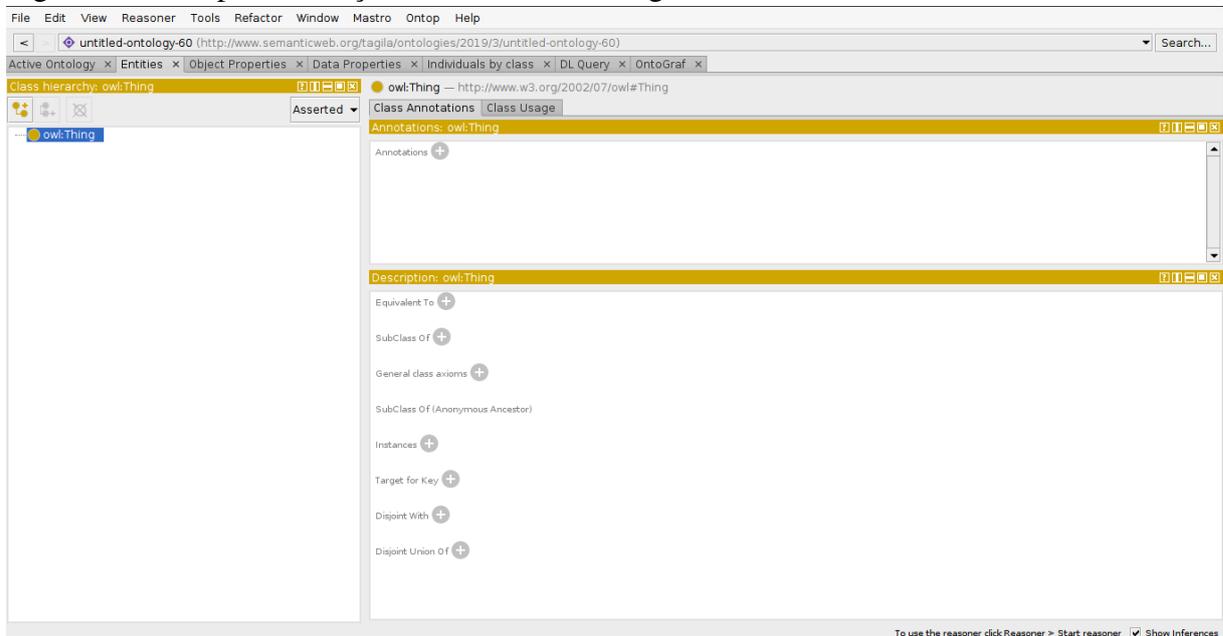
Fonte: (CARDOSO, 2007)

5.1.1 Protégé

O Protégé é uma ferramenta livre e de código aberto, para criar e gerenciar ontologias. Está sendo desenvolvido pela Universidade de Stanford (STANFORD, 2012) na

linguagem de programação Java e permite a inclusão de vários outros *plugins* para adição de novas funcionalidades.

Figura 3 – Tela para definição de classes no Protégé.



Fonte: Autor.

Possui uma interface razoavelmente fácil de se usar (ver Figura 3) incluindo classificações dedutivas para verificar a consistência da ontologia e inferir informações de acordo com o que foi analisado.

A construção da ontologia deste trabalho no Protégé se dará por meio da linguagem OWL (ver Código-fonte 1), que é o padrão mais recente para ontologias e é usada para publicação e compartilhamento de ontologias na W3C-World Wide Web Consortium, além disso, suporta tanto versões para desktop como também para web.

Código-fonte 2 – Hierarquia de classes

```

1 <owl:Class rdf:about="#Final">
2     <rdfs:subClassOf rdf:resource="#Partida"/>
3     ...
4 </owl:Class>

```

Fonte: Autor.

O Protégé além de auxiliar na construção e manipulação da ontologia ele também implementa meios para sua visualização, através de uma interface gráfica, o que nos permite um

maior entendimento do que já foi construído. Há suporte para rastrear possíveis inconsistências e possui uma enorme comunidade e documentações que juntos trazem uma composição de tudo que a ferramenta pode fazer.

5.1.2 *Reasoner*

Para que uma ontologia seja usada por algum sistema, é de grande importância que ela esteja coerente e consistente. Uma vantagem desenvolver ontologias em OWL-DL - tipo de linguagem OWL suportada pelo protégé e que se baseia nas lógicas de descrição - pois é possível a utilização de raciocinadores para verificar se uma classe pertence ou não a uma outra classe ou até mesmo a consistência da ontologia. Segundo Horridge *et al.* (2011) uma classe é tida como inconsistente se não puder ter instâncias.

O *reasoner* (termo em inglês) foi utilizado nesta pesquisa para validar a correteza da ontologia através de suas inferências. Assim ao executar esses mecanismos de inferências observamos possíveis erros escondidos para imediatamente resolvê-los. Fizemos uso dos *plugins Pellet e FaCT++*.

5.2 Levantamento de dados para o domínio

Para a construção de uma ontologia é necessário que seja definido para onde se pretende utilizá-la, assim esse levantamento foi realizado através de possíveis usuários que possuam algum tipo de relação com o tema abordado na ontologia. Os dados coletados expressarão o que é mais relevante saber sobre a Copa da Rússia 2018, como por exemplo:

- Qual o time vencedor da copa?
- Qual os jogadores do Brasil?

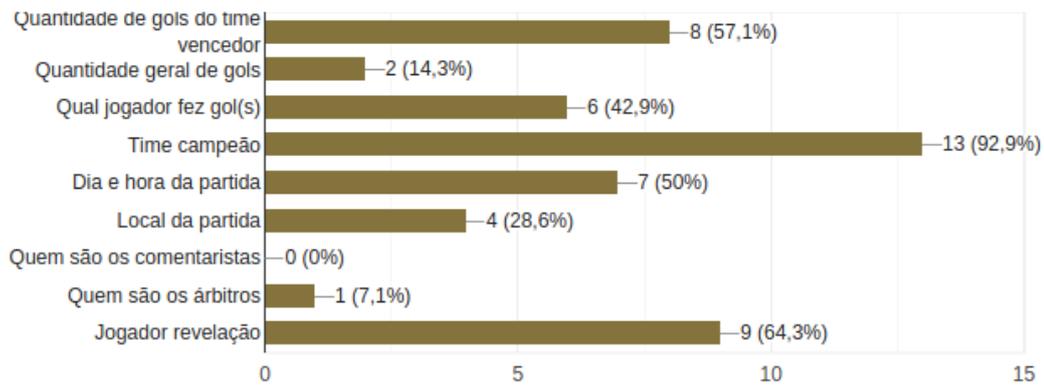
Esses dados foram coletados através de formulários online, com o objetivo de entender o que seria mais importante saber em relação ao tema da Copa da Rússia 2018. O formulário (Apêndice A) foi distribuído para 20 pessoas em que apenas 14 responderam. Os principais assuntos abordados em uma copa de futebol são as partidas dos jogos, os jogadores e quais as seleções que participaram, com isso, o formulário foi desenvolvido de acordo com cada um destes assuntos.

Os resultados obtidos através das respostas dos usuários serviram de base para se dar início na construção da ontologia. Foram coletados diferentes tipos de respostas para possíveis

perguntas que a ontologia poderá responder.

Em relação as partidas, pode-se observar na Figura 4 que a grande maioria das respostas está relacionado ao time campeão da partida, a quantidade de gols do time vencedor, qual o jogador que fez gol(s), horários e dias e qual o jogador que se destacou durante a partida. Com isso, obtivemos as informações necessárias para descrever uma partida na ontologia, como por exemplo: uma partida ela possui times, horários, dia e jogadores.

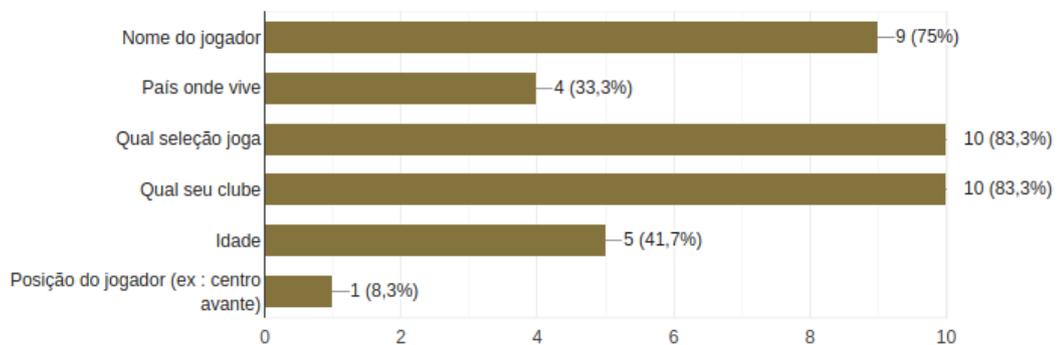
Figura 4 – Respostas relacionadas as partidas da Copa.



Fonte: Autor.

Além das partidas, podemos obter informações sobre os jogadores, seu nome, qual seleção joga, e seu clube, como também há dados que podem ser levados em consideração na construção da ontologia, como a posição do jogador (Figura 5), significando que um dos participantes optou por dar sua opinião, acrescentando um novo dado em sua resposta no formulário descrito no Apêndice A.

Figura 5 – Respostas relacionadas aos Jogadores da Copa.

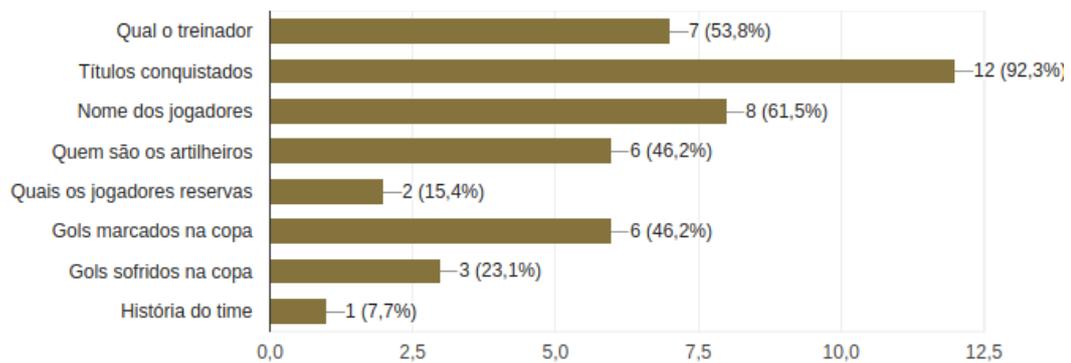


Fonte: Autor.

Como cada jogador ou um conjunto de jogadores tem no mínimo uma seleção além de outras características que os descrevem como jogadores de uma copa, as seleções também

têm suas próprias características. Foi observado através das respostas na Figura 6 que saber quem é o treinador, títulos de cada seleção e quais os nomes dos jogadores, ficaram entre as maiores respostas, assim essas informações já podem ser o suficiente para descrever uma seleção na ontologia.

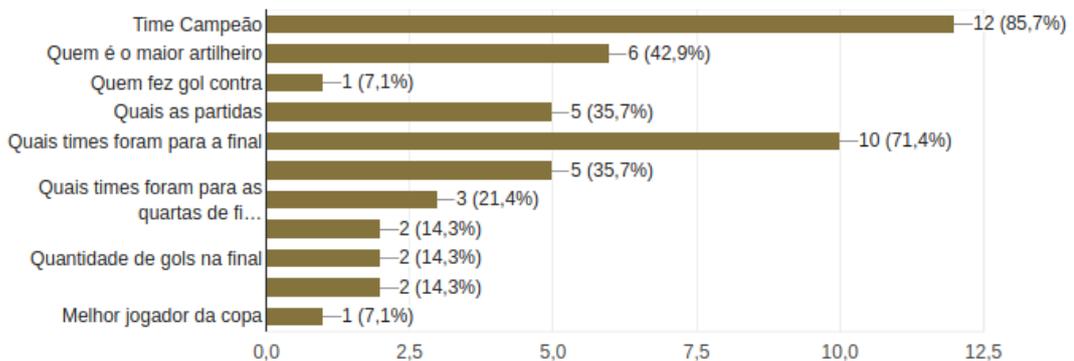
Figura 6 – Respostas relacionadas a cada Seleção da Copa.



Fonte: Autor.

Descartamos os dados não citados, como por exemplo a história do time, devido possuir um nível de detalhamento bem maior na ontologia.

Figura 7 – Respostas relacionadas a Copa em Geral.



Fonte: Autor.

Além das informações sobre os jogadores, partidas e seleções, também se pode observar e obter dados adicionais de grande importância, como, o time campeão, qual o maior artilheiro na copa em geral, quais os times foram para a final, ver Figura 7. Com isso, a ontologia foi desenvolvida e modelada a partir desses dados, de maneira que não fosse complexa, mas que garantisse respostas em suas buscas.

5.3 Construção da Ontologia

Após ser definido tudo o que a ontologia abrangerá o próximo passo é a sua construção utilizando ferramentas apropriadas (ver Seção 5.1), não existe uma forma geral para se definir como uma ontologia deverá ser desenvolvida, mas há termos práticos que uma ontologia deverá incluir, segundo Almeida (2006).

Definição das classes da ontologia; arranjar as classes em uma hierarquia de subclasses e superclasses; definir slots (slots, algumas vezes denominados "papéis" ou "propriedades") e descrever os valores permitidos para eles; e, preencher os valores para os slots com instâncias.

Pensando nisso, metodologias foram desenvolvidas para atender a necessidade de cada desenvolvedor no processo de construção da ontologia, sendo possível a combinação de mais de uma delas, como por exemplo temos: Ontology101 (NOY *et al.*, 2001), Methontology (FERNÁNDEZ-LÓPEZ *et al.*, 1997), On-To-Knowledge Methodology (OTKM) (SURE *et al.*, 2004), SENSUS (SWARTOUT *et al.*, 1996), entre outras.

Para a elaboração da ontologia desde trabalho foi utilizado o método 101 de Noy *et al.* (2001), que consiste em 7 etapas iterativas para o projeto da ontologia, o que permite vários ciclos de desenvolvimento até que se alcance a uma estrutura ontológica desejada.

Na primeira etapa do método é determinado o escopo e o domínio da ontologia, respondendo algumas perguntas básicas (NOY *et al.*, 2001):

- Qual é o domínio que a ontologia cobrirá?
- Para o que vamos usar a ontologia?
- Para quais tipos de perguntas as informações na ontologia devem responder?
- Quem vai usar e manter a ontologia?

O domínio se resume nas informações e dados estatísticos sobre a copa da Rússia 2018 e todo o conhecimento que se tem atualmente sobre ela, podendo a ontologia ser utilizada para auxiliar em aplicações que usem Processamento de Linguagem Natural voltados para o português ou até mesmo ser utilizada pela comunidade para o reuso, fazendo combinações de ontologias relacionadas a futebol. Já o escopo é limitado baseado nas informações obtidas dos formulários distribuídos (ver Seção 5.2), onde a ontologia deve ser capaz de responder algumas questões, conhecidas como questões de competência.

- Quais os jogadores e suas características devo considerar?
- Quais os jogadores do Brasil?
- Quais as partidas e suas características devo considerar?

- Quais os times da partida final?
- Quantidade de gols marcados na copa
- Quais os técnicos?
- Quem fez gols na partida final?
- Quem é o jogador revelação?
- Qual o time vencedor da copa?

Essa lista de questões não foi elaborada de forma exaustiva como pode ser visto acima, mas representa apenas um esboço para uma possível visualização do que deve conter na ontologia podendo ser alterada conforme o projeto se avança.

A segunda etapa consiste em considerar o reuso de ontologias existentes, mas para este trabalho nenhuma foi considerada, pois as bases ontológicas encontradas eram voltadas para o inglês e para domínios bem específicos sobre futebol, como por exemplo, detalhes sobre clubes.

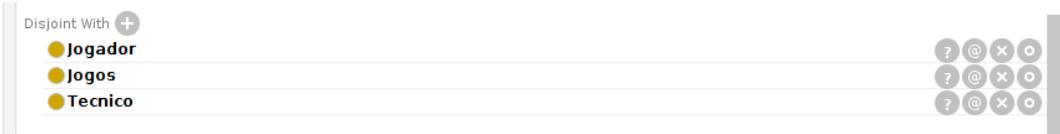
Com base nos assuntos referentes a copa, foi feito a enumeração de termos que possam ser importantes para ontologia sem se preocupar com seus conceitos ou até mesmo como eles irão se relacionar. Essa enumeração consiste na terceira etapa da construção. Os principais termos observados foram: partida, jogador, time, técnico, horários, localização e estádio.

Tendo em mãos os termos listados na etapa 3, selecionamos os que poderiam ser possíveis classes da ontologia e assim damos início a quarta etapa, onde foram definidas as classes e suas hierarquias. Utilizando-se de uma abordagem *top-down*, as classes foram definidas iniciando de classes mais gerais para as mais específicas.

Quando uma classe é implementada alguns conceitos importantes devem ser pensados. As classes *primitivas* possuem as condições necessárias para descrevê-las, ou seja, se um indivíduo pertence a uma classe **A** ele deve satisfazer as condições (HORRIDGE *et al.*, 2011). No entanto, isto não é suficiente para dizer que qualquer indivíduo aleatório que satisfaça estas condições pertence a **A**. Já as classes que são descritas como *definidas* também conhecidas como equivalentes, possuem condições necessárias e suficientes, ou seja, não são apenas condições necessárias para um indivíduo pertencer a **A**, mas também suficientes para determinar que algo que satisfaça essas condições é um membro de **A** (HORRIDGE *et al.*, 2011).

Além de descrever os tipos de classes, também é necessário especificar que as classes estão desconectadas uma das outras, ou seja, as classes não devem ter instâncias em comum. No Protégé, existe a opção *disjoint*, Figura 8.

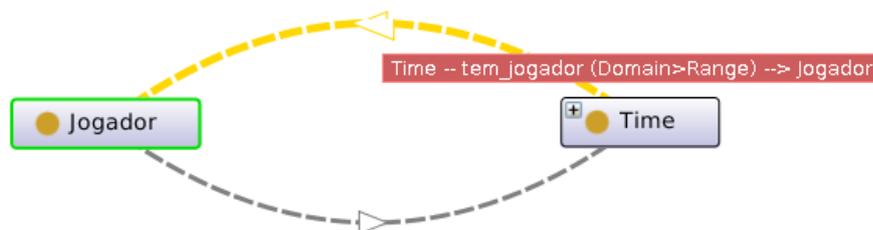
Figura 8 – Classes *disjoint* da classe time na ontologia.



Fonte: Autor. Retirado do software Protégé.

Após a criação das classes é definido as propriedades (*slots*) em que auxiliarão nas respostas para as questões de competência da etapa 1. As propriedades são divididas em Propriedades de dados (*Data Property*) que representam relações binárias entre indivíduos e valores de dados e as Propriedades de objeto (*Object Property*) que descreve relações binárias entre dois indivíduos, ver Figura 9. Cada propriedade é analisada, a fim de determinar qual a classe que melhor irá descrever esta propriedade e assim anexar a mesma a esta classe.

Figura 9 – Relação binária entre Jogador e Time, através da *Object Property* tem_jogador.



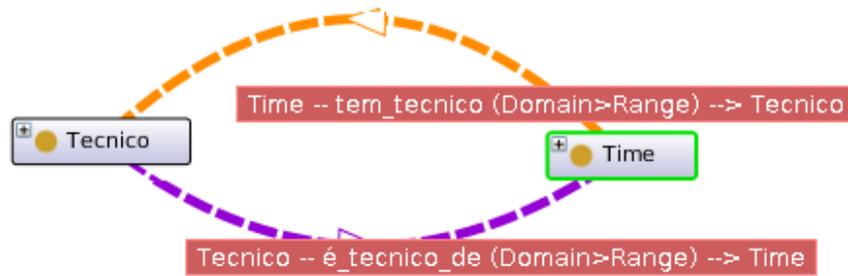
Fonte: Autor. Retirado do software Protégé, *plugin OntoGraf*.

Existe várias outras características para as propriedades, como inversa funcional, transitiva, simétrica, assimétrica, reflexiva e irreflexiva. Se atentarmos apenas nas definições das utilizadas neste trabalho. Uma propriedade inversa é a propriedade que especifica que se um indivíduo **a** se relacionar com um indivíduo **b** através de alguma propriedade, então sua propriedade inversa relacionará o indivíduo **b** ao indivíduo **a**. Um exemplo de propriedade inversa se encontra na Figura 10, onde a propriedade *é_tecnico_de* é inversa de *tem_tecnico*, especificando que o domínio de uma é imagem da outra.

A funcional, descreve que se uma propriedade é declarada como funcional para um determinado indivíduo, então existirá no máximo um indivíduo relacionado a este indivíduo. A Figura 11 mostra o indivíduo da classe Time se relacionando com um indivíduo da classe Tecnico, através da propriedade funcional *tem_tecnico*, trazendo para a linguagem natural por exemplo, o time Brasil só possuirá um técnico, ou seja, o time não se relacionará com dois ou mais técnicos através da propriedade *tem_tecnico*.

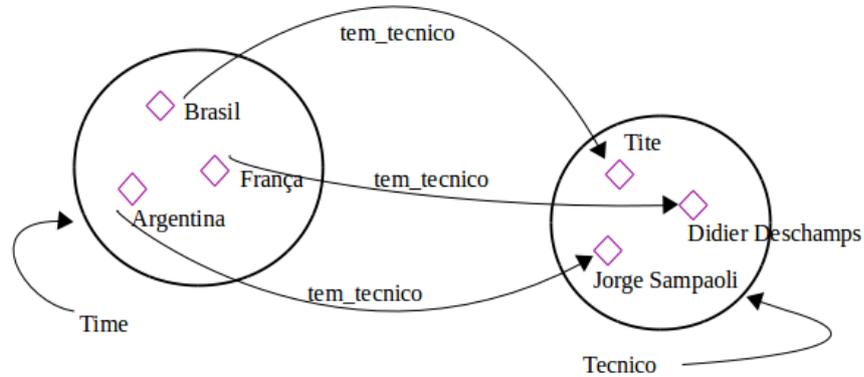
A etapa 6 da construção da ontologia se deu juntamente com a etapa 5 (etapa

Figura 10 – Exemplo de uma propriedade inversa da ontologia.



Fonte: Autor. Retirado do software Protégé, *plugin OntoGraf*.

Figura 11 – Exemplo de uma propriedade funcional da ontologia.



Fonte: Autor.

onde as propriedades são definidas e anexadas às classes-slots). Nestas etapas é observado a necessidade de especificar o domínio (*Domain*) e a imagem (*Range*) de uma propriedade, pois cada propriedade anexada a uma classe se tornará domínio do slot. O domínio e a imagem são axiomas usado pelo *Reasoner* para fazer inferência na ontologia.

Podemos ver na Figura 9 que o domínio da propriedade *tem_jogador* está definido como *Time* e a imagem como *Jogador*, em outras palavras, isto indica que um time ele possui algum jogador.

Com as classes e propriedades definidas podemos nos adentrar na criação de instâncias (*indivíduo*), etapa final do desenvolvimento. Cada classe foi selecionada, criada uma instância individual desta classe e atribuídos valores para os slots.

5.4 Definição e construção da Linguagem Controlada

Utilizando um português controlado proporcionamos facilidades tanto na construção de sentenças, devido ao fato de se trabalhar com um subconjunto da linguagem, como também para a elaboração das consultas SPARQL. Assim, a LNC foi construída para o português,

especificando os elementos terminais de algumas sentenças da nossa gramática, utilizando de gramáticas livre de contexto para especificar a estrutura da nossa linguagem.

Para a elaboração da gramática foram analisadas algumas sentenças sobre a copa da Rússia de acordo com os possíveis tipos de consultas que a ontologia poderia responder. Cada sentença recebeu um estado inicial e dividimos partes da sentença em sintagmas nominais, preposicionais e algumas vezes verbais, analisando cada elemento de sua estrutura.

Sintagmas designam uma sequência hierarquizada de elementos linguísticos, que compõem uma unidade na sentença. Um sintagma nominal refere-se a uma sequência ou agrupamentos de palavras em torno de pelo menos um substantivo, já os sintagmas que possuem preposições são chamados de sintagmas preposicionais e os sintagmas verbais possuem verbos em sua estrutura acompanhado de várias outras coisas como por exemplo um sintagma nominal (JURAFSKY; MARTIN, 2000).

5.5 Construção das Consultas SPARQL

Nesta fase é mostrado como pode ser feito a conversão de uma sentença em linguagem natural para a linguagem de consulta SPARQL em nossa ontologia, descrevendo sua estrutura e como é feito esse processo.

Há várias modelos na literatura que mostra a possibilidade de ter algo em linguagem natural transformado em consultas para ontologias usando PLN, mas, essa conversão vai depender de qual domínio a ontologia descreve e como ela é estruturada, pois o modelo deve se adequar a ontologia ou vice-versa. Luz (2013) desenvolve um modelo de conversão para ontologias já existentes com domínios voltados para o currículo *Lattes* e para conceitos de *smartphones*, mas, detalha uma linguagem controlada mostrando uma gramática e alterando as ontologias para se adequar ao seu modelo. A transformação de uma sentença em linguagem natural para uma linguagem de consulta pode incluir diversas fases, mas é de grande importância que se tenha o entendimento da sintaxe da linguagem que se deseja trabalhar.

Descrevemos os mapeamentos de uma LNC para o SPARQL baseando-se no trabalho de Luz (2013) e adequando para a nossa LNC e Ontologia.

5.6 Validação da ontologia

Verificamos se a ontologia foi desenvolvida como esperado e elaborando testes para saber se as buscas estão sendo efetuadas de forma correta. Estes testes foram executados efetuando algumas consultas a nossa base ontológica e analisando as repostas retornadas.

6 RESULTADOS

6.1 Base ontológica

Seguindo a metodologia descrita na Seção 5.3, na definição das classes foram criadas primeiramente as classes Jogador, Partida, Tecnico e Time e posteriormente a especialização de Jogador e Partida, ver Figura 12. Classificamos 5 classes como primitivas, sendo elas: Jogador, Jogos, Partida, Tecnico e Time. Já para as classes mais específicas na hierarquia de classes de Partida (Final, Semifinal, Quartas_de_final, Oitavas_de_final) definimos como classes definidas.

Figura 12 – Hierarquia de classes da ontologia.



Fonte: Autor.

A tabela 1 mostra todas as classes criadas na ontologia sobre a copa da Rússia 2018, incluindo suas descrições, tipo de classe e suas classes disjoint.

Para as propriedades de objetos foram criados 8 slots, em que, 6 deles são definidos como função inversa (*Inverse Of, no software Protégé*), 2 com características funcionais e 2 sem nenhuma característica à mais. As propriedades de objeto criadas para a ontologia são mostradas na Tabela 2, onde mostrará uma breve descrição, suas características, seu domínio e imagem.

A Tabela 3 mostra as propriedades de dados da ontologia a descrição de cada propriedade, suas características e domínio e imagem associados. Cada propriedade de dados foi definida para descrever os indivíduos da ontologia.

Tabela 1 – Classes da ontologia e suas características

Classes	Descrição	Classes disjoint	Tipo
Jogador	Descreve os jogadores da copa em que devem ser associados a pelo menos um time.	Time, Tecnico, Jogos	Primitiva
Tecnico	Descreve os técnicos dos times da copa em que devem ser associados a pelo menos um time.	Time, Jogador, Jogos	Primitiva
Time	Descreve os times que participaram da copa em que devem ser associados a 23 jogadores e ter exatamente um técnico.	Tecnico, Jogador, Jogos	Primitiva
Jogos	Descreve a classe que irá conter todos os jogos da copa.	Time, Jogador, Tecnico	Primitiva
Partida	Descreve a classe que irá conter as fases da copa e suas partidas.	Time, Jogador, Tecnico	Primitiva
Final	Descreve a fase final da copa, onde é necessário ter pelo menos um time e uma partida e é necessário e suficiente que seja instância de Partida e tenha exatamente 1 Partida e 2 Times.	Oitavas_de_final, Quartas_de_final, Semifinal	Definida
Semifinal	Descreve a fase das semifinais da copa, onde é necessário ter pelo menos um time e uma partida e é necessário e suficiente que seja instância de Partida e tenha exatamente 2 Partidas e 4 Times.	Oitavas_de_final, Quartas_de_final, Final	Definida
Quartas_de_final	Descreve a fase das quartas de finais da copa, onde é necessário ter pelo menos um time e uma partida e é necessário e suficiente que seja instância de Partida e tenha exatamente 4 Partidas e 8 Times.	Oitavas_de_final, Semifinal, Final	Definida
Oitavas_de_final	Descreve a fase das oitavas de finais da copa, onde é necessário ter pelo menos um time e uma partida e é necessário e suficiente que seja instância de Partida e tenha exatamente 8 Partida e 16 Times.	Quartas_de_final, Semifinal, Final	Definida

Fonte: Autor.

Para detalhar melhor um indivíduo na ontologia temos o seguinte exemplo, *o time do Brasil tem jogadores e possui títulos*. Esta afirmação ressalta que a ontologia deve possuir o indivíduo Brasil, que tenha relação com algum jogador e que possua dados sobre títulos ganhos. Na ontologia o indivíduo Brasil é relacionado com indivíduos da classe Jogador através da propriedade de objeto *tem_jogador* e os títulos pela propriedade de dados *títulos*, além disso, também existe relação com o ano do mundial ganho através da propriedade de dados *ano_mundial*. As Figuras 13, 14, 16 mostra exemplos das relações através das propriedades de algumas instâncias existentes na ontologia.

Tabela 2 – Propriedades de objeto (*Object Property*) da ontologia.

Propriedade	Descrição	Característica	Domínio	Imagem
tem_jogador	Indica que um time tem jogadores.	Inversa de é_jogador_de	Time	Jogador
tem_partida	Indica que em uma partida pode existir outras partidas.	-	Partida	Partida
tem_tecnico	Indica que um time tem algum técnico.	Funcional, Inversa de é_tecnico_de	Time	Tecnico
tem_time	Indica que uma partida tem algum time.	Inversa de é_time_da	Partida	Time
é_jogador_de	Indica que algum jogador é jogador de algum time.	Funcional, Inversa de tem_jogador	Jogador	Time
é_tecnico_de	Indica que um técnico é tecnico de algum time.	Funcional, Inversa de tem_tecnico	Tecnico	Time
é_time_da	Indica que um time participa de uma ou mais partidas.	Inversa de tem_time	Time	Partida
é_vencedor_da	Indica que um time é vencedor de uma ou mais partidas.	-	Time	Partida

Fonte: Autor.

Tabela 3 – Propriedades de dados (*Data property*) da ontologia.

Propriedade	Descrição	Característica	Domínio	Imagem
ano_mundial	Indica o ano que um time ganhou um mundial da copa.	-	Time	xsd:positiveInteger
data_partida	Indica a data que ocorreu uma determinada partida.	Funcional	Partida	xsd:string
gols_partida	Indica a quantidade de gols marcados em uma partida.	-	Partida	xsd:int
gols_vencedor	Indica a quantidade de gols do time vencedor em uma partida.	-	Partida	xsg:int
hora_partida	Indica o horário de uma determinada partida.	Funcional	Partida	xsd:string
local	Indica o local que ocorreu a partida.	Funcional	Partida	xsd:string
posição	Indica a posição de um jogador {Defesa, Goleiro, Atacante e Meio-campo}.	Funcional	Jogador	xsd:string
qtd_gols	Indica a quantidade de gols de uma partida ou de um jogador.	-	Jogador, Partida	xsd:int
títulos	Indica a quantidade de títulos de um time.	-	Time	xsd:int

Fonte: Autor.

Figura 13 – Relações da instância Tite da classe Tecnico.

Property assertions: Tite	
Object property assertions +	é_tecnico_de Brasil
Data property assertions +	

Fonte: Autor.

Figura 14 – Relações da instância Brasil da classe Time.

Property assertions: Brasil	
	tem_jogador Philippe_Coutinho
	tem_jogador Pedro_Geromel
	tem_jogador Miranda
	tem_jogador Taison
Data property assertions +	
	titulos 5
	ano_mundial 2002
	ano_mundial 1970
	ano_mundial 1958
	ano_mundial 1962
	ano_mundial 1994

Fonte: Autor.

Figura 15 – Relações da instância Neymar da classe Jogador.

Property assertions: Neymar	
Object property assertions +	é_jogador_de Brasil
Data property assertions +	
	posição "Atacante"^^xsd:string
	qtd_gols 4

Fonte: Autor.

Figura 16 – Relações da instância Final da classe Final subclasse de Partida.

Property assertions: Final	
Object property assertions +	tem_time França
	tem_time Croácia
Data property assertions +	
	data_partida "DOM 15/07/2018"^^xsd:string
	gols_partida "6"^^xsd:int
	gols_vencedor "4"^^xsd:int
	local "Estádio Lujniki, Moscou"^^xsd:string
	hora_partida "12:00"^^xsd:string

Fonte: Autor.

6.2 Conversão da LNC para o SPARQL

6.2.1 Tratando os tipos de consultas da LNC

Analisamos a estrutura das sentenças para o tipo *qual* de acordo com a Seção 5.4. A Tabela 4 e 5 mostra os elementos da sentença *qual o técnico do Brasil?* e o agrupamento em frases nominais ou preposicionais respectivamente. A partir da análise de cada elemento criamos os símbolos da nossa gramática. Atribuímos um símbolo inicial para a sentença *qual*, que chamaremos de *Set_Qual*.

Tabela 4 – Elementos sintáticos da sentença, *qual o técnico do Brasil?*.

Qual =	Pronome qual
o =	Artigo
técnico =	Substantivo comum
do =	Preposição (de + artigo o)
Brasil =	Substantivo próprio

Fonte: Autor.

Tabela 5 – Agrupamento dos sintagmas da sentença, *qual o técnico do Brasil?*.

Qual =	Pronome qual
o técnico =	frase nominal
do Brasil =	frase preposicional

Fonte: Autor.

A partir dos elementos analisados da sentença, construímos uma gramática inicial para representar a sentença mencionada anteriormente e sua respectiva derivação representada na Figura 17:

$Set_Qual \rightarrow Pron_qual\ Fnominal\ Fprepo$

$Fnominal \rightarrow Sub_comum \mid Sub_proprio \mid Art\ Sub_comum \mid Art\ Sub_proprio$

$Fprepo \rightarrow Prepo\ Fnominal$

$Pron_qual \rightarrow Qual$

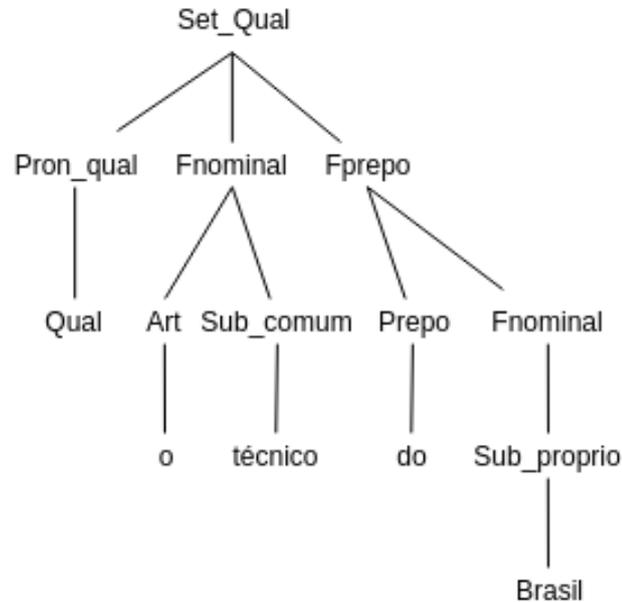
$Sub_comum \rightarrow técnico$

$Sub_proprio \rightarrow Brasil$

$Art \rightarrow o$

$Prepo \rightarrow do$

Figura 17 – Árvore de derivação para a sentença, *qual o técnico do Brasil?*



Fonte: Autor.

A próxima sentença a ser analisada (ver Tabela 6) é semelhante a anterior, mas envolve o verbo *ser* em sua estrutura.

Tabela 6 – Elementos sintáticos da sentença, *qual é a posição do jogador Neymar?*.

Qual =	Pronome qual
é =	Verbo ser na 3ª pessoa do singular no presente do indicativo
a =	Artigo
posição =	Substantivo comum
do =	Preposição (de + artigo o)
jogador =	Substantivo comum
Neymar =	Substantivo próprio

Fonte: Autor.

Da mesma forma podemos agrupar as partes da sentença como mostra a Tabela 7 em frases nominais e preposicionais.

Tabela 7 – Agrupamento dos sintagmas da sentença, *qual é a posição do jogador Neymar?*.

Qual =	Pronome qual
é =	Verbo ser na 3ª pessoa do singular no presente do indicativo
a posição =	Sintagma nominal
do jogador Neymar =	Sintagma preposicional

Fonte: Autor.

Dando continuidade a gramática incluindo novos símbolos não terminais e terminais

para reconhecimento do verbo *ser*, temos:

Para representar a palavra vazia é usado o símbolo lambda (λ).

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Fprepo$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$

$Fnominal \rightarrow Art\ Sub_comum \mid Art\ Sub_proprio$

$Fprepo \rightarrow Prepo\ Fnominal$

$Pron_qual \rightarrow Qual$

$Verbo_ser \rightarrow \acute{e}$

$Sub_comum \rightarrow \acute{t}cnico \mid posi\c{c}o\tilde{a}$

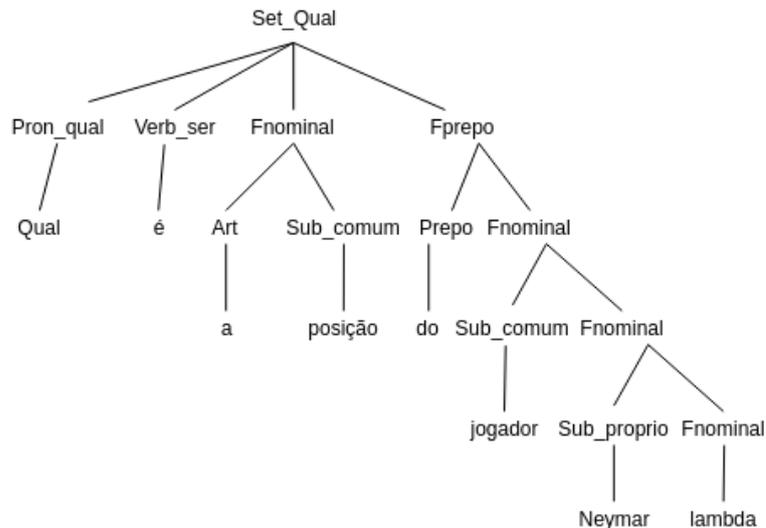
$Sub_proprio \rightarrow Brasil \mid Neymar$

$Art \rightarrow o \mid a$

$Prepo \rightarrow do$

A Figura 18 mostra a árvore de derivação para a sentença, *qual é a posição do jogador Neymar*.

Figura 18 – Árvore de derivação para a sentença *qual é a posição do jogador Neymar?*



Fonte: Autor.

Além das estruturas das sentenças representadas acima podemos também responder perguntas na ontologia que envolva o plural das palavras, onde no início das perguntas utilizaremos *Quais* e não *Qual*.

Tabela 8 – Elementos sintáticos da sentença, *quais os times que jogaram na partida final?*.

Quais =	Pronome quais
os =	Artigo (o + s)
times =	Substantivo comum
que =	Pronome relativo
jogaram =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
na =	Preposição (em + artigo a)
partida =	Substantivo comum
final =	Substantivo comum

Fonte: Autor.

A Tabela 9 mostra como a sentença ficará após a separação em sintagmas. Não foi feito um tratamento específico para a representação de palavras no plural na gramática, assim essas palavras foram atribuídas como substantivos.

Tabela 9 – Agrupamento dos sintagmas da sentença, *quais os times que jogaram na partida final?*.

Quais =	Pronome quais
os times =	Sintagma nominal
que =	Conjunção integrante
jogaram =	Sintagma verbal
na partida final =	Sintagma preposicional

Fonte: Autor.

A sentença *quais os times que jogaram na partida final*, possui um tratamento diferente, pois a sentença inclui uma oração subordinada em sua estrutura, ou seja: pode ser dividida em duas orações, *quais os times* - oração principal e *que jogaram na partida final* - oração subordinada. Assim, a gramática seguinte incluirá estes tratamentos.

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Fprepo$

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Fverbal\ Fprepo$

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Subordinada$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$

$Fnominal \rightarrow Art\ Sub_comum \mid Art\ Sub_proprio \mid \lambda$

$Subordinada \rightarrow ProRelQue\ Fverbal\ Fprepo\ Fnominal$

$Fprepo \rightarrow Prepo\ Fnominal \mid \lambda$

$Fverbal \rightarrow Verbo\ Fnominal$

$Pron_qual \rightarrow Qual \mid Quais$

ProRelQue \rightarrow *que*

Verbo_ser \rightarrow *é* | *são* | λ

Verbo \rightarrow *jogaram*

Sub_comum \rightarrow *técnico* | *posição* | *times* | *partida* | *final*

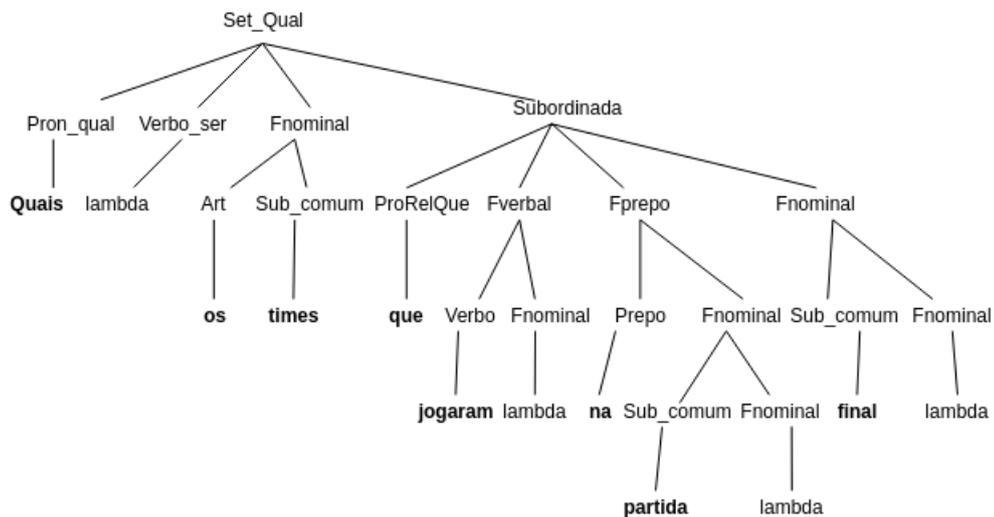
Sub_proprio \rightarrow *Brasil* | *Neymar*

Art \rightarrow *o* | *a*

Prepo \rightarrow *do* | *na*

Podemos observar sua derivação representada em árvore na Figura 19.

Figura 19 – Árvore de derivação para a sentença, *qual os times que jogaram na partida final?*



Fonte: Autor.

Existe mais de uma partida, por exemplo, nas oitavas de final. Pensando nisso para especificar cada partida definimos como *oitavas de final 1*, *oitavas de final 2*, *oitavas de final 3* e assim sucessivamente ou também partida entre França e Croácia por exemplo.

Para a sentença *quais os times das oitavas de final 1*, temos sua seguinte estruturação e agrupamento (ver Tabela 10 e 11, respectivamente):

As partidas, como é o caso das oitavas de final para uma consulta mais específica em linguagem natural deve ser descrita explicitamente qual a partida das oitavas estará na busca. Alterando nossa gramática temos:

Tabela 10 – Elementos sintáticos da sentença, *quais os times das oitavas de final 1?*.

Quais =	Pronome quais
os =	Artigo
times =	Substantivo comum
das =	Preposição (de + artigo as)
oitavas =	Substantivo comum
de =	Preposição
final =	Substantivo comum
1 =	numeral

Fonte: Autor.

Tabela 11 – Agrupamento dos sintagmas da sentença, *quais os times das oitavas de final 1?*.

Quais =	Pronome quais
os times =	Sintagma nominal
das oitavas =	Sintagma preposicional
de final 1 =	Sintagma preposicional

Fonte: Autor.

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Fprepo$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal$

$Fnominal \rightarrow Art\ Sub_comum \mid Art\ Sub_proprio \mid \lambda$

$Fprepo \rightarrow Prepo\ Fnominal \mid Prepo\ Spartida \mid \lambda$

$Spartida \rightarrow Fnominal\ Prepo\ Fnominal\ Numeral$

$Pron_qual \rightarrow Qual \mid Quais$

$Sub_comum \rightarrow times \mid partida \mid oitavas \mid final$

$Art \rightarrow o \mid a \mid os$

$Verbo_ser \rightarrow \acute{e} \mid s\tilde{a}o \mid \lambda$

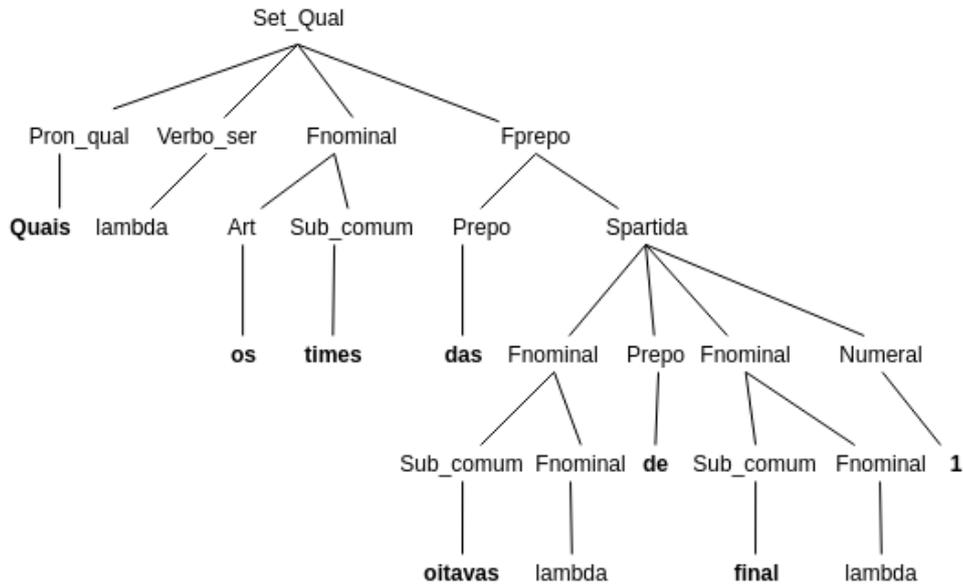
$Prepo \rightarrow do \mid na \mid das \mid de$

$Numeral \rightarrow 1$

De acordo com a gramática acima podemos observar a derivação da sentença, *quais os times das oitavas de final 1* na Figura 20

Pode-se ter também consultas em que não se sabe especificamente o número da partida e seja necessário dizer qual os times desta partida, como por exemplo *qual o time vencedor na partida entre França e Croácia?*, onde incluirá a conjunção (*e*). Podemos observar os elementos que compõe a sentença e os agrupamentos dos sintagmas respectivamente nas Tabelas 12 e 13.

Figura 20 – Árvore de derivação para a sentença, *quais os times das oitavas de final 1?*



Fonte: Autor.

Tabela 12 – Elementos sintáticos da sentença, *qual o time vencedor na partida entre França e Croácia?*

Qual =	Pronome qual
o =	Artigo
time =	Substantivo comum
vencedor =	Substantivo comum
na =	Preposição (em + a)
partida =	Substantivo comum
entre =	Preposição
França =	Substantivo próprio
e =	Conjunção
Croácia =	Substantivo próprio

Fonte: Autor.

Tabela 13 – Agrupamento dos sintagmas da sentença *qual o time vencedor na partida entre França e Croácia?*

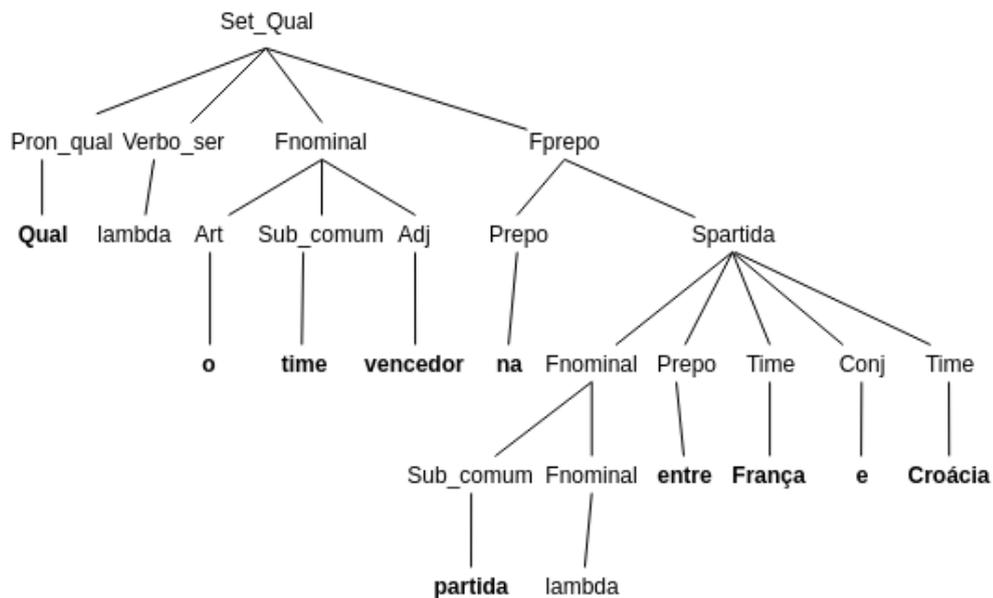
Qual =	Pronome qual
o time =	Sintagma nominal
vencedor =	Adjetivo
na partida =	Sintagma preposicional
entre França e Croácia =	Sintagma preposicional

Fonte: Autor.

A palavra *vencedor* na sentença, funciona como adjetivo para *time*, assim, além de incluir conjunções na nossa gramática devemos também especificar o tratamento de adjetivos.

$Set_Qual \rightarrow Pron_qual\ Verbo_ser\ Fnominal\ Fprepo$
 $Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$
 $Fnominal \rightarrow Art\ Sub_comum\ Adj \mid Art\ Sub_proprio\ Adj \mid \lambda$
 $Fprepo \rightarrow Prepo\ Fnominal \mid Prepo\ Spartida \mid \lambda$
 $Spartida \rightarrow Fnominal\ Prepo\ Fnominal\ Numeral$
 $Spartida \rightarrow Fnominal\ Prepo\ Time\ Conj\ Time$
 $Pron_qual \rightarrow Qual \mid Quais$
 $Sub_comum \rightarrow time \mid partida \mid final$
 $Time \rightarrow França \mid Croácia$
 $Conj \rightarrow e$
 $Art \rightarrow a$
 $Prepo \rightarrow na \mid entre$
 $Adj \rightarrow vencedor$

Figura 21 – Árvore de derivação para a sentença *qual o time vencedor na partida entre França e Croácia?*



Fonte: Autor.

Além das sentenças para o tipo *qual* a nossa LNC também aceita consultas para o tipo *quem*, *que*, *onde*, *quando* e *quanto*. A análise para esses tipos de consultas é semelhante a vista anteriormente (tipo *qual*) e poderão ser encontradas no Apêndice B deste trabalho, nesta seção apenas mostraremos algumas diferenças de cada tipo de consulta de acordo com a nossa ontologia.

As consultas do tipo *quem*, estão relacionadas a pessoas e nunca a coisas, sendo assim, para a ontologia esse tipo de consulta está relacionado a indivíduos das classes Jogador e Técnico.

Para consultas do tipo *que* se referem tanto a coisas como também a pessoas na ontologia, podendo incluir até mesmo preposições antes do pronome *que*, como por exemplo na sentença *em que dia aconteceu a partida final?*.

Já as consultas do tipo *onde*, referem-se a lugares. Em nossa ontologia, as únicas buscas possíveis são sobre o estádio ou região que ocorreu as partidas da copa da Rússia. Apenas consultas para saber o estádio ou região serão aceitas pela gramática.

Consultas do tipo *quando* remete a ideia de tempo. Com relação a ontologia apenas consultas sobre o dia que ocorreu as partidas podem ser respondidas. Sendo assim, a gramática só poderá aceitar sentenças referente a datas de partidas, por exemplo a sentença *quando aconteceu as oitavas de final 2?*.

Para consultas do tipo *quanto* nos referimos a quantidade das coisas. De acordo com a nossa ontologia poderemos apenas ter buscas relacionadas a quantidade de gols ou títulos de um time.

6.2.1.1 Limitações da LNC

A LNC para a nossa ontologia só poderá aceitar consultas para os tipos mencionados na Seção 6.2.1. Consultas que envolvem pronomes como *por quê* e *como* necessitam de respostas mais específicas, em que a nossa ontologia não está preparada para esses tipos de consultas. Portanto, a ontologia não foi modelada para suportar essas situações.

6.2.2 Gerando consultas SPARQL

O reconhecimento de uma sentença (em linguagem natural) dada em uma consulta a ontologia se dá pela análise sintática de através de *parses* que são algoritmos que recebem uma sentença e uma gramática e gera uma árvore de derivação. Dada uma árvore de derivação podemos mapeá-la para uma consulta SPARQL usando um dicionário que relacione termos da ontologia com a nossa LNC (ver exemplo na Tabela 14).

A Tabela 14 mostra um exemplo de um possível dicionário para serem usados na consulta SPARQL. Neste dicionário deve conter as palavras e suas variações presentes na ontologia. Por exemplo, o substantivo técnico é definido na ontologia como uma classe, mas se

Tabela 14 – Exemplo de dicionário para a ontologia.

Descrição	Classificação
técnico	_Class_
Brasil	_Individual_
técnico do	_Object_
títulos	_Data_

Fonte: Autor.

este for acompanhado com uma preposição passará a ser classificado como uma *Object Property*. Já a palavra títulos em nossa ontologia nos remete a *Data Property* relacionando a quantidade de títulos que um time possui. E assim, o dicionário pode ser desenvolvido atribuindo classificações para cada termo em uma sentença.

Cada tipo de consulta em linguagem natural também deve ser mapeado para os tipos de consultas em SPARQL. Para a ontologia descrita neste trabalho podemos associar as consultas umas às outras da seguinte forma: entradas como *qual*, *quem*, *que*, *onde* e *quando* para consultas do tipo **SELECT ?X**, já para entradas *quanto*, usamos **SELECT Count(?X)**.

(LUZ, 2013) apresenta em seu trabalho além de uma gramática para a sua LNC ele também descreve uma gramática para o SPARQL com o intuito de facilitar a sua construção.

Quando uma consulta em Linguagem Natural (LN) é aceita pela LNC é executado a classificação dos seus elementos (ver Tabela 14). Após isso, é feita uma outra análise sintática, Luz (2013) usa em seu modelo o algoritmo CKY (YOUNGER, 1967) no qual usa uma abordagem *bottom-up* ou seja, faz uma análise de cada elemento da sentença depois consulta a gramática, com isso é gerada uma árvore de derivação que será entrada para um algoritmo que fará toda a conversão para o SPARQL.

Adequando o modelo para o nosso trabalho podemos observar algumas árvores de derivação após a segunda análise sintática. Os símbolos terminais presentes na Figura 22 refere-se a gramática proposta por Luz (2013) para construção do SPARQL, então não adentraremos sobre sua definição formal, apenas o que é importante saber é como ficará a derivação final e o que ela nos mostrará.

S1: Qual é o técnico do Brasil?

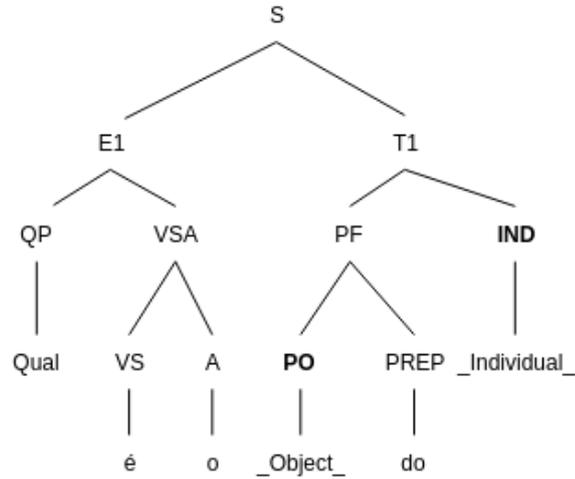
↓

Após a verificação de termos no dicionário.

S1: Qual é o _Object_ do _Individual_

↓

Figura 22 – Árvore de derivação para o SPARQL da sentença S1.



A sub árvore **E1** refere-se a consultas do tipo *qual, quem, quais*, referente a SELECT ?X, e a **T1** para sintagmas que possuem preposição em sua composição, se referindo a nossa tripla RDF no SPARQL onde c:_Individual_ p:_Object_ ?X.

Podemos adequar também perguntas quantitativas como a descrita na sentença abaixo.

S2: Quantos títulos tem o Brasil?

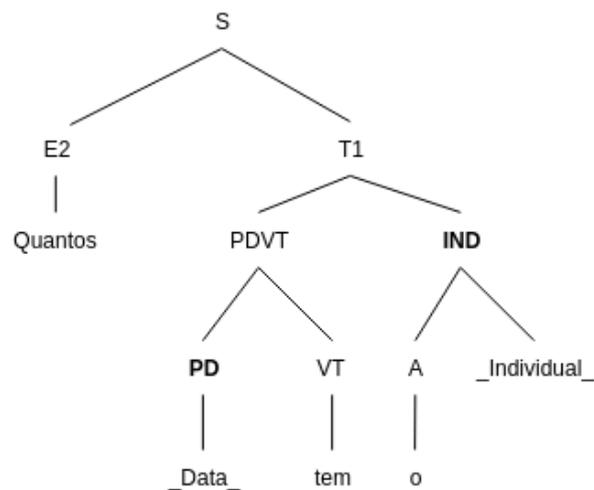


Após a verificação de termos no dicionário.

S2: Quantos _Data_ tem o _Individual_



Figura 23 – Árvore de derivação para o SPARQL da sentença S2.



O algoritmo de Luz (2013) percorre toda a árvore de derivação da raiz da esquerda para a direita deixando por final a raiz **S**. É identificado quais os nós que produzirão a estrutura, como por exemplo **SELECT ?X** concatenando-o em uma *string*. Após isso é verificado os nós que representam as triplas, gerando ao final uma string que é a concatenação das estruturas corretas em uma consulta SPARQL.

Para uma melhor visualização a árvore de derivação da Figura 22 seguirá os seguintes passos.

1. 1ª iteração T1, gera a *string*: { c:_Individual_ c:_Object_ ?X}.
2. 2ª iteração E1, gera a *string* final: SELECT ?X { c:_Individual_ c:_Object_ ?X}.

As palavras originais presentes na ontologia serão utilizadas para a consulta final para cada sentença, ou seja, para o Código-fonte 3 será utilizada a seguinte sintaxe do SPARQL.

Código-fonte 3 – Consulta SPARQL após as devidas alterações.

```

1 PREFIX c: <nossa_URI#>
2 SELECT ?X
3 WHERE {c:Brasil c:tem_tecnico ?X}

```

6.2.3 Considerações sobre a conversão

Como podemos observar é possível a conversão de uma consulta em LN para a nossa ontologia através do SPARQL, mas de acordo com o modelo é fato que a gramática de SPARQL deve ser construída ou até mesmo ao utilizar as existentes deve-se adequá-las a nossa ontologia. A construção de um dicionário, é bem relevante, mas dependendo da complexidade da ontologia seu uso possa se transformar em algo bem dispendioso. Além disso, um novo algoritmo deve ser desenvolvido para o auxílio das substituições dos termos na *string* final.

7 CONCLUSÕES E TRABALHOS FUTUROS

É notável o uso de ontologias hoje na web para o aprimoramento e rapidez nas buscas por informações e quando se fala em desenvolver para a língua portuguesa, ainda é algo que não é bem difundido, devido o português ser uma língua bem complexa de se trabalhar.

Apresentamos uma ontologia que mostrasse elementos sobre a copa da Rússia de 2018 para a língua portuguesa, mas para isso controlamos nosso português para diminuir a sua complexidade restringindo nossa ontologia para abranger apenas algumas estruturas da nossa linguagem. Com isso, a ontologia se mostrou favorável para se chegar ao nosso objetivo.

Mostramos também a possibilidade de conversão em LN para SPARQL em nossa ontologia, desde que se tenha um modelo adequado a sua ontologia e gramática, associando termos uma sentença para termos de uma ontologia, mas essa associação são problemas ainda conhecidos em PLN (LI *et al.*, 2003)

Podemos indicar como trabalhos futuros a nossa ontologia para que se possa incluir questões sobre os jogos das fases de grupo e assim popular ainda mais a base ontológica. O desenvolvimento de uma aplicação e de um modelo de conversão pode ser algo bem relevante para possíveis testes de PLN como também o aprimoramento da gramática, para aceitar mais consultas ou até mesmo desenvolvê-la por completo.

REFERÊNCIAS

- ALATRISH, E. Comparison some of ontology. **Journal of Management Information Systems**, v. 8, n. 2, p. 018–024, 2013.
- ALMEIDA, M. B. **Noções básicas para uso do Protégé**. 2006. Disponível em: <http://mba.eci.ufmg.br/onto_frames>.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação, Brasília**, SciELO Brasil, v. 32, n. 3, p. 7–20, 2003. Disponível em: <<http://www.scielo.br/pdf/ci/v32n3/19019/>>. Acesso em: 14 abr. 2019.
- BARBALHO, I. M.; SILVA, P. d. A.; FERNANDES, F. R. d. S.; NETO, F. M.; LEITE, C. R. An investigation on the use of ontologies for pattern classification-study applied to the monitoring of food intake. In: ACM. **Proceedings of the Euro American Conference on Telematics and Information Systems**. [S.l.], 2018. p. 4.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. *et al.* The semantic web. **Scientific american**, New York, NY, USA:, v. 284, n. 5, p. 28–37, 2001.
- CARDOSO, J. The semantic web vision: Where are we? **IEEE Intelligent systems**, IEEE, v. 22, n. 5, p. 84–88, 2007.
- CHOMSKY, N. Three models for the description of language. **IRE Transactions on information theory**, IEEE, v. 2, n. 3, p. 113–124, 1956.
- FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. American Association for Artificial Intelligence, 1997.
- FUCHS, N. E.; KALJURAND, K.; KUHN, T. Attempto controlled english for knowledge representation. In: **Reasoning Web**. [S.l.]: Springer, 2008. p. 104–124.
- GAMMA, E. **Padrões de Projetos: Soluções Reutilizáveis**. [S.l.]: Bookman editora, 2009.
- GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, Elsevier, v. 5, n. 2, p. 199–220, 1993.
- HORRIDGE, M.; KNUBLAUCH, H.; RECTOR, A.; STEVENS, R.; WROE, C. A practical guide to building owl ontologies using the protégé-owl plugin and co-ode tools edition 1.0. **University of Manchester**, 2011.
- JOSHI, A. K. **Natural Language Processing**. [S.l.]: Science, 1991. v. 253. 1242–1249 p.
- JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition**. 1st. ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2000. ISBN 0130950696.
- KLYNE, G.; CARROLL, J. J. **Resource Description Framework (RDF): Concepts and Abstract Syntax**. 2004. W3C Recommendation. Disponível em: <<http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>>.

LI, Y.; BANDAR, Z. A.; MCLEAN, D. An approach for measuring semantic similarity between words using multiple information sources. **IEEE Transactions on knowledge and data engineering**, IEEE, v. 15, n. 4, p. 871–882, 2003.

LUZ, F. F. **Consulta a ontologias utilizando linguagem natural controlada**. Tese (Doutorado) — Universidade de São Paulo, 2013.

NOY, N. F.; MCGUINNESS, D. L. *et al.* **Ontology development 101: A guide to creating your first ontology**. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05 and . . . , 2001.

PARALIC, J.; KOSTIAL, I. Ontology-based information retrieval. **Information and Intelligent Systems, Croatia**, p. 23–28, 2003.

PRUD'HOMMEAUX, E.; SEABORNE, A. **SPARQL Query Language for RDF**. 2008. W3C Recommendation. <<http://www.w3.org/TR/rdf-sparql-query/>>. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/>>.

SCHWITTER, R. Controlled natural languages for knowledge representation. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the 23rd International Conference on Computational Linguistics: Posters**. [S.l.], 2010. p. 1113–1121.

SILVA, B. C. D. da; MONTILHA, G.; RINO, L. H. M.; SPECIA, L.; NUNES, M. d. G. V.; JR, O. N. de O.; MARTINS, R. T.; PARDO, T. A. S. Introdução ao processamento das línguas naturais e algumas aplicações. **Série de Relatórios do Núcleo Interinstitucional de Linguística Computacional**, v. 3, 2007.

SIPSER, M. **Introdução à teoria da computação**. [S.l.]: Cengage Learning, 2007. 105–106 p.

STANFORD. **Protégé**. 2012. Disponível em: <<http://protege.stanford.edu/>>.

SURE, Y.; STAAB, S.; STUDER, R. On-to-knowledge methodology (otkm). In: **Handbook on ontologies**. [S.l.]: Springer, 2004. p. 117–132.

SWARTOUT, B.; PATIL, R.; KNIGHT, K.; RUSS, T. Toward distributed use of large-scale ontologies. In: **Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems**. [S.l.: s.n.], 1996. p. 138–148.

TORRES, C. E. A. **Transformación de lenguaje natural en Sparql para consultas de tipo Factoid**. Tese (Doutorado) — (Ingeniería de sistemas) - Universidad Nacional de San Agustín, Arequipa, Peru, 2017. Disponível em: <<http://repositorio.unsa.edu.pe/handle/UNSA/3434/>>. Acesso em: 20 mar. 2019.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da informação**, Brasília, v. 31, n. 2, p. 152–162, 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n2/12918/>>. Acesso em: 10 out. 2018.

YOUNGER, D. H. Recognition and parsing of context-free languages in time n^3 . **Information and control**, USA, v. 10, n. 2, p. 189–208, 1967. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S001999586780007X/>>. Acesso em: 16 abr. 2019.

**APÊNDICE A – QUESTIONÁRIO UTILIZADO PARA COLETAR DADOS SOBRE A
COPA DA RÚSSIA 2018**

Pesquisa Sobre a Copa da Rússia 2018

Este questionário tem como objetivo buscar qual a visão das pessoas sobre a Copa da Rússia 2018 abrangendo alguns tópicos referente a mesma. Se você é um amante por futebol, ou se interessa pelo assunto, por favor, responda o questionário abaixo. Será de grande ajuda e, vale lembrar, é tudo anônimo! Desde já, agradeço!

***Obrigatório**

1. O que mais lhe interessa saber sobre as partidas na Copa? Pode marcar mais de um item.

*

Marque todas que se aplicam.

- Quantidade de gols do time vencedor
- Quantidade geral de gols
- Qual jogador fez gol(s)
- Time campeão
- Dia e hora da partida
- Local da partida
- Quem são os comentaristas
- Quem são os árbitros
- Jogador revelação
- Outro: _____

2. Pra você é importante saber detalhes sobre cada jogador? *

Marcar apenas uma oval.

- Sim *Ir para a pergunta 3.*
- Não *Ir para a pergunta 4.*

Informações sobre jogadores

3. O que é interessante saber sobre cada jogador? *

Marque todas que se aplicam.

- Nome do jogador
- País onde vive
- Qual seleção joga
- Qual seu clube
- Idade
- Outro: _____

Informações sobre seleções

4. Pra você é importante saber detalhes sobre cada seleção? *

Marcar apenas uma oval.

- Sim *Ir para a pergunta 5.*
- Não *Ir para a pergunta 6.*

Informações sobre seleções

5. O que é interessante saber sobre cada seleção? *

Marque todas que se aplicam.

- Qual o treinador
- Títulos conquistados
- Nome dos jogadores
- Quem são os artilheiros
- Quais os jogadores reservas
- Gols marcados na copa
- Gols sofridos na copa
- Outro: _____

Sobre a copa em geral

6. Sobre a Copa da Rússia 2018 o que é crucial saber? Pode marcar mais de um item. *

Marque todas que se aplicam.

- Time Campeão
- Quem é o maior artilheiro
- Quem fez gol contra
- Quais as partidas
- Quais times foram para a final
- Quais times foram para a semi-final
- Quais times foram para as quartas de final
- Quais times foram para as oitavas de final
- Quantidade de gols na final
- Quantidade de gols na semi-final
- Outro: _____

Powered by



APÊNDICE B – CONTINUAÇÃO DAS ANÁLISES DA SUBSEÇÃO 6.2.1

B.0.1 *Tratando consultas do tipo Quem?*

A primeira sentença (ver Tabela 15) a ser analisada incluirá em nossa gramática o símbolo não terminal *Set_Quem*.

Tabela 15 – Elementos sintáticos da sentença, *quem é o técnico do Brasil?*.

Quem =	Pronome quem
é =	Verbo ser na 3ª pessoa do singular no presente do indicativo
o =	Artigo
técnico =	Substantivo comum
do =	Preposição
Brasil =	Substantivo próprio

Fonte: Autor.

A sua divisão em sintagmas é semelhante aos das consultas do tipo *qual* (ver Subseção 6.2.1), assim temos:

Tabela 16 – Agrupamento dos sintagmas da sentença, *quem é o técnico do Brasil?*.

Quem =	Pronome quem
é =	Verbo ser na 3ª pessoa do singular no presente do indicativo
o técnico =	Sintagma nominal
do Brasil =	Sintagma preposicional

Fonte: Autor.

A gramática a seguir mostra a estruturação da sentença *quem é o técnico do Brasil?* para seu reconhecimento na linguagem. Sua árvore de derivação pode ser vista na Figura 24.

$Set_Quem \rightarrow Pron_quem \ Verbo_ser \ FnominalPessoa \ Fnominal \ Fprepo$

$Fnominal \rightarrow Sub_comum \ Fnominal \mid Sub_proprio \ Fnominal \mid \lambda$

$FnominalPessoa \rightarrow Art \ Sub_tecnico \mid Art \ Sub_jogador$

$Fprepo \rightarrow Prepo \ Fnominal$

$Pron_quem \rightarrow Quem$

$Verbo_ser \rightarrow é$

$Sub_proprio \rightarrow Brasil$

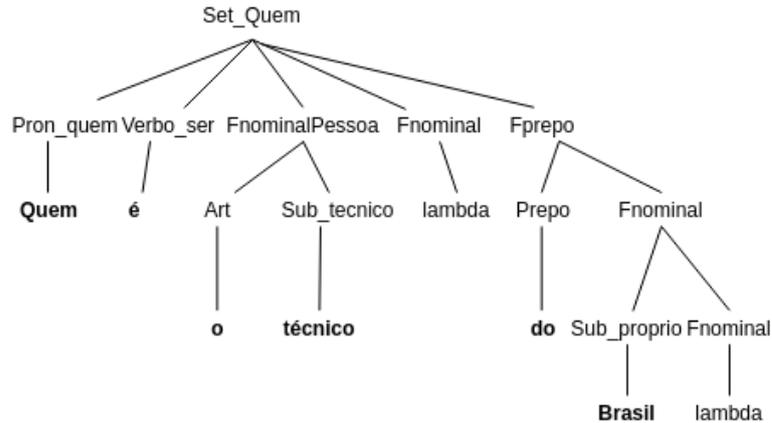
$Sub_tecnico \rightarrow técnico \mid técnicos$

Sub_jogador \rightarrow jogador | jogadores

Art \rightarrow o | os

Prepo \rightarrow do

Figura 24 – Árvore de derivação para a sentença, *quem é o técnico do Brasil?*



Fonte: Autor.

A seguinte sentença a ser analisada é *quem são os jogadores do Brasil?*, semelhante a vista anteriormente. Assim, podemos observar abaixo a gramática e sua respectiva árvore de derivação (ver Figura 25) para esta sentença.

Set_Quem \rightarrow Pron_quem Verbo_ser FnominalPessoa Fnominal Fprepo

FnominalPessoa \rightarrow Art Sub_tecnico | Art Sub_jogador

Fnominal \rightarrow Sub_comum Fnominal | Sub_proprio Fnominal

Fprepo \rightarrow Prepo Fnominal | λ

Pron_quem \rightarrow Quem

Verbo_ser \rightarrow é | são

Sub_proprio \rightarrow Brasil

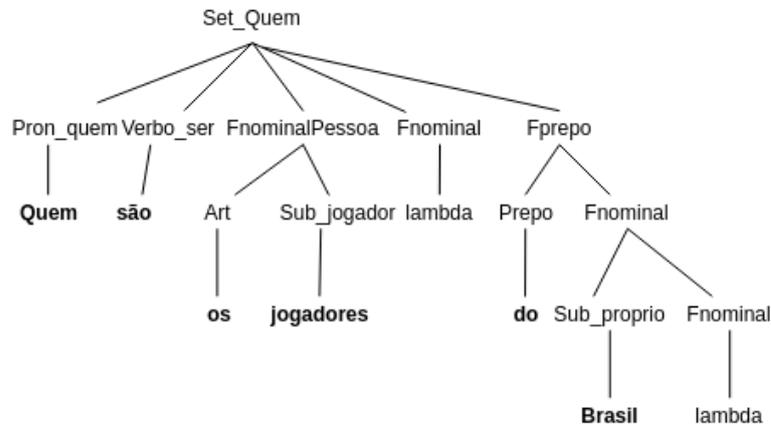
Sub_tecnico \rightarrow técnico | técnicos

Sub_jogador \rightarrow jogador | jogadores

Art \rightarrow o | os

Prepo \rightarrow do | na

Figura 25 – Árvore de derivação para a sentença, *quem são os jogadores do Brasil?*



Fonte: Autor.

B.0.2 Tratando consultas do tipo *Que?*

A primeira sentença a ser analisada *que horas aconteceu a partida final?* trata-se de um objeto *hora*. Podemos observar na Tabela respectivamente os seus elementos vistos separadamente e sua divisão em sintagmas.

Tabela 17 – Elementos sintáticos da sentença, *que horas aconteceu a partida final?*.

Que =	Pronome que
horas =	Substantivo comum
aconteceu =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
a =	Artigo
partida =	Substantivo comum
final =	Substantivo comum

Fonte: Autor.

Tabela 18 – Agrupamento dos sintagmas da sentença, *que horas aconteceu a partida final?*.

Que =	Pronome que
horas =	Sintagma nominal
aconteceu =	Sintagma verbal
a partida final =	Sintagma Fnominal

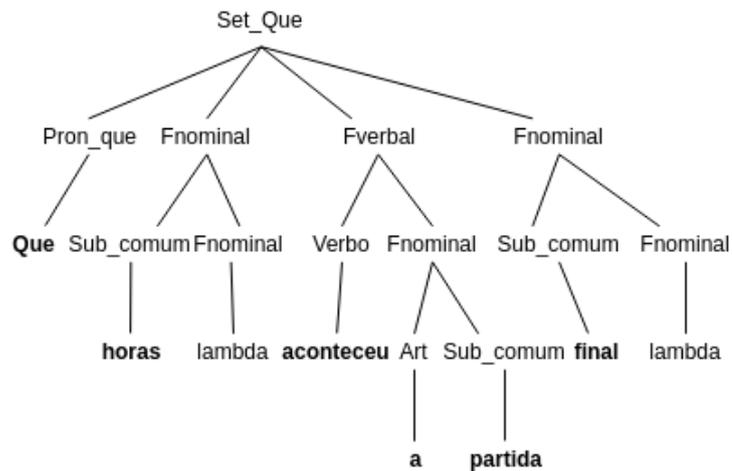
Fonte: Autor.

A gramática para esse tipo de consulta é semelhante as vistas anteriores, com a diferença que incluímos o símbolo não terminal *Set_que* em sua estrutura, vejamos abaixo.

$Set_Que \rightarrow Pron_que\ Fnominal\ Fverbal\ Fnominal$
 $Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$
 $Fnominal \rightarrow Art\ Sub_comum \mid Art\ Sub_proprio$
 $Fverbal \rightarrow Verbo\ Fnominal$
 $Pron_que \rightarrow Que$
 $Verbo \rightarrow aconteceu$
 $Sub_comum \rightarrow horas \mid partida \mid final$
 $Art \rightarrow a$

A Figura 26 mostra a árvore de derivação para a sentença acima.

Figura 26 – Árvore de derivação para a sentença, *que horas aconteceu a partida final?*



Fonte: Autor.

É possível também o uso de preposições antes do pronome *que*. Usamos a sentença *em que dia aconteceu a partida final?* para a nossa análise e atribuímos uma gramática para tal tratamento.

Tabela 19 – Elementos sintáticos da sentença, *em que dia aconteceu a partida da semifinal 1?*.

Em =	Preposição
que =	Pronome que
dia =	Substantivo comum
aconteceu =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
a =	Artigo
partida =	Substantivo comum
da =	Preposição (de + a)
semifinal =	Substantivo comum
1 =	Número

Fonte: Autor.

Tabela 20 – Agrupamento dos sintagmas da sentença, *em que dia aconteceu a partida da semifinal 1?*.

Em Que =	Preposição + pronome que
dia =	Sintagma nominal
aconteceu =	Sintagma verbal
a partida =	Sintagma Fnominal
da semifinal 1 =	Sintagma preposicional

Fonte: Autor.

$Set_Que \rightarrow Pron_que\ Fnominal\ Fverbal\ Fnominal$
 $Set_Que \rightarrow Prepo_em\ Pron_que\ Fnominal\ Fverbal\ Fprepo$
 $Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$
 $Fnominal \rightarrow Art\ Sub_comum \mid Art\ Sub_proprio$
 $Fverbal \rightarrow Verbo\ Fnominal$
 $Fprepo \rightarrow Prepo\ Fnominal \mid Prepo\ Ssemifinal$
 $Ssemifinal \rightarrow Fnominal\ Numeral$
 $Pron_que \rightarrow Que$
 $Prepo \rightarrow da$
 $Prepo_em \rightarrow em$
 $Verbo \rightarrow aconteceu$
 $Sub_comum \rightarrow horas \mid dia \mid partida \mid final$
 $Art \rightarrow a$

A Figura 27 mostra a árvore de derivação para a sentença, *em que dia aconteceu a partida da semifinal 1*.

B.0.3 Tratando consultas do tipo Onde?

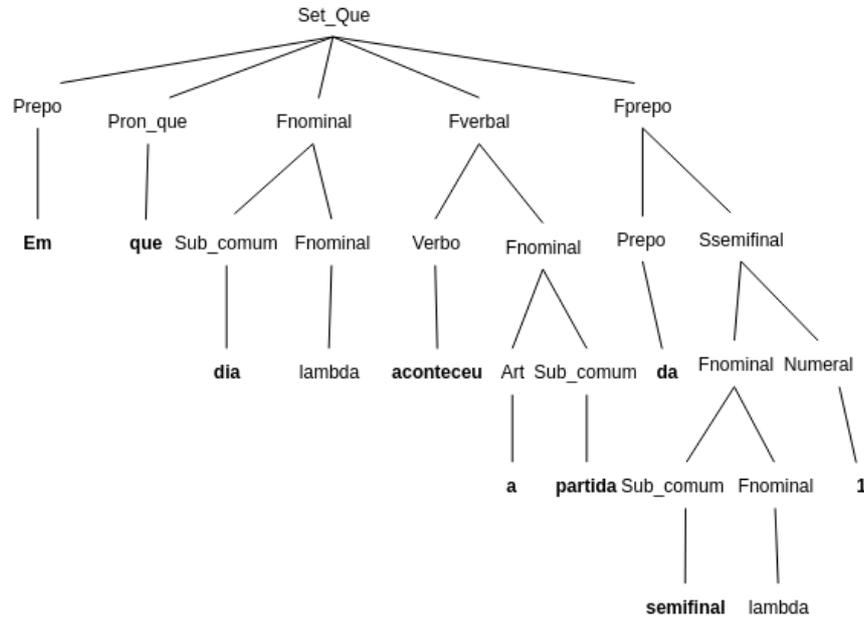
Na Tabela 21 podemos observar a estrutura para a sentença *onde aconteceu a partida final?*. A Tabela 22 mostra sua divisão em sintagmas para melhor representação na gramática.

Tabela 21 – Elementos sintáticos da sentença, *onde aconteceu a partida final?*.

Onde =	Pronome onde
aconteceu =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
a =	Artigo
partida =	Substantivo comum
final =	Substantivo comum

Fonte: Autor.

Figura 27 – Árvore de derivação para a sentença, *em que dia aconteceu a partida da semifinal 1?*



Fonte: Autor.

Tabela 22 – Agrupamento dos sintagmas da sentença, *onde aconteceu a partida final?*.

Onde =	Pronome onde
aconteceu =	Sintagma verbal
a partida final =	Sintagma nominal

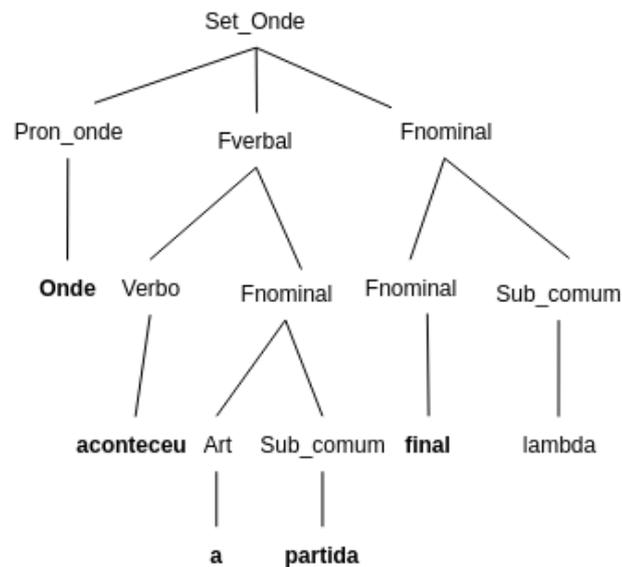
Fonte: Autor.

Adaptamos nossa gramática para consultas do tipo *onde*, em que incluímos um símbolo não terminal que chamaremos de *Set_Onde*. Assim, para esta sentença construímos a seguinte gramática.

$Set_Onde \rightarrow Pron_onde\ FVerbal\ Fnominal$
 $Fnominal \rightarrow Sub_comum\ Fnominal \mid Art\ Sub_comum \mid \lambda$
 $Fverbal \rightarrow Verbo\ Fnominal$
 $Pron_onde \rightarrow Onde$
 $Verbo \rightarrow aconteceu$
 $Sub_comum \rightarrow partida \mid final$
 $Art \rightarrow a$

Para a descrição de partidas como, quartas de final, oitavas de final e semifinal, a gramática é semelhante a da sentença mostrada na Tabela 10 da Subseção 6.2.1. Podemos observar a árvore de derivação para a sentença *onde aconteceu a partida final* na Figura 28.

Figura 28 – Árvore de derivação para a sentença, *onde aconteceu a partida final?*



Fonte: Autor.

B.0.4 Tratando consultas do tipo Quando?

A Tabela 23 detalha cada elemento de sua estrutura e logo após a gramática associada. O agrupamento em sintagmas é semelhante a sentença da Tabela 22, apenas incluindo a parte preposicional. A árvore de derivação para esta sentença é mostrada na Figura 29.

Tabela 23 – Elementos sintáticos da sentença, *quando aconteceu a partida de oitavas de final 2?*.

Quando =	Pronome quando
aconteceu =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
as =	Artigo
oitavas =	Substantivo comum
de =	Preposição
final =	Substantivo comum
2 =	Numeral

Fonte: Autor.

$Set_Quando \rightarrow Pron_quando\ Fverbal\ Fprepo$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Art\ Sub_comum \mid \lambda$

$Fprepo \rightarrow Prepo\ Fnominal \mid Prepo\ Spartida \mid \lambda$

$Fverbal \rightarrow Verbo\ Fnominal$

$Spartida \rightarrow Fnominal\ Prepo\ Fnominal\ Numeral$

Pron_quando → *Quando*

Verbo → *aconteceu*

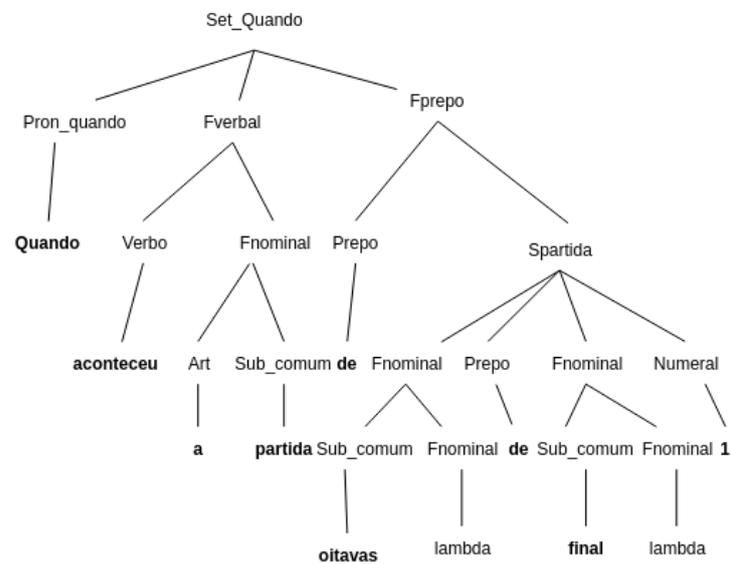
Sub_comum → *partida | oitavas | final*

Art → *a*

Prepo → *de*

Numeral → *2*

Figura 29 – Árvore de derivação para a sentença, *quando aconteceu a partida de oitavas de final 2?*



Fonte: Autor.

B.0.5 Tratando consultas do tipo Quanto?

A Tabela 26 e 27 mostra os elementos presentes na sentença e o agrupamento dos sintagmas, respectivamente.

Tabela 24 – Elementos sintáticos da sentença, *quantos gols a França fez na partida final?*.

Quantos =	Pronome quanto
gols =	Substantivo comum
a =	Artigo
França =	Substantivo próprio
fez =	Verbo na 3ª pessoa do pretérito perfeito do indicativo
na =	Preposição (em + a)
partida =	Substantivo comum
final =	Substantivo comum

Fonte: Autor.

Tabela 25 – Agrupamento dos sintagmas da sentença, *quantos gols a França fez na partida final?*.

Quantos =	Pronome quanto
gols =	Sintagma nominal
a França =	Sintagma nominal
fez =	Sintagma verbal
na partida final =	Sintagma preposicional

Fonte: Autor.

Definindo uma gramática para tratamento desses tipos de sentenças temos:

$Set_Quanto \rightarrow Pron_quanto\ Fnominal\ FVerbal\ Fprepo$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$

$Fnominal \rightarrow Art\ Sub_proprio \mid Art\ Sub_comum$

$Fverbal \rightarrow Verbo\ Fnominal$

$Fprepo \rightarrow Prepo\ Fnominal \mid \lambda$

$Pron_quanto \rightarrow Quantos$

$Verbo \rightarrow fez$

$Sub_proprio \rightarrow França$

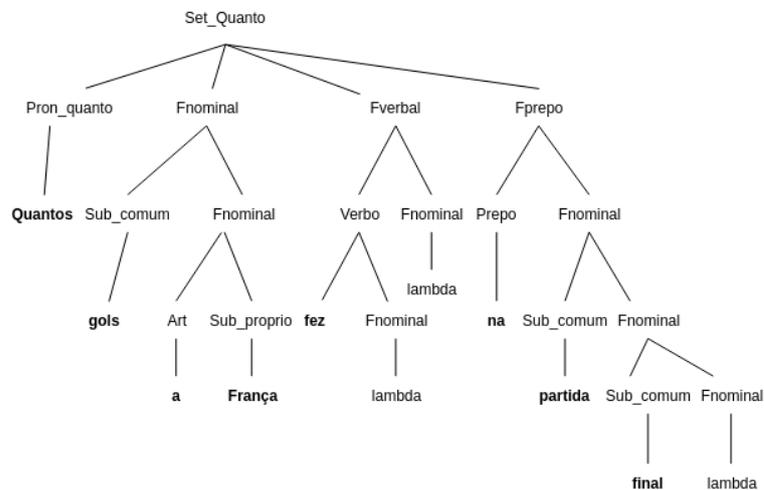
$Sub_comum \rightarrow partida \mid final \mid gols$

$Prepo \rightarrow na$

$Art \rightarrow a$

A árvore de derivação para a sentença acima poderá ser observada na Figura 30.

Figura 30 – Árvore de derivação para a sentença, *quantos gols a França fez na partida final?*



Fonte: Autor.

A segunda sentença a ser analisada é *quantos títulos tem o Brasil?*. Esta sentença trata do verbo *ter* conjugado na 1ª pessoa do presente do indicativo, apenas neste caso isso ocorrerá, pois como visto nas sentenças anteriores, os verbos estão conjugados no pretérito. Portanto fizemos um tratamento específico para essa situação em que nossa gramática poderá que aceitar.

Tabela 26 – Elementos sintáticos da sentença, *quantos títulos tem o Brasil?*.

Quantos =	Pronome quanto
títulos =	Substantivo comum
tem =	Verbo na 1ª pessoa do presente do indicativo
o =	Artigo
Brasil =	Substantivo próprio

Fonte: Autor.

Tabela 27 – Agrupamento dos sintagmas da sentença, *quantos títulos tem o Brasil?*.

Quantos =	Pronome quanto
títulos =	Sintagma nominal
tem =	Verbo na 1ª pessoa do presente do indicativo
o Brasil =	Sintagma nominal

Fonte: Autor.

A gramática abaixo mostra as produções para este tipo de tratamento e em seguida a árvore de derivação para a sentença *quantos títulos tem o Brasil?* (ver Figura 31).

$Set_Quanto \rightarrow Pron_quanto\ Ftitulos\ Fnominal$

$Fnominal \rightarrow Sub_comum\ Fnominal \mid Sub_proprio\ Fnominal \mid \lambda$

$Fnominal \rightarrow Art\ Sub_proprio \mid Art\ Sub_comum$

$Ftitulos \rightarrow Titulos\ Verbo_ter$

$Pron_quanto \rightarrow Quantos$

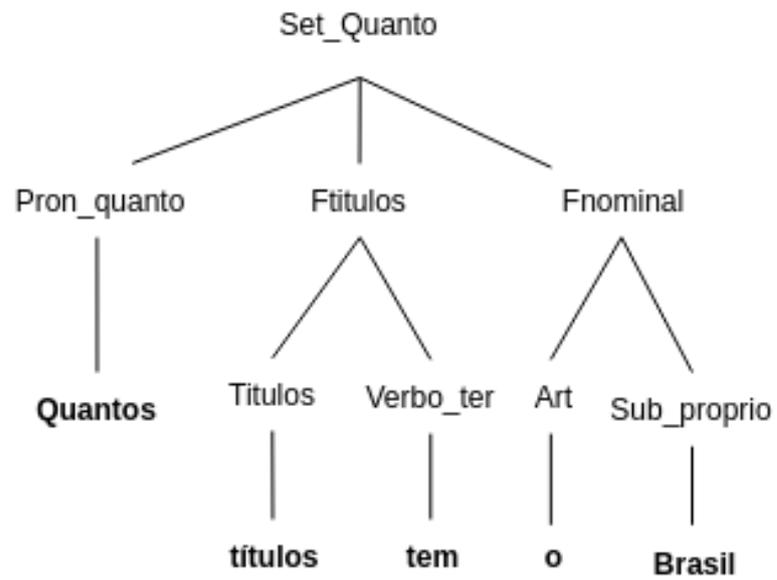
$Verbo_ter \rightarrow tem$

$Sub_proprio \rightarrow Brasil$

$Titulos \rightarrow titulos \mid mundiais$

$Art \rightarrow o$

Figura 31 – Árvore de derivação para a sentença, *quantos títulos tem o Brasil?*



Fonte: Autor.