

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS DA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
INFORMAÇÃO**

LUANA CARLA DE MOURA DOS SANTOS

**RECUPERAÇÃO DA INFORMAÇÃO EM ACERVOS DIGITAIS
DE JORNAIS: proposta para uso de ontologia no domínio do futebol**

DISSERTAÇÃO DE MESTRADO

Florianópolis, 2016

LUANA CARLA DE MOURA DOS SANTOS

**RECUPERAÇÃO DA INFORMAÇÃO EM ACERVOS DIGITAIS
DE JORNAIS: proposta para uso de ontologia no domínio do futebol**

Dissertação de mestrado apresentada à Banca Examinadora do Programa de Pós-Graduação em Ciência da Informação do Centro de Ciências da Educação da Universidade Federal de Santa Catarina, como requisito parcial para a obtenção do título de Mestre em Ciência da Informação, área de concentração Gestão da Informação, linha de pesquisa Organização, Representação e Mediação da Informação e do Conhecimento.

Orientadora: Prof^a Dra. Marisa Bräscher Basílio Medeiros.

Florianópolis

2016

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Santos, Luana Carla de Moura dos
RECUPERAÇÃO DA INFORMAÇÃO EM ACERVOS DIGITAIS DE
JORNAIS : proposta para uso de ontologia no domínio do
futebol / Luana Carla de Moura dos Santos ; orientadora,
Marisa Bräscher - Florianópolis, SC, 2016.
201 p.

Dissertação (mestrado) - Universidade Federal de Santa
Catarina, Centro de Ciências da Educação. Programa de Pós
Graduação em Ciência da Informação.

Inclui referências

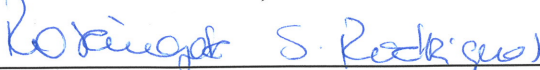
1. Ciência da Informação. 2. Recuperação da Informação. 3.
Acervo Digital. 4. Ontologia de Domínio. 5. Futebol. I.
Bräscher, Marisa. II. Universidade Federal de Santa
Catarina. Programa de Pós-Graduação em Ciência da Informação.
III. Título.

LUANA CARLA DE MOURA DOS SANTOS

**RECUPERAÇÃO DA INFORMAÇÃO EM ACERVOS DIGITAIS
DE JORNAIS: proposta para uso de ontologia no domínio do futebol**

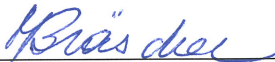
Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Informação do Centro de Ciências da Educação da Universidade Federal de Santa Catarina em cumprimento a requisito parcial para a obtenção do título de Mestre em Ciência da Informação.

APROVADA PELA COMISSÃO EXAMINADORA
EM FLORIANÓPOLIS, 07 DE JUNHO DE 2016

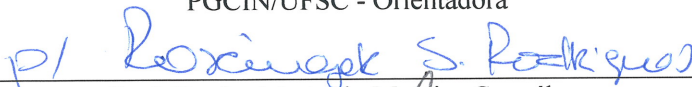


Prof.ª Dra. Rosângela Schwarz Rodrigues
Coordenadora do Curso

BANCA EXAMINADORA:



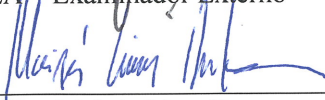
Prof.ª Dra. Marisa Bräscher Basílio Medeiros
PGCIN/UFSC - Orientadora



Prof. Dr. José Antonio Moreira González
Universidad Carlos III de Madrid – Examinador Externo



Prof. Dr. Cezar Karpinski
UNILA – Examinador Externo



Prof. Dr. Moisés Lima Dutra
PGCIN/UFSC – Examinador interno

AGRADECIMENTOS

Se fosse possível representar esses dois anos de mestrado em uma estrutura conceitual, certamente os principais conceitos seriam:

Conceito: Conhecimento

(Sinônimo: Aprendizado)

Definição/explicação:

Momentos em que foi possível apreender e compreender a ciência por meio de diferentes formas cognitivas. A troca intensiva de conhecimento, as críticas, a exposição de ideias divergentes e convergentes, a intuição, a experimentação, ou até mesmo a razão, modificaram o meu ser.

Elos de ligação/relacionamentos

Marisa Brascher: minha orientadora e sem dúvida alguma, importantíssima nessa trajetória. Obrigada por compartilhar o seu conhecimento, pelo seu comprometimento e acima de tudo, obrigada pela sensibilidade.

Professores do PGCin: possibilitaram que a minha essência investigativa e crítica fosse potencializada, por meio da promoção de discussões enriquecedoras, e no compartilhamento de leituras instigantes. Muito obrigada.

Colegas do grupo de pesquisa - Organização do Conhecimento e Gestão Documental (UFSC): pela troca intensiva de conhecimento, pelas considerações significativas ao meu trabalho e pelos momentos descontraídos.

Conceito: Gratidão

(Sinônimos: Agradecimento; Reconhecimento)

Definição/explicação

Atribuída àqueles que generosamente doaram um pouco de si para concretização deste trabalho.

Elos de ligação/relacionamentos

Rafael: sempre presente. Obrigada pelo incentivo, pelo carinho e pela compreensão nos momentos em me mantive ausente.

Família: obrigada pela torcida sem igual, e pela alegria demonstrada em cada etapa vencida, em especial à minha irmã, que dividiu alegrias e angústias e esteve sempre flexível para me ajudar em situações que me impediam de estar focada na pesquisa.

Marta e Vitor: foram uma surpresa grata que a vida me trouxe. Obrigada amigos, foi muito bom contar com vocês. Uma das qualidades que mais admiro na nossa área, que é bastante interdisciplinar, é a partilha, a colaboração, o compartilhamento do conhecimento. Tenho a plena certeza que uma pesquisa não se faz sozinha, e tenho também a convicção que a inteligência é construída de forma colaborativa.

Alíne: amiga que sempre esteve presente em todas as etapas, me incentivou, me cobrou e me divertiu. Você foi a mais sincera (e presente) amizade que a Biblioteconomia me trouxe.

Ana e Fran: Muito mais que colegas de trabalho. Meninas vocês foram muito importantes. Sei que posso contar com vocês e a recíproca é verdadeira.

Camila e demais colegas de trabalho: pela compreensão nos momentos que tive que me ausentar do trabalho.

Conceito: Memória

Sinônimos: (Lembrança; Recordação)

Definição/explicação

Referente às experiências vividas durante os dois anos de mestrado, que trazem boas lembranças de momentos e pessoas que foram fundamentais.

Elos de ligação/relacionamentos

Colegas do Pgcin: turma 2014/1, foi muito bom evoluir com vocês durante esses dois anos.

Aos professores de outras instituições: que sempre, generosamente, quando em contato, se mostraram solícitos no desenvolvimento científico.

Eventos da área: certamente são de grande valia para o aprendizado acadêmico. Possibilitam descobertas de novas temáticas e a socialização com grandes pesquisadores da área. Em especial aos organizadores do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB).

Revistas científicas: importantes bases de conhecimento, que viabilizam o contato de pesquisas, muitas vezes embrionárias, com pesquisadores renomados na área. Obrigada a todos envolvidos.

Por fim, entende-se que para a concretização de uma pesquisa, a junção de todos envolvidos é de extrema importância, não sendo viável de forma isolada, portanto, se fosse possível a formação de uma estrutura conceitual, certamente, todos esses conceitos estariam relacionados.

“Faça o melhor que puder até saber mais.
Então, quando souber mais, faça ainda melhor.”

(Maya Angelou)

RESUMO

Esta pesquisa visa a modelagem de uma ontologia do domínio do futebol para o tratamento de variações diacrônicas e sincrônicas da língua, com intuito de apoiar a recuperação da informação em acervos digitais de jornais. Com a oferta dos acervos de jornais em sítios na Internet é preciso que os acessos a essas informações sejam facilitados, pois esses ambientes reúnem documentos de diferentes épocas e culturas e também estão em constante atualização. Por essa razão, necessitam de uma organização que contemple a diversidade linguística e de conteúdo. Quando a organização da informação é conduzida para um domínio do conhecimento, as ambiguidades podem ser tratadas e entendidas conforme o contexto que se aplicam. Assim, o futebol tem o jornal como importante fonte de informação e também apresenta mudanças gramaticais e variações linguísticas que se mostraram pertinentes a este estudo. Trata-se de uma pesquisa aplicada, utilizando como base um acervo digital de jornal. Possui caracterização exploratória e descritiva e abordagem qualitativa. Utiliza a metodologia *OntoForInfoScience*, de Mendonça (2015) para desenvolver a ontologia do domínio do futebol. A coleta das informações foi realizada em materiais de referência do domínio e em notícias de jornais. Cronologicamente, o recorte estabelecido abrange terminologia utilizada entre 1900 a 2015, período que contempla a existência dos clubes de futebol no Brasil. A coleta foi realizada em jornais digitalizados e também nascidos digitais. A partir da amostra dos termos criou-se artefatos como Glossário de Conceitos, Dicionário de Conceitos, Glossário de Relações, Tabela de Conceitos e Valores e modelo de representação gráfica, em que foram tratados os aspectos linguísticos do domínio do futebol presentes nos acervos digitais. Com base do que foi modelado em linguagem natural foi formalizada a ontologia em linguagem lógica, com o auxílio do editor Protegé, etapa em que foram definidas classes, subclasses, propriedades de relação e equivalências. Como forma de avaliar a ontologia desenvolvida, elaborou-se questões de competência que foram executadas em linguagem SPARQL. Para atestar o uso da ontologia em ambientes compostos por jornais impressos e digitais foram realizadas buscas demonstrativas em um acervo real, analisando quais os resultados obtidos sem a ontologia e quais os prováveis resultados se ela fosse implementada. A análise dos resultados permitiu evidenciar que sem o uso da ontologia nos acervos digitais de jornais, a recuperação das informações é exaustiva e recupera documentos não relevantes, devido à ausência de relacionamentos entre os termos que

formam o domínio. Dessa forma, com a inclusão da ontologia no sistema, a pesquisa por informações não exigiria tanto letramento do usuário, visto que as ontologias tornam mais inteligentes os sistemas, pois suas relações são feitas com base na realidade. Assim, as principais contribuições científicas deste trabalho para a Ciência da Informação envolvem a análise e tratamento de documentos históricos do domínio do futebol, a fim de apoiar a recuperação da informação de acervos digitais de jornais, utilizando como recurso uma ontologia de domínio.

Palavras-chave: Recuperação da Informação. Acervo digital de jornal. Ontologia de domínio. Futebol.

ABSTRACT

This research proposes the making of a domain-ontology on soccer, to deal with the synchronic and diachronic variances of the language, and to aid information retrieval on digital files of newspapers. With the supply of newspaper archives in websites, it is necessary for the access to this information to be improved, because these sites group documents from different times and cultures and also are constantly under update. For that reason, they need an arrangement that encompasses the diversity of language and content. When the arrangement of the information is placed in a domain knowledge, all ambiguity can be dealt with and understood according to the context to which they apply. Thus, soccer has the newspaper as an important source of information, and also presents grammar changes and linguistic variations that are shown to be pertinent to this study. This is an applied research, based on a digital newspaper archive. It is exploratory and descriptive and a qualitative approach to its subject matter. Uses Mendonça's *OntoForInfoScience* methodology (2015) to develop the domain-ontology on soccer. The gathering of information was done based on reference material from the domain and in newspaper articles. The chronological interval that was established encompasses the terminology used between 1900 and 2015, period that covers the spawn of soccer clubs in Brazil. The data collection was from digitized and originally digital newspapers. Artifacts were created from the sample of terms, which are: glossary of concepts, dictionary of concepts, glossary of relationships, table of concepts and values, and graphic representation model, in which we deal with the linguistic aspects of the soccer domain, in the digital files. Based on what was modeled in natural language we formalized the ontology in logic language, with the help of the Protegé editor. At this stage, we defined categories, subcategories, relationship properties and equivalences. As a means to evaluate the developed ontology, questions of competence were created and executed in SPARQL language. In order to attest the usage of the ontology in environments with digital and printed newspapers some demonstrative searches were carried out in a real archive, analyzing what were the results without the ontology and the probable results if it were implemented. The analysis of the results allowed us to show that without the use of the ontology in the digital archive of newspaper the information retrieval is exhaustive and retrieves irrelevant documents, due to the lack of relationships among the terms that compose the domain. In this fashion, having the ontology included in the system, the research would not

require much specific knowledge of the user, since the ontologies make systems smarter because their relationships are based on reality. So the main scientific contributions of this work to the Information Science are about the analysis and management of historical documents on the soccer domain, in order to improve the information retrieval in digital archives of newspapers, using the domain ontology as a resource.

Keywords: Information retrieval. Digital newspaper archive. Domain ontology. Soccer. Football.

LISTA DE FIGURAS

Figura 1-Capa do Jornal O Globo de 17 de julho de 1950	44
Figura 2 - Capa jornal O Globo de 07 de junho de 1948.....	45
Figura 3 - Modelo Clássico de Busca.....	49
Figura 4 - Variantes da palavra (aspectos históricos).....	53
Figura 5 - Variantes da palavra (aspectos socioculturais)	54
Figura 6 - Variantes da palavra (contextos de uso).....	54
Figura 7 - Proposta para Construção de ontologia.....	73
Figura 8 - Aprendizado de ontologia em camadas.....	76
Figura 9 - Arquitetura da Web Semântica proposta em 2005.....	85
Figura 10 - Etapas que serão desenvolvidas nesta pesquisa	89
Figura 11 - Importar ontologia de fundamentação no Protégé 5.0	101
Figura 12 - Documento de referência para coleta de verbos	114
Figura 13 - Representação Gráfica com a amostra de conceitos da OntoFootballForNewspapers.....	122
Figura 14 - Representação esquemática das categorias básicas da UFO. .	124
Figura 15 - Classes da OntoFootballForNewspapers.....	127
Figura 16 - Classes principais definidas no Protégé	129
Figura 17 - Parte das classes e conceitos da hierarquia da OntoFootballForNewspapers.....	130
Figura 18 - Inclusão de sinônimos no Protégé.....	131
Figura 19 - Classes disjuntas	132
Figura 20 - Inclusão de Instâncias	133
Figura 21 - Propriedades de relação aplicadas no Protégé	137
Figura 22 – Representação gráfica das principais classes OntoFootballForNewspapers.....	138
Figura 23 - Execução da questão de competência 1 no SPARQL	139
Figura 24 - Execução da questão de competência 2 no SPARQL.....	140
Figura 25 - Execução da questão de competência 3 no SPARQL.....	141
Figura 26 - Execução da questão de competência 4 no SPARQL.....	142
Figura 27 - Recuperação no acervo Estadão do termo "Troféu"	144
Figura 28 - Demonstração da recuperação da Informação no acervo digital.....	147

Figura 29 - Demonstração da Recuperação da Informação - Estratégia de busca 1.....	148
Figura 30 – Texto 1 (Edição de 4 de dezembro de 1997, p.20).....	194
Figura 31 – Texto 2 (Edição de 17 de julho de 2000, p.24)	195
Figura 32 – Texto 3 (Edição de 20 de outubro de 1998, p.30)	196
Figura 33 – Texto 4 (Edição de 13 de junho de 1998, p.32)	197
Figura 34 – Texto 5 (Edição de 17 de janeiro de 2005, p.38)	198
Figura 35 – Texto 6 (Edição de 8 de dezembro de 1998, p.38).....	198
Figura 36 – Texto 7 (Edição de 21 de setembro de 1998, p.17).....	199
Figura 37 - Texto 8 (Edição de 14 de junho de 1981, p. 48)	200
Figura 38 – Texto 9 (Edição de 27 de junho de 1929, p.12)Termo: futebol;	200
Figura 39 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 40 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 41 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 42 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 43 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 44 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 45 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201
Figura 46 – Texto 9 (Edição de 27 de junho de 1929, p.12)	201

LISTA DE QUADROS

Quadro 1 - Comparação de alguns termos e expressões da Ciência da Informação e da Web semântica.....	55
Quadro 2 - Ontologias de fundamentação	63
Quadro 3 - Ciclo de construção de ontologias.....	66
Quadro 4 - Metodologias para construção de ontologias.....	68
Quadro 5 - Relações lógicas	75
Quadro 6 - Tipo de metadados quanto a sua função	80
Quadro 7 - Template de especificação da ontologia OntoFootballForNewspaper	90
Quadro 8 - Modelo de Coleta do Glossário de Conceitos	95
Quadro 9 - Modelo de Coleta do Glossário de Verbos	96
Quadro 10 - Modelo Dicionário de Conceitos.....	97
Quadro 11 - Modelo Tabela de Conceitos e Valores.....	98
Quadro 12 - Dicionário de Verbos.....	99
Quadro 13 - Modelo do Glossário de Relações	100
Quadro 14 - Template de Especificação da OntoFootballForNewspapers	109
Quadro 15 - Parte do Glossário de Conceitos da OntoFootballForNewspapers.....	112
Quadro 16 - Glossário de Verbos da OntoFootballForNewspapers	113
Quadro 17 - Parte do Dicionário de Conceitos da OntoFootballForNewspapers.....	116
Quadro 18 - Parte do Dicionário de Conceitos - Termo “Cera”	117
Quadro 19 - Tabela de Conceitos e Valores	118
Quadro 20 - Dicionário de Verbos da OntoFootballForNewspapers.....	120
Quadro 21 - Parte do Glossário das Relações da OntoFootballForNewspapers.....	121
Quadro 22 - Preenchimento das relações ontológicas das classes no Protégé.....	136
Quadro 23 - Termos recuperados Acervo Estadão	145
Quadro 24 - Resultados obtidos com o uso de operadores booleanos	146
Quadro 25 - Demonstração da Recuperação da Informação - Estratégia de busca 2	149
Quadro 26 - Demonstração da Recuperação da Informação com operadores booleanos Estratégia de busca 3	150
Quadro 27 - Demonstração da Recuperação da Informação - Estratégia de busca 3	151
Quadro 28 - Demonstração da Recuperação da Informação com operadores booleanos Estratégia de busca 3	152

LISTA DE SIGLAS E ABREVIATURAS

- ANJ** – ASSOCIASSÃO NACIONAL DO JORNAL
- DAML** - DARPA AGENT MARKUP LANGUAGE
- DC** - DUBLIN CORE
- DCMI** - DUBLIN CORE METADADA INITIATIVE
- FW** - FRAMEWORK DE WARWICK
- OIL** - ONTOLOGY INTERCHANGE LANGUAGE
- OWL** - WEB ONTOLOGY LANGUAGE
- RDF** - RESOURCE DESCRIPTION FRAMEWORK
- RDFS** - RESOURCE DESCRIPTION FRAMEWORK SCHEMA
- RI** - RECUPERAÇÃO DA INFORMAÇÃO
- SOC** – SISTEMA DE ORGANIZAÇÃO DO CONHECIMENTO
- SPARQL** - PROTOCOL AND RDF QUERY LANGUAGE
- SRI** - SISTEMA DE RECUPERAÇÃO DA INFORMAÇÃO
- TIC** - TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO
- UFO** - UNIFIED FOUNDATIONAL ONTOLOGY
- UNESCO** - ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA
- URI** - UNIFORM RESOURCE IDENTIFIER
- WS** – WEB SEMÂNTICA
- W3C** - WORLD WIDE WEB CONSORTIUM
- XML** - EXTENSIBLE MARKUP LANGUAGE

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Definição do problema	27
2	OBJETIVOS	30
2.1.1	Objetivo Geral	30
2.1.2	Objetivos específicos.....	30
2.2	Justificativa.....	30
3	FUNDAMENTAÇÃO TEÓRICA.....	33
3.1	Patrimônio digital	33
3.2	Patrimônio: O jornal do impresso ao digital.....	36
3.2.1	Informação, Futebol e Jornal	42
3.3	Recuperação da Informação	47
3.4	Busca Semântica.....	50
3.4.1	Web Semântica.....	55
3.4.3	Ontologias de domínio	58
3.4.4	Ontologias de fundamentação	62
3.4.5	Metodologia para construção de ontologias	66
3.4.6	Linguagens de representação descritiva	79
4	PROCEDIMENTOS METODOLÓGICOS.....	87
4.1	Descrição geral da metodologia de pesquisa	89
4.1.1	Avaliação da necessidade da ontologia	90
4.1.2	Especificação da ontologia	90
4.1.3	Aquisição de Informações	92
4.1.4	Conceitualização.....	96
4.1.5	Fundamentação ontológica.....	100
4.1.6	Formalização da ontologia	102
4.1.7	Avaliação da Ontologia	103

4.2	Demonstração do uso da ontologia na recuperação da informação em acervos digitais.....	104
5	APRESENTAÇÃO DOS RESULTADOS.....	106
5.1	Desenvolvimento da <i>OntoFootballForNewspapers</i>	106
5.1.1	Documento de especificação.....	109
5.1.2	Coleta de informações.....	111
5.1.3	Modelos conceituais.....	115
5.1.4	Reuso ontologia de fundamentação.....	123
5.1.5	Conteúdo ontológico.....	126
5.1.6	Elucidação das questões de competência.....	139
5.2	Mensuração da recuperação da informação.....	143
6	CONSIDERAÇÕES FINAIS.....	153
	REFERÊNCIAS.....	158
	APÊNDICE A – GLOSSÁRIO DE CONCEITOS.....	177
	APÊNDICE B – DICIONÁRIO DE CONCEITOS.....	179
	APÊNDICE C – TABELA DE CONCEITOS E VALORES.....	185
	APÊNDICE D – GLOSSÁRIO DAS RELAÇÕES.....	192
	ANEXO A – TEXTOS ANALISADOS PARA FORMAÇÃO DO GLOSSÁRIO DE VERBOS.....	194

1 INTRODUÇÃO

Assim como o surgimento da prensa de Gutenberg viabilizou a multiplicação dos documentos impressos, a expansão da Internet estimulou as publicações em meio digital. Nessa transição, do impresso para o digital, suscitada pelo avanço significativo das Tecnologias da Informação e Comunicação (TIC), foi evidenciada a necessidade de disponibilizar digitalmente os documentos que foram produzidos em suporte de papel, em décadas e séculos passados.

Essas iniciativas surgiram primeiramente em ambientes internos de bibliotecas, instituições e organizações. As informações eram armazenadas em bancos de dados privados, utilizados em princípio para suprir as carências informacionais de uma reduzida comunidade. Subsequentemente, com o intensivo uso da Internet foi evidenciada a necessidade da disponibilização desses bancos de dados na Internet, possibilitando assim que um maior número de pessoas tivesse acesso a essas informações.

No final do século XX e início deste século, jornais brasileiros que mantinham internamente seus bancos de dados, disponibilizaram seus acervos digitais na Internet. O jornal em ambiente digital oferece vantagens quando comparado à versão convencional. O acesso não é limitado por um único leitor por exemplar, conforme a versão em papel. Permite, dessa forma, que vários leitores possam fazer uso simultaneamente de um mesmo documento. A guarda e preservação são facilitadas pelos bancos de dados dos próprios jornais e a pesquisa é simplificada pelo auxílio de recursos que propiciam encontrar termos nos textos, por meio de uma navegação intuitiva, que permitem a busca por palavras-chave, em seções do jornal (esportivo, econômico, policial) e períodos de tempo (dia, mês, ano); (SANTOS; VIEIRA, 2015).

Essas transformações estão em consonância com a convergência midiática, termo cunhado por Henry Jenkins para designar a tendência de adaptação dos meios de comunicação à Internet. De acordo com Jenkins (2008, p. 41), “palavras impressas não eliminaram as palavras faladas. O cinema não eliminou o teatro” e, neste cenário, meios de comunicação são adaptados devido à necessidade de introdução de novas tecnologias.

No entanto, no processo de convergência de mídias, além do fluxo dos conteúdos serem alterados, o comportamento do usuário também é migratório, “a convergência envolve uma transformação tanto na forma de produzir quanto na forma de consumir os meios de comunicação” (JENKINS, 2008, p.44). Portanto, para acesso aos acervos

digitais de jornais, além de oferecer os conteúdos da antiga mídia, é preciso habilitar os recursos fundamentais praticados na nova mídia. Seus padrões de acesso, organização e, principalmente, recuperação da informação (RI) precisam ser adaptados.

Outra questão que se coloca do ponto de vista da RI em acervos digitais que reúnem documentos de diferentes épocas e culturas é a variação linguística que ocorre em função da forma diacrônica “ (dizer já dito em momentos diferentes) ” e sincrônica “ (dizer atual e simultâneo) ” (MOURA, 2009, p. 64), assim como aqueles referentes aos contextos de uso, marcados muitas vezes por aspectos geográficos e socioculturais.

Em ambientes digitais, geralmente, utiliza-se a linguagem natural como linguagem de busca. Por essa razão, explorar e direcionar as ferramentas automatizadas, para que os conceitos utilizados e interpretados na linguagem natural sejam atingidos, é determinante para um sistema de recuperação da informação (SRI). Nesse sentido, as ontologias são modelos de dados que podem apoiar a representação da informação e relacionar conceitos aos termos presentes nos acervos digitais, auxiliando a contornar as variações diacrônicas e sincrônicas.

O termo ontologia surgiu inicialmente na Filosofia, originado do estudo da metafísica. No entanto, a pesquisa sobre ontologias tem aparecido também com frequência na literatura das áreas de Ciência da Computação e Ciência da Informação. Na Ciência da Computação o termo passou a ser usado inicialmente nos anos 1960, em pesquisas relativas à Representação do Conhecimento, subárea da Inteligência Artificial e nos anos 1980 surgiram as primeiras pesquisas relacionadas à Engenharia de Software. Nos anos 1990 as publicações dos pesquisadores da Ciência da Informação demonstravam um crescente interesse sobre o assunto. Nessa visão as ontologias são identificadas como um tipo de Sistema de Organização do Conhecimento (SOC), com o propósito de organizar e representar o conhecimento. (SOERGEL, 1997; VICKERY, 1997; ALMEIDA, OLIVEIRA, COELHO, 2010; ALMEIDA, MENDONÇA, AGANETTE, 2013)

Na proposta da Web Semântica as ontologias são utilizadas como alternativa para agregar semântica aos recursos da web (BERNERS-LEE, HENDLER e LASSILA, 2001). De acordo com Breitman (2010, p.7) por meio de “modelos conceituais que capturam e explicitam o vocabulário utilizado nas aplicações semânticas” as ontologias podem contribuir para uma “comunicação livre de ambiguidade. ”. Ou seja, se a semântica for atribuída nos documentos digitais, alguns problemas comuns, como a polissemia e a sinonímia, podem ser evitados (SOUZA; ALVARENGA, 2004). Ao possibilitar que os termos sejam ‘compreendidos’ em seus

contextos de uso, é possível trazer significados próximos àqueles desejados pelos usuários e, dessa forma, proporcionar resultados mais satisfatórios nas buscas realizadas.

Ao direcionar as ontologias para um domínio, mais específica ela se torna, fazendo com que as ambiguidades daquele contexto sejam tratadas conforme as suas reais finalidades. O futebol, objeto de estudo deste trabalho, tem o jornal como uma importante fonte de informação e apresenta variações linguísticas pertinentes ao estudo proposto.

Assim, a consecução deste trabalho busca a elaboração de uma ontologia, aplicada ao domínio do futebol presente nos acervos digitais de jornais, a fim de apresentar um modelo viável, que consista na representação da linguagem natural, livre de ambiguidades e das questões que envolvem a variação linguística, como os aspectos diacrônicos e sincrônicos, característicos da mídia jornalística, com o objetivo de aprimorar a RI em acervos digitais de jornais. Para tal, é utilizado como base o Acervo digital Estadão¹, que disponibiliza edições do jornal Estado de S. Paulo. O recorte estabelecido para pesquisa contempla o surgimento dos primeiros clubes de futebol no Brasil em 1900, até novembro de 2015, período da coleta das informações do estudo.

1.1 Definição do problema

Geralmente, os sistemas de recuperação da informação representam seus documentos apenas com palavras-chave e, nesse sentido, a recuperação da informação (RI) pode ficar limitada, pois aspectos linguísticos, tais como sinonímia e homografia, podem interferir nos resultados de busca. Contudo, se as palavras-chave estiverem relacionadas com outras, com vínculos bem definidos, os significados das palavras serão entendidos conforme o contexto em que estão sujeitos. As ontologias podem criar essa ponte entre o conteúdo e o significado e fornecer vocabulário para descrever objetos e a forma como eles se relacionam. (SCHIESSL, 2015).

De acordo com Uschold e Jasper (1999, p.2), uma ontologia pode adotar vários formatos, mas, de maneira geral, elas contêm um vocabulário de termos e alguma especificação do seu significado. Ainda, segundo os mesmos autores, “deve abranger definições e uma indicação

¹ Acervo Estadão: Disponível em: <http://acervo.estadao.com.br/>. Acesso em: 21/01/2016.

de como os conceitos estão inter-relacionados, o que resulta na estruturação do domínio e nas restrições de possíveis interpretações de seus termos”.² Um exemplo que ilustra o papel da ontologia é o uso do termo “banheira”, que de acordo com o Dicionário da Língua Portuguesa - Aurélio³, pode tanto ser entendido como “tina para banho” e também como “posição irregular de um jogador quando, no momento do passe de um jogador da mesma equipe, se encontra no meio-de-campo adversário, à frente de todos os jogadores da outra equipe, à exceção de um”. Essa diferenciação é compreendida facilmente por humanos, mas máquinas não conseguem distinguir os termos semelhantes ou idênticos de acordo com o contexto que pertencem, se não tiverem o auxílio de recursos que permitam identificar esses significados.

No exemplo da frase “Paulo acertou uma bela bicicleta”, se a sentença tiver relação com o domínio do futebol, é possível concluir que a palavra “bicicleta” se refere à jogada em que o atleta chuta a bola com o corpo no ar. De acordo com Studer, Benjamins e Fensel (1998, p.184) “a ontologia pode ser compreendida como um modelo abstrato de algum fenômeno no mundo, esse fenômeno pode ser construído por meio da identificação de conceitos relevantes, geralmente um conhecimento do domínio”.

Ao orientar as ontologias ao cenário dos acervos digitais de jornais, ainda é necessário atentar ao cenário histórico que esses ambientes contemplam. Assim, as mudanças e variações linguísticas ocorridas durante a passagem das décadas precisam ser avaliadas, como no exemplo da palavra “Tropheo/ Tropheu”, que teve alteração na sua grafia e hoje se escreve Troféu, ou das palavras “escanteio”⁴ e “futebol”, que antigamente adotavam o termo em inglês “corner” e “foot-ball”.

² “An ontology may take a variety of forms, but necessarily it will include a vocabulary of terms, and some specification of their meaning. This includes definitions and an indication of how concepts are inter-related which collectively impose a structure on the domain and constrain the possible interpretations of terms.”(Uschold e Jasper, 1999, p.2)

³ Dicionário da Língua Portuguesa - Aurélio. Definição da palavra ‘banheira’. Disponível em <http://www.dicionariodoaurelio.com/banheira>. Acesso em: 23 de jan 2016.

⁴ No futebol, falta de um jogador que manda a bola para além da linha de fundo da sua equipe, "escanteio", in Dicionário Priberam da Língua Portuguesa [em linha], 2008-2013, <http://www.priberam.pt/DLPO/escanteio> [consultado em 07-12-2014].

Desse modo, é preciso criar relações para que as distintas formas de grafia sejam recuperadas, sem que seja preciso realizar inúmeras buscas no sistema.

O tratamento ontológico permite o relacionamento semântico. Portanto, a realização desse trabalho está norteada na modelagem de domínio para construção de ontologia, visando o aprimoramento da RI de acervos digitais de jornais. Essa modelagem do domínio levará em consideração, além da linguagem natural, as variações linguísticas vistas nos acervos digitais, provenientes de textos de períodos de tempo passado.

Assim sendo, a questão que motiva essa pesquisa é: Como a ontologia de domínio pode apoiar a recuperação da informação em acervos digitais de jornais em relação às variações diacrônicas e sincrônicas presentes no acervo e no domínio do futebol?

2 OBJETIVOS

Para responder à questão da pesquisa e orientar o desenvolvimento do trabalho foram definidos os seguintes objetivos:

2.1.1 Objetivo Geral

Modelar uma ontologia do domínio do futebol para o tratamento de variações diacrônicas e sincrônicas da língua, visando apoiar a recuperação da informação em acervos digitais de jornais.

2.1.2 Objetivos específicos

- a) Analisar o domínio do futebol no contexto dos acervos digitais de jornais impressos;
- b) Identificar os aspectos linguísticos do domínio do futebol, concernentes às variações sincrônicas e diacrônicas, que podem afetar a recuperação da informação nos acervos digitais de jornais impressos;
- c) Tratar os aspectos linguísticos identificados numa ontologia do domínio futebol;
- d) Demonstrar como a ontologia do futebol construída pode apoiar a recuperação da informação em acervos digitais de jornais no que diz respeito às implicações linguísticas tratadas.

2.2 Justificativa

De acordo com Dodebei (2011, p.6) “a passagem acelerada do patrimônio cultural para o território do ciberespaço” culminada pelos museus virtuais, bibliotecas digitais e, também pelos acervos digitais “implicou a mudança das mídias tradicionais para mídias digitais, o que resultou numa convergência que passa a ser a do objeto informacional. ”

Afirma ainda a autora, que o campo da informação enfrenta desafios que permeiam a passagem da memória para a modernidade e, ao lidarmos com o digital, o documento vivencia mudanças profundas e, suas características de texto, mídia e memória também são redefinidas.

Nessa transição dos objetos digitais, o patrimônio deixa ter o fundamento material, físico, impresso, assim, “o valor patrimonial

designado aos objetos indicados como memória parece ganhar supremacia ao valor documental, pelo caráter coletivo da herança social”. (DODEBEI, 2010, p.9). Dessa forma, diante do valor que esses objetos digitais representam, é preciso, além de disponibilizar os conteúdos de mídias impressas em formato digital, desenvolver padrões a fim de simplificar o acesso a essas informações.

À vista disso, os jornais são importantes fontes de informação, e em alguns casos, somente neles podemos encontrar determinadas informações, como no caso do domínio explorado, o futebol. Com o advento das TIC e, conseqüentemente, com a oferta dos acervos de jornais na Internet por meio de formatos digitais, é justificada a necessidade de direcionar os recursos de representação e RI na web, a fim de viabilizar que usuários possam obter essas informações de maneira mais precisa e simplificada.

As ontologias apresentam-se como um importante aliado na organização da informação e sua conseqüente recuperação. São utilizadas como a forma de atribuir significados mais exatos para determinadas especialidades, diminuindo assim, sistematicamente, o nível de ambigüidades. Para Moreiro González (2011, p.76) as ontologias permitem a “obtenção de respostas precisas a consultas concretas”, passando de um nível de recuperação da informação, para recuperação do conhecimento.

Assim, sabido que as ontologias não se limitam a estruturas pré-definidas de relações semânticas, mas permitem que essas relações sejam refinadas e incorporadas a sua estrutura de acordo com a necessidade de representação determinada pela própria dinâmica do conhecimento (Moreiro González, 2011), se aplicadas no contexto dos acervos digitais, também podem potencializar o acesso às informações.

O domínio a ser explorado na pesquisa é o do futebol. A escolha justifica-se devido à escassez de publicações em outras fontes que retornem informações detalhadas de jogos, jogadores, campeonatos, clubes e da história do futebol em geral. A construção de sistemas direcionados para domínios específicos, levando em consideração o vocabulário adotado nesses contextos, permite que a RI seja ampliada. Por esta razão e, também, pela importância que os jornais representam para a história do futebol, adotou-se o domínio para o desenvolvimento da pesquisa. Bonifácio e Heuser (2002) afirmam que a “conceitualização de um domínio específico de problema ou ontologias que forneçam um acordo comum de vocabulários para que os dados sejam referenciados é desejável”.

Como justificativa pessoal, envolve a possibilidade de dar continuidade a pesquisa iniciada no trabalho de conclusão do curso da graduação de Biblioteconomia, que avaliou o jornal impresso como uma das principais fontes de pesquisa para a reconstrução da história do futebol. Ainda no âmbito dos acervos digitais, a principal motivação está em explorar a estrutura de acervos digitais de jornais, fortalecer o SRI utilizado nessa fonte, otimizando o tempo de busca, aumentando o índice de satisfação nos resultados de pesquisas, a fim de fortalecer o potencial uso desses ambientes, para que mais pessoas se sintam atraídas em utilizá-los como fonte de pesquisa.

3 FUNDAMENTAÇÃO TEÓRICA

A revisão da literatura está estruturada em cinco grandes tópicos que abordam assuntos relacionados a: Patrimônio Digital; Jornal: do impresso ao digital; Informação, Futebol e Jornal; Recuperação da Informação e Busca Semântica

3.1 Patrimônio digital

Para compreensão inicial de patrimônio, que está diretamente ligado ao conceito de herança, precisamos definir os conceitos de memória, registro, testemunho e informação, que são componentes que interferem significativamente nas evoluções históricas, e são elementos que integram o Patrimônio digital.

O sentido original de memória refere-se à predisposição humana em guardar no cérebro as impressões das experiências vividas (LOPES, 1998). De acordo com Chapouthier (2006, p. 9) “memória é a capacidade que certos seres vivos têm de armazenar, no sistema nervoso, dados ou informações sobre o meio que os cerca, para assim modificar o próprio comportamento”.

No entanto, essa acepção de memória, em relação ao indivíduo, concebida de forma natural e particular, é imediata e vulnerável. Para Lévy (2000, p.177), essa afirmação pode ser entendida pelo trecho “[...] quando um ancião morre, é uma biblioteca que se queima”, pois com ele carrega todo as recordações armazenadas somente em sua memória.

Quando recorremos ao entendimento de “memória social” (LE GOFF, 1990), é possível estabelecer o raciocínio sobre a memória associada ao documento, que pode ser aplicada ao fundamento da memória sendo registrada, ou pautada na comunicação e na troca de informações (RUEDA; FREITAS; VALLS, 2011), que nesse caso, não necessariamente será registrada, mas permanecerá sendo disseminada devido aos testemunhos vivos que os perpetuam.

Para Le Goff (1990), a ação de perpetuação da história de uma sociedade pode ser compreendida como características de monumento e documento. O primeiro, o monumento, advém do entendimento que a preservação da memória, voluntária ou involuntária da sociedade, é um legado coletivo, em menor parcela de testemunhos escritos. E o segundo, o documento, ganha caráter de ‘prova’, sendo considerado o fundamento do fato histórico.

De fato, a memória é um conjunto de todos esses movimentos, da memória como testemunho e registro, e a informação será a extração desse conteúdo, que de maneira geral sobrevive devido àqueles que se dedicam a ciência e desenvolvem ações tendo como a motivação a preservação da história de uma sociedade para tempos futuros, constituindo os patrimônios históricos.

Para Dodebei (2006), patrimônio pode ser considerado como um atributo simbólico de um bem que se deseja preservar com indício de memória. Le Goff (1990) corrobora ao afirmar que desde a antiguidade, a ciência histórica reúne documentos escritos e faz deles testemunhos. Entretanto, com a transformação da sociedade, as formas de registramos as informações foram alteradas. Inicialmente tinha-se a forma oral, sucedidas pelas inscrições rupestres na pré-história, feitas geralmente em pedras, madeiras ou cavernas. Após ocorreu a revolução tecnológica em 1450, e o surgimento dos tipos móveis e a prensa de Gutenberg, aí veio a chamada “explosão documental” e atualmente ainda estamos em transição para o ciberespaço (RUEDA; FREITAS; VALLS, 2011).

Nessa transição, podemos considerar o jornal como um indicio de memória, tanto na forma testemunhada, por aqueles que compartilham suas impressões sobre assuntos pertinentes em uma determinada época, como na forma registrada, por meio da escrita, e em decorrência desses traços é evidenciada a importância do jornal para o fornecimento de informações sobre a história de uma sociedade.

Le Goff (2003, p. 25) afirma que “[...] o passado é uma construção e uma reinterpretação constante, e tem um futuro que é parte integrante e significativa da história”. Para tal, precisamos contemplar uma nova erudição, que responda de forma simultânea às exigências tecnológicas e a influência que elas detêm sobre a memória coletiva.

Assim, se as formas de registro da memória evoluíram, consequentemente, as formas de acessá-las também foram modificadas. Dessa maneira, ainda em passagem, os documentos que dispõem da memória registrada, que antes eram exclusivos de poucos pesquisadores, estão sendo migrados para ambientes digitais e assim sendo acessíveis a um número maior de pessoas. Esse cenário é possível com a introdução das TIC. Nesse contexto, fontes históricas continuam em desenvolvimento, acompanham as novas tecnologias e se tornam patrimônios digitais.

Por meio da Declaração Universal sobre a Diversidade Cultural, do Programa Memória do Mundo, a UNESCO (2003) lançou a Carta sobre a preservação do Patrimônio Digital. A carta manifesta a preocupação dos países sobre a salvaguarda da memória mundial

registrada em documentos digitais. Dentre as ações tratadas, estão iniciativas que devem ser consideradas que vão desde a criação destes documentos, até às políticas de acesso, as quais devem ser constantemente revistas.

No que diz respeito ao acesso, é preciso aplicar as práticas de indexação, arquivamento além da reprodução dos documentos em diferentes mídias. Porém, convergir o meio impresso para o ambiente digital, com a inclusão dos documentos “nascidos digitais”, a fim de tornar esses ambientes receptivos para usuários ubíquos, não é uma tarefa simples. Por esta razão, a intervenção contínua, com a atualização de tecnologias, vocabulários e métodos de organização da informação se faz necessária.

No contexto digital, o patrimônio, de acordo com Dodebei (2006), pode ser estudado sobre dois enfoques: o primeiro considera o patrimônio digital como um processo de constante produção, seja dos objetos digitalizados, ou dos objetos já que já nasceram digitais (*born-digital heritage*). O segundo refere-se ao patrimônio como produto de uma escolha e, nesse enfoque, um conjunto isolado de objetos é escolhido para representar, de forma simbólica, as ações sociais.

Os acervos digitais de jornais estão em conformidade com o primeiro enfoque, pois disponibilizam em seus sítios tanto edições de jornais que foram criados e distribuídos em papel, quanto edições que já nasceram digitalmente. Ambos os formatos se integram e formam uma linha do tempo de informações históricas que possuem data de início, assinalada pelos primeiros exemplares dos jornais, mas não possuem data final, pois a história continua e a reprodução dos jornais segue concomitante.

O segundo enfoque também é colocado em evidência e se relaciona com a salvaguarda da integridade dos acervos. No sentido digital, além da preservação, as formas de representação que garantem o acesso também devem ser contempladas.

Para tal, Dodebei (2011) enumera outras cinco questões que devem ser levadas em consideração na passagem da sociedade analógica para a digital:

1. salvar ainda a produção intelectual armazenada nas memórias documentárias
2. fazer os patrimônios digitais conversarem entre si;
3. considerar a leitura hipertextual dos recursos organizados em páginas, sites e portais na web;

4. indexar e recuperar os recursos informacionais com auxílio de programas lógicos e manter arquivos reprodutíveis;
5. proteger ou disseminar (como forma de preservação dinâmica) as memórias que circulam na web.

O cenário abordado neste trabalho está diretamente ligado à quarta questão, voltada à construção de modelos que incluam a lógica na sua representação, com o objetivo da recuperação da informação. Para Arellano (2004, p.18) “o dado e a mídia que suportam a informação devem possuir um nível de funcionalidade representacional que permita a sua reprodução a qualquer momento que a instituição mantenedora precisar recuperar o dado.”

No que se refere aos acervos digitais, de acordo com Dodebei (2006, p.9), a virtualização e a digitalização são demandas práticas e cotidianas no processo de transferência da massa documental física para ambientes digitais. Cavalcante (2007) afirma que disponibilizar uma parte do patrimônio histórico-cultural digitalmente, é uma forma de coletivizar acervos que antes eram destinados apenas a preservação, e não ao acesso, devido as fragilidades provocadas pela degradação física provocadas por séculos de existência.

Contudo, os novos suportes carecem de intervenção para o desenvolvimento de mecanismos que propiciem um acesso facilitado a essas mídias e, nesse sentido, explorar medidas que visem o uso contínuo de informações, para que elas não se tornem inacessíveis, também é fundamental.

Os acervos digitais se constituem de documentos históricos, por isso apresentam particularidades em seu vocabulário, que devem ser levadas em consideração quando organizados. Quando se trata de um contexto dinâmico é preciso aplicar métodos de representação que permitam contemplar essas particularidades, como as variações linguísticas com aspectos diacrônicos e sincrônicos, a fim de que essas informações sejam relacionadas e oportunizem uma recuperação da informação mais integrada, unindo termos que possuem o mesmo significado, mas formas de grafia diferenciada, devido às mudanças de ortografia ou aos estrangeirismos encontrados nos acervos de jornais.

3.2 Patrimônio: O jornal do impresso ao digital

O jornal, após o livro, é o veículo impresso mais longo e, assim como o livro, sobreviveu a todas as mudanças tecnológicas e sociais transcorridas na história. Conforme Dines (2001, p.26) o jornal resistiu

ao “obscurantismo político, à revolução industrial, às invenções para vencer distâncias, às guerras, às migrações, à recensão econômica, à TV” e todas essas ocorrências foram absorvidas e estão registradas nas páginas dos jornais diários.

De acordo com informações retiradas da cronologia da Associação Nacional dos Jornais (ANJ, 2008)⁵, a história do jornal no mundo inicia-se em 59 a.C., e sua invenção está atribuída ao Imperador Romano, Júlio César. Na ocasião, o Imperador ordenou que os principais acontecimentos políticos do dia fossem publicados e divulgados para seu império em um documento chamado de Acta Diurna.

Na Idade Média, o jornal teve o seu grande impacto tecnológico viabilizado pela prensa de Gutenberg, no ano de 1447, o que acarretou numa maior disseminação e audiência. A partir desse momento, surge a possibilidade da produção em massa, o que determinou maior acesso pela população, de forma mais rápida e ampla.

Após o surgimento da prensa de Gutenberg, importantes países tiveram a publicação dos seus primeiros jornais. No Brasil, o impresso nasceu no início do século XIX (1808) com finalidades políticas, diferentemente dos interesses de fortalecimento da classe mercantil, demonstrados no cenário Europeu, no início do século XVII, e nas Américas espanhola e inglesa, ao final do mesmo século (CALDAS, 2004, p.42).

No entanto, no ano de 1844, outra importante contribuição para o jornal e para a comunicação foi inventada, o telégrafo. O aparelho revolucionou a forma de se comunicar à distância, viabilizando a transmissão de mensagens de um ponto a outro. Consequentemente, a imprensa teve uma maior abrangência e a distribuição de notícias à sociedade foi facilitada.

Os jornais brasileiros dispunham, além do frescor dos acontecimentos diários, o espírito crítico da corrente literária. Famosos autores como Machado de Assis, Inglês de Souza, Aluísio de Azevedo, Júlio Ribeiro e Raul Pompéia descreveram suas visões sobre os aspectos sócio econômicos, culturais e políticos vivenciados na época (FERRARI, 2006, p.26). Tal era a importância atribuída ao jornal, que em um artigo de Machado de Assis, publicado originalmente no Correio Mercantil, no ano de 1859, repercutia as incertezas do autor em relação a sobrevivência

⁵ Linha do tempo com a cronologia da história do jornal no Mundo. Disponível em: <http://www.anj.org.br/cronologia>. Acesso em: 22 maio 2014.

do livro perante ao triunfo do jornal - “O jornal matará o livro?” - foi uma das perguntas levantadas pelo autor. O artigo também colocava em questão a revolução contida na imprensa e que o jornal podia ser considerado um “banco intelectual”:

O jornal, operando uma lenta revolução no globo, desenvolve esta indústria monetária, que é a confiança, a riqueza e os melhoramentos. O crédito tem também a sua parte no jornalismo, onde se discutem todas as questões, todos os problemas da época, debaixo da ação da idéia sempre nova, sempre palpitante. O desenvolvimento do crédito quer o desenvolvimento do jornalismo, porque o jornalismo não é senão um grande banco intelectual, grande monetização da idéia, como diz um escritor moderno (ASSIS, 1859).

Ambos, livro e jornal, suportaram as transformações ocorridas na evolução da sociedade, porém, necessitaram acompanhar as mudanças, principalmente as advindas da tecnologia. Em 1880 as primeiras fotografias foram associadas as páginas dos jornais, até então a mídia era somente constituída de textos.

No início do século XX, o jornal ainda era o principal veículo de comunicação da sociedade, fato que culminou o título de anos dourados à mídia (ANJ, 2008), mas, nos anos 1920, surge o rádio, e os jornais tiveram que se remodelar para compreender o seu real papel na sociedade.

Com a popularização do rádio, ocasionada pela sua facilidade de consumo, comparada ao jornal, que precisava de uma atenção maior do leitor, foi repercutido que os editoriais da mídia impressa seriam extintos. No entanto, os editores se adaptaram às novas tecnologias e apresentaram um produto com mais textos e imagens, a fim de inovar os formatos já existentes para continuar atraindo o público (ANJ, 2008).⁶

Logo após o surgimento do rádio, ainda nos anos 1920, surge a televisão, potente mídia de massa, que além de contar com o sonoro do rádio, ainda emitia imagens em ação, o que configurava um aparelho de forte entretenimento para sociedade. Neste período, o jornal tinha se readaptado ao surgimento do rádio, porém, a eclosão da televisão fez com que novas estratégias fossem adotadas para que a mídia impressa não se

⁶ Jornais: Breve História. Disponível em: <http://www.anj.org.br/jornais-breve-historia/page-2?showall=>. Acesso em: 08 jun. 2014

tornasse obsoleta. A adoção de textos curtos, coloridos e diretos, foram aplicados como alternativa para a concorrência com a televisão.

Na década de 1960 surgiu a Internet. Uma rede de computadores chamada primeiramente de Arpanet, desenvolvida pelo Departamento de defesa dos Estados Unidos para fins militares. (CASTELLS, 2003, p.13). Posteriormente, no final da década de 1980 e início da década de 1990, surge a World Wide Web, desenvolvida por Tim Berners- Lee. Com o surgimento do “www” a Internet se difundiu globalmente de forma mais acessível e organizada, o que permitiu ser considerada também como um sistema de comunicação. (CASTELLS, 2003, p.8;18)

A Internet, também intitulada “rede” passou por uma imensa transformação entre os anos 60 e 90 do século passado (VAZ, 2001), no entanto, o “boom” da Internet ocorreu nos anos 1990, com o progresso da Web.

Com a eclosão da Internet, os jornais precisaram adentrar nesse contexto digital, assumindo assim, também, outras características, mas não perdendo a sua identidade social e cultural, de informar. De acordo com Tellaroli e Albino (2007, p.13), a cada transformação tecnológica, a produção na imprensa precisava se transformar também, para seguir a evolução que a sociedade exigia:

Nota-se que a imprensa transforma-se a cada inovação tecnológica; muda a maneira de produção de conteúdo e altera a forma de recepção. Exemplos: o veículo impresso caracteriza-se por exigir do público o conhecimento da escrita; o rádio atinge um número maior de pessoas por usar apenas o som (audição); a televisão atrai um número também elevado de pessoas pelo apelo visual e auditivo; já o veículo on-line agrega em si todas as características dos outros veículos e gera uma mudança de paradigma da tecnologia da informação.

Com a introdução dos jornais na web, um dos seus traços principais é perdido - o papel. No ambiente digital é preciso salientar a diferença de duas perspectivas do jornal: O *on-line* e o digital. A dimensão *on-line*, de acordo com Palacios (2003) apresenta cinco características: Multimídia/ convergência, Interatividade, Hipertextualidade, Personalização e Memória. Nessa perspectiva, ainda é possível incluir a característica da instantaneidade.

A multimídia e convergência estão relacionadas à conversão de imagem, som, texto das notícias jornalísticas para diferentes sistemas. A interatividade está voltada à possibilidade de o leitor participar de forma colaborativa no processo de criação das notícias ou na inclusão de críticas em notícias já publicadas. A hipertextualidade traz ao leitor a possibilidade de percorrer o texto de uma forma não linear, viabilizando, por meio dos hiperlinks, que o leitor siga caminhos diferentes e chegue a informações distintas. A personalização contempla as necessidades de cada leitor, oportunizando assim que os jornais se adaptem, de acordo com as individualidades de cada leitor. A memória surge com a compreensão de que tudo o que for criado digitalmente, será preservado para acesso a longo prazo e será possível sua recuperação. Por fim, a instantaneidade permite a atualização dos conteúdos disponibilizados na web, conforme o ritmo dos acontecimentos, isso possibilita o acompanhamento contínuo dos assuntos de maior interesse pelos leitores.

Na dimensão digital, termo que adotamos aqui para discorrer sobre o jornal disponibilizado nos acervos digitais, as características são as mesmas da versão impressa, com publicação de nova edição feita a cada 24 horas, sendo essa, disponível primeiramente apenas para assinantes. Após um período estipulado pela gestão dos acervos, geralmente 30 dias, a edição do jornal é disponibilizada no banco de dados composto pelas versões mais antigas.

Os acervos digitais de jornais fundamentados na dimensão digital, são compostos tanto por jornais digitalizados e jornais que já nasceram digital, os chamados “patrimônio nascido digital” (*born-digital heritage*), abordados nessa pesquisa como “nascidos digitais”, juntos os dois formam os objetos de estudo desta pesquisa. As características abordadas na dimensão *on-line* não farão parte das pesquisas desse estudo, exceto a característica referente à memória, que também é aplicada à dimensão digital.

Um dos traços mais importantes que possibilitaram o surgimento dos acervos digitais de jornais, foi a digitalização. De acordo com Schiwingel e Zanotti (2010, p. 86) a prática da digitalização no jornalismo ocasionou uma transformação profunda:

A referência à digitalização como fenômeno específico da expressão tecnológica sobre o jornalismo procura demarcar o início de uma transformação profunda na atividade jornalística em que os registros captados ou gerados na organização jornalística são configurados em um

formato digital, o que garante uma conversibilidade e trânsito dos conteúdos em diferentes plataformas computacionais.

A possibilidade de transferir tudo o que foi criado em papel, anteriormente à propagação da Internet, para as telas de computador foi uma conquista para os acervos digitais. No caso dos jornais, foi possível deixar acessível séculos de história. Para Montenegro e Silva (2005), um dos benefícios que a evolução das TIC permitiu foi a possibilidade de utilização de informações e imagens dos documentos originais por meio da digitalização, além da facilidade de armazenar, recuperar e distribuir em meio digital essas informações, sem que os documentos originais estejam em constante manuseio, procedimento que prolonga a sua existência e auxilia a preservação.

Os acervos digitais são compostos em sua maioria por documentos digitalizados, pois a produção de jornais em ambiente digital teve início em meados da década de 90 do século XX, e os jornais já vinham sendo publicados no Brasil desde o século XIX. Com a junção dos documentos nascidos digitais, os acervos de jornais compõem uma base de dados de contínuo crescimento, fortalecendo a importância dessa fonte para pesquisa histórica.

Dessa forma, é importante assegurar que os procedimentos e recursos de organização da informação levem em consideração as características intrínsecas dos acervos digitais, permitindo a relação semântica de forma mais abrangente que a organização tradicional, por palavras-chave. Quando essa organização levanta conceitos relevantes para determinado domínio, pode contribuir para que a RI seja mais precisa.

O domínio do futebol, explorado nessa pesquisa, apresenta muitas das particularidades expostas nos acervos digitais, como as mudanças ortográficas, e também traz consigo as características derivadas do seu vocabulário, com aspectos da sinonímia, polissemia, homografia, palavras derivadas do inglês que foram modificadas para o português no passar dos anos, expressões que necessitam de inferência para serem compreendidas, além dos neologismos, que são responsáveis por atribuir novos sentidos a palavras que já existem na língua. Portanto, o domínio do futebol apresenta um cenário diferente nos acervos digitais, do que quando exposto em outros SRI, o que evidencia a necessidade de representação de diferentes tipos de relações entre os conceitos do domínio, por vezes não contempladas em vocabulários controlados.

3.2.1 Informação, Futebol e Jornal

Neste tópico não abordaremos como se constrói a informação esportiva, quais são os seus usuários e fontes de coleta. Mas sim, como se constituem as informações derivadas do futebol, contextualizadas nas notícias esportivas provenientes de jornais impressos, a fim de realçar a importância dessas informações para o campo de pesquisa.

O Brasil é conhecido popularmente como o “País do futebol”. Seus consumidores, chamados de torcedores, acompanham diariamente o esporte, discutem, comentam, lotam estádios e tornam-se clientes de suas marcas.

De acordo com Guterman (2009) o futebol foi introduzido no Brasil por Charles Miller há mais de 120 anos, um esporte de origem inglesa, praticado inicialmente por homens da elite e brancos. Os primeiros clubes a surgirem no país, praticavam um esporte amador, sem regras bem definidas, estilista, excludente e racista (JORGE, 2013).

Conforme a cronologia de Máximo (1999, p.2), o período que contempla o nascimento dos primeiros clubes no Brasil, vai de 1896 a 1905:

Em 1896, o São Paulo Athletic Club, fundado oito anos antes, seria o primeiro a aderir ao novo esporte, logo seguido do Sport Club Germania (1889), de Mackenzie Athletic Association (1898), Sport Club Internacional (1898), Clube Atlético Paulistano (1900), já com nome aportuguesado. Em Campinas, fundou-se a Associação Atlética Ponte Preta (1900). No Rio de Cox, o Fluminense Foot-ball club (1902), o Rio Foot-ball Club (1902), o Botafogo Foot-ball Club, o América Foot-ball Club, o Bangu Athletic Club (os três últimos em 1904). Flamengo e Vasco da Gama já existiam desde o fim do século, ambos dedicando-se ao remo: o primeiro, só criaria seu departamento de futebol em 1911; o segundo, em 1923. Em Porto Alegre, foi fundado o Esporte Clube Rio Grande (1900); em Minas, o Sport Club Belo Horizonte (1904); em Recife, o Club Náutico Capeberibe (1901); em Salvador, o Vitória Foot-ball Club (1905).

Os clubes mais antigos do país, ainda em atividade, são o Sport Club Rio Grande, do Rio Grande do Sul e a Associação Atlética Ponte Preta, ambos fundados em 1900.

Em 1933, o futebol brasileiro passou a ser profissional. A regulamentação foi assinada pela legislação social e trabalhista do Governo de Getúlio Vargas, e assim foi criada a profissão de jogador de futebol (CARRAVETTA, 2006).

No ano de 1950, o Brasil sediou a sua primeira Copa do Mundo, tendo como campeã a seleção do Uruguai. A partida que deu o título à seleção do Uruguai, ficou conhecida como “Maracanazo” devido ao triunfo uruguaio ter acontecido em solo brasileiro, no estádio do Maracanã, na cidade do Rio de Janeiro, diante de 200 mil torcedores que torciam a favor da seleção brasileira e que ficaram inconsoláveis com a derrota. A capa do Jornal o Globo, de 17 julho de 1950 (Figura 1), avaliou o fracasso com uma “decepção amarga” em uma “tarde infeliz dos jogadores patricios”.

Figura 1-Capa do Jornal O Globo de 17 de julho de 1950



Fonte: Acervo jornal O Globo

Com a evolução do futebol no país, o tema ganhou mais destaque nas páginas de jornais. Um exemplo da importância do uso do jornal para o domínio do futebol é a reconstrução da história do jogador Heleno de Freitas, ídolo do clube Botafogo, do Rio de Janeiro, na década de 1940. A história foi mostrada em uma cinebiografia intitulada Heleno, de 2012. Os produtores do filme contaram com poucas fontes de coleta para a pesquisa e utilizaram principalmente relatos jornalísticos da época e fotografias de jornais (BARROS, 2012). Um dos fatos dramatizados no longa metragem, de acordo com informações retiradas do Acervo O Estado de S. Paulo e O Globo (2015), foi a transferência do jogador para o futebol argentino, considerada a maior transação do futebol brasileiro naquela época. Apesar da recepção dos argentinos ter sido grande, como mostrado no título de uma

matéria do jornal O Globo, publicada no dia 07 de junho de 1948 (Figura 2), a estadia de Heleno no país foi curta, pois retorna em abril de 1949, em uma negociação com o time Vasco da Gama, do Rio de Janeiro.

Figura 2 - Capa jornal O Globo de 07 de junho de 1948



Fonte: Adaptado acervo O Globo

No título e na chamada da notícia ainda é perceptível a grafia da palavra “crack”, que atualmente se escreve “craque”. As mudanças de ortografia estão presentes nos acervos digitais de jornais, são importantes e interferem diretamente na recuperação da informação, portanto a representação da informação nesses sistemas, não deve desconsiderar essa característica fundamental.

É notória a expressividade que representa o esporte ao cenário brasileiro. De acordo com Guterman (2009, p.9) o futebol é uma “construção histórica, gerado como parte indissociável dos desdobramentos da vida política e econômica do Brasil. O futebol, se lido corretamente, consegue explicar o Brasil”

Hoje, diferentemente de como ocorreu na sua origem no país, o futebol contempla todos os públicos e engloba diferentes faixas etárias, gêneros e condição social, se tornando uma hegemonia. É sabido que o esporte é movido pela paixão, mas também pode ser compreendido como um negócio, nesse sentido, suas informações são amplamente pesquisadas.

É evidente a riqueza de informações sobre o futebol passíveis de localização nos jornais. No surgimento dos primeiros clubes de futebol no Brasil, ao final do século XIX e até meados da década de 20 do século XX, o jornal ainda era a principal mídia de comunicação e ainda permaneceu sendo predominante como fonte de informação textual até a intensificação do uso da Internet. Com a chegada da Internet, as fontes de informação são ampliadas, reduzindo o público do jornalismo impresso. No entanto, conforme Caldas (2004, p.14), a sua relevância consiste de não ter a pretensão de inventar nem revolucionar nada, e que “leitores de jornal e usuários da Internet tem interesses e curiosidades diferentes”, os leitores

dos jornais buscam textos com profundidade, originalidade, interpretativos e analíticos, com suas implicações e possíveis repercussões, posicionados dentro de um contexto mais amplo, baseados em pesquisa e opinião, distintos da instantaneidade e objetividade do jornalismo on-line.

Considerando esse cenário e também sustentando que os acervos digitais de jornais são fontes ininterruptas, que acompanharam o futebol desde o seu início no Brasil, o caderno esportivo, parte integrante da mídia jornalística, se configura como importante diário dos acontecimentos do esporte. Por meio das informações retiradas do jornal é possível recontar a história dos clubes, ou recriar biografias dos ídolos do esporte. Valentim (2006) evidencia a carência de informações voltadas diretamente ao esporte, que contribuam para auxiliar as organizações e os indivíduos do âmbito esportivo a desempenhar suas atividades.

Jorge e Valentim (2015) apontam a importância da organização esportiva. Segundo os autores, a sistemática deve considerar todos os tipos de informação que estão inseridos nesse cenário, independente do suporte e fluxo (formal ou informal). Com tais características, as informações publicadas nos jornais, contam com diferentes olhares, editadas em épocas, estilos e por jornalistas diferentes, com a contribuição de torcedores, colaboradores, jogadores, especialistas, entre outros.

Com advento das TIC e com a oferta dos jornais na internet por meio dos acervos digitais, é justificada a necessidade de direcionar as tecnologias da web para representação das informações, possibilitando assim que os documentos contidos nessas bases sejam encontrados.

Belluzo (2014, p.29) expõe a necessidade atual de informação no esporte, bastante relacionada com as TIC. Para isso, destaca três iniciativas que devem ser investigadas:

- a) Existe a necessidade de acompanhar os avanços tecnológicos e absorver as suas potencialidades;
- b) É importante que se agregue valor a estes conhecimentos e que sejam desenvolvidas novas metodologias para acessar e tornar acessível e compreensível a massa de informação disponibilizada em redes;
- c) Capacidade de migrar serviços convencionais ao novo meio e gerar novos serviços e produtos de informação, bibliotecas digitais e outros meios e suportes que possibilitem a disseminação da informação e a geração do conhecimento.

O jornal, importante fonte de informação para o futebol, acompanha os avanços tecnológicos salientado pela autora Belluzo, tendo migrado de um suporte em papel, para sistemas informatizados, alimentando assim os acervos digitais. No entanto, conforme a autora, além de tornar disponível essa massa documental, é preciso criar metodologias que assegurem o seu acesso. Para Moreiro González (2011), os sistemas de organização e representação utilizados na atualidade, sugerem da necessidade de estruturar a web, e essas linguagens foram adaptadas de linguagens documentais pré-existentes. Mas, para alcançar um nível de estrutura superior, que considere as transformações que a web propiciou e a precisão que esses vocabulários carecem, é preciso aliar às linguagens pré-existentes, como tesouros, com a introdução de ontologias. Para Moreiro González (2011, p.34):

A nova web, na sua caracterização semântica, requer linguagens de representação do conhecimento baseadas em ontologias para localizar e acessar recursos. Linguagens multicamadas que interajam entre modelos e reutilizem recursos com novos modelos de processamento para estabelecer relações de confiança e provas de autenticidade, para garantir aos usuários a disponibilidade de recursos e tornar seu funcionamento transparente.

Para garantir esse funcionamento transparente em que Moreiro González se refere, os sistemas que são baseados em correntes de signos – aqui entendidos como uma sequência de letras – precisam fazer a recuperação de conceitos e os seus contextos de uso.

3.3 Recuperação da Informação

O termo Recuperação da Informação (RI) foi cunhado por Calvin Mooers, em 1951. Conforme o autor, a RI surgiu para tratar dos “aspectos intelectuais da descrição da informação e sua especificação para busca, e também de qualquer sistema, técnicas ou máquinas que são empregadas para realizar esta operação” (MOOERS, 1951).

Conforme Manning e Schütze (1999, p.1), a RI consiste em “encontrar material (usualmente documentos) de natureza não estruturada

(usualmente textos) que satisfaz uma necessidade de informação dentro de grandes coleções (usualmente armazenadas em computadores) ”.

Baeza-Yates e Ribeiro-Neto (1999) apontam que a RI dedica-se aos recursos que englobam representação, armazenamento, organização e acesso aos itens de informação. Para Lopes (2002, p.41) a estratégia de busca, parte integrante da RI “precisa refletir a necessidade de informação do usuário” e “pode ser definida como uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados”

Com a explosão informacional, as discussões em torno da RI se tornaram mais relevantes e constantes. Entretanto, as questões evidenciadas por Mooers (1951) relacionadas a como descrever intelectualmente e especificar a busca da informação, além da investigação de métodos, sistemas, técnicas e máquinas necessárias para tornar mais satisfatória e simples a localização de informações, ainda são propósitos pertinentes e que envolvem esforços de pesquisadores que visam o aprimoramento de Sistemas da Recuperação da Informação (SRI).

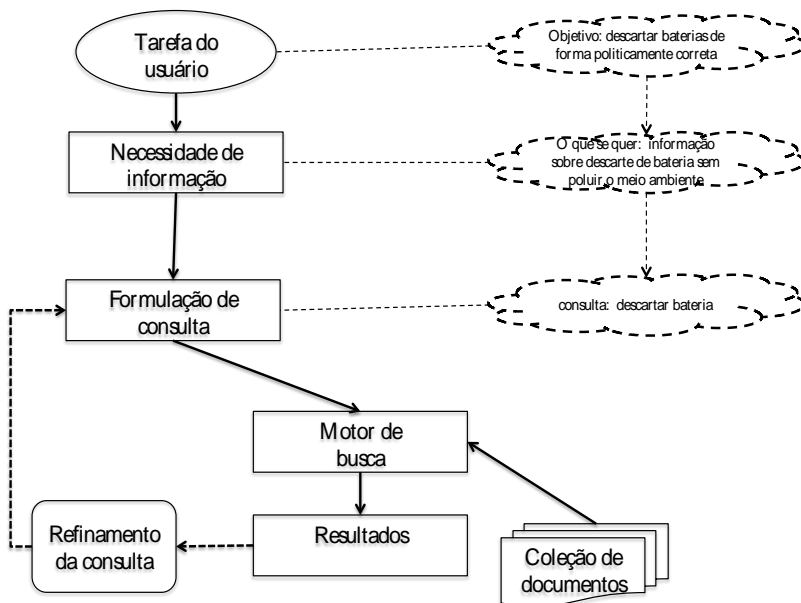
Na visão de Saracevic (1996, p.45), os estudos sobre RI devem ter como foco os seguintes aspectos:

[...] o comportamento humano frente à informação; a interação homem-computador, com ênfase no lado humano da equação; relevância, utilidade, obsolescência e outros atributos do uso da informação juntamente com medidas e métodos de avaliação dos sistemas de recuperação da informação; economia, impacto e valor da informação, dentre outros.

Ambas as problemáticas, evidenciadas por Mooers e Saracevic, visam tornar mais próxima a informação do usuário. No entanto, na perspectiva de Mooers, essa interação se estabelecia na relação homem-informação, já para Saracevic essa relação passa a envolver a interação homem-computador-informação. O modelo clássico de busca da informação (Figura 3) intermediada pelo computador inclui o raciocínio lógico do usuário, evidenciado na estratégia de busca, assim como a forma como os conteúdos estão organizados na base de dados, isso inclui o

tratamento da informação. O modelo contempla ainda a lógica computacional, dedicada a possibilitar que a necessidade de informação do usuário, armazenada nas bases de dados, seja solucionada.

Figura 3 - Modelo Clássico de Busca



Fonte: Schiessl (2015, p.120)

O usuário tem papel determinante nos modelos de buscas atuais. A maneira como é formulada a consulta pelo usuário determina quais serão os resultados obtidos e a coleção de documentos recuperados. Conforme Vieira e Garrido (2011), para que a RI aconteça efetivamente, é fundamental, além da visão lógica dos documentos adotados pelo SRI, a estratégia de busca adotada pelos usuários. Belkin (1996) do mesmo modo, compartilha desta afirmação, ao apontar que a inteligência em um SRI não está aplicada somente ao que é construído para dar suporte ao usuário, mas também no seu usuário.

Métodos com o intuito de aproximar o usuário da informação em SRI, utilizando técnicas baseadas no aprendizado de máquina, despontam como possibilidades que contribuem para minimizar os esforços do usuário. De acordo Monard e Baranaukas (2003, p.39), o Aprendizado de Máquina “é uma área de Inteligência Artificial cujo objetivo é o

desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática”. Assim sendo, a aplicação desse conceito está associada ao desempenho de alguma tarefa e, pode envolver a prática da linguagem natural, o aprimoramento de motores de busca e os processos da RI.

Os agentes inteligentes também empregam técnicas da inteligência artificial. De acordo com Souza e Alvarenga (2004, p.132) essas ferramentas possibilitam um estilo complementar de integração, chamado de gerência indireta, no qual o computador se torna uma atividade ativa, responsável por auxiliar os usuários no desempenho de suas atividades, conforme os seus interesses. Para os autores, o uso de agentes, no âmbito da recuperação e uso dos documentos, está relacionado com os mecanismos de busca e inferência (*Inference Engine*), que executarão o *harvesting* (colheita) de informações nos documentos.

No entanto, de acordo com Bräscher (1999) a RI vai além da verificação de uma sequência de caracteres, representados nos documentos buscados pelos usuários e também, as “ferramentas de busca que utilizam palavras como pontos de acesso ao conteúdo têm se mostrado ineficientes” (BRÄSCHER, 2002). Ainda assim, concentrar esforços apenas em algoritmos computacionais, não é a solução mais eficaz para a RI.

Para Corrêa e Bertocchi (2012) falta um olhar mais sensível. Falta explicar à máquina, que um mesmo termo pode ter inúmeros significados. Dessa forma, é preciso apontar quais são os sentidos existentes e em quais contextos se aplicam. De acordo com Bräscher (2002), existem alguns tipos de ambiguidades que podem ser sanadas automaticamente, para isso é preciso aplicar conhecimentos, passíveis de modelização, para desambiguá-las.

A junção de sistemas que suportem a comunicação entre modelos de descrição da informação, como ontologias, sustentadas em linguagens entendíveis pelos computadores, é fundamental para criar redes mais inteligentes. Para Moreira González (2011, p.76), “as ontologias supõem um salto qualitativo: do uso de motores de busca (com palavras-chave), passa-se a usar agentes de software inteligentes (com conceito).”

3.4 Busca Semântica

As problemáticas da CI estão muito relacionadas ao seu âmbito gerencial, que envolve a sua disseminação, organização e armazenamento, e também ao emprego das tecnologias que resolvam

estas questões. A semântica explora o significado das palavras e, suas implementações na web buscam solucionar as carências que envolvem a organização e recuperação da informação.

De acordo com Gonzalez e Lima (2003), “enquanto a sintaxe corresponde ao estudo de como as palavras agrupam-se para formar estruturas em nível de sentença”, a semântica abrange além do significado de cada palavra, também o conjunto resultante delas. Manning e Schütze (1999) dividem este processo em semântica lexical – responsável por investigar o significado individual das palavras – e em como essas palavras combinadas com outras formam o significado de construções frasais ou sentenças.

Para Kobashi e Tálamo (2003, p.10) a Teoria da Comunicação se aproxima muito da CI, pois ambas operam com o sentido. A semântica procura considerar o sentido que as palavras representam em determinado contexto. De acordo com Souza (2006, p.171) os SRIs geralmente utilizam palavras isoladas para indexar os documentos, mas, embora para a recuperação da informação esta aplicação seja bastante útil, falha em grande parte por não considerar o “contexto informacional implícito em toda consulta porque não estão preparados para lidar com a forma com que estas palavras ou conceitos estão relacionados”. A definição desses relacionamentos, quando colocada em prática, determina as especificidades de cada documento, possibilitando assim informações mais precisas e, a recuperação de um maior número de documentos relevantes sobre o assunto pesquisado.

Para Gonzalez e Lima (2003), um dos maiores desafios da semântica é o processamento de linguagem natural (PLN), pois envolve de um lado a morfologia e a estrutura sintática e de outro, em alguns casos, a introdução de informações da pragmática.

Fogl (1979, apud BRÄSCHER; CAFÉ, 2008, p.4) analisa a informação de acordo com seu aspecto semântico (cognitivo) e pragmático (real). Para o autor o aspecto semântico está relacionado ao “conteúdo do conhecimento e os juízos de valor fixados na informação, sem relação com as necessidades e interesses do sujeito, que avalia a informação em termos de sua veracidade, confiabilidade, conhecimento, adequação dos juízos de valor e assim por diante”, e ainda acrescenta que a informação no seu aspecto semântico deve ser vista como “algo subjetivo, relativo e dependente da interpretação do receptor”.

No aspecto pragmático, Fogl (1979, apud BRÄSCHER; CAFÉ, 2008, p.4) relaciona o conceito de informação com “a utilidade dos itens de conhecimento e dos juízos de valor registrados na informação para o

sujeito que avalia a informação”. Portanto, nesse aspecto a informação depende do significado dado pelo receptor.

Para Morris (1964 apud SIQUEIRA, 2008, p.92) a informação semântica é “aquela que possui significado contextual para um sujeito” e a informação pragmática é “aquela que está codificada e preparada para uso. É uma informação manipulada por um sujeito com fins de utilidade planejada.”

Podemos, portanto, compreender que a informação no seu sentido semântico, levando em consideração os conceitos aplicados por Fogl (1979, apud BRÄSCHER; CAFÉ, 2008, p.4) e Morris (1964), pode ser interpretada de diferentes formas, modificando-se conforme o usuário que a consome, nesse sentido, a informação aqui está em constante transformação.

Já no sentido pragmático, analisando os conceitos dos mesmos autores, a informação depende do significado que é introduzido pelo receptor e este pode ser modificado conforme os interesses de quem faz o seu uso.

Robredo (2003, p.111) pontua a capacidade de a informação pragmática oportunizar um “intercâmbio informacional (novos canais), ou de gerar uma nova informação pragmática”. Isso acontece devido a possibilidade de manipulação do significado da informação pragmática apontada por Morris (1964).

De acordo com Baranow (1983, p.28) o enfoque pragmático, possibilita também que um problema recorrente na semântica e sintática, a ambiguidade, seja minimizado quando colocado em questão a recuperação da informação. Afirma ainda o autor que muitas das preocupações linguísticas com o aspecto pragmático estão ligadas com a comunicação em linguagem natural e que esta “depende de condições que vão além da linguagem”, necessitando assim também o conhecimento do conteúdo dessa informação e a sua conexão com enunciados anteriores. O foco da pragmática está voltado para o uso e não para o significado.

Guiraud (1975) ressalta três principais ordens de problemas em nível semântico:

a) a ordem dos problemas psicológicos, que relaciona os estados fisiológicos e psíquicos dos interlocutores nos processos de comunicação de signos;

b) a ordem dos problemas lógicos, que estabelece as relações dos signos com a realidade no processo de significação;

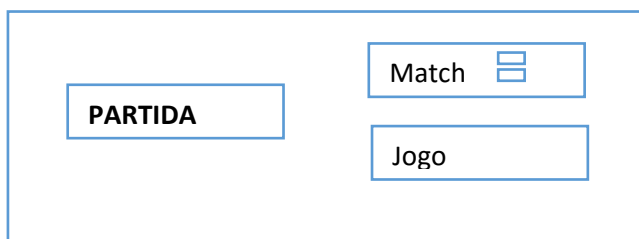
c) a ordem dos problemas linguísticos, que estabelece a natureza e as funções dos vários sistemas de signos.

Os desdobramentos da web projetam a solução contínua dos problemas que envolvem a segunda e a terceira ordem, relacionadas aos problemas lógicos de significação, o contexto em que estão inseridos e os problemas linguísticos. Quanto à primeira ordem, que envolve os problemas psicológicos, fisiológicos e psíquicos de quem consome a informação, esses seriam atribuições da Inteligência Artificial, e não da WS.

Portanto, o caminho para a estruturação de SRIs que levem em conta essas especificidades se torna um desafio. O tratamento da informação deve abranger de forma simultânea tanto aspectos morfológicos, lexicais, sintáticos, semânticos e pragmáticos para que assim possa obter uma recuperação da informação efetiva.

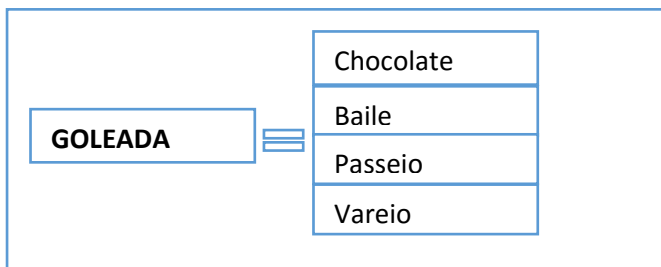
Em acervos de jornais, é preciso considerar ainda os aspectos sincrônicos e diacrônicos, que permitem analisar os processos evolutivos da língua. A sincronia refere-se às variações da língua no momento atual. A diacronia permite estabelecer comparações entre diferentes períodos (MARTELOTTA, 2011). É possível entender, dessa forma, que a diacronia é formada por diversas sincronias e que ambas não são opostas, mas sim complementares. Além disso, como destacam Melo e Bräscher (2011), o universo conceitual e a diversidade sociolinguística também devem ser considerados. Na Linguística Cognitiva, os aspectos históricos e socioculturais e as variabilidades do uso da língua são partes integrantes na estrutura linguística (SILVA, 2009). Nos estudos de Bourne (1977 apud LOPES, 2002), estas “formas variantes de palavras” influenciam diretamente nas buscas. Considerando os exemplos apresentados nas figuras 4, 5 e 6, é possível entender de forma resumida algumas dessas questões:

Figura 4 - Variantes da palavra (aspectos históricos)



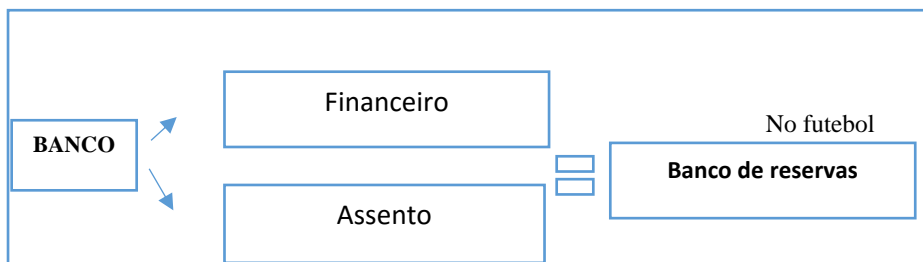
Fonte: Elaborado pela autora.

Figura 5 - Variantes da palavra (aspectos socioculturais)



Fonte: Elaborado pela autora.

Figura 6 - Variantes da palavra (contextos de uso)



Fonte: Elaborado pela autora.

A figura 4 mostra a variação da palavra partida, no contexto que considera as particularidades do acervo histórico e do domínio por trazer palavras que primeiramente eram escritas em inglês, devido ao futebol ser de origem inglesa. Essa análise pode também representar o aspecto diacrônico, pois no processo evolutivo da língua, muitas palavras do vocabulário inglês eram utilizadas no Brasil e com o passar do tempo foram aportuguesadas e introduzidas na língua portuguesa. A figura 5 mostra a variação da palavra goleada, que pode ser apresentada de diferentes formas conforme a localidade geográfica em que se aplica, mas representa o mesmo significado no domínio do futebol. Assim como palavra banco, mostrada na figura 6, que isoladamente é ambígua, mas se estiver relacionada com outras palavras, como “reservas” e “jogadores”, por exemplo, que possuem vínculo com o domínio do futebol, revela sua semântica (SCHIESSL, 2015).

Estas questões devem ser levantadas na organização de uma base de dados com textos históricos e contemporâneos e fundamentados na linguagem natural, como é o caso dos acervos digitais de jornais.

3.4.1 Web Semântica

O termo Web Semântica (WS) surgiu em 2001, com a proposta de Berners-Lee, Hendler, Lassila (2001), os primeiros a anunciar o mundo da WS e que a definiram como uma extensão da web atual, em que a informação estaria dotada de um significado bem definido que permitiria uma cooperação entre humanos e máquinas.

Souza e Alvarenga (2004, p.134) indicam que para atingir o propósito da WS é preciso uma padronização entre tecnologias, linguagens e metadados descritivos. O objetivo maior da WS é tornar a informação disponibilizada na Internet acessível e compreensível tanto para humanos e máquinas.

Apesar da WS ter sido idealizada no início do século 21, Robredo (2010, p.33) mostra que alguns conceitos e princípios da WS, já são utilizados pela CI há mais tempo, mas com outras nomenclaturas. Conforme o autor, na WS essas práticas se aperfeiçoam e são representadas, classificadas, organizadas, processadas, transferidas e aplicadas com novos códigos:

Quadro 1 - Comparação de alguns termos e expressões da Ciência da Informação e da Web semântica

Web / Web semântica	Ciência da Informação
Agentes de recomendação	Políticas de aquisição; Lei de Bradford
Ambiguidade / desambiguação	Sinonímia; polissemia. (nos tesouros: use; veja também) [para tal termo use tal outro]
Compartilhamento	Catálogos coletivos
Formato de comunicação e intercâmbio de dados	Norma ISO 2709, velha de mais de 40 anos, é atualizada periodicamente
Identificadores / Localizadores (URLs, URIs...)	Número de chamada, etc.
Inferência	Curvas estocásticas sobre séries históricas.
Interoperabilidade	OCLC, etc.

Linguagens de marcas; <i>hyperlinks</i>	Remissivas
Metadados	Dublin Core (metadados e qualificadores)
Ontologias	<i>Clusters</i> temáticos; métricas da informação
<i>Open systems</i> / Sistemas abertos	OPACs
<i>Parsers</i>	Indexação automática
Relações entre conceitos	Descritores compostos; adjacência; proximidade
Reúso	Pesquisa bibliográfica; este trabalho
<i>Reverse file</i>	Arquivo invertido
Tags	Etiquetas; tags; campos de dados; Unesco/Unisist; CDS/ISIS; MARC
Taxonomias	Sistemas de classificação; tesouros
Template	Planilha / Folha de entrada

Fonte: (ROBREDO, 2010, p. 33)

Evidentemente que as aplicações da WS atingem um contexto mais amplo, diversificado e complexo. Mas a CI, com a compreensão adquirida, tem total afinidade para contribuir e tentar preencher as lacunas que perduram e tornam a WS inacessível.

Em geral os sites de busca baseiam seus resultados em palavras-chaves. Quando o usuário inclui um termo de busca na ferramenta, as páginas com maiores ocorrências de palavras-chaves são consideradas mais relevantes, e conseqüentemente, aparecem nas primeiras posições dos resultados. Essa relevância é determinada geralmente pela frequência que as palavras-chave se repetem, pelo tempo que a página está no ar e, pelo número de outras páginas que remetem *link* para a mesma (CORRÊA e BERTOCCH, 2012). A recuperação da informação é determinada por algoritmos que não estão programados para atribuir significados e criar relações com a realidade (contexto) em que estão inseridos. Falta a semântica e a pragmática das linguagens naturais.

As transformações com a introdução da WS são significativas, muitas já estão acontecendo. No entanto, a evolução tem sido difundida de forma segmentada. Para os autores Corrêa e Bertocch (2012) algumas das razões para a lenta expansão da WS é a visão equivocada que se tem sobre ela:

A ideia de que a rede terá uma inteligência autônoma; de que será um repositório do conhecimento do mundo; de que conseguirá reconhecer os padrões individuais de uso e personalizar conteúdos para cada usuário de forma a fortalecer essa inteligência universal; um sistema de alavancagem do marketing digital; uma extensão da memória do usuário e, ao mesmo tempo, um arquivo individualizado de seus conhecimentos, entre outros.

O cenário que constitui a WS prevê o desenvolvimento de padrões tecnológicos que permitam e facilitem a troca de informações. A linguagem que será estabelecida nesse processo também é fundamental para garantir o compartilhamento de dados e a conversação entre os sistemas de informação. A WS inclui na sua arquitetura soluções como os metadados, linguagem de marcação XML e a arquitetura *Resource Description Framework* (RDF) a fim de garantir a interoperabilidade de dados semânticos, estruturais e sintáticos (VARELA E BARBOSA, 2013, p. 277).

A linguagem descrita na WS, quando formalizada e modelada “possibilita que agentes inteligentes possam entender o significado de um termo e estabelecer inferências sobre esses significados”, as características do termo são definidas pela atribuição de conceitos, que se relacionam com outros conceitos, formando um entendimento semântico (CAMPOS, 2010, p.134).

Um dos modelos de tratamento da informação adotados pela WS é a ontologia. Para Campos (2010) o compromisso ontológico é utilizado com o propósito de preencher a lacuna semântica entre o homem que define conceitos do mundo e aquilo que é possível representar por meio de uma linguagem de representação para recuperação da informação. Para a autora esse desdobramento pode ser dividido em: “as coisas cuja existência se admite e o seu significado pretendido”. Além de associar um termo a um conceito, é possível também definir a distinção entre ambos, ampliando as possibilidades de interpretação do sistema que irá fazer a leitura da estrutura organizacional. De acordo com Moreiro González (2011, p.35), a ontologia na visão da WS continua sendo responsável por permitir uma “organização da realidade” ou por envolver a “soma de uma série de conceitos relevantes que representam o conhecimento compartilhado por membro de

determinado domínio”, o complemento é que na WS a semântica é utilizada para “criar ontologias dirigidas a aplicações de software”.

3.4.3 Ontologias de domínio

O surgimento do termo ontologia vem da Filosofia. O objetivo da disciplina na área tem como finalidade o fornecimento de sistemas de categorização para organizar a realidade (BREITMAN, 2010, p.30).

A Ciência da Informação tem direcionado estudos sobre ontologias, com o intuito de esforçar-se para “diminuir as ambiguidades presentes na linguagem natural através do consenso terminológico, proporcionando semântica aos termos”, nos sistemas de recuperação da informação (SANTOS; CORREA; SILVEIRA, 2013, p. 1).

Conforme Schiessl e Bräscher (2011, p.310) destacam os benefícios com a evolução dos estudos de ontologias:

O estudo de ontologias estabelece importante linha de pesquisa para o incremento da Web Semântica. Seu desenvolvimento e popularização incentivam a busca de soluções que tragam qualidade às estruturas de informações dispersas no mundo digital da Web. As grandes bases de dados, especialmente as bases textuais, são beneficiadas com a evolução de tecnologias que buscam a tradução de textos para uma linguagem que a máquina “compreenda”.

Corrêa e Bertocchi (2012, p.127) apontam a importância do investimento em ontologias e as diferenças quando comparados aos dicionários, tesouros e taxonomias:

Por que a web semântica precisa de ontologias? Não bastaria um dicionário ou uma taxonomia bem estruturada para explicar à máquina as nuances entre os termos? O dicionário tem como propósito definir conceitos. O formato permite apresentar mais de uma definição por termo. Há dicionários de sinônimos, antônimos, de termos específicos de uma área, entre outros. O Tesaurus trabalha na lógica do sinônimo, aproximando conceitos similares. A taxonomia coloca os conceitos em

hierarquia. Mas é a ontologia que cria relações semânticas entre os termos – e, de fato, com ontologias, um número indeterminado delas pode ser criado, como um mapa mental.

De acordo com Pickler (2007, p.71), por meio das ontologias é possível construir uma “enorme rede de conhecimento humano” em cooperação com as máquinas, melhorando assim os serviços oferecidos na web, “sobretudo os serviços de busca e recuperação de dados”.

O compromisso ontológico, é destacado por Campos (2010, p.134) como um espaço de mediação que une a conceituação, orientada no plano da semântica entendível pelo humano e a representação, traduzida no plano do processamento pela máquina,

Quando classificadas as ontologias quanto a sua generalidade, os autores (HAAV; LUBI, 2001; GUARINO, 1998; BREITMAN, 2010), definem-nas como:

- a) Ontologias de nível superior – descrevem conceitos muito genéricos, tais como espaço, tempo e eventos. Estes seriam, a princípio, independentes de domínio e poderiam ser reutilizados na confecção de novas ontologias. Conforme Mendonça (2015, p.56) “na literatura da área, também chamadas de ontologias de fundamentação, ontologias fundacionais ou ainda meta-ontologias”.
- b) Ontologias de domínio – descrevem o vocabulário relativo a um domínio específico através da especialização de conceitos presentes na ontologia de alto nível.
- c) Ontologias de tarefas – descrevem o vocabulário relativo a uma tarefa genérica ou atividade através da especialização de conceitos presentes na ontologia de alto nível.
- d) Ontologias de aplicação – são as ontologias mais específicas. Conceitos em ontologias de aplicação correspondem, de maneira geral, a papéis desempenhados por entidades do domínio no desenrolar de alguma tarefa.

Quanto ao grau de formalidade ou grau de rigorosidade as ontologias Gómez-Pérez (2004) classificam-nas como:

- a) Ontologias leves (light-weight) são ontologias com pouco rigor formal, pois não se preocupam em fazer a descrição detalhada de cada conceito representado. Conforme

Guimarães e Diniz (2014, p.6) “ a principal ênfase das ontologias leves é definir a taxonomia que representa a relação hierárquica entre conceitos”. Aqui são definidas “classes que são facilmente compreensíveis e relações mais comuns entre estas classes”. (MENDONÇA, 2015, p.56)

- b) Ontologias pesadas (heavy-weight) são ontologias com alto rigor formal. São compostas pela junção das classes e relações comuns, com relações especiais e axiomas lógicos com alto grau de formalização (MENDONÇA, 2015).

A ontologia a ser utilizada nesta pesquisa é classificada como ontologia de domínio com formalidade leve. As ontologias de domínio possibilitam a utilização de termos específicos de cada área. De acordo com Bhogal, Macfarlane e Smith (2007), as ontologias são mais úteis se aplicadas em tarefas de recuperação de informações especializadas. Os autores ainda apontam que ontologias, independentes de domínio, apresentam ampla cobertura, ocasionando assim termos ambíguos e, por esta razão a não escolha de um domínio pode se tornar problemática se o objetivo principal da tarefa de busca for encontrar informações precisas.

Uma ontologia, conforme Moreira, Alvarenga e Oliveira (2004), pode apresentar diferentes formas, mas necessariamente contará como parte integrante um vocabulário de termos e algumas especificações de seus significados. Isso também incluirá definições e indicações de como se relacionam os termos, possibilitando assim uma estrutura no domínio com a limitação de interpretações dos termos, diminuindo assim a ambiguidade.

Para Chandrasekaran et al. (1999), a ontologia é um vocabulário de representação, frequentemente especializada em algum domínio ou assunto e, também usada para conceituar os termos do vocabulário do respectivo domínio. Conforme Moreiro González (2011, p.77), “a ontologia é uma descrição explícita e formal de conhecimentos em um domínio”.

No âmbito da RI, as ontologias podem ser aplicadas em diversas tarefas, que podem incluir sumarização temática, desambiguação do sentido da palavra, indexação, classificação de textos, recuperação de imagem, recuperação de informação multilíngue e, para a expansão de consulta (BHOGAL; MACFARLANE; SMITH, 2007).

A sumarização temática, a indexação e a classificação estão relacionadas aos processos de organização da informação que envolvem a representação da informação por meio de termos, atribuição de notação

numérica, índices e sumários que forneçam orientações capazes de direcionar o usuário até a informação solicitada na estratégia de busca.

A desambiguação do sentido das palavras, está associado à semântica aplicada aos termos de um vocabulário controlado. A escolha de um domínio para o processo de desambiguação é fundamental para definição de conceitos dos termos sem que haja dupla interpretação. As tarefas de recuperação de imagem, envolvem técnicas de indexação baseadas no conteúdo ou no conceito. “A indexação por conteúdo opera no nível sintático, enquanto a indexação por conceitos opera no nível semântico” (MANINI; LIMA-MARQUES; MIRANDA, 2007). E a recuperação multilíngue é quando a consulta e os documentos estão em idiomas diferentes, mas podem ser " acessados na língua nativa ou preferencial do usuário” (CARRASCO; THALLER; VIDOTTI, p.209). Por fim, dentre as tarefas que podem ser aperfeiçoadas com o auxílio de ontologias citadas por Bhogal, Macfarlane e Smith (2007), é apresentada a expansão de consulta.

A expansão de consulta, também chamado de “aumento da pergunta”, conforme os autores acima citados, tem como objetivo principal adicionar novos termos à consulta inicial. Este processo de expansão pode ser realizado de três formas: manual, automático ou assistido. A expansão manual baseia-se no letramento do usuário, que necessita tomar decisões para a escolha de termos que possam ser submetidos à uma nova consulta. Na expansão automática, o sistema realiza ponderações com todos os termos, define pesos, e no final as ponderações são calculadas e os termos que apresentaram os maiores pesos são adicionados à consulta inicial. Na expansão assistida o sistema gera possíveis relações e apresenta os termos que podem ser selecionados pelos usuários para expandir a consulta.

Conforme Krovetz e Croft (1992), os melhores resultados sobre determinado assunto são obtidos por meio de consultas individuais. Ou seja, é preciso que o usuário refaça inúmeras vezes a busca no sistema para conseguir informações suficientes para suprir sua necessidade de informação. Ainda, de acordo com os autores, o usuário não está interessado na recuperação de documentos que obtenham exatamente os mesmos termos incluídos na pesquisa, mas sim com o conceito que esses termos representam, por esta razão os sistemas devem expandir suas consultas, utilizando relações entre sinônimos ou apresentando os conceitos para que seja possível identificar os termos polissêmicos.

Para Breitman (2010, p.36, 37) os tipos de relacionamentos a serem sistematizados entre os termos contemplados em uma ontologia são

infinitos, mas precisam ser bem definidos. Os vocabulários controlados também podem formar relacionamentos, entretanto, nesse tipo de linguagem, os relacionamentos tendem a ser limitados, não atingindo aspectos do mundo real. Em muitas vezes é preciso criar relacionamentos do tipo “parte-de-membro-conjunto, fase-processo, lugar-região, material-objeto, causa-efeito” entre muitos outros delimitados conforme as necessidades do domínio modelado, que são suportados em uma ontologia.

A informação nesse estágio é avaliada em fragmentos e os dados não são tratados individualmente, mas de acordo com o contexto que integram, usando como base os relacionamentos definidos previamente em uma ontologia. Feitosa (2006, p.68) designa esse nível estrutural como inteligente, flexível, independente e parte constituinte de um sistema de informação mais amplo.

Os relacionamentos descritos e formalizados permitem que agentes inteligentes entendam o significado implícito em cada termo, e estabeleçam inferências sobre esses significados, para a partir das características dos conceitos se relacionar e formar o entendimento semântico. Conforme Moreiro González (2011), as ontologias permitem a definição de relações semânticas mais complexas, que as existentes em outros SOC. Essas relações não privilegiam apenas uma definição, pois os conceitos “são de níveis diversos (conceitos gerais e específicos) e de natureza categorial diversa (objeto, propriedade, processo)” (CAMPOS 2010, p.136), nesse sentido, mesmo que a ontologia opere em domínios do conhecimento, as relações estabelecidas devem contemplar os diferentes elementos que formam esse domínio.

3.4.4 Ontologias de fundamentação

Geralmente, utilizam-se ontologias de fundamentação como ponto de partida para a construção de ontologias de domínios (MENDONÇA, 2015). De acordo com Degen et al. (2001), toda ontologia de domínio deve utilizar como referência alguma ontologia de fundamentação, adotando suas teorias e princípios. Para Campos, M. L. A., e Campos, L. M., (2012. p.6) “as ontologias de fundamentação permitem que a construção de uma teoria sobre o domínio possibilite testar e validar um modelo conceitual”.

As ontologias de fundamentação, também chamadas de ontologias de nível superior ou alto nível, tem forte sustentação na Filosofia, permitem uma representação semântica baseada no mundo real,

com a restrição das interpretações dos conceitos que formam o domínio (CAMPOS, M. L. A.; CAMPOS, L. M., 2012). Para Guizzardi, Giancarlo; Falbo, Ricardo A.; Guizzardi (2008, p.244) as ontologias de fundamentação são como “sistemas de categorias filosoficamente bem fundamentados e independentes de domínio que têm sido utilizados com sucesso para melhorar a qualidade de linguagens de modelagem e modelos conceituais”.

Conforme Campos, M. e Campos, L. (2012, p.7), as ontologias de fundamentação permitem diferenciar os tipos de elementos que fazem parte de um domínio, permitindo a representação e o estabelecimento da posição de cada elemento na estrutura do domínio. Essa organização é construída de forma a evitar proposições errôneas, a fim de garantir a qualidade da modelagem conceitual do domínio.

De acordo com Campos (2004) as ontologias de fundamentação são baseadas em ontologias formais. Suas estruturas representam e formalizam o conhecimento existente em domínio, por meio de conceitos e categorias. Mendonça (2015) ressalta que, ainda que o objetivo das ontologias de fundamentação seja o mesmo, elas apresentam princípios ontológicos que as diferem, como o conjunto de entidades, meta propriedades, relações e axiomas. Mendonça (2015), destaca algumas ontologias que surgiram nos últimos anos:

Quadro 2 - Ontologias de fundamentação

ONTOLOGIAS DE FUNDAMENTAÇÃO	
CYC (CYC PROJECT) – 1990	Seu objetivo principal é a integração de aplicações computacionais que possuem dados com variações estruturais. Além desse enfoque computacional, a CYC esforçou-se também em ser uma ontologia unificada que pudesse descrever os objetos do mundo. O projeto da CYC destaca-se mais como um esforço de unificação e integração de algumas ontologias do que propriamente como uma ontologia de fundamentação, que propõe princípios formais. A representação do conhecimento da CYC é operacionalizada por uma linguagem formal denominada de Cyc Language (CycL), um vocabulário de termos e um conjunto de declarações que relacionam os termos. Ao todo, possui 47.000 termos e 306.000 declarações, numa linguagem formal de representação, as quais relacionam esses termos e restringem o seu significado. Essas declarações são agrupadas sem microteorias que representam um domínio particular de conhecimento, um nível específico

	<p>de detalhes, um intervalo de tempo, etc. A CYC também possui uma versão de código aberto conhecido como projeto Open-Cyc. A partir do site do Open-Cyc18 obtém-se um software para instalação do servidor Cyc, o qual dá acesso a uma página de vocabulário e a uma interface de busca por termos e relações.</p>
<p>DOLCE (GUARINO) -1991</p>	<p>Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), é a ontologia de fundamentação originada do projeto WonderWeb19, que tem por objetivo principal o desenvolvimento de uma biblioteca de ontologias de fundamentação, denominada de WonderWeb Foundational Ontologies Library (WFOL). A DOLCE está disponível em domínio público, desde sua primeira versão, como forma de estimular outros grupos de pesquisa que trabalham com ontologias formais WonderWeb é um projeto que se originou na Victoria University of Manchester com o propósito de desenvolver uma biblioteca de ontologias de fundamentação, sistematicamente relacionadas umas com as outras, e capazes de justificar as preferências que fundamentam as escolhas ontológicas, tornando-as mais explícitas possível (MASOLO et al., 2003). A DOLCE é considerada uma ontologia de particulares e que possui uma rica axiomatização, incluindo ao todo 37 categorias básicas, 7 relações básicas, 80 axiomas, 100 definições e 20 teoremas.</p>
<p>Sowa's Ontology – 1999</p>	<p>Knowledge Representation Ontology ou simplesmente ontologia de Sowa é composta por categorias básicas e distinções derivadas de fontes variadas da lógica, da filosofia e da inteligência artificial. A ontologia de Sowa não é baseada em uma estrutura fixa de categorias, mas em um conjunto de distinções a partir das quais uma hierarquia pode ser gerada automaticamente, decisões ontológicas a serem tomadas no processo de desenvolvimento.</p> <p>Quanto à estrutura da ontologia de Sowa, ela é composta por níveis denominados alto-nível, processos, relações, causalidade, agentes e papéis temáticos. Cada um dos níveis desta ontologia é formado por um conjunto de termos específicos, usados para a representação e descrição dos objetos de um domínio.</p>
<p>Sumo (IEE) - 2003</p>	<p>Já a Standard Upper Merged Ontology (SUMO) é uma ontologia de fundamentação, que foi disponibilizada pela associação internacional <i>Institute of Electrical and Electronics Engineers</i> com o objetivo de para pesquisa sobre busca de informações, sobre linguística e sobre deduções automáticas. A SUMO é composta por um conjunto de ontologias bem diversificado, que incluem a própria</p>

	<p>ontologia SUMO, a MidLevel Ontology (MILO), além de ontologias de comunicação, de países e regiões, de economia e finanças, de engenharia, de geografia, de pessoas, de transportes, entre outras. Essa característica da SUMO provoca questionamentos justificados sobre sua natureza de ontologia de fundamentação, já que inclui ontologias de domínio. No entanto, pode-se considerar que ela representou uma iniciativa de cobrir o maior número possível de domínios do conhecimento, através da unificação de diversas ontologias para descrever os objetos do mundo real. Sabe-se, portanto, que a SUMO não pode ser considerada uma ontologia de fundamentação em sua essência.</p>
BFO (Smith) - 2004	<p>Basic Formal Ontology (BFO) se propõe a ser uma ontologia estritamente formal ou de nível superior, com o propósito principal de dar suporte ao desenvolvimento de ontologias de domínio, especialmente na área biomédica. Nesse sentido, a BFO auxilia nas tomadas de decisão ontológicas do desenvolvedor e ajuda também a tornar o mais claro possível esse processo de escolha. Por ser considerada uma ontologia estritamente formal, termos específicos dos domínios físicos, químicos ou biológicos não estão na BFO, mas sim categorias fundamentais da realidade, sendo que seu domínio é neutro.</p>
GFO (Herre) - 2006	<p>General Formal Ontology (GFO) é uma ontologia de fundamentação que integra objetos e processos. Possui uma arquitetura metaontológica formada por três camadas: (i) um nível básico, que inclui todas as categorias e relações GFO relevantes; (ii) um meta-nível, chamado de nível nuclear, que contém as meta-categorias²⁴ sobre o nível básico; e (iii) um nível superior, que inclui as chamadas meta-meta-categorias.</p>
UFO (Guizzardi) - 2009	<p>Unified Foundational Ontology (UFO) foi desenvolvida com o propósito principal de prover uma teoria para resolver os problemas clássicos de modelagem conceitual. Seu desenvolvimento baseou-se numa série de teorias da Ontologia Formal, da Lógica Filosófica, da Filosofia de Linguagem, da Linguística e da Psicologia Cognitiva. Inclui teorias sobre partes e todos, tipos, estruturas taxonômicas, relacionamentos, atributos e atributos de espaço, papéis e qualidades individuais, além de outras coisas.</p>

Fonte: Mendonça (2015, p.75-95)

O propósito desta pesquisa não está ligado às ontologias de fundamentação, no entanto, à medida que elas podem ser utilizadas como ponto de partida para construção de ontologias domínio, considerou-se importante trazer uma breve contextualização do seu uso.

3.4.5 Metodologia para construção de ontologias

De forma geral, não existe um modelo único, ou um roteiro que deva ser seguido para construção de ontologias. Schiessl e Bräscher (2011) afirmam que não existe um consenso entre pesquisadores sobre a melhor forma de construir uma ontologia. Para Fernández-López (et al., 1999) há diversas estratégias para o desenvolvimento de ontologias com abordagens e características diversas, sendo direcionadas para propósitos e aplicações diferentes. No entanto, devido ao notável crescimento da pesquisa de ontologias e da WS, algumas metodologias já são consolidadas e utilizadas constantemente para auxílio na modelagem do conhecimento, na formalização da linguagem e na implementação da ontologia.

Para Schiessl e Bräscher (2011), as etapas que envolvem a construção de ontologias são muito influenciadas pelas atividades de engenharias de software. De acordo com Silva, Souza e Almeida (2008) este fato pode ser justificado devido as ontologias e sistemas de informações serem objetos de estudo proeminentes da Ciência da Computação.

Porém, muitas das dificuldades desse ciclo estão nos gargalos que envolvem a aquisição do conhecimento e também no tempo investido para construí-las e integrá-las a outros domínios e aplicações. Os autores destacam o ciclo de construção de ontologias, (conforme o quadro 3), citado abaixo, baseado nas etapas de identificação do domínio, descrição dos conceitos e relações, formalização em linguagem formal, aplicação da ontologia e manutenção, que propõe a contínua atualização dos termos e também correção da ontologia. Além destes passos os autores ainda destacam atividades paralelas ao ciclo de vida das ontologias, que devem ser executadas.

Quadro 3 - Ciclo de construção de ontologias

ETAPA	PROCEDIMENTO
Especificação	Identifica o propósito e o âmbito da ontologia;
Conceitualização	Descreve, em modelo conceitual, a ontologia a ser construída, de forma que atenda às especificações do passo anterior. O modelo conceitual de ontologia consiste no domínio de conceitos e as relações entre eles. As relações reforçam as conexões mais fortes entre

	grupos de conceitos. Os grupos de conceitos fortemente relacionados geralmente correspondem a diferentes módulos (subontologias) em que o domínio pode ser decomposto;
Formalização	Transforma a descrição conceitual em modelo formal, isto é, a descrição do domínio no passo anterior é representada em linguagem formal, ainda que não seja a forma final. Conceitos são normalmente definidos através de axiomas que delimitam as interpretações possíveis para o significado desses conceitos. Conceitos são geralmente organizados hierarquicamente através de uma relação estruturante, tal como “é-um” (classe-superclasse, instância-classe) ou “parte-de”;
Aplicação	Implementa a ontologia formalizada em linguagem de representação de conhecimento. Para isso, escolhe-se uma linguagem para representação e escreve-se o modelo formal na linguagem escolhida;
Manutenção	Atualiza e corrige a ontologia aplicada.
Reforça-se que o encadeamento das atividades não pretende ser o melhor recurso, pois não é este o foco. Elos são uma opção viável e consolidada por estudos anteriores e, como tudo em ciência, são passíveis de melhoramentos. Logo, esses passos não esgotam o tema, pois há atividades paralelas ao ciclo de vida que podem e devem ser realizadas.	
Aquisição de conhecimento	Adquire o conhecimento sobre o assunto utilizando técnicas de dedução junto aos especialistas de domínio ou por referência à bibliografia relevante. Várias técnicas podem ser utilizadas para a aquisição de conhecimentos, tais como o brainstorming, entrevistas, questionários, análise de textos análise e técnicas de indução;
Avaliação	Julgamento técnico, baseado em técnicas disponíveis, da qualidade da ontologia;
Documentação	Relata o que foi realizado, como foi feito e o porquê. A documentação associada aos termos representados na ontologia é particularmente importante não apenas para melhorar sua clareza, mas também para facilitar a manutenção, a utilização e a reutilização.

A construção de forma manual de ontologias se desenvolveu nas últimas décadas com certa maturidade, entretanto, demanda de um intenso trabalho, que consome além dos recursos financeiros, a dedicação de especialistas para o seu desenvolvimento (SCHIESSL; BRÄSCHER, 2011). Esta é uma das razões pela qual, os métodos de engenharia de software se destacam, pois, visam ampliar a automatização dessas tarefas. O grande volume de informações disponibilizadas na web também requer que lógicas e algoritmos de programação sejam introduzidos nas estruturas ontológicas.

Mendonça e Almeida (2014) citam alguns exemplos de metodologias “maduras” empregadas na engenharia de software e na construção de tesouros, referindo-se à USCHOLD; GRUNINGER (1996); JONES; BENCH-CAPON; VISSER (1998); FERNANDÉZ-LOPEZ (1999); FERNANDÉZ-LOPEZ; CORCHO (2004); BREITMAN (2005); CARDOSO (2007); SILVA; SOUZA; ALMEIDA (2008).

Além desses exemplos, Mendonça e Almeida (2014) elaboraram um quadro (quadro 4) com análise das principais metodologias em uso, para construção de ontologias:

Quadro 4 - Metodologias para construção de ontologias

Metodologia	Descrição (objetivo e domínio)	Etapas (processos e atividades)
Metodologia de Gruninger e Fox – TOVE (GRUNINGER e FOX, 1995)	Criada no projeto <i>Toronto Virtual Enterprise</i> (TOVE), cujo objetivo era criar um modelo de senso comum compartilhado corporativo. Domínio de aplicação: Negócios (empresarial).	1. Elaboração de cenários de motivação; 2. Questões de competência informal; 3. Concepção da terminologia formal; 4. Questões de competência formal; 5. Especificação de axiomas formais; 6. Verificação de teoremas completos.

<p>Metodologia de Uschold e King ENTERPRISE (USCHOLD e KING, 1995)</p>	<p>Desenvolvido com base na prática da construção da ontologia Enterprise Ontology, que descreve conhecimento no domínio corporativos. Domínio de aplicação: Negócios (empresarial).</p>	<ol style="list-style-type: none"> 1. Propósito, grau de formalismo e usuário; 2. Construção da ontologia em três etapas; <ol style="list-style-type: none"> a) conceitualização; b) implementação; c) integração com ontologias já existentes; 3. Avaliação; 4. Documentação.
<p>Methontology GÓMEZPEREZ, FERNANDEZLOPES e VICENTE, 1996)</p>	<p>Possibilita a construção de ontologias por reengenharia, utilizando-se do conhecimento de domínio. Domínio de aplicação: Diversos.</p>	<ol style="list-style-type: none"> 1. Especificação; 2. Aquisição de Conhecimento; 3. Conceitualização; 4. Integração; 5. Implementação; 6. Avaliação; 7. Documentação.
<p>Método <i>Kactus</i> (BERNARAS, LARESGOTT, CORERA, 1996)</p>	<p>Método derivado do projeto <i>Kactus</i> Que permite reutilização desconhecimento em sistemas técnicos, tal como redes elétricas. Domínio de aplicação: Sistemas de complexidade técnica.</p>	<ol style="list-style-type: none"> 1. Lista de necessidades ou requisitos; 2. Identificação de termos relevantes; 3. Criação de modelo preliminar; 4. Estruturação refinamento; 5. Integração e reutilização.
<p>Método <i>Sensus</i> (SWARTOUT et al., 1996).</p>	<p>Método derivado da ontologia <i>Sensus</i>, desenvolvida para de processamento de linguagem natural. Domínio de aplicação: Diversos</p>	<p>Identificar termos-chave do domínio; Ligar os termos-chave à ontologia <i>Sensus</i>; Adicionar novos termos para o domínio; Adicionar subárvores completas.</p>

<p>Método 101 (NOY e GUINNESS, 2001)</p>	<p>Método usado no desenvolvimento de exemplos práticos utilizando o editor de ontologias Protégé. Domínio de aplicação: Diversos</p>	<ol style="list-style-type: none"> 1. Determinar escopo; 2. Considerar o reuso de termos; 3. Enumerar termos; 4. Definir classes; 5. Organizar as classes em uma taxonomia; 6. Definir propriedades e suas restrições; 7. Adicionar valores de instâncias.
<p>Método CYC (REED, LENAT, 2002)</p>	<p>Método usado na construção da ontologia CYC, que busca abranger o conhecimento consensual do mundo. Domínio de aplicação: Diversos.</p>	<ol style="list-style-type: none"> 1. Extração manual do conhecimento; 2. Extração auxiliada por computador; 3. Desenvolvimento de representação; 4. Representação do conhecimento de diferentes domínios usando primitivas.
<p><i>On-toKnowledge Methodology</i> (OTKM) (SURE, STAAB e STUBER, 2003)</p>	<p>Metodologia desenvolvida para a construção de ontologias para aplicações de gestão do conhecimento Domínio de aplicação: Gestão do conhecimento empresarial.</p>	<ol style="list-style-type: none"> 1. Estudo de viabilidade; 2. Kickoff; 3. Refinamento; 4. Avaliação; 5. Aplicação e evolução; 6. Geração de conhecimento; 7. Captura de conhecimento; 8. Recuperação e acesso.
<p><i>Metodologia UP for ONtology</i> (UPON) (DE NICOLA, MISSIKOFF e</p>	<p>Metodologia de construção de ontologias derivada em padrão de engenharia de software</p>	<ol style="list-style-type: none"> 1. Workflow de Requisitos; 2. Workflow de Análise;

NAVIGLI, 2009)	Processo Unificado. Domínio de aplicação: Negócios (ebusiness).	3.Workflow de Desenvolvimento; 4.Workflow de Implementação; 5.Workflow de Teste.
<i>Metodologia NeON</i> (SUARÉZFIGUEROA,2010)	Metodologia para construção de redes ontológicas baseado em um desenvolvimento colaborativo Domínio de aplicação: Diversos.	1. Gerenciamento de processos e atividades; 2.Desenvolvimento orientado de processos e atividades; 3. Suporte aos processos e atividades.
Metodologia MFPPFO (LIM, LIU e LEE, 2011)	Metodologia de construção de ontologia multi-facetada, anotada semanticamente, para a modelagem de uma família de produtos. Domínio de Aplicação: Domínios que possuem uma família de produtos.	1.Construção de uma taxonomia da família de produtos; 2. Extração de entidades; 3.Identificação do conceito e geração da unidade facetada; 4. Anotação semântica e modelagem faceta; 5.Construção de uma ontologia de família de produtos multi-facetada e anotada semanticamente; 6. Avaliação e validação da ontologia.
Ciclo de Vida de Schiessl e Bräscher (2011)	Embora não seja uma metodologia, o ciclo de vida ontológico inclui todas as etapas necessárias no processo	1. Especificação; 2.Conceitualização; 3. Formalização; 4. Aplicação; 5. Manutenção;

	de construção de ontologias. Domínio de Aplicação: Diversos.	6. Atividades paralelas: aquisição de conhecimento, avaliação, documentação.
--	--	--

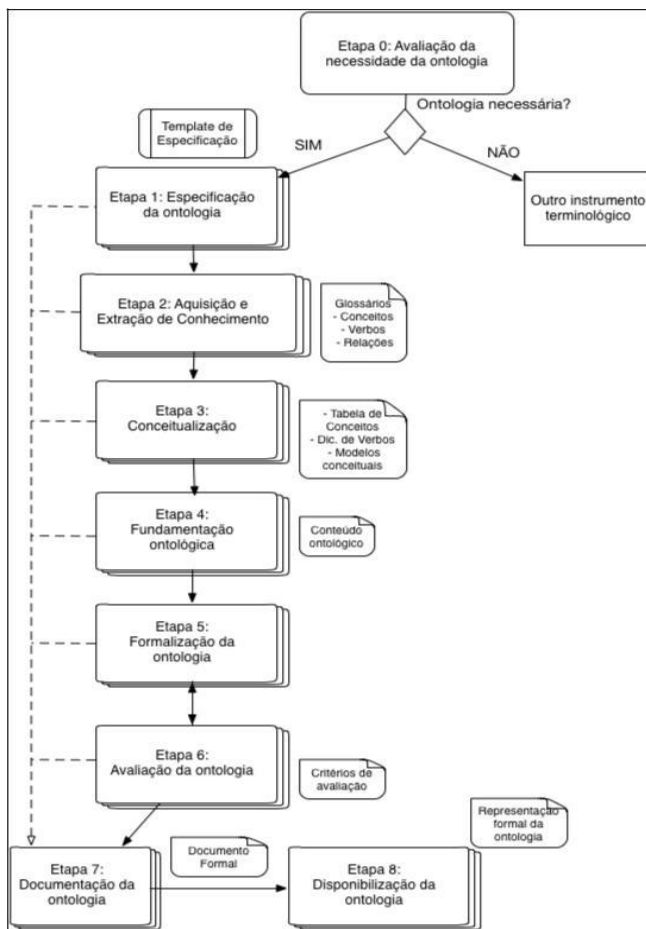
Fonte: Mendonça e Almeida (2014)

Conforme os autores, as metodologias analisadas demonstram semelhanças entre os principais aspectos das etapas previstas para construção de ontologias, entretanto, é possível distinguir algumas diferenças, que são percebidas principalmente quando as metodologias são direcionadas para domínios de aplicação distintos.

A fim de unificar as etapas de algumas metodologias atuais, suprir limitações que cada uma apresenta e também descrever mais detalhadamente cada etapa, Mendonça (2015) propôs um conjunto de princípios metodológicos para construção de ontologias – a *OntoForInfoScience*. Para isso, utilizou metodologias que têm se destacado no desenvolvimento de ontologias de domínio, tais como: Methontology (FERNÁNDEZ et al.,1997); Metodologia NeON (SUARÉZ-FIGUEROA, 2010) e Método 101 (NOY e GUINNESS, 2001).

Os princípios metodológicos propostos por Mendonça (2015) estão divididos em nove etapas. Na proposta, as etapas básicas das metodologias usadas para fundamentação são mantidas, porém é apresentado “subetapas”, com passos mais detalhados e adaptações que considerem o domínio estudado, itens que não são muito explorados nas metodologias atuais. As etapas consideradas na proposta estão estruturadas na figura 7.

Figura 7 - Proposta para Construção de ontologia



Fonte: Mendonça (2015)

Inicialmente a metodologia *OntoForInfoScience*, sugere que seja feita uma avaliação prévia da necessidade de construção da ontologia. Caso seja evidenciada tal necessidade, são seguidas as próximas etapas da metodologia, caso contrário, recomenda-se o uso de outros instrumentos terminológicos.

A primeira etapa da metodologia, recomenda que sejam incluídas informações a respeito do propósito da ontologia, objetivos específicos, tipo da ontologia desenvolvida (domínio, nível superior, aplicação, de tarefas (BREITMAN, 2010, p.39), sua cobertura, público que se destina e grau de formalidade.

A segunda etapa ocupa-se com adoção de métodos para a aquisição do conhecimento. Para criar esse corpus podem ser analisados textos de documentos de referência do domínio, entrevistas estruturadas ou semi-estruturadas, aplicação de questionários e *brainstorming* com especialistas da área. Conforme os autores Mendonça e Almeida (2014, p.441), esta atividade “na prática, traduz-se na escolha dos materiais de referência mais relevantes e representativos acerca do domínio”. A identificação dos termos pode ser formada também usando como base dicionários de termos, glossários e tesouros específicos do domínio. Nessa etapa da *OntoForInfoScience* são gerados os glossários de conceitos, glossários de verbos e glossários de relações.

A terceira etapa diz respeito à conceitualização, cujo objetivo é descrever a ontologia, de uma forma conceitual com seus conjuntos de termos e relacionamentos do domínio. De acordo com Mendonça (2015, p.197), essa é uma das fases mais importantes para uma ontologia ser bem-sucedida, nessa etapa acontecerá “a organização e estruturação do conhecimento por meio de artefatos de representação, que irão se transformar em relações, propriedades e restrições de uma ontologia”. Para a consecução dessa etapa são utilizados como base os artefatos gerados na etapa anterior - glossários de conceitos, verbos e relações, para conceber novos artefatos: dicionário de conceitos; tabela de conceitos e valores; tabela de conceitos e propriedades; dicionário de verbos, conjunto de relações conceituais e modelos conceituais gráficos.

Esses artefatos permitem originar um modelo conceitual do domínio, a partir dos conceitos, relações e propriedades que formam o domínio. Conforme Mendonça (2015), nessa etapa o conceito é associado à sua definição textual, obtidos a partir de materiais de referência do domínio. A identificação de sinônimos também é feita nessa etapa, pois é nesse momento que as informações do domínio passam a ser tratadas em nível semântico.

Nas relações formadas nessa etapa, Shamsfard e Barforoush (2003) definem que em uma ontologia um relacionamento conecta dois ou mais conceitos. Dahlberg (1978) aponta que quando conceitos diferentes possuem características em comum e há comparação entre eles, então há relação entre tais conceitos e estes podem ser designados da seguinte forma:

- a) Hierárquicas: também chamado de relação de gênero e espécie. O conceito superior é o mais genérico e o inferior é o mais específico, que herda as características do conceito mais amplo. Exemplo: mamífero – gato.
- b) Partitivas: existe entre um todo e suas partes. Exemplo: carro – pneus, direção, bancos, portas, etc.
- c) De oposição: se estabelece com base em contradições. Exemplo: alto e baixo; gordo e magro.
- d) Funcionais: é aplicada aos processos. Exemplo: produto – produtor – comprador.

Os relacionamentos baseados em características comuns, também podem ser representados de uma forma lógica, conforme o quadro 5.

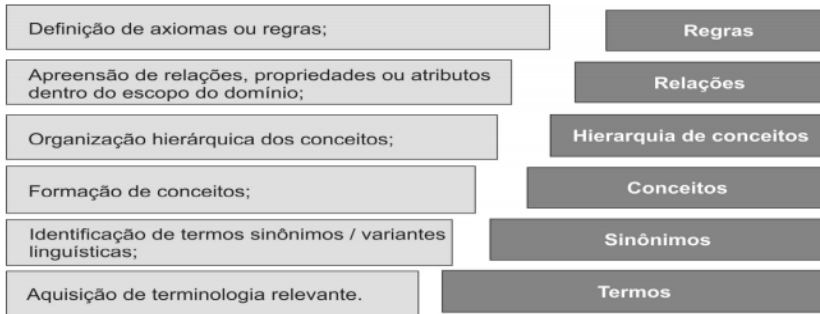
Quadro 5 - Relações lógicas

Identidade	A (x,x,x)	B (x,x,x)	As características são as mesmas
Implicação	A (x,x)	B (x,x,x)	O conceito A está contido em B
Intersecção	A (x,x,o)	B (x,o,o)	Os dois conceitos coincidem algum elemento
Disjunção	A (x,x,x)	B (o,o,o)	Os conceitos se excluem mutuamente. Nenhuma característica em comum.
Negação	A (x,x,o)	B (o,x,o)	O conceito A contém uma característica cuja negação se encontra em B.

Fonte: (DAHLBERG, 1978).

As atividades e termos utilizados na aquisição e extração do conhecimento, na conceitualização e na definição de relacionamentos e hierarquias da ontologia podem ser compreendidas de acordo com a figura 8.

Figura 8 - Aprendizado de ontologia em camadas



Fonte: Schiessl e Bräscher (2011)

A figura 8 deve ser compreendida da parte inferior para o topo, conforme o grau de complexidade das etapas. Ou seja, a primeira etapa a ser executada é a aquisição dos termos, a camada seguinte se dará conforme seja construída a camada anterior. Conforme Schiessl e Brascher (2011) o núcleo fundamental da construção de ontologias está na definição de conceitos e nas relações obtidas entre eles. Para tanto, demanda também da aquisição de conhecimentos linguísticos, para explorar eventuais sinônimos dos termos. Além das inferências, sinalizadas pelas regras ou axiomas, demandadas quando há fatos que não estão claramente expostos nos textos e dependem de uma dedução.

Ainda na etapa da conceitualização são definidas as propriedades específicas que que formam os conceitos dos termos identificados no domínio. Conforme Mendonça (2015, p.203), nessa etapa o conhecimento ainda está em nível conceitual, o nível ontológico será definido na etapa de formalização:

Propriedades de um conceito ou classe são características que permitem diferenciar um conceito do outro no domínio sob estudo e são uma herança dos trabalhos de Aristóteles que tratava as propriedades sob o nome de “differentia”. Incluir as propriedades das classes em qualquer desenvolvimento ontológico é indispensável para uma representação fidedigna do domínio sob estudo. Para que a ontologia consiga responder as questões de competência relativas a ela, as

propriedades das classes precisam estar definidas. Nesta etapa da metodologia, que trata o conhecimento a nível conceitual e ainda não ontológico, a definição das propriedades das classes pode ser feita de forma descritiva em formato textual, não sendo necessário especificar explicitamente propriedades tais como: tipo de valores das classes (numérico, string, booleano), domínio e imagem, cardinalidade. Essas propriedades específicas de uma ontologia deverão ser definidas na etapa de formalização da ontologia, usando um editor de ontologias para tal.

Por fim, a última fase da etapa de conceitualização proposta na *OntoForInfoScience* são os modelos conceituais gráficos, representados por meio de taxonomias, hierarquias, grafos conceituais, ou outras estruturas que consigam representar graficamente o conhecimento modelado. Em relação à construção de taxonomias relacionada a classes do domínio ou a construção de hierarquias, Mendonça e Almeida (2014, p.442) chamam atenção para alguns pontos:

[..] tais estruturas devem conter classes relacionadas unicamente através da relação ontológica *is_a* (*é_um*), de maneira que se defina formalmente o tipo de relação *is_a* utilizado na construção das taxonomias, uma vez que existe mais de um tipo dessa relação. A mesma regra se aplica para as partonomias, que devem conter classes relacionadas unicamente através da relação ontológica *todo_parte*, que também deve ter seu tipo especificado formalmente. Por fim, o que se convencionou chamar de estruturas gerais de representação são estruturas que englobam classes relacionadas através de diferentes relações ontológicas, como, por exemplo, as relações *produces*, *participates_in*, *has_agent* da ontologia Relation Ontology (RO) e as próprias relações *is_a* e *part_of* das taxonomias e partonomias, respectivamente.

A quarta etapa refere-se à fundamentação da ontologia e envolve a seleção da abordagem filosófica a ser utilizada no desenvolvimento da ontologia, assim como a avaliação do uso e aplicação de ontologias de

fundamentação a serem utilizadas como ponto de partida para a etapa seguinte, de formalização da ontologia. Na metodologia *OntoForInfoScience* defende-se que “a adoção de princípios filosóficos de uma determinada linha auxilia nas decisões ontológicas e permite justificá-las pelo seu embasamento teórico”. (MENDONÇA, 2014, p.211, 212)

A quinta etapa destina-se à formalização da ontologia que foi modelada nas etapas anteriores. Nesta etapa a ontologia é descrita formalmente, sendo capaz de representar a conceitualização do domínio modelada anteriormente, nesse momento, o conhecimento tratado em nível conceitual, passa para o nível ontológico-formal. Toda etapa de formalização pode ser desenvolvida com a utilização de softwares editores de ontologia, como o Protégé.

A sexta etapa refere-se à avaliação do conteúdo ontológico desenvolvido. As atividades dessa etapa envolvem:

É essencial é a adoção de critérios, métodos ou técnicas para avaliação do conteúdo da ontologia. Assim como ocorre com outros aspectos essenciais, geralmente, todas as metodologias de construção incluem uma etapa para a avaliação do conteúdo da ontologia. O que diferencia uma metodologia da outra quanto à avaliação é o nível de rigorosidade como é feita essa avaliação, que varia desde um conjunto de critérios básicos até um framework formal de avaliação, e também os tipos de critérios, métodos ou técnicas empregados (MENDONÇA; ALMEIDA, 2014, p.153, 154).

A *OntoForInfoScience* estabeleceu parâmetros e critérios de avaliação da ontologia desenvolvida. Essa validação analisa aspectos que envolvem a adequação ao domínio, referentes às etapas de especificação e conceitualização, além da consulta de especialistas para validação do conhecimento do domínio. Os parâmetros também verificam a correção ontológica, comprometida com a análise da etapa de formalização, com aspectos de integridade e completude, consistência, precisão e documentação.

A sétima etapa constitui a parte da documentação da ontologia desenvolvida nas partes anteriores. Os elementos que devem ser incluídos nessa documentação são:

(i) elementos de especificação da ontologia (propósito geral, escopo de cobertura, público-alvo, etc.); (ii) questões de competência para definição do escopo e também para validação da ontologia; (iii) relação e breve descrição das ontologias reutilizadas no projeto ontológico; (iv) relação dos termos reaproveitados dessas ontologias; (v) dicionários de classes; (vi) dicionário de relações; (vii) hierarquias e outras estruturas de representação da ontologia; (viii) conjunto de critérios, métodos e técnicas usados na avaliação e validação; (ix) relação e descrição dos materiais de referência utilizados na aquisição do conhecimento. (MENDONÇA; ALMEIDA, 2014, p.443)

Por fim, na oitava etapa é disponibilizada a ontologia. Na ontologia do domínio do futebol, desenvolvida neste trabalho, optou-se pelo uso da metodologia *OntoForInfoScience*, por esta apresentar uma unificação de metodologias já difundidas na área e também por disponibilizar etapas mais detalhadas, o que permitem um melhor entendimento para o desenvolvimento da ontologia. Contudo, nesta pesquisa, devido ao reduzido prazo para execução, serão contempladas as seis primeiras etapas. As etapas que envolvem a documentação e disponibilização da ontologia serão consideradas em pesquisas futuras.

3.4.6 Linguagens de representação descritiva

A representação descritiva determina o apontamento das características que compõem um documento, permitindo a sua individualização. Esse procedimento tem o intuito de tornar os documentos únicos e mais fáceis de serem encontrados. Nessa etapa são definidos os pontos de acesso para recuperação da informação e as possibilidades de relacionamentos com outros documentos. Para Catarino e Souza (2012) ao aproximar as práticas da representação descritiva da WS, vislumbra-se a expectativa de reutilizar a atuação do profissional da informação com os princípios da organização dos recursos na web, que permitiram, conforme a regras de catalogação, o compartilhamento e a preservação dos dados bibliográficos na Internet.

Há centenas de anos bibliotecários, museólogos, arquivistas e editores, fazem o uso de metadados para o gerenciamento de coleções. A

adoção de metadados é utilizada para a identificação, descrição e localização de recursos (BREITMAN, 2010, p.16).

Conforme Breitman (2010), os metadados podem ser extraídos automaticamente de bases de dados ou ser fornecidos por humanos. Podem ser criados, modificados ou até mesmo descartados. Segundo Berners-Lee, Hendler e Lassila (2001), os computadores precisam de coleções estruturadas de informações formadas por dados e metadados, além do estabelecimento de um conjunto de regras que permita a inferência e ajude no processo do raciocínio automático.

O *World Wide Web Consortium - W3C*, principal organização que promove ações para padronizar a web, tem desenvolvido pesquisas que envolvem metadados e tecnologias que permitem acrescentar semântica nas informações, de modo que seja possível entender o que a informação significa (BERNERS-LEE, HENDLER e LASSILA, 2001).

De acordo com Bonifacio e Heuser (2002) a inclusão de semântica nos metadados pode ser denominada de metadados semânticos, estes “fornecem informações sobre o significado dos dados disponíveis e seus relacionamentos semânticos”. Indicados para descrever o conteúdo dos documentos, por meio de informações como título, autores, localização, idioma, formato, etc., os metadados podem ser adaptados conforme o contexto em que se aplicam.

A utilização de metadados aplica-se a diferentes aspectos da organização da informação e, portanto, há diferentes tipos de metadados, conforme exposto no quadro 6.

Quadro 6 - Tipo de metadados quanto a sua função

Tipo	Definição	Exemplos
Metadados administrativos	Utilizados na gerência de recursos de informação	i. Aquisição de informação ii. Registros de direitos e reprodução iii. Documentação dos requisitos legais de acesso iv. Informação de localização v. Critérios de seleção para a digitalização vi. Controle de versão

Metadados descritivos	Utilizados para descrever e identificar recursos de informação	<ul style="list-style-type: none"> i. Registros de catalogação ii. Auxílio para a procura de informação iii. Índices especializados iv. Utilização de hiperlinks entre recursos v. Anotações
Metadados de Preservação	Relacionados ao gerenciamento dos recursos de informação	<ul style="list-style-type: none"> i. Documentação sobre a condição física dos recursos ii. Documentação sobre as ações tomadas nos documentos físicos e digitais iii. Atualização e migração
Metadados Técnicos	Relacionados à funcionalidade do sistema e ao comportamento dos metadados	<ul style="list-style-type: none"> i. Documentação sobre o hardware e software ii. Informações referentes à digitalização, formatos, compressão, rotinas de escalonamento. iii. Registro do tempo de resposta do sistema iv. Autenticação de dados, senhas, criptografia
Metadados de Utilização	Relacionados ao nível e tipo de uso dos recursos	<ul style="list-style-type: none"> i. Registro de exibição ii. Registro dos usos e dos usuários iii. Reutilização do conteúdo e de versões

Fonte: Gilliland-Swetland (1997, p.5)

Na definição dos metadados, para representação descritiva, é preciso estar em conformidade com algum formato de padronização, a fim de garantir a interoperabilidade entre diferentes sistemas. Alguns formatos obtêm maior impacto com a introdução da WS, conforme Breitman (2010, p.17) são eles: Dublin Core, Framework e o RDF.

O padrão *Dublin Core* (DC) é destacado pela sua simplicidade na forma de descrever os recursos. Seu conjunto é composto de 15 elementos (BREITMAN, 2010, p.18):

- 1) Assunto (subject): Assunto do conteúdo do objeto
- 2) Título (title): Nome dado ao objeto que está sendo descrito
- 3) Criador (creator): Responsáveis pela produção do conteúdo do objeto
- 4) Descrição (description): descrição do conteúdo do objeto
- 5) Editor (publisher): Agente, agência, entidade ou editora responsável por disponibilizar o objeto
- 6) Outro agente (contributor): Outras pessoas e/ou entidades que fizeram contribuições para o objeto
- 7) Data (date): Data de publicação
- 8) Tipo de Objeto (type): Inclui termos que descrevem as categorias/funções que o objeto está relacionado, ex.: ficção, novela, poema ou dicionário
- 9) Formato (format): Formato físico do objeto, ex.: texto o PDF.
- 10) Identificador (identifier): Referência (não ambígua) para identificar aquele recurso, ex.: número de identificação.
- 11) Relacionamento (relation): Direcionamento para outros objetos que podem ser relacionados
- 12) Fonte (source): Outros Objetos dos quais este foi originado (se aplicável)
- 13) Linguagem (language): Idioma do conteúdo
- 14) Cobertura (coverage): Extensão ou escopo do conteúdo
- 15) Direitos (rights): Direitos de uso do conteúdo do objeto

A iniciativa *Dublin Core Metadada Initiative* (DCMI) é formada por integrantes que se dedicam à adoção de padrões de metadados e à construção de vocabulários especializados para a descrição de recursos, com a finalidade de expandir a RI. Nenhum dos padrões de metadados DC é de uso obrigatório, entretanto, é comum a adoção de um conjunto mínimo de elementos, que seja capaz de descrever objeto, fornecendo dados suficientes para RI (DCMI, 2007).

O padrão *Framework de Warwick* (FW) surgiu da necessidade de ampliar o padrão DC, considerado muito simples, devido a disponibilização de apenas um formato para a descrição dos objetos. Portanto, nesse padrão, não foram modificados os padrões básicos do DC, mas implementados outros com termos, condições e responsabilidades (BREITMAN, 2010, p.19).

Os elementos adicionados no padrão FW, conforme Breitman (2010) foram:

- 1) Descrições específicas do domínio do documento (objeto);
- 2) Termos e condições de uso do documento;
- 3) Rótulos e gradação do documento;
- 4) Informações de segurança. Autenticidade, assinaturas;
- 5) Origem do fornecedor;
- 6) Conjunto de *containers* (engloba todos os elementos presentes) para documentos compostos e ponteiros para todas as manifestações e instâncias ou versões do documento;
- 7) Responsável por armazenar o documento, além do;
- 8) Conjunto de descritores do Dublin Core do documento.

No padrão FW, assim como no padrão DC, não é necessário que todos os elementos estejam presentes na descrição do documento. A escolha dos metadados que irão compor a descrição deve ser estabelecida conforme o contexto em que se aplica, mas precisa englobar os padrões básicos de descrição do padrão DC (BREITMAN, 2010, p.20).

O *Resource Description Framework* (RDF), pode ser entendido como uma implementação do padrão FW, pois propõe algumas soluções para problemas evidenciados nesse último. Para Hernández-Pérez (2013, p.202) o RDF pode ser definido como “um modelo de dados que se baseia na ideia de fazer declarações sobre os recursos (em especial, recursos web) em forma de expressões sujeito-predicado-objeto”. Os objetos descritos pelo RDF podem ser de qualquer formato, incluindo textos, vídeos imagens, etc., entretanto, é necessário que estejam em um ambiente digital e conttenham um identificador universal (URI ou URL, por exemplo).

De acordo com Breitman (2010, p.20) o RDF é uma linguagem declarativa, que possibilita por meio a representação dos metadados no formato de sentenças entre propriedades e relacionamentos entre os itens da web. A autora ainda acrescenta, que uns dos objetivos do RDF é incluir semântica nos recursos da web e torná-los acessíveis por máquinas. A ideia desse padrão é representar os conceitos por meio de triplas, formadas por sujeito, propriedade (ou predicado) e objeto.

O W3C⁷ apresenta ainda o RDF Schema (RDFS) sendo uma extensão semântica do RDF, referido muitas vezes como *RDF Vocabulary Description Language*. O RDFS permite a definição de classes (todos os tipos de dados são classes) e a descrição de recursos

⁷ Disponível em: http://www.w3.org/TR/2014/REC-rdf-schema-20140225/#ch_introduction. Acesso em: 01/07/2016

relacionados e as relações entre eles. Além de possibilitar estabelecer propriedades e suas restrições, hierarquia entre classes (subclasses e superclasses) e propriedades (subpropriedades e superpropriedades).

Resumindo, o RDF (S) é um modelo simples de dados para modelagem semântica. Possibilita a descrição dos recursos por meio das suas classes, propriedades e valores. Usa um vocabulário formado por URI e sintaxes XML.

Outro formato para descrição semântica é o *Web Ontology Language* (OWL). Assim como o RDF pode ser considerado uma evolução do padrão FW. O OWL foi desenvolvido para estruturas de classes e propriedades mais complexas que o RDF (S), sendo resultante de uma combinação de OIL e DAML, uma linguagem de marcação semântica que utiliza RDF (S) como base, porém, possui mais vocabulário para descrever as classes, suas características, relacionamentos e instâncias.

De acordo com o WRC⁸, o OWL não é uma linguagem de programação, é declarativa, sendo responsável por descrever o “estado das coisas” de uma maneira lógica. Conforme (SCHIESSL, 2015) o OWL possui uma estrutura que permite descrever os objetos que se deseja representar, com possibilidade de incluí-las em categorias ou dizer algo sobre suas relações. Para entender como o conhecimento é representado no OWL, são destacadas algumas noções básicas (MOREIRO GONZÁLEZ, 2011, p.77, 78):

- a) **Classes:** são aspectos gerais, que unem os conceitos do domínio;
- b) **Conceitos:** ideias básicas que se tenta formalizar, podem envolver classes de objetos, métodos, planos, estratégias, processos de raciocínio, etc.;
- c) **Relações:** representam a ligação entre os conceitos do domínio;
- d) **Axiomas:** declarações básicas de uma ontologia OWL. São afirmações verdadeiras sobre os conceitos;
- e) **Entidades:** elementos utilizados para se referir a objetos do mundo real;
- f) **Expressões:** descrições de entidades para formar combinações desde as mais básicas, até as mais complexas
- g) **Instâncias:** representam objetos de um determinado conceito;

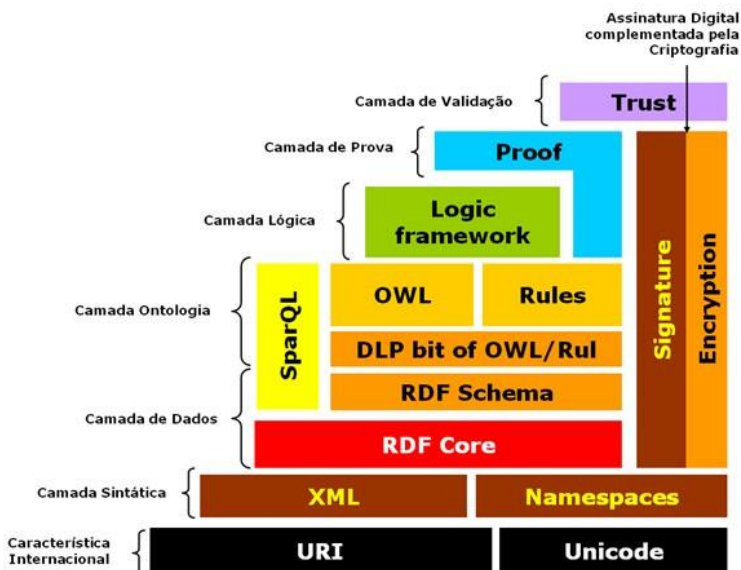
⁸Disponível em: http://www.w3.org/TR/owl2-primer/#What_is_OWL_2.3F
Acesso em: 01/07/2016

- h) **Propriedades:** ou atributos que as classes, conceitos ou instâncias podem ter;

Para recuperar e manipular dados modelados em uma ontologia, existem as linguagens das consultas, como o SPARQL *Protocol and RDF Query Language*. O SPARQL é baseado no modelo de triplas (sujeito-predicato e objeto) e é capaz de retornar resultados conforme as consultas incluídas no seu protocolo de acesso (BREITMAN et al. ([20--]; W3C, 2013).

Deste modo, de acordo com Santos e Alves (2009) a arquitetura da WS, fornece um conjunto de várias ferramentas tecnológicas, visando uma melhor estruturação e representação dos dados. Essa estrutura pode ser observada conforme a arquitetura da WS proposta por Berners-Lee (2005), apresentado na figura 9:

Figura 9 - Arquitetura da Web Semântica proposta em 2005



Fonte: Berners-Lee (2005); Santos e Alves (2009)

Em suma, a aplicação da WS na web, torna os conteúdos mais inteligentes, reduzindo a curva de aprendizado dos usuários. Este breve

resumo apresentou uma noção geral das tecnologias que fazem parte da WS, embora, nesta pesquisa, não sejam utilizadas todas as tecnologias da WS, constatou-se a importância de se compreender a abrangência do tema. Para desenvolver esta pesquisa serão utilizados além do editor de ontologias Protegé, a linguagem descritiva OWL, o formato RDF e a linguagem de consulta SPARQL.

4 PROCEDIMENTOS METODOLÓGICOS

Este capítulo discute os caminhos, os métodos e os recursos utilizados para o desenvolvimento da pesquisa. Para Pádua (2007, p.37) nos procedimentos/técnicas que dão suporte à pesquisa, é necessário identificar o tipo da pesquisa, reconhecer os diferentes aspectos, enfoques, níveis de aprofundamento e recursos que a pesquisa pode admitir. A direção a ser seguida dependerá dos objetivos almejados e da capacidade do pesquisador em desenvolvê-los.

Segundo Silva e Menezes (2005, p.28), levando em consideração que existe mais de uma maneira de raciocínio que pode ser aplicado nas investigações científicas, é ideal também que mais de um método seja utilizado na pesquisa, conseqüentemente, as possibilidades de análise e obtenção dos resultados serão ampliadas.

Nesse sentido, a escolha dos métodos define como será o caminho de investigação para a consecução da pesquisa. Como forma de traçar a tipologia da pesquisa, optou-se pelo uso dos critérios definidos por Lakatos e Marconi (1991) classificados como: (i) objetivos da pesquisa; (ii) natureza da pesquisa; (iii) procedimentos adotados; e (iv) forma de abordagem do problema de pesquisa.

A pesquisa é caracterizada quanto aos objetivos como aplicada, haja visto a sua orientação para a resolução de problemas voltados à aplicação prática. Quanto aos procedimentos técnicos caracteriza-se como uma pesquisa exploratória e descritiva.

De acordo com Gil (2010) uma pesquisa exploratória consiste em numa visão geral aproximativa de um determinado assunto, com análises de exemplos que estimulem a compreensão do problema. Envolve levantamento bibliográfico e documental, entrevistas não padronizadas ou estudos de caso.

Já a pesquisa descritiva, para o mesmo autor têm como propósito “a descrição das características de determinada população” (GIL, 2010, p.27). Este tipo de pesquisa irá descrever um fato ou fenômeno após a primeira aproximação, que acontece por meio da pesquisa exploratória (SANTOS, 2000).

Quanto à natureza, Lakatos e Marconi (1991) afirmam que uma pesquisa pode equivaler-se a um trabalho científico original ou em um resumo de assunto. O primeiro está relacionado a uma pesquisa inédita, que contribua para a geração de novas pesquisas para evolução do conhecimento científico. O resumo é um trabalho que dispensa a originalidade, mas ainda assim exige todo o rigor científico compreendido

em uma pesquisa. Com base nesses critérios, supõe-se que este trabalho, mesmo tratando-se de uma pesquisa de dissertação, que não exige ineditismo, constitui-se de um trabalho original, pois corresponde ao desenvolvimento de uma ontologia do futebol, que considere as características dos acervos digitais de jornais impressos, no que dizem respeito principalmente a linguagem adotadas nos textos jornalísticos e provindos da evolução histórica e que esses fatores, unificados, contribuam para o aprimoramento da recuperação dos acervos digitais impressos.

Em relação aos procedimentos adotados, Lakatos e Marconi (1991) classificam-nos em “de campo”, que são coletas feitas diretamente no local de ocorrência das pesquisas, e de “fonte de papel”, que envolvem a pesquisa bibliográfica e documental. A pesquisa em questão é bibliográfica e adotou-se como fonte principal o portal da Capes, por este conter inúmeras outras bases das mais variadas áreas do conhecimento que disponibilizam inúmeros estudos que deram subsídios para o desenvolvimento da pesquisa. De acordo com e Lakatos e Marconi (2012, p.57), “a pesquisa bibliográfica, ou de fontes secundárias, abrange toda bibliografia já tornada pública em relação ao tema de estudo, desde publicações avulsas, boletins, jornais, revistas, livros, pesquisas, monografias, teses, material cartográfico, etc.”. Também, devido aos objetivos da pesquisa, foram consultados dicionários e glossários especializados sobre o futebol para a formação do corpus de termos e a conceitualização do conhecimento, além do uso do Acervo Estadão⁹ utilizado para coleta das informações e demonstração da recuperação da informação com a implementação da ontologia.

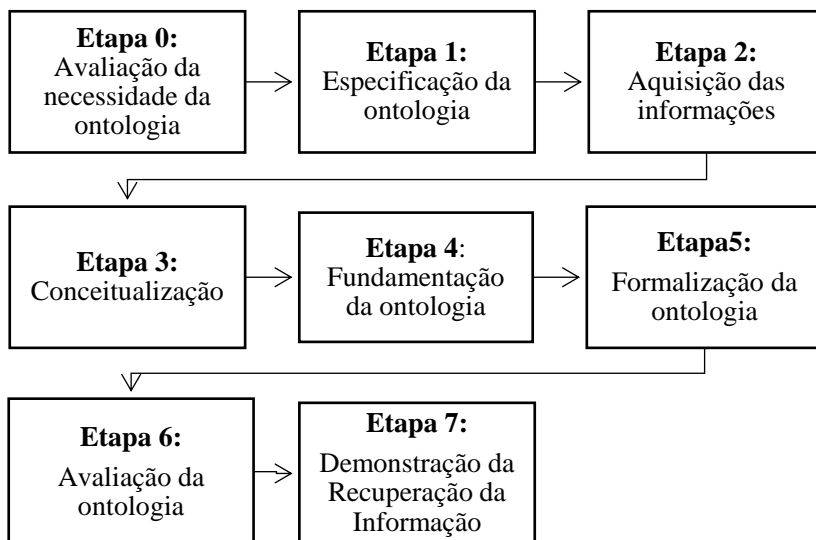
Quanto à forma de abordagem a pesquisa é qualitativa. Para Triviños (2010, p. 125), a pesquisa qualitativa possibilita ao pesquisador compreender o “contexto do fenômeno social que se estuda, privilegia a prática e o propósito transformador do conhecimento que se adquire da realidade que se procura desvendar em seus aspectos essenciais e acidentais”. Tal pesquisa é construída essencialmente com base em análises de materiais de referência do domínio, o estudo de acervos digitais, objeto desta pesquisa, além de considerações interpretativas feitas por este autor.

⁹ Acervo Estadão: Disponível em: <http://acervo.estadao.com.br/>. Acesso em: 21/01/2016.

4.1 Descrição geral da metodologia de pesquisa

Nas etapas consideradas para o alcance dos resultados desta pesquisa que é a construção da ontologia do futebol para uso em acervos digitais de jornais, será utilizada a metodologia OntoForInfoScience, proposta por Mendonça (2015). As etapas consideradas são: Avaliação da necessidade da ontologia; Especificação da ontologia; Aquisição de Informações; Conceitualização; Fundamentação da ontologia, Formalização da ontologia e Avaliação da Ontologia. As etapas que envolvem a documentação e disponibilização da ontologia, também contempladas na OntoForInfoScience, não serão consideradas nesta pesquisa, devido ao curto período disposto para consecução do trabalho. A etapa 7 – Demonstração da recuperação da informação, não faz parte da metodologia proposta por Mendonça (2015), porém, faz parte das etapas consideradas nesta pesquisa. A ontologia desenvolvida na pesquisa será nomeada de *OntoFootballForNewspapers*. A seguir são mostradas as etapas envolvidas neste trabalho, desde os procedimentos e artefatos considerados na construção da ontologia, assim como as etapas que envolvem a sua validação e a demonstração do seu uso no acervo digital:

Figura 10 - Etapas que serão desenvolvidas nesta pesquisa



Fonte: Elaborado pela autora.

4.1.1 Avaliação da necessidade da ontologia

Nesta etapa é feita a avaliação da necessidade de construir a ontologia. Assim, antes de iniciar as etapas propostas *OntoForInfoScience* deve ser respondida a seguinte questão: - “O projeto a ser desenvolvido e seu contexto necessitam ou demandam a construção de uma ontologia? Ou a criação de outro instrumento de representação, tal como um tesouro, seria suficiente?” (Mendonça, 2015)

Identificado a necessidade de desenvolver a ontologia, tem se a sequência de etapas do processo de construção da ontologia, em caso negativo, recomenda-se o uso de outros instrumentos terminológicos, como tesouros, taxonomias e vocabulários controlados. Para responder à questão que orienta esta etapa, realizou-se pesquisas bibliográficas com objetivo de justificar a escolha.

4.1.2 Especificação da ontologia

Nesta etapa são definidos o propósito geral e os objetivos específicos da ontologia, tipo da ontologia, escopo geral, possibilidades de uso, público-alvo, grau de formalidade e a delimitação da cobertura, em que são elaboradas as questões de competência que a ontologia deve responder e que serão utilizadas também na etapa de avaliação da ontologia. A ontologia desenvolvida é denominada de *OntoFootballForNewspapers*. As informações foram organizadas e detalhadas conforme o quadro abaixo (quadro 7), proposto por Mendonça (2015):

Quadro 7 - *Template* de especificação da ontologia *OntoFootballForNewspaper*

Domínio/Escopo Geral
Especificar o domínio do conhecimento que a ontologia irá representar, detalhando quais partes deste domínio serão tratadas na ontologia e quais não serão, isto é, delimitar o escopo geral de cobertura da ontologia desenvolvida.
Propósito Geral
Explicar o motivo e o objetivo da construção da ontologia, ou seja, porquê e para que a ontologia em questão será desenvolvida.
Classes de usuários
Especificar todos os possíveis usuários e os seus objetivos em relação à ontologia.
Uso pretendido

Explicar o cenário de aplicação para o uso da ontologia. Apontar os problemas no ambiente atual que a ontologia aplicada visa resolver.
Tipo da ontologia
Classificar a ontologia desenvolvida quanto à: (i) estrutura ou nível de conhecimento: ontologia de alto nível, ontologia de domínio ou ontologia de tarefa; (ii) grau de formalidade: ontologia leve, ontologia com médio rigor formal ou ontologia pesada; (iii) propósito de criação (opcional): ontologia de sistemas de informação ou ontologia para sistemas de informação. Tal classificação é útil na definição do grau de formalidade da ontologia, em suas classes de usuários e seu uso pretendido
Grau de formalidade
Definir o grau de formalidade (pesada, leve ou médio rigor formal) requerido para a ontologia com base em seu propósito geral, cenários de uso ou aplicação, classes de usuários e tipo da ontologia.
Delimitação do escopo de cobertura
<p>A fim de assegurar que a delimitação do escopo de cobertura da ontologia seja clara e explícita para todas as classes de usuários, recomenda-se ao desenvolvedor especificar os seguintes itens, considerados complementares e indispensáveis nessa tarefa:</p> <ul style="list-style-type: none"> - Ponto de partida da ontologia: corresponde ao conjunto de classes iniciais a partir das quais a ontologia começa a ser desenvolvida. Geralmente, esse conjunto deve ser formado por um conjunto de classes mais gerais reaproveitadas de alguma ontologia de fundamentação ou alto nível e mais uma ou duas classes específicas do domínio tratado. - Limite do domínio coberto: corresponde ao nível de conhecimento máximo do domínio que será representado na ontologia. - Questões de competência: é o método mais comum utilizado em metodologias de construção para auxiliar na definição do escopo de cobertura de modo a especificar os requisitos que a ontologia deverá ser capaz de atender e os problemas que a ontologia pode solucionar antes mesmo de sua construção. O desenvolvedor, portanto, deve elaborar um conjunto de questões de competência que sejam anteriores às etapas propriamente de construção de ontologias de forma que elas deixem claro o propósito geral e os objetivos específicos que a ontologia é capaz de atender. Tais questões de competência podem ser definidas tanto em linguagem natural quanto em uma linguagem semiformal ou ambas.

Fonte: Adaptado de Mendonça (2015)

4.1.3 Aquisição de Informações

Nesta etapa são coletadas todas as informações modeladas na ontologia. Mendonça (2015) define essa etapa como Aquisição do Conhecimento, no entanto, por entendermos que nessa etapa as informações ainda não são tratadas de maneira a obter conhecimento, mas são coletadas por conveniência afim de obter a amostra da pesquisa, optamos por utilizar Aquisição de Informações.

Existem algumas técnicas comuns da área ontológica que são utilizadas para captura do conhecimento, são elas:

i) *análise informal de textos* em materiais de referência do domínio: permite o estudo e entendimento dos conceitos nos materiais utilizados;

(ii) *análise formal de textos* nos documentos de referência do domínio: consiste na identificação de estruturas textuais do domínio, tais como definição e afirmação, e o tipo de conhecimento que tais estruturas podem representar na ontologia: conceitos (classes), propriedades, instâncias, relações, etc.

(iii) *entrevistas estruturadas, semi-estruturadas ou não-estruturadas* com especialistas da área: possibilitam o entendimento de conceitos do domínio por parte do desenvolvedor e a construção de árvores de classificação de conceitos do domínio.

(iv) *brainstorming ou grupos focais* com especialistas da área: é uma técnica mais colaborativa que a realização de entrevistas para a compreensão de conceitos do domínio, porém pode se tornar ineficiente quando não se obtém consenso sobre os termos da área, necessitando de intervenção contínua do mediador desta atividade.

(v) *emprego de técnicas estatísticas e/ou métodos linguísticos* que possibilitem a extração de conceitos e relações relevantes do domínio a partir

da análise dos textos dos materiais de referência utilizados.

(vi) *uso de ferramentas automatizadas* que realizam a extração de conceitos e relações relevantes do domínio. De maneira geral, tais ferramentas são usadas a partir de uma abordagem semi-automática, tal que o desenvolvedor ou especialista do domínio avaliem e validem os termos candidatos extraídos de maneira automática.

Como primeiro passo são coletados os conceitos que serão considerados no domínio desta ontologia, essa coleta será organizada em um glossário de conceitos. Na formação deste artefato optou-se pela coleta das informações em materiais de referência do futebol, sendo eles: i) dicionários e glossários específicos do futebol; ii) teses; iii) notícias jornalísticas.

A amostra dos termos formada é não probabilística, por conveniência, regida pelos critérios estabelecidos na pesquisa, que envolvem: a) termos sinônimos, polissêmicos e homógrafos; b) termos com variação diacrônica e sincrônica; que envolvem também - c) termos estrangeiros que tiveram aportuguesamento. Essas questões podem ser compreendidas também da seguinte forma:

- a) Termos do futebol que eram de origem inglesa e tiveram tradução para o português no período de existência dos jornais que compõem os acervos digitais;
 - a. Exemplo: Partida – *Match*.
- b) Expressões do futebol que possuem interpretação de duplo sentido;
 - a. Exemplo: Chegada (falta).
- c) Termos aplicados ao futebol que tiveram alteração na grafia no período de existência dos jornais que compõem os acervos digitais;
 - a. Exemplo: Troféu (Tropheu).
- d) Termos do futebol que possuem outro significado em outros domínios;
 - a. Exemplo: Caneco (taça), chocolate (goleada).
- e) Termos do futebol que possuem variação conforme a Região do Brasil em que estão inseridos.
 - a. Exemplo: Goleada (vareio);

Cronologicamente, o recorte estabelecido abrangeu terminologia utilizada entre 1900 a 2015, período que contempla a existência dos clubes de futebol no Brasil. A coleta foi realizada em jornais digitalizados e também os nascidos digitais. O acervo utilizado como base para estudo nessa pesquisa é o Acervo Estadão. Apesar do acervo ter publicações desde 1875 e o futebol no Brasil ter seu surgimento no ano de 1895 (GUTERMAN, 2009), nas consultas no acervo, por meio dos termos “*football*”, e “*futebol*” não foram localizados documentos anteriores a década de 1900, do século XX.

Existem outros acervos de jornais disponíveis para pesquisas na Internet, no entanto, a escolha do Acervo Estadão justifica-se por este contemplar todo o período de existência do futebol no Brasil. Também pela facilidade de consulta das informações no sistema e pelo fato do acervo ser o único disponível na Internet, que sinaliza dentro do texto recuperado, as palavras-chave incluídas na busca. Esse item facilita a coleta dos textos para análise, na medida que os documentos dos jornais são constituídos com páginas que possuem diversas notícias, sobre variados assuntos, assim sendo, quando o termo solicitado na estratégia de busca é destacado dentro do texto, é possível dirigir-se diretamente para o assunto solicitado na busca. As pesquisas no acervo são de acesso livre, no entanto, o acesso ao documento na íntegra é restrito a assinantes do jornal impresso ou digital. Considerando essa restrição, utilizamos uma conta de assinante para realizar esta pesquisa.

A opção de usar um acervo como base para pesquisa sucedeu-se em função da etapa de formação do corpus de pesquisa e também da etapa que demonstra o uso da ontologia em acervos digitais. Dessa maneira, o corpus de termos definido por conveniência, está sujeito a problemas que se mostraram pertinentes no acervo, durante o período de recorte da pesquisa. A coleta dos termos pela autora dessa pesquisa foi realizada de setembro a novembro de 2015.

A metodologia *OntoForInfoScience* sugere que a etapa de aquisição das informações, compreendida também como a “pré-conceitualização” da ontologia, ou seja, uma fase precedente a etapa de conceitualização, seja desenvolvida com a combinação de métodos diferentes na exploração das fontes – análise manual e extração automática dos termos. Para Mendonça (2015, p.196) a identificação dos termos no método manual é semelhante ao automático, nesse sentido “verbos podem indicar uma relação conceitual e nomes (substantivos, pronome, adjetivos) podem representar um conceito”. A grande diferença entre os dois, conforme Mendonça, é que o método manual tende a ser

mais criterioso, já o automático é mais produtivo, por processar um maior número de dados em um tempo relativamente menor.

No entanto, nesta pesquisa, devido aos acervos digitais de jornais serem compostos por textos digitalizados, históricos, com mais de 100 anos de existência, a utilização de ferramentas para coleta automática dos termos não obteve sucesso. Muitos textos não foram processados por *Optical Character Recognition* (OCR), que é uma tecnologia para reconhecer caracteres dentro do texto, sendo estes documentos imagens, manuscritos ou impressos, impossibilitando assim a coleta automática. Portanto, na formação do glossário de conceitos, optou-se por ser todo coletado em materiais de referências listados acima. Para o glossário foram levantados 56 termos, mas com o desenvolvimento da ontologia foi necessário expandir e o glossário alcançou 65 termos. Esta amostra consegue representar os problemas no contexto do domínio que queremos problematizar, entretanto, em uma ontologia implementada em um ambiente de produção, para tratar de problemas pertinentes ao ambiente, muito possivelmente esta amostra precisaria ser ampliada. Os termos foram coletados conforme o quadro abaixo:

Quadro 8 - Modelo de Coleta do Glossário de Conceitos

	Glossário de Conceitos
ID	Termo (conceito)
1	
2	
3	

Fonte: Adaptado de Mendonça (2015)

Posterior ao glossário de conceitos, a metodologia *OntoForInfoScience* sugere o desenvolvimento do glossário de verbos. Esta etapa foi desenvolvida com o reaproveitamento de alguns verbos utilizados ontologia sobre os componentes do sangue humano – Hemonto, desenvolvida como aplicação da metodologia *OntoForInfoScience* (Mendonça, 2015). Para localização de outros verbos mais específicos da ontologia do futebol, foi feita análise manual de textos recuperados no acervo digital Estadão. A seleção dos textos foi realizada a partir da década de 1900, e os termos de consulta aplicados nos acervos digitais para recuperar os textos foram “*football*”, e “futebol”. Para chegar aos textos para análise foram utilizadas categorias: a) notícias que possuíam um ou mais termos do glossário de conceitos - b) notícias que possuíam

um dos termos do glossário de conceitos no título ou chamada do texto. Para chegar nos textos analisados, após a consideração das categorias, foi feita seleção aleatória. Desta maneira, foram analisados 10 textos, a fim de extrair verbos que fizessem ligações entre os conceitos coletados. O seguinte modelo exposto no quadro 9 foi utilizado para a coleta.

Quadro 9 - Modelo de Coleta do Glossário de Verbos

Glossário de Verbos	
ID	Verbo
1	
2	
3	

Fonte: Adaptado de Mendonça (2015)

Ainda nesta etapa, a metodologia *OntoForInfoScience* sugere o desenvolvimento do Glossário de Relações, no entanto, na metodologia este momento é feito de forma automática e na etapa seguinte ele é avaliado de forma manual para a verificação das relações formadas. Como neste trabalho o desenvolvimento está sendo de forma manual, e considera uma reduzida amostra de conhecimento de domínio, optou-se por desenvolver este passo já na fase de conceitualização, visto que nesse instante já estamos analisando o conhecimento a fim de modelá-lo para o desenvolvimento da ontologia e na etapa de aquisição apenas coletamos o conhecimento sem inferências.

4.1.4 Conceitualização

Após a coleta das informações o que será considerado na modelagem do domínio, segue-se para a etapa de conceitualização, conforme Mendonça (2015), neste momento são feitas a organização e estruturação do conhecimento por meio de artefatos de representação que se transformarão em relações, propriedades e restrições de uma ontologia. Os artefatos propostos na *OntoForInfoScience* são: Dicionário de conceito; Tabela de conceitos e valores; Tabela de conceitos e propriedades; Dicionário de verbos, Conjunto de Relações e Modelos Conceituais Gráficos. Para formação desses artefatos foram utilizados os conjuntos obtidos na etapa anterior (glossários de conceitos e verbos), os materiais de referência do domínio utilizados na aquisição das

informações, etapa anterior da modelagem da ontologia e o conhecimento do pesquisador/desenvolvedor sobre o domínio do futebol. Como resultado final dessa etapa tem-se um modelo conceitual que engloba conceitos, suas relações e suas propriedades, de forma que esse modelo seja utilizado como base para a etapa de formalização da ontologia, em que todo esse conhecimento modelado em língua natural será ponto de partida para a construção da ontologia em linguagem lógica formal.

No primeiro passo da conceitualização foi construído o Dicionário de Conceitos, feito a partir do Glossário de Conceitos desenvolvido na etapa anterior. Neste estágio é associada a definição textual de cada termo e também a identificação de sinônimos, que de acordo com Mendonça (2015, p.199) é melhor realizar nessa etapa, do que na anterior “porque agora as informações do domínio passam a ser tratadas a nível semântico”. Os sinônimos e as definições referentes a cada conceito foram retirados de obras de referências do domínio (dicionários e glossários). O modelo utilizado para formação do Dicionário de Conceitos obedeceu a estrutura do quadro 10.

Quadro 10 - Modelo Dicionário de Conceitos

Dicionário de Conceitos			
ID	Conceito	Sinônimos	Definição
1			
2			
3			

Fonte: Adaptado de Mendonça (2015)

O próximo passo foi a construção da Tabela de Conceitos e Valores. Este artefato utiliza o Dicionário de Conceitos e acrescenta para cada entrada valores possíveis que o conceito pode assumir, que na formalização da ontologia podem se tornar instâncias ou classes. Segundo Mendonça (2015, p.201) os valores possíveis são:

[...] Exemplos, casos reais ou ocorrências de um determinado conceito do domínio. Pode-se afirmar também que os valores possíveis são equivalentes aos particulares de um determinado universal e que poderão se tornar instâncias das classes da ontologia. Em uma ontologia da árvore genealógica de uma determinada família, por exemplo, podemos ter como conceitos Avô, Avó, Pai, Mãe, Idade, Sexo e Cor, e como seus valores possíveis os seguintes: Avô = {João}, Avó = {Maria}, Pai =

{José}, Mãe = {Ana Maria}, Idade = {1 a 120 anos}, Sexo = {M, F} e Cor = {Branca, Negra, Parda, Mulata, Mestiça, Amarela}.

No domínio do futebol, por exemplo, em uma classe chamada “time”, há uma subclasse com os nomes dos times, e os valores nessa tabela, que posteriormente poderão se transformar em instância, são os nomes dos jogadores, os salários, entre outros. A entrada desses valores também interfere no nível de formalização da ontologia. Quanto mais instâncias são definidas em uma ontologia, mais específica ela se torna. É possível ainda que não sejam definidos valores para os conceitos, tornando a ontologia mais geral, no entanto, este passo se faz necessário para a análise da necessidade dessa inclusão e do olhar do domínio sobre esta perspectiva. O modelo utilizado para preenchimento dos valores é apresentado no quadro 11, seguir.

Quadro 11 - Modelo Tabela de Conceitos e Valores

Tabela de Conceitos e Valores				
ID	Conceito	Sinônimos	Definição	Valores
1				
2				
3				

Fonte: Adaptado de Mendonça (2015)

O próximo passo proposto da conceitualização da metodologia *OntoForInfoScience* é a Tabela de Conceitos e Propriedades. Nele, as propriedades específicas do conceito, que são características que os diferenciam de outros conceitos ou classes no domínio são definidas. Essas propriedades são herdadas dos trabalhos de Aristóteles, nomeadas como “*differentia*”, sendo elementos que completam o conceito ou qualidades próprias, como por exemplo gênero e família. (MENDONÇA, 2015). Chamadas de propriedades descritivas ou propriedades de anotação, elas são incluídas no campo anotações no momento de formalização da ontologia, diferente das propriedades de objeto e dados, entendidas nessa metodologia como propriedades de relações, definidas mais à frente. Neste trabalho optou-se pelo não desenvolvimento da Tabela de Conceitos e Propriedades, devido ao escopo que pretendemos resolver na ontologia e pelo nível de complexidade do levantamento dessas informações não ser necessário para apoiar o SRI, do ponto de vista que a RI vai exigir.

O próximo passo desta etapa é a criação do Dicionário de Verbos que utiliza o Glossário de Verbos e acrescenta em cada verbo, sinônimos, definição informal (descritiva) e exemplo de uso. Mendonça (2015, p. 205) aponta os critérios para identificação dos verbos no Dicionário de Verbos:

Essa definição para cada verbo auxilia também na identificação de verbos que não possuem o mesmo rótulo (nome) mas possuem o mesmo significado, ou seja, são verbos sinônimos nesse contexto. Tais verbos devem ser representados em uma única entrada do Dicionários de Verbos, com o rótulo mais comum (preferencial) anotado com seus sinônimos

A identificação dos sinônimos dos verbos que já tinham sido levantados na etapa de aquisição das informações utilizou ferramentas de busca de sinônimos para atribuir os verbos que desempenham o mesmo propósito quando relacionados com os conceitos. A definição (informal) é descrita pelo próprio autor. O quadro 12 apresenta o modelo utilizado para a formação do Dicionário de Verbos.

Quadro 12 - Dicionário de Verbos

Dicionário de Verbos			
ID	Verbo	Sinônimo	Definição
1			
2			
3			

Fonte: Adaptado de Mendonça (2015)

No passo seguinte é desenvolvido o Glossário de Relações. São utilizados os verbos coletados no Glossário de Conceitos e no Glossário de Verbos, sendo o “conjunto formado por todos os padrões de relacionamento extraídos dos textos na forma <nome1> <verbo> <nome2>, tal que os nomes podem ser substantivos, pronomes ou adjetivos e o verbo representa um relacionamento existente entre os nomes” (MENDONÇA, 2015, p.197). Diante disso, alguns critérios para formação desse artefato foram seguidos: i) os <nome1> e <nome2> só podem ser formados por conceitos presentes no Dicionário de Conceitos e estes devem estar contidos como rótulos preferenciais e não os seus sinônimos; ii) O item <verbo> só pode ser adicionado no

Glossário de Relações caso faça parte do Dicionário de Verbos como rótulo principal, caso seja sinônimo, ele deve ser substituído verbo preferencial. O modelo do quadro 13 foi utilizado para a formação do glossário de relações.

Quadro 13 - Modelo do Glossário de Relações

Glossário das Relações			
ID	Termo 1	Verbo	Termo 2
1			
2			
3			

Fonte: Adaptado de Mendonça (2015)

Por fim, na etapa de conceitualização, são gerados os modelos conceituais. Essa fase utiliza os artefatos gerados anteriormente como base para representar o domínio através de estruturas de representação gráficas, como taxonomias, paronômias, grafos conceituais, ou estruturas de representação gerais ou mistas. Nesta pesquisa optou-se pelo desenvolvimento de estruturas de representações mistas, geradas a partir do Glossário de Conceitos e Relações e o Dicionário de Verbos. A ferramenta utilizada foi a *violetumeditor-2.1.0*, que permite o desenvolvimento de estruturas de representação gráfica.

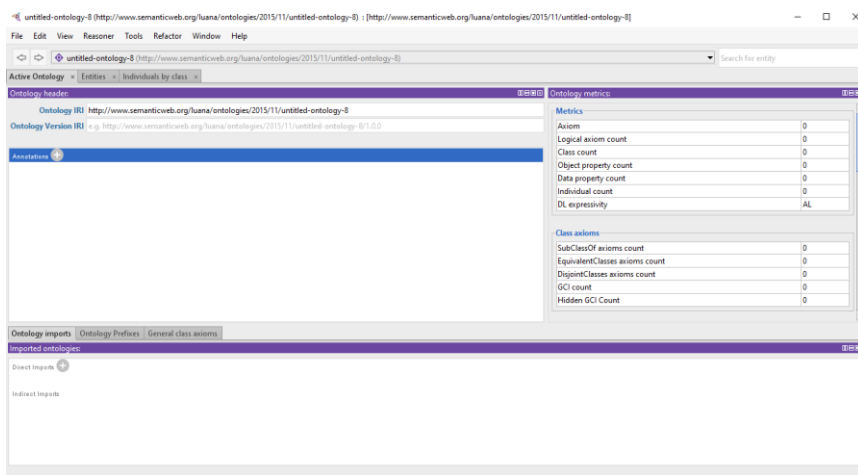
4.1.5 Fundamentação ontológica

No decorrer da pesquisa foi apresentada a importância da implementação de ontologias de fundamentação no desenvolvimento de ontologias de domínio. As ontologias de fundamentação seguem uma corrente teórica-filosófica, portanto, ao definir qual ontologia de fundamentação será utilizada na pesquisa, adota-se essa corrente filosófica. O primeiro passo dessa etapa é pesquisar na literatura, ontologias de fundamentação relevantes que se encaixem nos propósitos de construção desta ontologia de domínio. A *Unified Foundational Ontology* (UFO) foi identificada, as justificativas de escolhas estão apresentadas nos resultados da pesquisa, que descreve o reuso da ontologia de fundamentação (item 5.1.4).

Após a escolha da ontologia de fundamentação que será utilizada, é necessário fazer uma pesquisa na Web, para verificar se o arquivo OWL

da ontologia está disponível, e assim fazer o seu carregamento em algum editor de ontologia. Ao importar a ontologia de fundamentação para o Protegé o desenvolvedor terá a sua disposição a estrutura base da ontologia, composta por “classes (aba “classes”), suas relações (aba “object properties”) e as propriedades de classe e relação encontradas em forma de campos descritivos e também sob a forma de axiomas lógicos dentro de cada classe e relação” (MENDONÇA, 2015, p.213). Nesta pesquisa utiliza-se o Protegé 5.0. Nesta versão na aba *Active ontology*, aparece a opção *Ontology imports* e dentro dela a opção *Import Direct* que permite a importação para o reuso de ontologias. A figura 11 mostra a tela de acesso para importação de ontologias pelo Protegé.

Figura 11 - Importar ontologia de fundamentação no Protegé 5.0



Fonte: Elaborado pela autora.

Foi realizada pesquisa exaustiva na Internet, a fim de recuperar o arquivo OWL da ontologia para que fosse importado no editor Protegé, e assim, a partir da estrutura base adquirida fosse possível dar início à etapa de formalização. Entretanto, a pesquisa não obteve sucesso e o arquivo não foi localizado. De acordo com Mendonça (2015, p.214), a etapa de fundamentação ontológica não se limita apenas à importação do arquivo da ontologia de fundamentação em um editor. Para seguir o desenvolvimento da ontologia, é preciso manter o compromisso ontológico com os princípios de fundamentação para todo processo de desenvolvimento da ontologia.

Seguidos estes princípios, é assegurado e justificado o uso da ontologia de fundamentação. Assim, optou-se por utilizar a abordagem filosófica da UFO como forma de justificar algumas decisões ontológicas adotadas no desenvolvimento da ontologia.

4.1.6 Formalização da ontologia

Na era da WS uma das principais características das ontologias é a capacidade de serem interpretadas por máquinas. Nessa etapa de formalização, o conhecimento tratado anteriormente em nível conceitual, passa a ser tratado em nível ontológico-formal. De acordo com Andrade, Ferreira e Pereira (2010), a finalidade da formalização pode ser justificada pois a “linguagem natural muitas vezes introduz incertezas e ambiguidades. Por isso, a necessidade de utilização de uma linguagem formal, que, por meio de seus símbolos não ambíguos e formulações exatas, conduza a uma maior clareza e correção nas deduções”.

Portanto, nessa fase as estruturas conceituais da etapa de conceitualização são adaptadas de forma que atendam às necessidades ontológicas e formais da ontologia. Para tal, alguns itens relativos a ontologias são estabelecidos, como classes, sub-classes, instâncias e propriedades. Para desenvolver esta etapa foi utilizado o editor de ontologias Protégé 5.0.

As etapas consideradas para formalização atenderam a seguinte disposição:

- a) Definir classes da ontologia: nesse passo são definidas as classes gerais da ontologia. Aqui devem ser levados em consideração a representação do conhecimento conforme proposto na ontologia de fundamentação determinada na etapa anterior. Pode ser considerado também o reuso de classes já definidas em outras ontologias de domínio aplicadas ao futebol.
- b) Introduzir os conceitos nas classes: nesse passo os conceitos levantados na etapa de aquisição das informações e estruturados no glossário de conceitos são distribuídos nas classes que foram definidas. Aqui podem ser definidas subclasses para a organização dos conceitos em uma taxonomia geral da ontologia.
- c) Introduzir os sinônimos: o levantamento dos termos sinônimos foi feito no Dicionário de Conceitos desenvolvido na etapa de conceitualização deste trabalho. Nessa etapa eles devem ser associados no Protégé.

- d) Definir as classes disjuntas: as classes disjuntas são aquelas que não possuem indivíduos em comum.
- e) Criar instâncias: a definição de instâncias geralmente é um dos últimos passos a ser desenvolvido na ontologia, elas não são obrigatórias, e conforme a sua adesão irão definir o quão específica é a ontologia. A Tabela de Conceitos e Valores desenvolvida na etapa de conceitualização deste trabalho deve ser utilizada como ponto de partida para identificar que conceitos possuem instâncias.
- f) Definir propriedades de relação das classes: nesse passo são identificados, de forma mais aproximada, as relações das classes do domínio, conforme as ligações apresentadas no Glossário de Relações. Essas propriedades são nomeadas como propriedades de objeto ou dados.

Não foram definidos axiomas nesta ontologia por entender que as consultas são terminológicas, e não serão obtidos resultados por meio dos axiomas em um ambiente como os acervos digitais de jornais. Sendo assim, optou-se pelo mapeamento dos conceitos e a definição de relações, que serão capazes de suprir e absorver o conhecimento necessário para retornar resultados que irão de fato ser úteis ao usuário, que é o foco da modelagem desta ontologia.

4.1.7 Avaliação da Ontologia

Após o desenvolvimento da ontologia é feita a avaliação do conteúdo ontológico. A metodologia da *OntoForInfoScience* propõe um conjunto de parâmetros e critérios avaliativos, no entanto, devido ao prazo para o desenvolvimento desta pesquisa de mestrado, dentre os critérios estabelecidos na metodologia *OntoForInfoScience*, optou-se pelo uso do critério avaliativo que verifica se todas as questões de competência estabelecidas no item da especificação da ontologia foram respondidas. Este critério, conforme Mendonça (2015, p.246) é um “forte indicativo de que a ontologia foi capaz de apreender o conhecimento no domínio explorado.”

As questões de competência são desenvolvidas em linguagem natural na etapa de especificação da ontologia. Nesta etapa, de avaliação do conteúdo ontológico, elas serão elucidadas com o auxílio de *reasoners* (raciocinadores de inferência). Essas ferramentas permitem fazer consultas, por meio de regras lógicas. De acordo com Filho, Lóscio e De

Campos (2008, p. 4) os raciocinadores ou motores de inferência, são capazes de ratificar a consistência de ontologias:

Os motores de inferência ou raciocinadores são ferramentas de software que mapeiam uma base de conhecimentos existente (coleção de conceitos e relações entre estes conceitos, fatos e regras), inferindo conhecimentos adicionais e mostrando informações implícitas. Como exemplo de inferência, podemos citar a classificação (computação de todas as classes as quais um determinado indivíduo pertence) e a realização (encontrar as classes mais específicas no qual um indivíduo pertence) de indivíduos. Além da inferência de informações, os raciocinadores têm a capacidade de responder consultas em cima do conhecimento provido e inferido, bem como checagem de consistência de ontologias.

Para esta etapa será utilizado o raciocinador *SPARQL Protocol and RDF Query Language*. Ele está disponível junto com as ferramentas do software Protégé 5.0. O SPARQL é uma linguagem de consulta que possibilita o acesso aos dados modelados no editor Protégé.

4.2 Demonstração do uso da ontologia na recuperação da informação em acervos digitais

O desenvolvimento da ontologia orienta-se para sua importância para a recuperação da informação em acervos digitais de jornais. Sendo assim, para comprovar a aplicabilidade do seu uso, compara-se de forma demonstrativa e qualitativa, quais os resultados obtidos nos acervos digitais sem a ontologia, e quais seriam os prováveis resultados obtidos se ela fosse aplicada.

Essa etapa é realizada no Acervo Estadão, disponibilizado online e composto por todas as edições já impressas do jornal O Estado de São Paulo desde 1875. No ambiente é possível pesquisar os textos por palavras-chave, ou data, subdividida em ano, mês e dia. Ainda é possível filtrar a busca por cadernos, como econômico, esporte, político, agrícola, entre outros.

Para mensurar o uso da ontologia, foram incluídos no buscador do acervo os termos empregados na *OntoFootballForNewspapers*. Com os documentos recuperados foi feita uma análise, com base nos relacionamentos formados na *OntoFootballForNewspapers*, para mensurar demonstrativamente qual seria o resultado obtido se ela fosse utilizada no sistema de recuperação no acervo.

Para apoiar esta etapa foram elaboradas três estratégias de busca utilizadas como ponto de partida para análise dos relacionamentos *OntoFootballForNewspapers* na base de documentos disponibilizada pelo Acervo Estadão, são elas:

1. Campeões no futebol no Brasil.
2. Goleadas no futebol brasileiro.
3. Zagueiros artilheiros no futebol brasileiro.

Por motivos de tempo e escopo, não foi possível desenvolver um sistema informatizado para a implementação da *OntoFootballForNewspapers* para a validação da ontologia no acervo. No entanto, como o propósito desta pesquisa é atestar a eficácia do uso da ontologia em acervos digitais de jornais, optou-se pela análise demonstrativa do seu uso com problemas reais levantados no acervo do Estadão.

5 APRESENTAÇÃO DOS RESULTADOS

A apresentação dos resultados foi dividida em duas partes, a primeira está direcionada o desenvolvimento da ontologia, sendo a modelagem conceitual, a construção da ontologia no editor de ontologias Protegé, assim como a avaliação por meio da elucidação das questões de competência. A segunda parte contempla a demonstração no acervo digital do uso da ontologia para aprimorar a recuperação da informação desses ambientes.

5.1 Desenvolvimento da *OntoFootballForNewspapers*

Antes de iniciar as etapas de desenvolvimento da ontologia, é preciso avaliar a sua necessidade de construção. Mendonça (2015, p.187) propõe a resolução da seguinte questão, a fim do desenvolvedor compreender o real objetivo do seu projeto e identificar o instrumento de organização e representação da informação mais adequado para a sua demanda: “O projeto a ser desenvolvido e seu contexto necessitam ou demandam a construção de uma ontologia? Ou a criação de outro instrumento de representação, tal como um tesouro, seria suficiente? ”

Primeiramente é preciso definir o que é tesouro e ontologia. Para Arano (2005) um tesouro é uma ferramenta documental utilizado para representar e recuperar informação. Sua estrutura conceitual permite uma organização mediante o estabelecimento das relações conceituais e as restrições do significado e dos termos que os representam. Suas relações conceituais podem ser do tipo hierárquicas, associativas e de equivalência. Os tesouros podem ser utilizados tanto pelos profissionais da informação responsáveis por organizar a informação nos sistemas de informação, quanto pelos usuários que buscam a informação, ou seja, os usuários finais.

A ontologia, já discutida nos capítulos anteriores, mas abordada aqui brevemente para elucidar sua definição, é uma representação formal e explícita de uma estrutura conceitual de um campo do conhecimento. É um suporte semântico das palavras descritas como objetos linguísticos em uma base de dados lexicais ou terminológicos. Suas relações conceituais podem ser representadas de forma variada, e levam em consideração o campo do conhecimento que fazem parte. Ao ser formalizada (por meio de linguagens lógicas), pode ser interpretada tanto por pessoas, como por programas de computador (ARANO, 2005). Para Boccato, Ramalho e

Fujita (2008), as ontologias precisam “descrever formalmente recursos informacionais disponíveis em ambientes digitais, a partir da utilização de linguagens lógicas que possibilitem a realização de inferências automáticas por programas computacionais”.

Tesouro e ontologia, de uma maneira geral, possuem o mesmo propósito, “minimizar os obstáculos da representação e da recuperação da informação e, conseqüentemente, reduzir os ruídos conceituais da comunicação técnico e científica” (SALES; CAFÉ, 2008). Boccato, Ramalho e Fujita (2008) afirmam que entre as linguagens documentárias existentes, o tesouro é o que mais se aproxima das ontologias. Os dois são formados a partir de linguagens documentárias e operam em níveis semelhantes - o epistemológico. No entanto, de acordo com Arano (2005) a diferença entre as duas ferramentas de representação e recuperação da informação, está baseada no nível de abstração que as relações conceituais são estabelecidas. Nos tesouros o conceito é abordado sob uma ótica mais teórica (abstrata), já nas ontologias é tratado sob uma ótica mais aplicada.

As ontologias possibilitam um desenvolvimento semântico mais profundo, na medida que fornecem a descrição lógica e formal da informação, permitindo a interpretação por humanos e softwares. Os tesouros não possuem a expressividade semântica adquirida pelas ontologias, e suas estruturas não são construídas concomitantemente com os recursos tecnológicos. Deste modo, conforme Pedraza-Jiménez et al. (2007), a web, sendo um sistema descentralizado e heterogêneo, é totalmente distinto dos cenários vistos nas disciplinas vinculadas à documentação. Portanto, mesmo que os tesouros sejam aperfeiçoados conforme as necessidades dos ambientes digitais, os instrumentos tecnológicos que são formulados e concebidos inicialmente em contextos computacionais, como as ontologias, apresentam resultados mais significativos para a organização e recuperação de informações em ambientes digitais (BOCCATO; RAMALHO; FUJITA, 2008).

Compreendidas as diferenças entre ontologias e tesouros, Mendonça (2015, p. 188), apresenta orientações, a fim de determinar qual instrumento de organização e representação da informação é o mais adequado:

Se o objetivo do projeto consiste unicamente na indexação e recuperação de informações em um domínio específico, o uso de um tesouro consistente é recomendado, uma vez que a partir dele é possível formalizar a padronização de termos da linguagem específica do domínio, eliminando

ambiguidades no uso dos termos e, assim, auxiliar o usuário no refinamento de suas buscas e localização da informação desejada. Nesses casos, não se faz necessário a construção de uma ontologia, já que um tesouro consegue atender bem a demanda. Por outro lado, se o objetivo do projeto corresponde a indexação e recuperação de informação em um contexto dinâmico para descrição de recursos de um dado domínio do conhecimento a utilização de ontologias é fundamental, pois nestes casos alguma dedução e inferência automatizada é necessário, além da necessidade do uso de relações extensíveis, o que é conseguido com ontologias. Outras situações que demandam o uso de ontologias referem-se à necessidade de representação de aspectos e objetos do mundo real, a necessidade de mais relações entre os objetos de um domínio além daquelas contidas em vocabulários controlados e o uso de formalismos lógicos para representação da informação e inferência automatizada.

Nesse sentido, levando em consideração todos os aspectos analisados, que consideram a abrangência de tesouros e ontologias, e também as orientações apontadas por Mendonça (2015), identificamos a ontologia como o instrumento de organização e representação da informação ideal para o problema que desejamos tratar. Alguns dos pontos extraídos do texto que confirmam essa decisão, são:

- a) possuir contexto dinâmico para descrição de recursos de um dado domínio do conhecimento – haja visto, que o domínio do futebol tem a inserção de expressões novas e voláteis de forma contínua;
- b) dispor de dedução e inferência automatizada – por precisar representar aspectos e objetos do mundo real, necessita que a informação seja entendida conforme o contexto em que se insere;
- c) apresentar a necessidade de mais relações entre os objetos de um domínio além daquelas contidas em vocabulários controlados – o objetivo da *OntoFootballForNewspapers* expõe, além da demanda de resolução dos problemas do domínio futebol, a intersecção da linguagem inerente aos textos jornalísticos, portanto, a junção desses dois contextos, mostra a necessidade de relações mais específicas;

- d) possibilitar o suporte semântico das palavras descritas de objetos linguísticos em uma base de dados lexicais ou terminológicos;
- e) proporcionar que suas relações conceituais possam ser representadas de forma variada, e levem em consideração o campo do conhecimento que fazem parte;
- f) tratar a informação sob uma ótica mais aplicada;
- g) permitir a interpretação por humanos e softwares;
- h) analisar o conhecimento de um domínio.

5.1.1 Documento de especificação

A especificação da ontologia reúne o propósito da *OntoFootballForNewspapers*, os usuários que ela visa atender e o seu uso pretendido, assim como aspectos relacionados a ontologia como o tipo de ontologia e o grau de formalidade que ela envolve. Essas informações estão detalhadas no quadro 14.

Quadro 14 - *Template* de Especificação da *OntoFootballForNewspapers*

<p>Domínio/Esopo Geral</p> <p>A <i>OntoFootballForNewspapers</i> é uma ontologia de domínio do futebol que representa o conhecimento relativo ao futebol brasileiro e internacional que estão contidos em publicações de jornais disponibilizados em acervos digitais. Portanto, além da introdução dos termos do futebol, as linguagens adotadas nos textos jornalísticos que tiveram variação durante a passagem do tempo, também serão consideradas, como as evoluções diacrônicas e sincrônicas, que envolvem também estrangeirismo, além de termos sinônimos, polissêmicos e homógrafos.</p>
<p>Propósito Geral</p> <p>A <i>OntoFootballForNewspapers</i> tem como propósito geral facilitar a recuperação da informação de usuários que buscam informações sobre futebol em jornais disponíveis em acervos digitais. Esse apoio visa unificar as informações referentes ao futebol, para que a recuperação da informação não dependa tanto do letramento do usuário, mas sim que os sistemas simplifiquem esse processo, unindo termos que possuem relações, como é o caso dos sinônimos e também dos que tiveram variação linguística, ocorridas com o passar dos anos.</p>

Classes de usuários
A ontologia <i>OntoFootballForNewspapers</i> é destinada a usuários de acervos digitais interessados por futebol, historiadores e estudantes. Além de pesquisadores e profissionais de diversas áreas que se interessam pelo tema.
Uso pretendido
O uso geral pretendido com a ontologia <i>OntoFootballForNewspapers</i> é no suporte de informações sobre o futebol contidas em acervos digitais de jornais, tal que essa organização seja capaz de relacionar as informações publicadas em diferentes épocas, representadas com diferentes formas e aplicadas em diferentes contextos. Um cenário aplicado na <i>OntoFootballForNewspapers</i> é nas terminologias do futebol que tiveram aportuguesamento, como o caso de <i>corner</i> (escanteio) e <i>goal</i> (gol). Outro está relacionado com as expressões usadas no esporte, como “fazer cera” que se refere à quando o jogador demora na reposição da bola, ou retarda sua substituição, ou delonga quando sofre falta, a fim de tirar proveito no gasto do tempo da partida. Outro exemplo é o termo “chocolate”, que no futebol é utilizado quando um time ganha a partida com grande diferença de gols. Esses termos/expressões precisam fazer sentido e estar relacionados com o contexto em se aplicam e não somente com as suas definições no dicionário, determinando assim, a necessidade de uso da <i>OntoFootballForNewspapers</i>
Tipo da ontologia
A <i>OntoFootballForNewspapers</i> é classificada como uma ontologia de domínio (futebol), representada com leve rigor formal e ontologia para sistemas de informação.
Grau de formalidade
A ontologia do futebol apresenta um grau de leve rigor formal. Inclui classes, subclasses e propriedades da relação da ontologia. A linguagem utilizada é a OWL.
Delimitação do escopo de cobertura
Ponto de partida: foram criadas classes conforme os princípios prescritos na UFO e também nos trabalhos de Reis e Lau (2002) - “ <i>Coach Unilang: a Standard Language for Coaching a (Robo) Soccer Team</i> ” e de Faria (2009) - “Definição de uma Ontologia Aplicada ao Futebol”. Depois de criadas as classes principais, os conceitos considerados do domínio modelado foram inseridos nas classes correspondentes. Seguidament da definição das hierarquias no editor Protegé foram incluídos os sinônimos, as classes disjuntas e as propriedades de relações dos conceitos.

Limite do domínio coberto: Os conceitos foram coletados conforme uma peça da realidade do domínio do futebol considerada em textos jornalísticos impressos e disponibilizados em acervos digitais de jornais.

Questões de competência: as questões de competências são geradas como forma de prover condições para avaliação da ontologia. São elaborados de acordo com os problemas identificados no domínio que a ontologia busca selecionar e das necessidades dos usuários. As questões de competência elaboradas são:

1. Impedimento e Escanteio é atribuição de quem? Resposta: Bandeirinha e seus sinônimos
2. Quem senta no banco de reservas do futebol? Resposta: Jogadores e sinônimos
3. O que é uma bomba? Resposta: Chute.
4. Quais times possuem jogadores que participaram da Copa do Mundo de Futebol?

Fonte: Elaborado pela autora com Mendonça (2015)

5.1.2 Coleta de informações

Nesta etapa o conhecimento do domínio considerado na ontologia foi coletado. Esse conhecimento foi utilizado como base para a etapa seguinte de conceitualização, em que acontece a modelagem do domínio. As obras utilizadas nesta etapa, tanto para formação do glossário de conceitos, verbos e relações foram:

- a) BARROS, Geraldo Monteiro de. Dicionário ilustrado de futebol. [Publicado em fascículos pela Revista Placar]. São Paulo: Placar-Abril, 1980
- b) RIBOLDI, Ari. Cabeça de bagre: termos, expressões e gírias do futebol. São Paulo: AGE, 2008
- c) SILVA, George Batista da. Além das quatro linhas: dicionário futebolístico. Rio de Janeiro: Clube de Autores, 2014
- d) QUEIROZ, João Machado de. Vocabulário do futebol na mídia impressa: o glossário da bola. 904 f. Tese (Doutorado) - Curso de Letras, Faculdade de Ciências e Letras de Assis, Universidade Estadual Paulista, Assis, 2005
- e) Notícias jornalísticas coletadas no acervo Estadão.

A parte da coleta dos conceitos, levando em consideração os aspectos contidos em textos impressos de jornais do domínio do futebol que apresentavam implicações de sinonímia, polissemia e homografia; b) variação diacrônica e sincrônica; e c) termos estrangeiros que tiveram aportuguesamento, formam o Glossário de Conceitos. Parte do glossário é apresentado no Quadro 15 e o glossário completo no apêndice A deste trabalho:

Quadro 15 - Parte do Glossário de Conceitos da *OntoFootballForNewspapers*

	Glossário de Conceitos (coleta manual)
1.	Armação
2.	Assistência
3.	Avenida
4.	Banco
5.	Bandeirinha
6.	Banheira
7.	Banho de bola
8.	Banho-Maria
9.	Barreira
10.	Bicicleta

Fonte: Elaborado pela autora.

O glossário de conceitos foi extraído de forma manual nas obras acima mencionadas, de acordo com as implicações linguísticas consideradas. Foram levantados inicialmente 56 termos para a amostra do domínio do futebol. Mas com o desenvolvimento das outras etapas da ontologia, a amostra foi finalizada com 65 termos. O glossário abrange substantivos, pronomes e adjetivos que serão utilizados para a construção do dicionário de conceitos e posteriormente irão constituir as classes da ontologia.

Subsequente, o glossário de verbos reutilizou os verbos *is_a* (é um) e *part_of* (parte_de) verbos empregados na ontologia sobre os componentes do sangue humano – Hemonto, desenvolvida como aplicação da metodologia *OntoForInfoScience*. De acordo com Mendonça (2015, p.101) estes verbos formam “relações estritamente formais desta ontologia e, geralmente, são as mais comuns na representação das ontologias. Os outros verbos foram extraídos de documentos de referências contidos no acervo digital Estadão, ambiente

de pesquisa deste trabalho. O Glossário de verbos está exposto no Quadro 16:

Quadro 16 - Glossário de Verbos da *OntoFootballForNewspapers*

Glossário de Verbos	
ID	Verbo
1	É_um
2	Parte_de
3	Defender
4	Faz
5	Causar
7	Permitir
8	Conter
10	Resultar
11	Fazer
12	Utilizar
13	Auxiliar
14	Indicar
15	Conquistar

Fonte: Elaborado pela autora.

Para coleta dos verbos foram analisados 10 textos. Um dos textos analisados está exposto na Figura 12 e os restantes estão disponibilizados no anexo A ao final deste trabalho.

Figura 12 - Documento de referência para coleta de verbos
(Edição de 14 de janeiro de 1917, p.9)



Fonte: Captura de tela acervo Estadão

A figura 12 integrou a edição de 14 de janeiro de 1917 (p.9) do acervo Estadão e contém termos que fazem parte do glossário de conceitos da pesquisa - Futebol (*Football*), Partida (*Matches*) e Jogadores (*Footballers*). O verbo retirado do texto foi "Constituir", sinalizado no glossário de verbos como `parte_de`, mas incluído no dicionário de verbos como seu sinônimo.

Esta etapa levantou o conhecimento a ser considerado na modelagem da ontologia. Muitas ontologias guiam essa parte a partir da reutilização de instrumentos já construídos, como taxonomias e tesouros. No entanto, nesta pesquisa optou-se pela coleta das informações, pois trata-se de uma amostra muito específica, que remete ao domínio do futebol, mas que este esteja presente também nas implicações de textos jornalísticos de acervos digitais.

5.1.3 Modelos conceituais

A etapa de conceitualização é uma das mais importantes na construção da ontologia. Nela que o desenvolvedor se envolve mais com o contexto de aplicação do domínio. Aqui são feitas as análises, a organização e estruturação do conhecimento e a representação em artefatos que vão subsidiar os passos que definem as relações, propriedades e restrições da ontologia.

O primeiro passo desta etapa é a elaboração do Dicionário de Conceitos. Neste momento foram agregados aos conceitos que fazem parte do Glossário de Conceitos, desenvolvido na etapa anterior, de Aquisição das informações, suas definições e sinônimos.

Para a formação do Dicionário de Conceitos foram utilizados tais documentos de referência: i) RIBOLDI, Ari. Cabeça de bagre: termos, expressões e gírias do futebol. São Paulo: AGE, 2008; ii) SILVA, George Batista da. Além das quatro linhas: dicionário futebolístico. Rio de Janeiro: Clube de Autores, 2014; iii) Acervo Estadão. Parte do Dicionário de Conceitos está exposto no quadro 18 e o artefato completo está no apêndice C ao final deste trabalho.

As definições incluídas nessa etapa também foram analisadas sob duas perspectivas – quem executa a ação e quem sobre a ação, como no exemplo do conceito ‘banho de bola’, que pode se referir ao time que aplica um banho de bola no adversário e também ao time que perde levando um banho de bola do adversário.

Os sinônimos incluídos nesta ontologia, substituirão as relações de equivalência, mas apresentarão o mesmo propósito afirmado por Dodebei (2002, p.91) em que “um conceito pode ser representado por vários símbolos distintos ou quando se quer reduzir, por questões pragmáticas, os níveis de implicação conceitual”. Segundo Currás (1995), existem algumas variações de sinonímia que podem resultar nas relações de equivalência, como palavras com ortografias diferentes, termos de origem linguística diferente e derivados da linguagem popular. A norma ANSI/NISO Z39.19 (2005), considera que as relações de equivalência são formadas por sinônimos, variantes lexicais e sinônimos aproximados, como os formados por ortografias diferentes. Tendo em vista esse entendimento, o Dicionário de Conceitos foi formado e parte dele está exposto no Quadro 17. O artefato completo está disponível no apêndice C.

Quadro 17 - Parte do Dicionário de Conceitos da *OntoFootballForNewspapers*

Dicionário de Conceitos			
ID	Conceito	Sinônimos	Definição
1.	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.
2.	Assistência	Passê; <i>Shoot</i> ; Toque	Lançamento dado em boas condições para o companheiro prosseguir a jogada; Ação pela qual um jogador passa a bola para o outro; Jogador que recebe a bola do companheiro em boas condições.
3.	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.
4.	Banco	-	Lugar de permanência dos jogadores reservas ou aqueles que foram substituídos no decorrer do jogo e também da comissão técnica.
5.	Bandeirinha	Bandeira; Auxiliar de arbitragem; juiz auxiliar; fiscal de linha; juiz de linha; árbitro assistente.	Indivíduos que se encontram às margens do campo, carregando consigo uma pequena bandeira, para sinalizar as saídas da bola pela lateral ou linha de fundo, marcação de impedimentos e escanteios, além de auxiliar o juiz nos lances do jogo.
6.	Banheira	-	Quando o jogador fica somente no ataque, esperando a bola chegar nele; ficar em situação de impedimento.
7.	Banho de bola	Baile	Perder de goleada; ganhar de goleada sem oferecer reação ao adversário.
8.	Banho-Maria	Jogo morno	Jogo em que os times não atacam, jogo morno e desinteressante.
9.	Barreira	-	Grupo de jogadores que se colocam em frente à bola para dificultar a cobrança de uma penalidade com direção ao gol.

10.	Bicicleta	-	O Jogador atira-se para o ar, de costas para a goleira, e acerta a bola como se estivesse pedalando uma bicicleta.
-----	-----------	---	--

Fonte: Elaborado pela autora.

É comum também nesta etapa alguns conceitos serem unificados, devido à análise mais minuciosa sobre os significados. Assim, alguns conceitos que apresentam a mesma definição são unificados, nesse caso deve se adotar um termo preferencial, eliminar o seu sinônimo na entrada (célula) do Dicionário de Conceitos e incluir o sinônimo no lado do termo preferencial, na entrada respectiva, permanecendo apenas um conceito no Dicionário de Conceitos. Esse exemplo pode ser visto no termo “Cera” que possui a seguinte definição “recurso utilizado com o fim de reter a bola, retardar o jogo” (RIBOLDI, 2008, p.18) e o termo “Cozinhar” que possui a definição “enrolar o jogo para ganhar tempo” (RIBOLDI, 2008, p.18). Esses termos podem ser utilizados como sinônimos, sendo assim, o termo “Cozinhar” foi excluído como termo preferencial no Dicionário de Conceitos (apresentado no apêndice C ao final do trabalho) e incluído como termo sinônimo de “Cera”, conforme o quadro 18.

Quadro 18 - Parte do Dicionário de Conceitos - Termo “Cera”

23.	Cera	Cozinhar	Recurso utilizado com o fim de reter a bola, retardar o jogo; Enrolar o jogo para ganhar tempo.
-----	------	----------	---

Fonte: Elaborado pela autora.

Outro caso no desenvolvimento do Dicionário de Conceitos, foi o termo “Boleiro” que segundo Riboldi (2008, p.16) e Priberam¹⁰ significa Jogador de futebol ou Jogador profissional, desta maneira, nesta fase de desenvolvimento, o termo preferencial foi substituído por “jogador” e o termo “boleiro” foi anotado como sinônimo, juntamente com o termo “*footballer*”. O termo “passe” também foi introduzido como sinônimo do termo “assistência”.

O próximo passo é o desenvolvimento da Tabela de Conceitos e Valores. Para realizar esta tarefa, foi utilizado o Dicionário de Conceitos

¹⁰ Dicionário Priberam. Definição da palavra Boleiro. Disponível em: <<https://www.priberam.pt/dlpo/boleiro>>. Acesso em: 15 mar 2016.

e a partir de cada entrada, definido valores que o conceito pode apresentar, que podem se tornar instâncias da ontologia, ou inclusive classes e subclasses, de acordo com a especificidade do domínio. Neste estágio o conhecimento ainda é tratado em nível textual e descritivo. Parte da Tabela de Conceitos e Valores é apresentada no quadro 19, a tabela completa está no apêndice D deste trabalho.

Quadro 19 - Tabela de Conceitos e Valores

Tabela de Conceitos e Valores				
ID	Conceito	Sinônimos	Definição	Valores
1.	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.	Time
2.	Assistência	Passe; <i>Shoot</i> ; Toque	Lançamento dado em boas condições para o companheiro prosseguir a jogada; Ação pela qual um jogador passa a bola para o outro; Jogador que recebe a bola do companheiro em boas condições.	-
3.	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.	Adversário
4.	Banco	-	Lugar de permanência dos jogadores reservas ou aqueles que foram substituídos no decorrer do jogo e também da comissão técnica.	-
5.	Bandeirinha	Bandeira; Auxiliar de arbitragem; juiz auxiliar; fiscal de linha; juiz de linha; árbitro assistente.	Indivíduos que se encontram às margens do campo, carregando consigo uma pequena bandeira, para sinalizar as saídas da bola pela lateral ou linha de fundo, marcação de impedimentos e escanteios, além de auxiliar o juiz nos lances do jogo.	Bandeira

6.	Banheira	-	Quando o jogador fica somente no ataque, esperando a bola chegar nele; ficar em situação de impedimento.	-
7.	Banho de bola	Baile	Perder de goleada; ganhar de goleada sem oferecer reação ao adversário.	Bola.
8.	Banho-Maria	Jogo morno	Jogo em que os times não atacam, jogo morno e desinteressante.	Time; Ataque; Atacante
9.	Barreira	-	Grupo de jogadores que se colocam em frente à bola para dificultar a cobrança de uma penalidade com direção ao gol.	-
10.	Bicicleta	-	O Jogador atira-se para o ar, de costas para a goleira, e acerta a bola como se estivesse pedalando uma bicicleta.	-

Fonte: Elaborado pela autora.

A partir desses valores, na etapa de formalização, será preciso decidir se eles serão instâncias, classes ou subclasses. Quanto mais instâncias são definidas, mais especializada a ontologia se torna. A identificação de instâncias não é obrigatória no desenvolvimento de uma ontologia, muitas vezes essa etapa é umas das últimas a ser realizada. No entanto, mesmo que não sejam incluídas instâncias na ontologia, esta etapa é importante pois poderá determinar novas classes ou subclasses.

O passo seguinte foi a formação do Dicionário de Verbos. Para este artefato foi utilizado o Glossário de Verbos desenvolvido na etapa de Aquisição das Informações. Além dos verbos, nesse estágio são incluídos sinônimos e definições para cada entrada preferencial. A definição corresponderá à uma descrição informal da utilidade do verbo para auxiliar no desenvolvimento do Glossário de Relações, passo seguinte da modelagem do conhecimento. Não há necessidade utilizar todos os verbos coletados no Glossário de Verbos, mas sim aqueles essenciais para a formação das relações da ontologia.

Quadro 20 - Dicionário de Verbos da *OntoFootballForNewspapers*

Dicionário de Verbos			
ID	Verbo	Sinônimo	Definição
1	É_um	Corresponde a	Representa uma relação entre dois conceitos no sentido gênero-espécie.
2	Parte_de	Integrar	Indica uma relação entre dois conceitos, de forma que um conceito é parte que integra o outro conceito.
3	Defender	Proteger Impedir Bloquear	Representa uma relação em que o conceito-agente evita ou dificulta uma ação no conceito-paciente.
4	Causar	Provocar	Representa uma relação em que o conceito-agente interfere na ação do conceito-paciente.
5	Permitir	Possibilitar	Indica uma relação em que o conceito-agente tem responsabilidade no desfecho do conceito-paciente.
6	Conter	Incluir,	Representa uma relação em que o conceito-agente abrange o conceito-paciente.
7	Utilizar	-	Indica uma relação em que o conceito-agente faz uso do conceito-paciente.
8	Auxiliar	-	Representa uma relação em que o conceito-agente contribui para a ação do conceito-paciente
9	Indicar	Sinalizar	Indicar uma relação que o conceito-agente é responsável pelo evento do conceito-paciente.
10	Conquistar	-	Indica uma relação em que o conceito-agente

Fonte: Elaborado pela autora.

A partir da definição de sinônimos e do entendimento da finalidade de cada verbo é organizado o Glossário das Relações, em que as relações devem ter a forma <termo 1> <verbo> <termo2>.

As relações aqui formadas evidenciam as ligações entre os conceitos da amostra do domínio. O verbo refletiu a ação entre a eles. Além das relações estabelecidas no Glossário de Relações, todos os termos possuem ligação direta com o termo futebol.

Quadro 21 - Parte do Glossário das Relações da *OntoFootballForNewspapers*

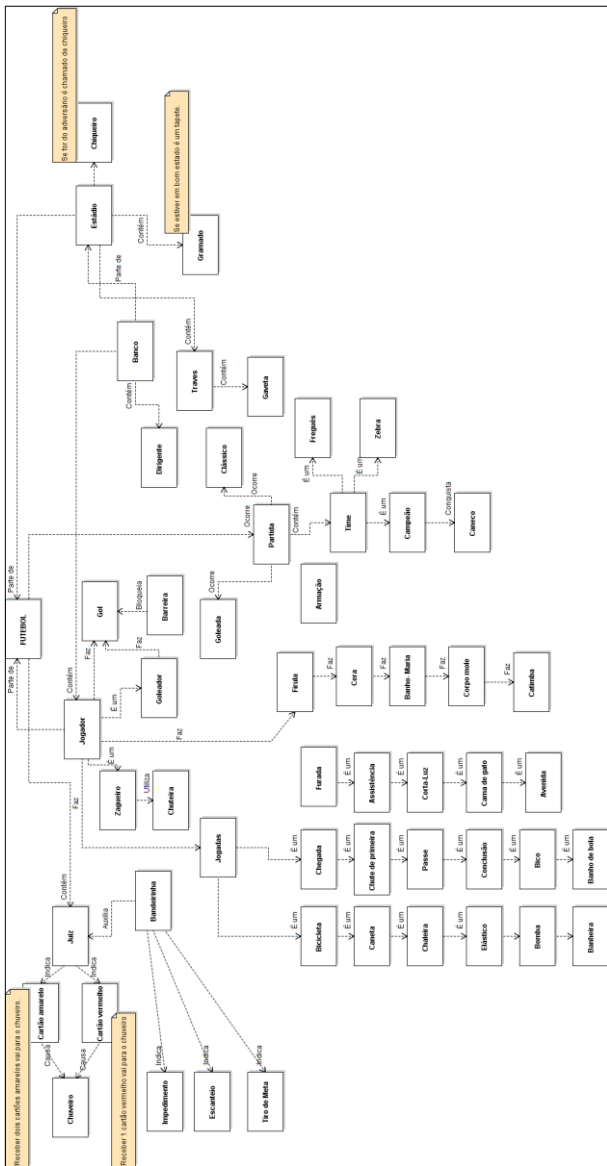
Glossário das Relações			
ID	Termo 1	Verbo	Termo 2
1.	Assistência	É uma	Jogada
2.	Banco	Contém	Jogador
3.	Bandeirinha	Indica	Escanteio
4.	Bandeirinha	Auxilia	Juiz
5.	Banheira	Causa	Impedimento
6.	Banho de bola	Permite	Gol
7.	Barreira	Protege	Gol
8.	Barreira	Bloqueia	Bola
9.	Bicicleta	É um	Jogada
10.	Bico	É um	Chute

Fonte: Elaborado pela autora.

Por fim, na etapa de conceitualização são formados os instrumentos de representação gráfica. Para formar esses modelos foram utilizados os conceitos do domínio e suas relações. As relações foram criadas a partir do conceito macro (futebol) modelando os outros conceitos que apresentavam relações hierárquicas ou associativas com esse termo macro. A abordagem combina a estratégia *top-down* com a estratégia *bottom-up* iniciando com o conceito macro e relacionando os conceitos micros. Para desenvolver esta etapa foi utilizado como ferramenta o editor gráfico *Violet 2.1.0*.

Esta representação demonstra as relações que os conceitos podem envolver na amostra do domínio. Todavia, é possível formar outras relações ou continuar a hierarquia apresentada, no entanto, devido ao escopo desta pesquisa, a representação considerou a seguinte forma, exposta na figura 13.

Figura 13 - Representação Gráfica com a amostra de conceitos da *OntoFootballForNewspapers*



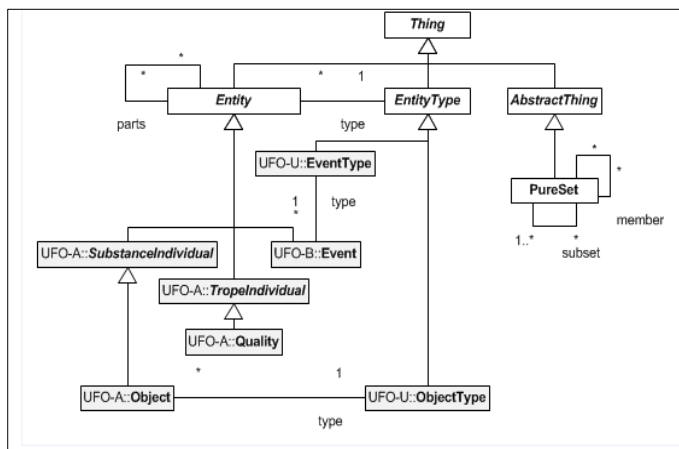
Fonte: Elaborado pela autora.

Nesta etapa, em que os conceitos são organizados de forma visual, é possível identificar outras relações que não foram preenchidas primeiramente no Glossário de Relações e que podem ser agregadas na etapa de formalização da ontologia. É importante ressaltar que os artefatos desenvolvidos, glossários e dicionários de conceitos, dicionários de verbos e glossário de relações quando implementados em um ambiente de produção, devem ser sujeitos a um processo de lematização, considerando todas as flexões dos termos, tempo verbal, singular, plural, gênero, etc.

5.1.1 Reuso ontologia de fundamentação

Para desenvolvimento da *OntoFootballForNewspapers* foi selecionada a *Unified Foundational Ontology* (UFO). De acordo com Guizzardi, Giancarlo; Falbo, Ricardo A.; Guizzardi (2008, p.244) a UFO tem sido desenvolvida nas áreas de Ontologias Formais, Lógica Filosófica, Filosofia da Linguagem, Lingüística e Psicologia Cognitiva e aplicadas para “avaliar, (re)projetar e integrar os modelos de linguagens de modelagem conceitual, bem como para prover semântica de mundo real (*real-world semantics*) para seus elementos de modelagem”. A *OntoFootballForNewspapers* é baseada no pragmatismo, abordagem que considera os conceitos com base em seu contexto prático, assim sendo, a UFO fornece estrutura semântica que possibilita a modelagem conceitual da *OntoFootballForNewspapers*. Para Campos (2012), a UFO permite um melhor detalhamento e flexibilidade na modelagem de determinados aspectos do conhecimento do domínio. A Figura 14 a seguir, ilustra a estrutura taxonômica da UFO que inclui a coleção de categorias básicas:

Figura 14 - Representação esquemática das categorias básicas da UFO.



Fonte: xygen¹¹, 2015.

A UFO é composta de três principais partes, UFO-A, UFO-B e UFO-C. As diferenças entre as três está nos endurantes e perdurantes. De acordo com Campos, M. L. A., e Campos, L. M., e Medeiros (2011, p.149) “endurantes são entidades que não mudam sua essência ao longo do tempo, preservando sua identidade durante as mudanças que ocorrem durante o período de sua existência”, já os perdurantes são compostos por partes temporais, entidades que possuem “diferentes partes em diferentes instantes, como um processo” (MOREIRA, 2012, p.117).

A UFO-A é formada por endurantes, de acordo com Guizzardi, Giancarlo; Falbo, Ricardo A.; Guizzardi (2008, p.245) uma parte fundamental dessa ontologia é a sua divisão nas categorias Indivíduo (Particular) e Universal (Tipo), sendo que “indivíduos são entidades que existem na realidade, possuindo uma identidade única. Universais, por sua vez, são padrões de características que podem ser instanciados em um número de diferentes indivíduos.”

A UFO-B é formada por perdurantes, uma ontologia de eventos (ocorrências). Para os autores, Guizzardi, Giancarlo; Falbo, Ricardo A.; Guizzardi (2008, p.246) eventos “são indivíduos compostos de partes temporais. Eles acontecem no tempo no sentido de se estenderem no

¹¹ The Unified Foundational Ontology. Disponível em: <https://oxygen.informatik.tu-cottbus.de/drupal7/ufo/?q=node/2>. Acesso em: 25/02/2016

tempo acumulando partes temporais. São exemplos de eventos: uma conversa, uma partida de futebol, a execução de uma sinfonia e um processo de negócio”.

A UFO-C é constituída tanto por endurantes, como perdurantes, é considerado uma ontologia de entidades sociais (objetos e eventos) baseada na UFO-A e UFO-B. Suas distinções ainda podem ser feitas entre agentes (Agent) e objetos não-agentivos (Objetct), os agentes são capazes de realizar ações, com alguma finalidade, já os objetos somente participam de eventos, conforme explicam os autores Guizzardi, Giancarlo; Falbo, Ricardo A.; Guizzardi (2008, p.246, 247):

Um agente é um substancial que cria ações (Action), percebe eventos (Event) e que pode possuir tipos especiais de modos, chamados de intencional modes. Os agentes podem ser físicos (Physical Agent) (por exemplo, uma pessoa) ou sociais (Social Agent) (por exemplo, uma organização). Um objeto, por outro lado, é um substancial incapaz de perceber eventos ou ter intencional modes. Objetos também podem ser divididos em objetos físicos (Physical Object) (por exemplo, um livro, um carro) e objetos sociais (Social Object) (por exemplo, dinheiro, língua). A descrição normativa (Normative Description) é um tipo de objeto social, que define uma ou mais regras/normas reconhecidas por, pelo menos, um agente social e que pode definir entidades sociais como universais (p.ex., tipos de comprometerimentos sociais) e papéis sociais, tais como presidente, ou pedestre.

A metodologia *OntoForInfoScience* sugere que o arquivo OWL localizado na Internet seja importado no editor Protegé, e a partir da estrutura base adquirida seja iniciada a etapa de formalização. No entanto, após exaustiva pesquisa na Internet, não foi possível localizar o arquivo para importação da UFO. Segundo Mendonça (2015) o reuso da ontologia de fundamentação deve ir além da sua importação em um editor de ontologias, é preciso manter o comprometimento ontológico com os princípios de fundamentação. O autor ainda enfatiza que em um projeto em que se deseja desenvolver uma ontologia leve de domínio, a etapa de fundamentação pode se tornar opcional. No entanto, optou-se ainda por

utilizar a abordagem filosófica como forma de justificar algumas decisões ontológicas adotadas no desenvolvimento da ontologia.

Para Campos, M., Campos, L e Medeiros (2011, p.146), no estabelecimento do compromisso ontológico com uma ontologia de fundamentação, existem diversos modos de representar esta visão de mundo e também, ao conduzir esse conhecimento para a linguagem das máquinas, devido as suas limitações lógicas, a representação conforme os princípios de fundamentação, será sempre aproximada.

Assim, o estabelecimento de compromisso ontológico revela uma tomada de posição em relação aos objetos que se reconhecem como fundamentais em uma visão de mundo e que possibilitam o diálogo sobre esse mundo, e ainda em relação à maneira pela qual esses objetos são caracterizados. Vale ressaltar que existem diversos modos possíveis de um acordo sobre esta visão de mundo. Deve-se ter em mente, entretanto, que nas ontologias, principalmente nas chamadas ontologias de fundamentação, devido às restrições dos mecanismos de representação do conhecimento para as máquinas, a representação dos conceitos e de suas relações que constituem o seu modelo conceitual é sempre limitada e, portanto, aproximada.

À vista disso, a utilização dos princípios de fundamentação da UFO, serão empregados na etapa seguinte, de formalização, para a definição das classes principais da ontologia e inclusão dos conceitos da amostra do domínio. Para tal, utilizam-se como base as distinções fundamentais da UFO-C, apresentada nesta seção.

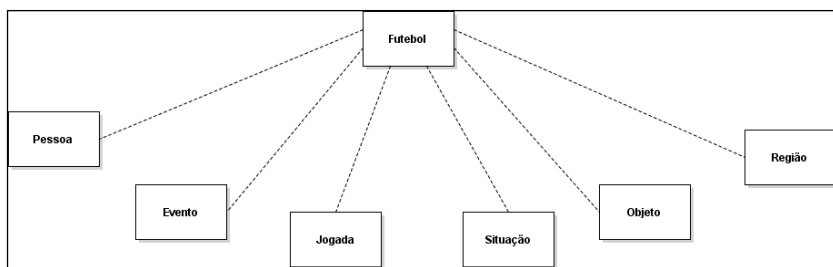
5.1.2 Conteúdo ontológico

Nesta etapa, o conhecimento modelado até agora em nível conceitual, passa a ser representado em nível ontológico-formal. Para tal, é utilizado o editor Protégé 5.0. O domínio deste trabalho é o futebol, mas o propósito desta ontologia não é representar todos os conceitos e termos associados ao futebol em geral, mas sim uma peça da realidade do domínio do futebol, expressa na linguagem dos textos jornalísticos disponibilizados em acervos digitais de jornais, a fim de demonstrar a

utilidade de ontologias nesses ambientes. A amostra considera conceitos distintos. Portanto, neste trabalho, os conceitos foram agrupados a áreas conceituais, que aqui são identificadas como classes principais, em que estes conceitos possam ser introduzidos.

De acordo com Mendonça (2015, p.212) as “ontologias de fundamentação ou de alto nível são ontologias que descrevem as entidades mais gerais dos diversos domínios do conhecimento”. Sendo assim, o primeiro passo dessa etapa foi a definição das classes. Elas foram baseadas nas categorias da UFO - Objeto, Evento e Situação, e também nos trabalhos de Reis e Lau (2002) “*Coach Unilang - A Standard Language for Coaching a (Robo) Soccer Team*” e Faria (2009) “*Definição de uma Ontologia Aplicada ao Futebol*, com a adaptação das classes - Pessoa, Jogada e Região. As classes principais foram representadas conforme a figura 15:

Figura 15 - Classes da *OntoFootballForNewspapers*



Fonte: Elaborado pela autora.

As classes são definidas no singular pois cada conceito, introduzido posteriormente, pode representar um ou mais de um comportamento. As classes reutilizadas da UFO e das ontologias aplicadas ao futebol, podem ser entendidas da seguinte forma:

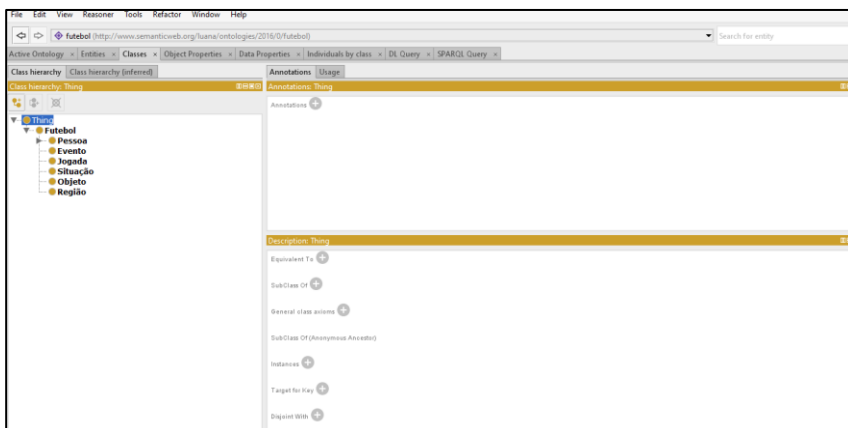
- I. **Futebol:** No editor Protégé a classe principal “thing” liga todos os conceitos, no entanto, foi incluída uma classe futebol por esta representar um conceito da tabela de conceitos e este não se encaixar especificamente em nenhuma outra classe, pois está relacionada com todas as outras.
- II. **Pessoa:** São os agentes da UFO. Podem ser físicos, como por exemplo uma pessoa, ou sociais, como uma organização. No

futebol, podem ser os jogadores, o arbitro, ou o time, a comissão de arbitragem, por exemplo;

- III. **Evento:** É passível de transformação. Podem altear o estado de coisas da realidade. Eventos são entidades ontologicamente dependentes, pois para existirem dependem da existência de seus participantes. No futebol, as regras do futebol podem ser consideradas nessa classe;
- IV. **Jogada:** É constituída por ações típicas do jogo, como ataque, defesa, drible;
- V. **Situação:** É uma entidade constituída por vários objetos ou várias situações. É tratada como sinónimo de “estado de coisas”, transmite uma porção da realidade. Em uma partida de futebol, pode incluir situações do jogo, como acordos e conflitos;
- VI. **Objeto:** Está presente em qualquer instante do tempo. São substancias inanimadas, que são modelizadas, como por exemplo, bola, taças, traves etc.
- VII. **Região:** São as representações da área do campo ou das regiões que fazem parte do futebol, como uma parte do estádio – vestiário, ou definições de lugares mais precisos, como ângulos, assim como o estádio, que contempla todos essas regiões.

Após definir as divisões das classes, foi necessário inseri-las no editor Protégé, para então dar início à taxonomia geral da ontologia, conforme as Figuras 16 e 17:

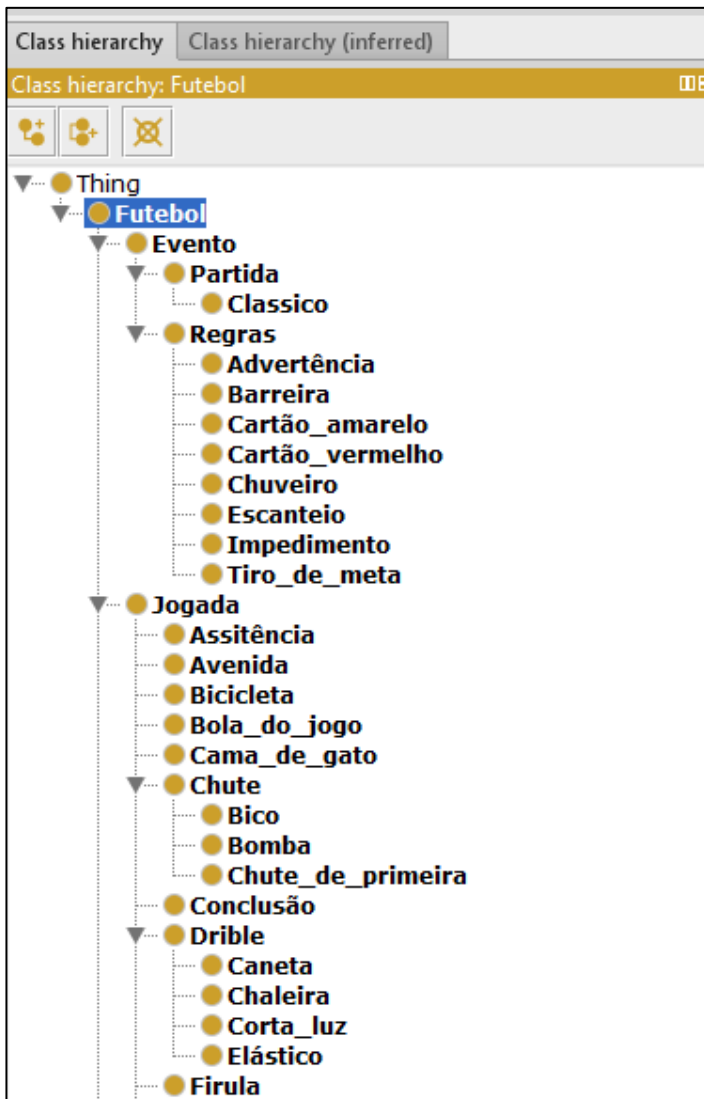
Figura 16 - Classes principais definidas no Protégé



Fonte: Captura de tela do Protégé da *OntoFootballForNewspapers*.

Subsequentemente foram introduzidos os conceitos coletados, presentes no Dicionário de Conceitos, elaborado na etapa de conceitualização desta pesquisa. Os conceitos foram divididos conforme as definições das classes principais e também incluíram as tratativas da ANSI/NISO Z39.19 (2005), que estabelecem as relações hierárquicas como genéricas ou gênero/espécie e partitivas. Parte da estrutura hierárquica está representada na Figura 17:

Figura 17 - Parte das classes e conceitos da hierarquia da *OntoFootballForNewspapers*

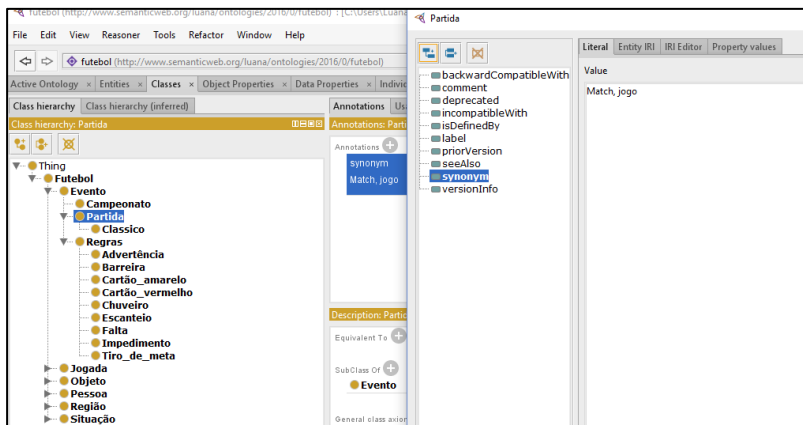


Fonte: Captura de tela do Protégé da *OntoFootballForNewspapers*

A partir desse momento inicia-se o estabelecimento das relações da ontologia, primeiramente com a introdução dos sinônimos. O

levantamento dos sinônimos já foi realizado na etapa de conceitualização e incluído no Dicionário de Conceitos. Neste momento é necessário introduzi-los no editor Protege, em que esta atividade é incluída nas classes e subclasses por meio da propriedade de anotação “*Synonym*” conforme ilustrado na figura 18.

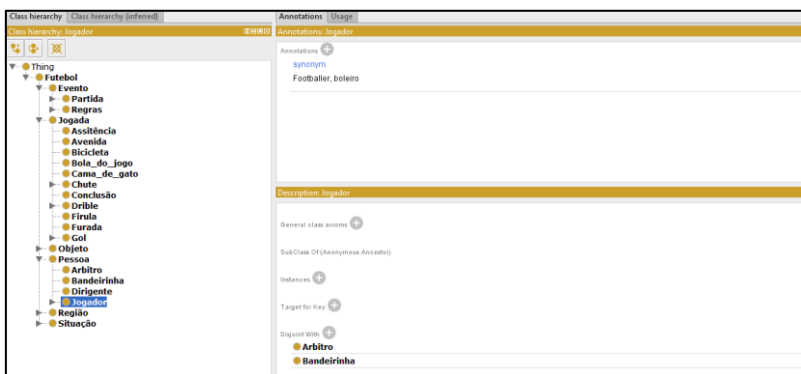
Figura 18 - Inclusão de sinônimos no Protégé



Fonte: Captura de tela do Protégé da *OntoFootballForNewspapers*

Depois da inclusão dos sinônimos foram definidas as classes disjuntas, aquelas que não possuem indivíduos em comum. Nesta ontologia foram definidas como disjuntas as classes jogador, árbitro e bandeirinha, ou seja, nenhum membro da classe jogador pode ser membro da classe árbitro ou bandeirinha e assim vice-versa, assim como nenhum dirigente pode ser membro da classe árbitro e bandeirinha e vice-versa. Para tornar uma classe disjunta no editor Protege é necessário selecionar a classe e no item “*Disjoint With*” inserir as classes que não possuem interseção, conforme mostra a figura 19:

Figura 19 - Classes disjuntas



Fonte: Captura de tela do Protégé da *OntoFootballForNewspapers*

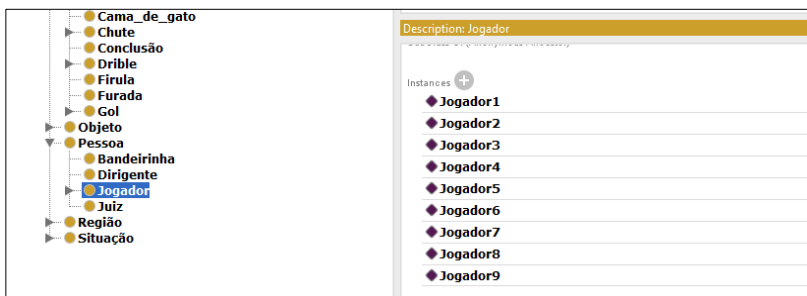
No próximo passo foi analisada a Tabela de Conceitos e Valores e definido se os valores ali introduzidos seriam subclasses ou instâncias na ontologia. De acordo com Mendonça (2015, p.232) instâncias podem ser definidas como “exemplos, ocorrências, casos particulares e reais das classes usadas para representar objetos específicos da realidade de um dado domínio” e podem ser representadas quanto se tornam os itens mais específicos da árvore de conceitos (nós-folha), não existindo, abaixo deles, continuidade da ontologia, ou seja, a ontologia acaba. A definição de instâncias geralmente é um dos últimos passos a ser desenvolvido em uma ontologia, não sendo obrigatório. A introdução vai depender do escopo da modelagem do domínio que se deseja cobrir.

Diante desse entendimento, foi evidenciado que a ontologia não possuía instâncias, pois seus valores representam conceitos mais gerais, e não representações mais específicas da realidade. No entanto, no interesse de criar inferências automatizadas por meio das instâncias, foram criadas instâncias sem as condições reais de um ambiente em produção, como exemplo, na subclasse Jogador foram definidas as instâncias – Jogador 1, Jogador 2, Jogador 3, etc. Em um ambiente de produção, com um conjunto de conceitos ampliados, seriam utilizados nomes de jogadores existentes.

Da mesma maneira foram criadas instâncias para a subclasse Time – Time A, Time B, Time C, e para a subclasse Campeonato, que foi uma subclasse incluída depois da análise da Tabela de Conceitos e Valores, foi

incluída a instância– Copa do Mundo. A figura 20 mostra um exemplo da criação de instâncias no Protégé:

Figura 20 - Inclusão de Instâncias



Fonte: Captura de tela do Protégé da *OntoFootballForNewspapers*

Ainda nessa etapa, os valores levantados na Tabela de Conceitos e Valores que não foram definidos como instância, foram incluídos como subclasses na taxonomia geral da ontologia. Esses valores foram avaliados antes da inclusão das propriedades das classes, pois são fundamentais para a formação de algumas relações. Sendo assim, os valores considerados como subclasses e incluídos na ontologia, além de Campeonato, já mencionado anteriormente, foram: adversário; advertência; atacante; campeonato; chute; estádio; falta; gol-contra; gol-pró; goleiro; time. Esses conceitos foram incluídos também no Dicionário de Conceitos, disponível no Apêndice B ao final do trabalho.

No entendimento que somente a definição das classes e suas hierarquias por meio da relação *is_a* (é_um), das definições das relações de equivalência por meio da atribuição dos sinônimos e a inclusão de instâncias, não seriam suficientes para responder as questões de semântica e elucidar as questões de competência, foram incluídas relações associativas a fim de ampliar as relações entre conceitos. As relações associativas podem incluir ligações como causa/efeito, produtor/produto, etapas de um processo, entre outros. A norma ANSI/NISO Z39.19 (2005) mostra alguns exemplos de relações associativas:

Quadro 22 - Possibilidades de relações semânticas entre os termos

RELAÇÕES ASSOCIATIVAS	EXEMPLOS
Causa / Efeito	Acidente / lesão
Processo / Agente	Medição de velocidade / velocímetro
Processo / Contra-agente	Fogo / retardadores de chama
Ação / Produto	Escrever / publicação
Ação / Propriedade	Comunicação / habilidades de comunicação
Ação / Alvo	Ensino / aluno
Conceito ou Objeto / Propriedade	Liga de aço / resistência à corrosão
Conceito ou Objeto / Origem	Água / poço
Conceito ou Objeto / Unidade ou Mecanismo de medida	Cronômetro / minuto
Matéria-prima / Produto	Uvas / vinho
Disciplina ou Campo / Objeto ou Praticante	Neonatologia / infantil

Fonte: Norma ANSI/NISO Z39.19 (2005, p.42).

As aplicações das relações associativas na *OntoFootballForNewspapers* foram feitas por meio das propriedades de relações entre as classes. As propriedades são formadas por relações entre dois indivíduos, sendo formadas por *Object Properties* (Propriedades de Objeto), responsável por conectar um indivíduo a outro e *Data Type Properties* (Propriedades de Tipos de Dados), que faz a ligação de um indivíduo a um valor, que pode ser números, datas ou texto. Para relacionar as classes será utilizada as propriedades de objeto. As propriedades utilizam a teoria de conjuntos e podem ter característica funcional, quando se relacionam apenas com um outro indivíduo. Funcional inversa, quando a relação inversa é funcional, sendo apenas essa possível de relação. Transitiva, normalmente utilizadas em relações do tipo *part_of*, como no exemplo “se bola tem relação com campo e campo tem relação com jogador, então bola tem relação com jogador”. Simétrica, sendo uma relação binária em que A tem relação com B, então B tem relação com A. Assimétrica não é excludente da simétrica, em que “se A está contido em B e B está contido em A, então A = B”. Reflexiva em que todos os elementos se relacionam com si próprio e irreflexiva em que nenhum elemento se relaciona com si próprio (MELO; CHAGAS; FREITAS, ([20--]). No editor Protege após escolher o tipo de propriedade Object ou DataType e selecionar as características, é preciso também

definir o domínio em que a propriedade estará presente e também as classes que são permitidas na propriedade, marcadas no editor em “*Domains*” e “*Ranges*”. As características das propriedades podem ser entendidas também conforme o quadro 22.

Quadro 22 - Características das propriedades de objeto

CARACTERÍSTICA DA PROPRIEDADE	FORMA DE PREENCHIMENTO
Funcionalidade	Uma relação é funcional para um determinado indivíduo (particular) quando só pode haver no máximo um indivíduo a ser relacionado através desta relação.
Reflexividade	Uma relação é reflexiva se as entidades relacionadas são idênticas e a relação é válida entre elas.
Transitividade	Uma relação é transitiva quando é possível inferir essa mesma relação em uma cadeia de entidades relacionadas.
Simetria	Uma relação é simétrica quando ela é verdadeira entre as entidades relacionadas nos dois sentidos (direita-esquerda, esquerda-direita).
Relação inversa (<i>Inverse Of</i>)	Esse atributo deve ser preenchido com o nome (label) da relação que é inversa à relação que está se definindo.
Domínio (<i>Domains</i>)	Correspondem às classes da ontologia que podem ser relacionadas através desta relação, desde que tais classes façam parte do domínio da relação.
Imagens (<i>Ranges</i>)	Correspondem às classes da ontologia que podem ser relacionadas através desta relação, desde que tais classes façam parte da imagem da relação.

Fonte: Adaptado de Mendonça (2015)

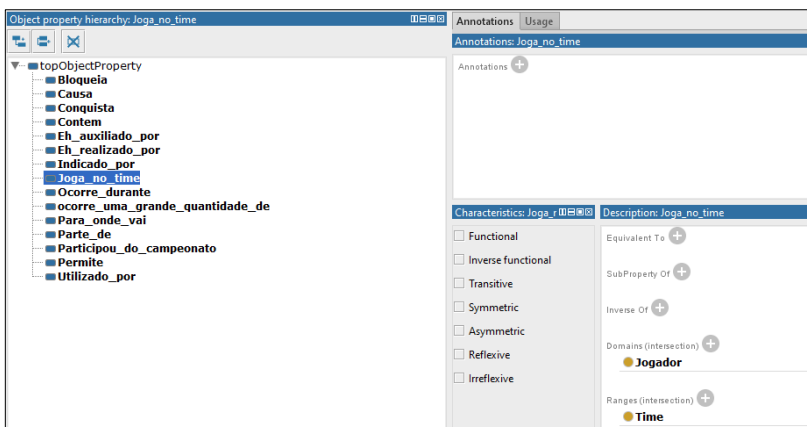
Para apoiar essa etapa foram utilizados o Dicionário de Verbos, Glossário de Relações e a representação gráfica desenvolvidos na etapa de conceitualização. As propriedades de relações estão representadas no quadro 22 e tela de inserção no Protégé está exposta na figura 21.

Quadro 22 - Preenchimento das relações ontológicas das classes no Protégé

<i>PROPERTY</i>	<i>DOMAINS</i>	<i>RANGE</i>	<i>CARACTERÍSTICA</i>
Bloqueia	Barreira Goleiro Zagueiro	Gol	Funcional
Causa	Cama_de_gato Banheira	Impedimento Cartão_amaro	-
Conquista	Jogador Time	Caneco	-
Contem	Banco Time	Jogador	-
Eh_auxiliado_por	Juiz	Bandeirinha	-
Eh_realizado_por	Cera Corpo_mole e Bicicleta	Jogador	-
Indicado_por	Regras	Bandeirinha Juiz	Simétrica
Joga_no_time	Jogador	Time	-
Ocorre_durante	Cera	Partida	-
Ocorre_uma_grande_quantidade_de	Goleada	Gol	-
Para_onde_vai	Geladeira Chuveiro	Expulso	-
Parte_de	Trave Gramado Gaveta	Estádio	Transitiva
Participou_do_campeonato	Jogador	Campeonato	-
Permite	Banho_de_bola	Goleada	-
Utilizado_por	Chuteira	Jogador Juiz Bandeirinha	-

Fonte: Elaborado pela autora.

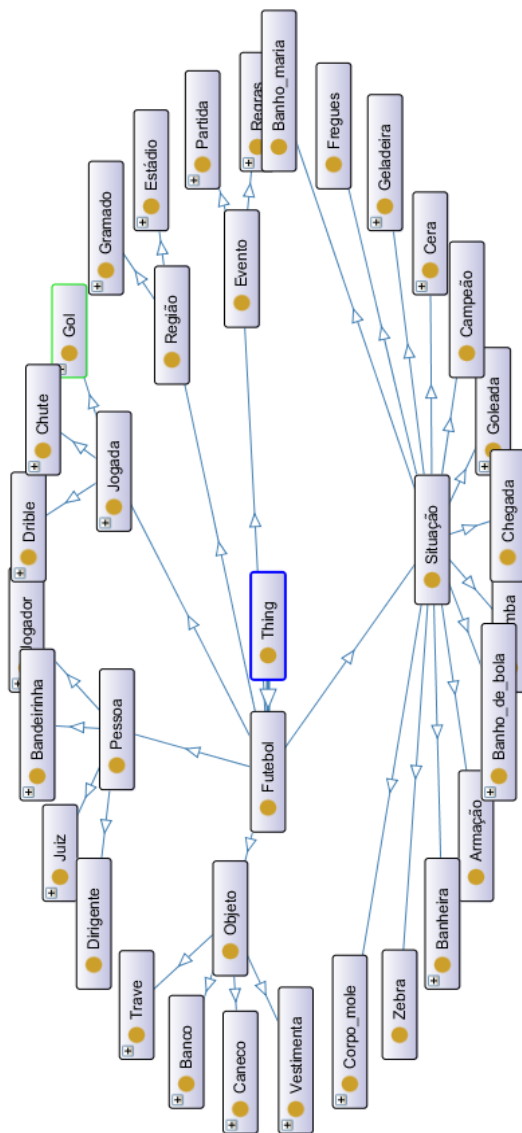
Figura 21 - Propriedades de relação aplicadas no Protégé



Fonte: Captura de tela Protégé da *OntoFootballForNewspapers*.

Na figura 22 é possível visualizar de forma gráfica, por meio do suporte OntoGraf, disponibilizado no Protégé, as principais relações formadas na *OntoFootballForNewspapers*:

Figura 22 – Representação gráfica das principais classes *OntoFootballForNewspapers*



Fonte: Imagem exportada pela autora da ferramenta OntoGraf disponível no Protegé.

5.1.3 Elucidação das questões de competência

Para responder às questões de competência foi utilizado o SPARQL, que permite realizar consultas e obter resultados em documentos semânticos. O raciocinador é dividido em três partes: linguagem de consulta, formato dos resultados e protocolo de acesso. Para realizar as consultas, as principais cláusulas utilizadas são: “*Select*”- busca valores e “*Where*”- utiliza padrões de triplas (*subject, predicate e object*) que serão localizados na base. As consultas foram feitas transformando as questões de competência na linguagem lógica exigida no SPARQL.

A primeira questão de competência - Impedimento e escanteio é atribuição de quem? Desejava-se encontrar como resposta: Juiz (sinônimos: Arbitro; 1ª árbitro); Bandeirinha (sinônimos: Bandeira; Auxiliar de arbitragem; juiz). A execução no SPARQL está exposta na figura 23:

Figura 23 - Execução da questão de competência 1 no SPARQL

The screenshot shows a SPARQL query execution window. The query is as follows:

```

PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX : <http://www.semanticweb.org/luana/ontologies/2016/0/futebol#>

SELECT DISTINCT ?o as ?Atribuição ?os as ?Sinônimos
WHERE
{
  ?i :indicado_por rdf:type domain :Escanteio .
  ?i :indicado_por rdf:type domain :Impedimento .
  ?i :indicado_por rdf:type range ?o .
  ?o :synonym ?os .
}

```

The results are displayed in a table with two columns: 'Atribuição' and 'Sinônimos'.

Atribuição	Sinônimos
Bandeirinha	"Bandeira, auxiliar_de_arbitragem, juiz_auxiliar, fiscal_de_linha, juiz_de_linha,"@
Juiz	"Juiz, 1_1árbitro"@

Fonte: Captura de tela do Protegé.

Para chegar nessas respostas foram incluídas as seguintes sentenças:

- a) Sujeito: Impedimento and Escanteio (*domain*);
- b) Predicado/relação: indicado_por (*range*);
- c) Objeto/Propriedade: é a resposta que buscamos;
- d) E os sinônimos associados a resposta.

A relação formada pela propriedade “indicada_por”, conseguiu relacionar a subclasse “Regras”, juntamente às subclasses de “Pessoa” - “Juiz” e “Bandeirinha”, que tiveram seus sinônimos associados a eles. A segunda questão de competência - Quem senta no banco de reservas do futebol? Desejava-se encontrar como resposta: Jogadores (sinônimos: *Footballer*, boleiro). A execução no SPARQL está exposta na figura 24:

Figura 24 - Execução da questão de competência 2 no SPARQL

SPARQL query:		
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>		
PREFIX : <http://www.semanticweb.org/luana/ontologies/2016/0/futebol#>		
SELECT DISTINCT (?o as ?QuemSentaNoBanco) (?os as ?Sinónimos)		
WHERE		
{		
	:Sentado_por rdfs:domain :Banco .	
	:Sentado_por rdfs:range ?o .	
	?o :synonym ?os	
}		
	QuemSentaNoBanco	Sinónimos
Jogador		"Footballer, boleiro"Ⓞ

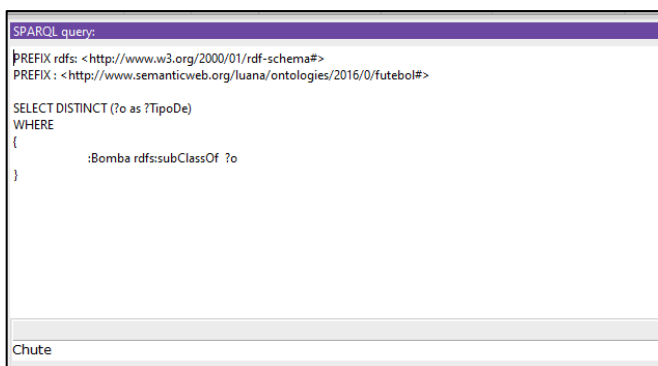
Fonte: Captura de tela do Protégé.

Para chegar a essas respostas foram incluídas as seguintes sentenças:

- a) Sujeito: Banco
- b) Predicado/relação: Contém.
- c) Objeto/Propriedade: é a resposta que buscamos;
- d) E os sinônimos associados a resposta.

A relação formada com a propriedade “Contém” conseguiu relacionar a subclasse de “Pessoa” - “Jogador” juntamente aos sinônimos associados a ele. A terceira questão de competência – O que é uma bomba: Desejava-se encontrar como resposta: Chute.

Figura 25 - Execução da questão de competência 3 no SPARQL



```
SPARQL query:
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX : <http://www.semanticweb.org/luana/ontologies/2016/0/futebol#>

SELECT DISTINCT (?o as ?TipoDe)
WHERE
{
    :Bomba rdfs:subClassOf ?o
}

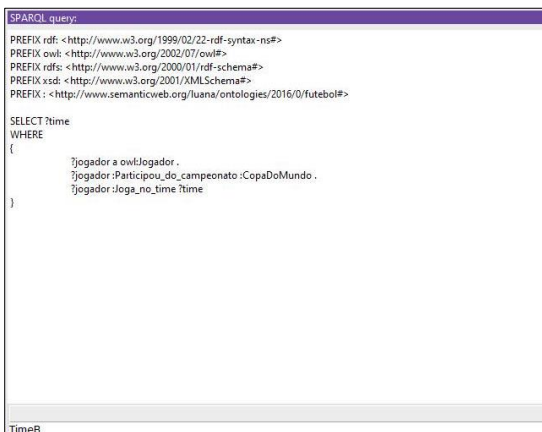
Chute
```

Fonte: Captura de tela do Protégé

Na questão 3, apenas a hierarquia conseguiu responder o problema semântico, afirmando que “bomba” que é um conceito ambíguo, remetendo para outros significados, como o hidráulico ou o artefato explosivo, no domínio do futebol, bomba é um tipo de chute, dado com força.

A quarta questão - Quais times possuem jogadores que participaram da Copa do Mundo de Futebol? Buscou obter como resposta o time ou os times que possuíam jogadores que participaram da Copa do Mundo. A execução no SPARQL está exposta na figura 26:

Figura 26 - Execução da questão de competência 4 no SPARQL



```
SPARQL query:
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX : <http://www.semanticweb.org/luana/ontologies/2016/0/futebol#>

SELECT ?time
WHERE
{
    ?jogador a owl:Jogador.
    ?jogador :Participou_de_campeonato :CopaDoMundo.
    ?jogador :joga_no_time ?time
}
```

Fonte: Captura de tela do Protégé.

Na quarta questão as instâncias criadas para as subclasses jogadores (jogador 1, jogador 2, jogador 3...), time (time A time B, time C) e campeonato (Copa do Mundo), foram ligadas as propriedades “Joga_no_time” e “Participa_do”Campeoanto”, para inferir que o time B era o único que tinha jogadores participantes na copa do mundo, conforme mostrado na figura 25. Essa questão mostra como as ontologias podem relacionar objetos do mundo real e deduzir, por meio de inferência automatizada, resposta mais inteligentes para o usuário.

As consultas feitas em linguagem SPARQL no Protégé conseguiram responder as questões de competência, ou seja, com esse resultado é possível afirmar que o conhecimento modelado conseguiu responder as questões de semântica que se buscou resolver na amostra do domínio do futebol.

É importante ressaltar que as consultas foram respondidas de acordo com o que foi modelado na ontologia, mas em um ambiente real de produção, isso seria automatizado, levando-se em consideração os requisitos do sistema e o comportamento do usuário perante àquela base de dados, conforme os termos de buscas mais utilizados.

De acordo com Sure, Staab e Stuber (2003) as avaliações de ontologia podem ser distinguidas em três tipos distintos: (i) focada em tecnologia; (ii) focada no usuário; (iii) focada na ontologia. Assim, nesta etapa da pesquisa, a avaliação consistiu na ontologia, no entanto, avaliou de forma mais reduzida, se ocupando na resolução das questões de competência. A próxima etapa preocupa-se com a avaliação focada no usuário, mostrando, de forma demonstrativa, como a ontologia seria utilizado em um sistema como os acervos digitais. Em pesquisas futuras, espera-se realizar uma avaliação focada na tecnologia, implementando assim, a ontologia em um contexto automatizado de um acervo digital.

5.2 Mensuração da recuperação da informação

Para certificar a eficácia do uso da ontologia para a recuperação da informação, foi feita de forma demonstrativa, a comparação entre os resultados obtidos no acervo digital sem o uso da ontologia, e os resultados prováveis que seriam alcançados com o uso da ontologia aqui desenvolvida. O acervo Estadão, ambiente de pesquisa deste trabalho, também foi utilizado nesta etapa. É importante lembrar que este trabalho não possui nenhuma restrição quanto a sua aplicabilidade a demais

ambientes, todavia, o acervo Estadão é utilizado como base para a coleta das informações e aplicabilidade desta pesquisa, assim sendo, esta pesquisa pode ser submetida a ambientes semelhantes aos acervos digitais de jornais, tais como bibliotecas, instituições, museus, arquivos, entre outros.

Para guiar este estágio foram desenvolvidas estratégias de busca que serão utilizadas para apoiar na demonstração da recuperação da informação, são elas:

- 1) Campeões no futebol no Brasil.
- 2) Goleadas no futebol brasileiro.
- 3) Zagueiros artilheiros no futebol brasileiro.

No desenvolver da primeira questão foi necessário localizar no Dicionário de Conceitos, os conceitos e seus sinônimos, nesse caso, para obter resultados com a estratégia de busca “Campeões no futebol no Brasil” foi preciso atribuir na busca do acervo os termos “futebol”, “*football*”, “caneco”, “taça”, “troféu”, “tropheu”, “trophéo”, conforme a figura 27:

Figura 27 - Recuperação no acervo Estadão do termo "Troféu"



Fonte: Captura de tela do resultado de busca no Acervo Estadão.

Os resultados obtidos no acervo do Estadão, com notícias datadas desde 1875 até dezembro de 2015, para os termos adicionados, estão apresentados no quadro 23:

Quadro 23 - Termos recuperados Acervo Estadão

	Termo	Resultado	Período
1.	Futebol	19.4248	1920 a 2015
2.	<i>Football</i>	3.128	1900 a 2015
3.	Taça	76.253	1.875 a 2015
4.	Troféu	27.196	1875 a 2015
5.	Tropheu	26.967	1875 a 2015
6.	Tropheo	26.967	1875 a 2015
7.	Caneco	1.458	1879 a 2015
8.	Campeão	104.128	1875 a 2015

Fonte: Elaborado pela autora.

Nos resultados obtidos é possível identificar que os termos considerados como sinônimos no Dicionário de Conceitos, apresentam resultados distintos, como é o caso do termo “*Football*” que traz resultados nas duas primeiras décadas do século XV, período em que o esporte ainda tinha a escrita conforme a sua origem inglesa. E também o termo “Futebol”, que traz resultados a partir da década de 20, do século passado. Ambos os termos possuem o mesmo significado, mas não estão relacionados no acervo.

Ao analisar os termos “Troféu”, “Tropheu” e “Tropheo”, que possuem o mesmo significado, no entanto, possuem formas de escritas diferentes devido às mudanças ortográficas ocorridas, os resultados obtidos são semelhantes, pois ao contrário dos termos “*Football*” e “Futebol” que não estão relacionados, o termo troféu consegue recuperar as distintas formas de escritas com apenas uma busca e adotando apenas um dos termos.

Os resultados obtidos com a inclusão do termo “Taça” na busca do acervo, trouxe também outros resultados que envolviam, por exemplo, termos como “Raça” e “Faça”. Esse problema pode ser resolvido com as relações formadas na *OntoFootballForNewspapers*, que relacionam o termo “Taça” à “Caneco”, “Troféu”, “Tropheu” e “Trophéo”, por meio dos sinônimos, e atribui a propriedade “Conquista” para relacionar os termos “Jogador” e “Time”. Assim como, os resultados obtidos com o termo “Caneco”, que em sua maioria, apresentaram documentos que se referiam ao recipiente destinado para armazenar líquidos, nas relações

formadas na ontologia é possível reunir as informações referentes ao futebol.

O acervo possui alguns filtros, que possibilitam selecionar os resultados obtidos nas buscas conforme os cadernos do jornal. É o caso do caderno Esporte, sessão do jornal em que são reunidas notícias esportivas. No entanto, o filtro se mostrou ineficiente, pois nas análises realizadas no acervo, são associadas notícias a partir da década de 1950, nesse caso, as notícias anteriores a esse período não são contempladas por este filtro.

Visto que os termos isolados não supriam a necessidade de informação evidenciada, foram adicionados, para reunir os diferentes termos, os operadores booleanos “AND” e “OR”. Os operadores booleanos são utilizados para combinar termos ou expressões em uma pesquisa. O operador “AND” (E), é usado para delimitar a pesquisa, ou seja, quando usado combina todos os termos presentes na estratégia de busca. O operador “OR” (OU), amplia a pesquisa, quando usado na estratégia de busca os resultados recuperados devem conter um termo ou outro. O termo “Brasil” não faz parte do Dicionário de Conceitos, no entanto, foi adicionado juntamente para tornar a busca mais completa. Os resultados obtidos com a combinação dos termos estão expostos no quadro 24:

Quadro 24 - Resultados obtidos com o uso de operadores booleanos

	Expressão	Resultados	Período
1	Campeao AND Futebol AND Brasil	39.804	1920 a 2015
2	Campeao AND <i>Football</i>	742	1903 a 2015
3	Caneco AND <i>Football</i>	4	1917 a 2011
4	Caneco AND Futebol	344	1921 a 2015
5	Caneco AND Futebol AND Brasil	269	1922 a 2014
6	Trofeu AND Futebol AND Brasil	8.876	1920 a 2015
7	Taça OR Caneco	154	1920 a 2015

Fonte: Elaborado pela autora.

Algumas questões podem ser observadas com os resultados obtidos na busca. Primeiramente, foi necessário incluir operadores booleanos nas buscas, assim sendo, para ter melhores resultados nas buscas, é preciso que o usuário tenha um letramento em pesquisa, e conheça dos mecanismos de buscas utilizando operadores booleanos, conforme a Figura 28.

Figura 28 - Demonstração da recuperação da Informação no acervo digital
Estratégia de busca 1



Fonte: Captura de tela Acervo Estadão.

Em relação aos resultados nas buscas não foi possível obter informações suficientes para levantar os “Campeões no futebol no Brasil” pois a pesquisa se tornou exaustiva, visto que os termos mesmo que combinados com os operadores booleanos, não estão relacionados, ou seja, eles estão presentes dentro do texto, mas são apresentados de maneira desconecta, muitas vezes tratam assuntos distintos.

O motor de busca que utiliza palavras-chave é automático, executando somente o rastreamento de palavras nos textos. Ao utilizar os operadores booleanos, muitas vezes, a combinação dos termos é feita na pesquisa, e não em um mesmo documento, como no exemplo da inclusão dos termos “Caneco AND *Football*”, em que os resultados apresentaram documentos que continham somente o termo “Caneco” ou o termo

“Football”. Em poucas situações foi possível analisar documentos que possuíam os dois termos juntos, como mostra a figura 29, exercendo assim a função do operador “AND”. Na maioria das vezes, mesmo ao incluir o operador AND, foi obtido a função do operador “OR”, que tem como finalidade ampliar os resultados de uma pesquisa, tornando a busca pelas informações desejadas um processo exaustivo.

Figura 29 - Demonstração da Recuperação da Informação - Estratégia de busca 1

*O capitão
Carlos
Alberto
Torres com a
taça (ou
caneço, na
intimidade),
na Copa do
Mundo de
1970:
reafirmação
da
supremacia
do craque
brasileiro*



Fonte: Acervo Estádio (14/05/1998, p.73)

ttu g Foram incluídas nas buscas estratégias mais frequentes utilizadas pelos usuários, como as expressões “campeões do futebol no Brasil” e “campeões do futebol brasileiro”, todavia, essas estratégias de busca não obtêm sucesso em ambientes como os acervos digitais, pois só recuperam os documentos que possuem a expressão exatamente igual à incluída na busca ou os termos isolados, segmentando a pesquisa e

recuperando muitos documentos que não atendem a necessidade de informação manifestada. Ao utilizar a ontologia é possível localizar os termos e os sinônimos que foram associados a ele, como por exemplo para o termo ‘campeão’ a expressão ‘colocar a mão no caneco’, assim como as relações que fazem o sistema entender que o time campeão é também o time que venceu o campeonato. Esse entendimento é de fácil compreensão por humanos, mas em um SRI é preciso indicar essas relações, e a ontologia permite essas relações mais complexas, que demandam da inteligência humana.

Na resolução da segunda estratégia de busca “Goleadas no futebol brasileiro”, foram incluídos os seguintes termos no buscador do acervo “Goleada”, “Chocolate”, “Baile”, “Passeio”, “Vareio”, “Gol” e “Goal”, conforme os sinônimos do Dicionário de Conceitos. Os termos “Futebol” e “Football” já foram adicionados na busca anterior. Os resultados obtidos estão apresentados no quadro 25.

Quadro 25 - Demonstração da Recuperação da Informação
Estratégia de busca 2

	Termo	Resultado	Período
1.	Goleada	12.280	1910 a 2015
2.	Chocolate	34.649	1875 a 2015
3.	Baile	46.274	1875 a 2015
4.	Passeio	84.732	1875 a 2015
5.	Vareio	1.556	1875 a 2015
6.	Gol	116.772	1875 a 2015
7.	Goal	1.928	1886 a 2015

Fonte: Elaborado pela autora.

Pode se observar que os termos sinônimos do mesmo domínio, possuem resultados diferentes e em períodos diferentes. O termo principal do Dicionário de Conceitos - “Goleada”, conseguiu recuperar assuntos referentes a goleadas do futebol, no entanto, algumas vezes, o termo é utilizado como “metáfora” para situações adversas, como o exemplo da figura 28, em que o termo é utilizado para contextualizar uma derrota fora de campo, nesse caso na política:

Figura 28- Notícia com o termo goleada

Essa goleada de projetos do Executivo provoca reclamações, revoltas e decepções que vão levar alguns parlamentares a desistir de concorrer a um novo mandato.

Fonte: Acervo Estadão (19/05/2002)

Além disso, os sinônimos de goleada, listados no Dicionário de Conceitos -“Chocolate”, “Baile”, Passeio” e “Vareio”, não estão relacionados com o domínio do futebol. O termo “chocolate”, por exemplo, traz resultados, em sua grande maioria, relacionados ao alimento derivado do cacau. Assim como o termo “baile” traz resultados relacionados à eventos festivos, como “baile de máscara”, o termo passeio relaciona ainda publicidades anunciadas nas edições dos jornais, como de agências de viagens, o termo “vareio” recupera documentos que contém assuntos relacionados a “varejo”, assim como o termo gol, recupera documentos com os termos “gel” e “sol” e o termo “goal”, que recupera também documentos que se referem ao contexto do futebol americano. Em tais casos, na ontologia, com o envolvimento dos conceitos sinônimos para afirmar que uma goleada = chocolate = baile = passeio = vareio e que uma goleada ocorre uma grande quantidade de gols, e todos esses conceitos estão subordinados ao conceito futebol, refinando assim os resultados obtidos na estratégia de buscas.

Quando combinados os termos com operadores booleanos, os resultados não foram diferentes, ainda continuaram trazendo as ambiguidades retratadas nas pesquisas isoladas, pois os termos não estavam relacionados com o domínio, assim como o operador AND permanecia assumindo a função de OR.

Quadro 26 - Demonstração da Recuperação da Informação com operadores booleanos - Estratégia de busca 3

	Expressão	Resultado	Período
1.	Goleada AND Futebol	9.550	1923 a 2015
2.	Goleada AND Brasil	8.587	1910 a 2015
3.	Chocolate AND Futebol	4.150	1920 a 2015
4.	Baile AND Futebol	7.404	1920 a 2015
5.	Passeio AND Futebol	10.521	1920 a 2015
6.	Vareio AND Futebol	127	1921 a 2014

Fonte: Elaborado pela autora

Ao pesquisar informações para responder a estratégia de busca - “Zagueiros artilheiros no futebol brasileiro” foram localizados no Dicionário de Conceitos os sinônimos para ‘Zagueiro’, são eles: Beque e *Beck*. No entanto, os resultados dos 3 termos foram bem distintos, e em períodos diferentes. Os resultados podem ser observados no quadro 27.

Quadro 27 - Demonstração da Recuperação da Informação
Estratégia de busca 3

	Termo	Resultado	Período
1.	Zagueiro	38.097	1905 a 2015
2.	Beque	931	1878 a 2015
3.	<i>Beck</i>	6.382	1875 a 2015
4.	Goleador	3.844	1940 a 2016
5.	Artilheiro	17.486	1875 a 2016
6.	Brocador	0	0

Fonte: Elaborado pela autora

O termo Zagueiro, mais atual, teve resultados a partir do século XX, já os termos Beque e Beck, mais antigos, mas que no domínio possuem o mesmo significado - jogador que se posiciona na parte defensiva, entre a linha média e as goleiras do campo, tiveram resultados no século XIX. A justificativa dos resultados dos termos “beque” e “beck” no decorrer do século XX é devido as buscas terem recuperado também muitos textos em que os termos eram associados à sobrenomes ou outras circunstâncias que não tinham relação com o futebol.

Os termos “goleador” e seus sinônimos “brocador” e “artilheiro”, incluídos no Dicionário de Conceitos, também apresentaram resultados distintos e em datas diferentes. O termo “brocador” não resultou em nenhum documento, assim como quando buscado isoladamente no buscador, resultando apenas quatro resultados, que faziam referência ao jogador Hernane Vidal de Souza, mais conhecido como Hernane Brocador, apelido concedido à jogadores que fazem muito gols.

Os resultados das combinações dos termos com operadores booleanos são apresentados no quadro 28.

Quadro 28 - Demonstração da Recuperação da Informação com operadores booleanos - Estratégia de busca 3

	Expressão	Resultado	Período
1.	Zagueiro AND Goleador	1.816	1940 a 2015
2.	Beque AND Goleador	33	1968 a 2014
3.	Beck AND Goleador	0	0
4.	Beque AND Artilheiro	85	1960 a 2014
5.	Beck AND Artilheiro	25	1910 a 2016
6.	Artilheiro AND Goleador	1.652	1960 a 2016

Fonte: Elaborado pela autora

As junções dos termos mostram resultados diferentes, quando comparados com a busca de forma isolada, apresentando uma redução dos documentos recuperados. No entanto, essa diminuição não é justificada devido à utilização do operador AND, em que deveriam ser retornados, em um mesmo documentos, ambos os termos solicitados, pois em muitos documentos apenas um dos termos está presente no texto. Assim, é possível afirmar que os operadores booleanos não tiveram tanta eficiência no acervo, visto que eles não combinam os termos como esperado.

Para um melhor desempenho das ontologias em um SRI, é importante aliar estudos preliminares sobre as características do sistema, o tipo de documento, o perfil do usuário, além de fazer o levantamento estatístico que permita identificar quais os termos mais buscados e os documentos mais acessados no sistema. Essas informações irão permitir que a ontologia seja modelada de forma mais assertiva, de acordo com as especificidades do SRI.

É importante ressaltar que a elaboração de ontologias não segue um modelo pré-definido, sendo construídas de forma distintas e obtendo resultados diferentes, mesmo que sejam alvo de um mesmo domínio. A qualidade e a consistência da ontologia irão depender dos seus desenvolvedores e o conhecimento que demandam sobre o conteúdo analisado. Esse conhecimento também será modelado de forma diferente dependendo do propósito da construção da ontologia, do seu público-alvo e do ambiente que será implementado. Por esta razão, na construção de ontologias que possuem um grau de rigorosidade mais elevado, faz-se necessário o envolvimento de especialistas do domínio e profissionais da Ciência da Informação e Computação trabalhando em conjunto, a fim de propor um melhor modelo que atenda ao sistema de informação que se deseja beneficiar.

6 CONSIDERAÇÕES FINAIS

Como visto, os acervos estudados na pesquisa são compostos de documentos que nasceram impressos e foram digitalizados e também de documentos já nascidos digitais, mas que possuem as mesmas características do formato impresso, com textos apresentados de forma linear e divididos em cadernos de grandes assuntos. Diante desse cenário, a pesquisa envolveu o uso de uma ontologia para resolver as implicações linguísticas presentes nesses ambientes e no domínio do futebol, do ponto de vista da RI. Assim, o objetivo geral da pesquisa foi modelar uma ontologia do domínio do futebol para o tratamento de variações diacrônicas e sincrônicas da língua, visando apoiar a recuperação da informação em acervos digitais de jornais.

A partir desse cenário e a fim de alcançar o objetivo geral, o primeiro objetivo específico foi analisar o domínio do futebol no contexto dos acervos digitais de jornais impressos. Como contribuição dessa etapa, foi evidenciado o envolvimento entre a fonte de informação jornalística e o esporte, e a relevância que uma representa para a outra. Destaca-se que esta pesquisa é continuidade do trabalho de graduação da autora, que avaliou o jornal impresso como fonte histórica para o futebol e atestou a sua importância, devido à longevidade e ao detalhamento dos acontecimentos pertinentes ao domínio presentes nos jornais (SANTOS, 2012). Ampliando a pesquisa, neste trabalho foi investigado como o jornal impresso e o domínio do futebol se comportam em ambientes digitais, sob o enfoque da organização e recuperação da informação.

Visto que o jornal impresso no Brasil teve seu surgimento em 1808 e os clubes mais antigos do futebol nasceram em 1900, é possível constatar que a história do futebol brasileiro está contemplada nas páginas dos folhetins. Além disso, o esporte também dispõe de significativo espaço nas edições diárias, os chamados cadernos esportivos. Os textos são trabalhados com profundidade e originalidade, possibilitando que informações detalhadas sobre jogos, jogadores, clubes e eventos sejam encontradas e reconstruídas.

Alguns dos principais jornais do país disponibilizam a memória brasileira descrita nas publicações jornalísticas por meio da Internet, possibilitando assim que maior número de usuários possa acessar esses documentos. Além do Acervo Estadão utilizado nesta pesquisa, que possui documentos a partir de 1875, podemos destacar também

como importantes o Acervo Folha¹², do jornal Folha de São Paulo, com publicações a partir de 1921, o Acervo O Globo¹³, com publicações desde 1925, o Jornal do Brasil¹⁴, com documentos desde 1891, ou até mesmo a Biblioteca Nacional Digital¹⁵, que dispõe de versões de folhetins portugueses datados de 1740, período que antecede o surgimento dos jornais no Brasil. Esses acervos reúnem muita informação e são importantes bases históricas, considerados patrimônios digitais e, além da história, também são fontes atualizadas, pois continuam sendo alimentadas pelas edições atuais.

Essas constatações defendem a importância do desenvolvimento e atualização de métodos que possibilitem manter as informações dos patrimônios digitais, como os acervos de jornais, acessíveis aos seus usuários. E quando nos referimos em acesso, precisamos contextualizar os elementos além da digitalização e armazenamento, mas os mecanismos de tratamento para que os documentos sejam encontrados pelos usuários. E com a visão dos acervos digitais, também é preciso considerar a pluralidade dos documentos digitalizados e os nascidos digitais e a transitoriedade do vocabulário inerente aos textos jornalísticos, que testemunham às transformações ocorridas na sociedade e as diferenças socioculturais e geográficas na evolução dos anos.

Sabido desse cenário, foi desenvolvido o segundo objetivo, que identificou os aspectos linguísticos do domínio do futebol que poderiam afetar a recuperação da informação nos acervos digitais de jornais impressos. Para alcançá-lo foram analisadas obras de referência do futebol, como dicionários, glossários especializados e textos. Encontramos dificuldades para utilizar ferramentas para a coleta automática de termos. Verificou-se que na maioria das vezes isso se deve pela baixa qualidade nas digitalizações e no processamento OCR, mas também por se tratarem de textos históricos, com muitos anos de existência. Diante dessa constatação, sugere-se

¹² **Acervo Folha.** 2016. Disponível em: <<http://acervo.folha.uol.com.br/>>. Acesso em: 31 mar 2016

¹³ **Acervo O Globo.** 2016. Disponível em: <<http://acervo.oglobo.globo.com/>>. Acesso em: 31 mar 2016

¹⁴ **Acervo Jornal do Brasil.** 2016. Disponível em: <<http://www.jb.com.br/paginas/news-archive/>>. Acesso em: 31 mar 2016.

¹⁵ **Hemeroteca Digital:** Biblioteca Nacional Digital. 2016. Disponível em: <<http://bndigital.bn.br/hemeroteca-digital/>>. Acesso em: 31 mar 2016

que trabalhos futuros possam identificar ou propor tecnologias para esse tipo de documento histórico. Nesta pesquisa optamos por realizar a coleta de termos de forma manual, apesar de mais criteriosa, é menos produtiva em termos quantitativos.

Tendo em mãos a amostra que iria compor a pesquisa, o terceiro objetivo buscou tratar os aspectos linguísticos identificados numa ontologia do domínio futebol. A partir da metodologia para construção de ontologias *OntoForInfoScience*, foram estruturados artefatos que abrangeram as variações dos conceitos. Grande parte dos problemas tratados envolveu a diacronia, referentes às mudanças ocorridas durante a existência dos acervos, principalmente relacionados à gramática e às diferentes formas de escrita, bem como aos termos estrangeiros que foram modificados para o português. A sincronia também se mostrou pertinente, principalmente no tratamento de termos polissêmicos, como aqueles oriundos de diferentes regiões geográficas.

A partir dessa modelagem foi possível definir os relacionamentos semânticos que a ontologia adotou para minimizar os problemas causados pela ambiguidade. Essa fase que antecede a formalização é fundamental, e apesar de ser feita em linguagem natural e informal, é a mais complexa, pois exige muito conhecimento, técnica e clareza do desenvolvedor. Tendo essa parte modelada, a inclusão em editor de ontologias é facilitada.

De maneira geral, tal ontologia não é considerada como totalmente finalizada, em razão do seu escopo de cobertura, pelo objetivo dessa pesquisa ser voltado para o seu desenvolvimento, a fim de defender a sua eficiência na recuperação da informação em acervos digitais de jornais, e também pela necessidade de as ontologias serem periodicamente atualizadas. Dessa forma, sugere-se também para trabalhos futuros, a ampliação da ontologia, de forma que considere mais termos dos contemplados na amostra, além da condução para validação por especialistas e a implementação em um acervo digital de jornal, de maneira que o conteúdo modelado possa ser utilizado para solucionar questões concretas (reais), aplicadas em um ambiente de produção.

Quanto à metodologia utilizada para construir a ontologia, ressalta-se que ela foi finalizada recentemente, no primeiro semestre de 2015, e teve como objetivo servir como ferramenta aos pesquisadores e profissionais da Ciência da Informação envolvidos com a Organização do Conhecimento. Diante disso, tendo em vista um ciclo de

desenvolvimento das pesquisas da área, destaca-se a importância de tal uso nessa pesquisa, pois serve como medida para avaliação da aplicabilidade prática da metodologia em um outro domínio.

O último objetivo buscou demonstrar como a ontologia do futebol construída poderia apoiar a recuperação da informação em acervos digitais de jornais no que diz respeito às implicações linguísticas. Observou-se que o acervo digital Estadão utilizado como base para a coleta das informações e aplicabilidade desta pesquisa, não conseguiu responder às necessidades de informações solicitadas na busca do sistema, exigindo um conhecimento em estratégias de pesquisa. No entanto, ainda com a introdução dessas estratégias, os resultados obtidos foram exaustivos e com muitos documentos irrelevantes, devido à ausência de relacionamentos entre os termos que formam o domínio. Com a inclusão da ontologia, a pesquisa por informações pode prescindir tanto letramento do usuário, pois com os relacionamentos formados, não é necessário realizar inúmeras buscas para recuperar conceitos e expressões equivalentes.

Um contraponto que deve ser levado em consideração é que o acervo digital do Estadão não facilita pesquisas que precisam fazer extração de termos ou até mesmo que necessitem reunir dados para desenvolvimento de métricas. Assim, o processamento dos resultados obtidos é feito de forma mais lenta, pois é preciso trabalhar com pequenas amostras, diferentemente de bases de dados que permitem a transferência de registros dos documentos, e facilitam o manuseio de grandes quantidades de dados e a utilização de ferramentas para estruturação.

Para trabalhos futuros sugerem-se estudos que avaliem as carências de outros sistemas de informação, como os repositórios institucionais, periódicos científicos e também revistas impressas - aquelas em formato *flip* que podem ser visualizados conforme um livreto impresso. Tais iniciativas podem contribuir para que esses ambientes ofereçam mais eficiência na recuperação das informações, possibilitando assim que sejam mais utilizadas como fontes de pesquisa. Propõe-se também, que outros domínios sejam modelados com os aspectos dos acervos digitais, como economia, política, e outros tradicionalmente presentes em jornais.

Por fim, pode-se afirmar que a presente pesquisa mostrou como principal contribuição para a Ciência da Informação o estudo de uma fonte de informação de documentos da imprensa tradicional, com a finalidade de simplificar a busca por informações nos acervos digitais que armazenam esses documentos, e tornar mais fácil aos usuários a

busca por informações. Tais iniciativas para disponibilizar a memória de forma organizada, e de fácil acesso, possibilitam que pesquisadores, estudantes, jornalistas e público em geral tenham uma ferramenta mais inteligente para a pesquisa da história de uma sociedade.

REFERÊNCIAS

ACERVO O ESTADO DE S. PAULO. **Acervo**. Disponível em: <http://acervo.estadao.com.br/>. Acesso em: 24 fev. 2016.

ACERVO O ESTADO DE S. PAULO (Ed.). **Heleno já pode jogar no Vasco**. 1949. Disponível em: <<http://acervo.estadao.com.br/pagina/#!/19490430-22684-nac-0011-999-11-not/busca/Heleno>>. Acesso em: 17 ago. 2015

ACERVO O GLOBO (Ed.). **Heleno vira alvo**. 1948. Disponível em: <<http://acervo.oglobo.globo.com/busca/?tipoConteudo=pagina&ordenacaoData=relevancia&allwords=HELENO+FOI+ALVO&anyword=&noword=&exactword=&decadaSelecionada=1940&anoSelecionado=1948>>. Acesso em: 17 ago. 2015

_____. **Campeão o Uruguai**. 1950. Disponível em: <<http://acervo.oglobo.globo.com/busca/?tipoConteudo=pagina&ordenacaoData=relevancia&allwords=campeão+o+uruguai&anyword;=&noword;=&exactword;=&decadaSelecionada=1950&anoSelecionado=1950&mesSelecionado=7>>. Acesso em: 17 ago. 2015.

ALMEIDA, M. B.; OLIVEIRA, V.; COELHO, K. Estudo exploratório sobre ontologias aplicadas a modelos de sistemas de informação: perspectivas de pesquisa em Ciência da Informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, set. 2010. Disponível em: <http://www.periodicos.ufsc.br/index.php/eb/article/view/10987/14369>>. Acesso em: 13 set. 2015.

ALMEIDA, M. B.; MENDONÇA, F. M.; AGANETTE, E. C. Interfaces entre ontologias e conceitos seminais da Ciência da informação: em busca de avanços na organização do conhecimento. In: **ENANCIB**, v. 4, 2013, Florianópolis. Anais eletrônicos... Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/2339/INTERFACES_ENTRE_ONTOLOGIAS.pdf?sequence=1>. Acesso em: 13 set. 2015.

ANDRADE, Maria Teresinha Tamanini; FERREIRA, Cristiano Vasconcelos;

PEREIRA, Hernane Borges de Barros. Uma ontologia para a Gestão do Conhecimento no Processo de Desenvolvimento de Produto. **Gestão & Produção**, São Carlos, v. 17, n. 3, p.537-551, jan. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-530X2010000300008>. Acesso em: 28 jun. 2016

ANJ. **Linha do tempo com a cronologia da história do jornal no Mundo**. 2008. Disponível em: <<http://www.anj.org.br/cronologia>>. Acesso em: 22 maio 2014.

ANSI/NISO Z39.19-2005. **Guidelines for the construction, format, and management of monolingual thesauri**. 2005.

Disponível em:

<http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf>. Acesso em: 23 jun.2016

ARANO, Silvia Los tesauros y las ontologías en la Biblioteconomía y la Documentación. **Thesauruses and ontologies**. Hipertext.net, 2005, n. 3.

ARELLANO, Miguel Angel. Preservação de documentos digitais. **Ci. Inf.**, Brasília, v. 33, n. 2, p.15-27, 2004. maio/ago. Disponível em: <<http://www.scielo.br/pdf/ci/v33n2/a02v33n2.pdf>>. Acesso em: 28 maio 2014.

ASSIS, Machado de. **O jornal e o livro**. Correio Mercanti. Rio de Janeiro, jan. 1859. Disponível em:

<<http://machado.mec.gov.br/images/stories/pdf/cronica/macr13.pdf>>.

Acesso em: 13 ago. 1989.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. 1st ed. Boston, MA: Addison-Wesley Longman Publishing Co., Inc., 1999.

BARANOW, Ulf Gregor. Perspectivas na contribuição da lingüística e de áreas afins à ciência da informação. **Ciência da Informação**, Brasília, v.12, n.1, p.23 35 jun., 1983. Disponível em:

<<http://revista.ibict.br/cienciainformacao/index.php/ciinf/article/view/1527/1145>>. Acesso em: 2 Set. 2014

BARROS, Ernesto. Craque Heleno de Freitas renasce no cinema. **Jornal do Commercio**. São Paulo, p. 1-1. 14 mar. 2012. Disponível em: <<http://jconline.ne10.uol.com.br/canal/cultura/cinema/noticia/2012/03/14/craque-heleno-de-freitas-renasce-no-cinema-35671.php>>. Acesso em: 17 ago. 2015

BARROS, Geraldo Monteiro de. Dicionário ilustrado de futebol. [Publicado em fascículos pela Revista Placar]. São Paulo: Placar-Abril, 1980.

BELKIN, Nicholas J. Intelligent information retrieval: whose intelligence? **SiteSeer**: 1996. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.145.2825&rep=rep1&type=pdf>> Acesso em: 22 jan 2015

BELLUZZO, Regina Celia Baptista. **O acesso e uso da informação no esporte**: uma competência cidadã. 2014. Disponível em: <<http://www.revistadocseme.prefeitura.sp.gov.br/wp-content/uploads/2014/05/oacessoeusodainformacaonoesporteapresentacao.pdf>>. Acesso em: 13 ago. 2015

BERNERS-LEE, Tim; HENDER, James; LASSILA, Ora. **The semantic web**: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. 2001. Disponível em: <<http://www.cs.umd.edu/~golbeck/LBSC690/SemanticWeb.html84A9809EC588EF21>>. Acesso em: 02 set. 2014.

BERNERS-LEE, T. Semantic Web - XML2000. [2005?]. Disponível em: <<http://www.w3.org/2000/Talks/1206-xml2k-tbl/Overview.html>>. Acesso em: 28 fev 2016.

BHO GAL, J.; MACFARLANE, A.; SMITH, P.. A review of ontology based query expansion. **Information Processing & Management**, v. 43, n. 4, p.866-886, jul. 2007. Disponível em: <<http://www.sciencedirect.com.ez46.periodicos.capes.gov.br/science/article/pii/S0306457306001476#>>. Acesso em: 01 fev. 2015.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesouros na construção de ontologias como instrumento de organização e recuperação da informação. **IBERSID: revista de sistemas de información y documentación**, número monográfico, p. 199-209, 2008.

BONIFÁCIO, A. S.; HEUSER, C. A. **Metadados Semânticos para Buscas em Bibliotecas Digitais**, 2002. Disponível em: <<http://www.uel.br/pessoal/ailton/Trabalhos/SemanaAcad-Ailton.html>>. Acesso em: 02 set. 2014.

BRASCHER, Marisa. **Tratamento automático de ambiguidades na recuperação da informação**, 1999. Brasília, DF: Universidade de Brasília.

_____. A Ambigüidade na Recuperação da Informação. **DataGramZero**, v.3, n.1, fev., 2002. Disponível em: <http://www.dgz.org.br/fev02/Art_05.htm>. Acesso em: 14 ago. 2014.

BRASCHER, Marisa; CAFÉ, Lígia. **Organização da Informação ou Organização do Conhecimento?** In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 9., 2008, São Paulo. Anais. São Paulo: Ancib, 2008. p. 1 - 14. Disponível em: <<http://enancib.ibict.br/index.php/enancib/ixenancib/paper/viewFile/3016/2142>>. Acesso em: 17 ago. 2015

BREITMAN, Karin et al. **Fundamentos de SPARQL**. Rio de Janeiro: [20--]. Color. Disponível em: <<http://www.w3c.br/cursos/dados-abertos/curso/Parte-3-Modulo-1-SPARQL.pdf>>. Acesso em: 28 fev. 2016

BREITMAN, Karin Koogan. **Web semântica: a internet do futuro**. Rio de Janeiro: LTC, 2010. xvii, 190 p

CALDAS, Alvaro. **Deu no jornal: o jornalismo impresso na era da Internet**. 2. ed. São Paulo: PUC-Rio; São Paulo: Loyola, 2004. 207p

CASTELLS, Manuel. **A galáxia da internet: reflexões sobre a internet, negócios e sociedade**. Rio de Janeiro: Zahar, 2003. 243p.

CAMPOS, M. L. A. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, jan./abr. 2004. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/download/56/53>>. Acesso em: 12 ago. 2015.

CAMPOS, M. L. A., CAMPOS, Linair Maria e MEDEIROS, Jackson da Silva. A representação de domínios de conhecimento e uma teoria de representação: a ontologia de fundamentação. **Inf. Inf**, Londrina, v. 3, n. 16, p.140-164, jan. 2011. Semestral. Disponível em: <https://chasqueweb.ufrgs.br/~jacksonmedeiros/pubs/2011_onto_fund_inf.pdf>. Acesso em: 25 jan. 2016.

CAMPOS, M. L. A., O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 15, p.220-238, abr. 2010. Disponível em: <<http://www.scielo.br/pdf/pci/v15n1/13.pdf>>. Acesso em: 24 fev. 2016.

CAMPOS, M. L. A.; CAMPOS, Linair Maria. **A organização do conhecimento e suas teorias de representação: a ontologia de fundamentação como um modelo teórico para a representação de domínios**. In: encontro nacional de pesquisa em ciência da informação, 13., 2012, Recife: Fiocruz, 2012. p. 1 - 18. Disponível em: <<http://200.20.0.78/repositorios/bitstream/handle/123456789/648/4.pdf?sequence=1>>. Acesso em: 24 fev. 2016.

CARRASCO, Laís Barbudo; THALLER, Manfred; VIDOTTI, Silvana Aparecida Borsetti Gregório. Ontologia Cidoc CRM no contexto dos ambientes digitais de patrimônios culturais. **Liinc em Revista**, Rio de Janeiro, v. 1, n. 11, p.208-222, maio 2015. Disponível em: <<http://revista.ibict.br/liinc/index.php/liinc/article/view/789/538>>. Acesso em: 24 fev. 2015.

CARRAVETTA, E. S. P. **Modernização da gestão no futebol brasileiro: perspectivas para a qualificação do rendimento competitivo**. Porto Alegre: Age, 2006.

CATARINO, Maria Elisabete; SOUZA, Terezinha Batista de. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, v. 2, n. 24, p.77-90, maio 2015. Trimestral. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/766/746>>. Acesso em: 13 ago. 2015

CAVALCANTE, L. E. . Patrimônio digital e informação: política, cultura e diversidade. **Encontros Bibli**, v. 23, p. 152-170, 2007. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/viewFile/1518-2924.2007v12n23p152/401>>. Acesso em: 01 ago 2015.

CHANDRASEKARAN, B. et al. What Are Ontologies, and Why Do We Need Them? **Ieee Intelligent Systems**, v. 14, n. 1, p.20-26, jan. 1999. Disponível em: <<http://www.csee.umbc.edu/courses/771/current/papers/chandrasekaranetal99.pdf>>. Acesso em: 01 fev 2015.

CHAPOUTHIER, Georges. **Registros evolutivos**. Viver Mente & Cérebro: Memória, São Paulo, n. 2, p. 8-13, jul. 2006

CORRÊA, Elizabeth Saad; BERTOCCHI, Daniela. A cena cibercultural do jornalismo contemporâneo: web semântica, algoritmos, aplicativos e curadoria. **Matrizes**, São Paulo, v. 2, n. 5, p.123-144, jan. 2012. Semestral. Disponível em: <<http://www.matrizes.usp.br/index.php/matrizes/article/view/340/pdf>>. Acesso em: 13 ago. 2015

CURRÁS, Emilia. **Tesouros, linguagens terminológicas**. Tradução de Antônio Felipe Corrêa da Costa. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1995.

DAHLBERG, I. Teoria do Conceito. **Ciência da Informação**, v. 7, n. 2, p. 101–107, 1978

FERNÁNDEZ, M. et al. Building a chemical ontology using methontology and the ontology design environment. **Intelligent Systems**, v. 14, n. 1, p. 37-46, Jan./Feb.1999

DCMI - DUBLIN CORE METADATA INITIATIVE. **Dublin Core Qualifiers**. 2007 Disponível em: <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>. Acesso em: 01 fev 2015.

DEGEN, W.; HELLER, B.; HERRE, H.; SMITH, B. GOL: **Toward an axiomatized upperlevel ontology**. In: Proceedings of the 2nd Int. Conf. on Formal Ontology in Information Systems. New York, USA: ACM, 2001. p.34–46. Ogunquit, USA, 2001. Disponível em: <<http://ontology.buffalo.edu/smith/articles/gol2001-report.pdf>>. Acesso em: 30/11/2015

DINES, Alberto. **O papel do jornal: uma releitura**. 7. ed. São Paulo: Summus, 2001. 157p.

DODEBEI, Vera. **Patrimônio digital virtual: Herança, documento e informação**. Trabalho apresentado na 26ª. Reunião Brasileira de Antropologia, realizada entre os dias 01 e 04 de junho, Porto Seguro, Bahia, Brasil. 2006. Disponível em: <http://www.abant.org.br/conteudo/ANAIS/CD_Virtual_26_RBA/grupos_de_trabalho/trabalhos/GT%2037/vera%20dodeber.pdf> Acesso em: 09 ago. 2014.

_____. Cultura digital: novo sentido e significado de documento para a memória social? **DataGramZero**, v. 12, n. 2, 2011. Disponível em: <http://www.brapci.ufpr.br/brapci/repositorio/2011/04/pdf_15b7c5a842_0016272.pdf>. Acesso em: 09 ago. 2014.

_____. **Tesouro: linguagem de representação da memória documentaria**. Rio de Janeiro: Intertexto, 2002.

FARIA, Mário Joaquim Firmino Leite. **Definição de uma Ontologia Aplicada ao Futebol**. 2009. 147 f. Dissertação (Mestrado) - Curso de Engenharia Electrotécnica e de Computadores, Faculdade de Engenharia da Universidade do Porto, Porto, 2009. Disponível em: <<http://paginas.fe.up.pt/~ee99157/FariaTese.pdf>>. Acesso em: 24 fev. 2016.

FEITOSA, Ailton. Organização da informação na Web: das tags à web semântica. Brasília, DF: Thesaurus, 2006. 131p

FILHO, Fernando Wagner B. H; LÓSCIO, Bernardete F; De Campos, Gustavo A. L. Inferência sobre Ontologias no contexto da Web Semântica. **Info Brasil**. 2008. Disponível em:

<<http://www.infobrasil.inf.br/userfiles/Infer%C3%A2ncia%20sobre%20Ontologias%20no%20contexto%20da%20Web%20Sem%C3%A2ntica.pdf>>. Acesso em: 23/02/2016.

FERNÁNDEZ-LÓPEZ, M. et al. Building a chemical ontology using methotology and the ontology design environment. [S. l.]: **IEEE** Intelligent Systems & their Applications, p. 37-46, Jan./Feb. 1999

FERRARI, Pollyana. **Jornalismo digital**. São Paulo: Contexto, 2006. 120p.

FOGL, J. **Relations of the concepts 'information' and 'knowledge'**. International Fórum on Information and Documentation, The Hague, v.4, n.1, p. 21-24, 1979

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184p.

GILLILAND-SWETLAND, A. **Defining metadata**. In: Metadata: Pathways to Digital Information. Murtha Baca., ed. Getty Information Institute, 1997. Pp.1-8.

GÓMEZ-PÉREZ, A. **Ontology Evaluation. Handbook on Ontologies**. S. Staab and R. Studer Editors. Springer. International Handbooks on Information Systems. Pp: 251-274. 2004.

GONZALEZ, M.; LIMA, V. L. S. **Recuperação de informação e processamento da linguagem natural**. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 23., 2003, Campinas. Anais da Jornada de Mini-Cursos de Inteligência Artificial, 3. Campinas: [s.n.], v. 3, p.347-395. Disponível em:

<<https://www.inf.pucrs.br/~gonzalez/docs/minicurso-jaia2003.pdf>>.

Acesso em: 01 fev 2015.

GUARINO, N. **Formal Ontology in Information Systems**.

Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15, 1998.

GUIMARÃES, Caroline Burle dos Santos; DINIZ, Vagner. Dados Abertos Conectados: m busca da Web do Conhecimento. São Paulo: Novatec, 2015. Disponível em: <<http://ceweb.br/livros/dados-abertos-conectados/capitulo-3/>>. Acesso em: 24 fev. 2016

GUIRAUD, Pierre. **A semântica**. 2. ed. São Paulo: DIFEL, 1975. 133p.

GUIZZARDI, Giancarlo; FALBO, Ricardo A.; GUIZZARDI, Renata S.S. A importância de Ontologias de Fundamentação para a Engenharia de Ontologias de Domínio: o caso do domínio de Processos de Software. **Ieee Latin America Transactions**, v. 3, n. 6, p.244-251, jul. 2008.

Mensal. Disponível em:

<<http://www.inf.ufes.br/~gguizzardi/IEEE2008.pdf>>. Acesso em: 24 fev. 2016.

GUTERMAN, Marcos. **O futebol explica o Brasil: uma história da maior expressão popular do país**. São Paulo: Contexto, 2009. 270 p.

HAAV, Hele-Mai; LUBI, Tanel-Lauri. **A Survey of Concept-based Information Retrieval Tools on the Web**.2001. Disponível em:

<<http://www.mii.vu.lt/ADBIS/local2/haav.pdf>>. Acesso em: 24/02/2016.

HERNÁNDEZ-PÉREZ, Tony. **Bibliotecas digitais: de la web social a la web semántica**. In: CIANCONI, Regina de Barros; CORDEIRO, Rosa Inês Novais; ALMEIDA, Carlos Henrique Marcondes de (Org.). Gestão do Conhecimento, da Informação e de Documentos em Contextos Informacionais. 3. ed. Niterói: Editora da Uff, 2013. Cap. 3. p. 191-209.

JENKINS, Henry. **Cultura da convergência**. São Paulo: Aleph, 2008. 380p.

JORGE, Carlos Francisco Bitencourt. **Gestão da informação esportiva no contexto da inteligência competitiva em clubes de futebol: um estudo de caso no marília atlético clube.** 2013. 322 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/jorge_cfb_me_mar.pdf>. Acesso em: 13 ago. 2015.

JORGE, Carlos Francisco Bitencourt; VALENTIM, Marta Lígia Pomim. **Informação e esporte: a informação esportiva e sua relação com clubes de futebol.** Inf. Inf, Londrina, v. 1, n. 20, p.183-208, jan. 2015. Trimestral. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/19712/pdf_50>. Acesso em: 13 ago. 2015

KOBASHI, Nair Yumiko; TÁLAMO, Maria de Fátima Gonçalves Moreira. Informação: fenômeno e objeto de estudo da sociedade contemporânea. **Transinformação**, Campinas, v. 15, n. 1, p.7-21, set. 2003. Trimestral. Disponível em: <<http://www.scielo.br/pdf/tinf/v15nspe/01.pdf>>. Acesso em: 13 ago. 2015

KROVETZ, Robert; CROFT, William Bruce. Lexical Ambiguity and Information Retrieval. **ACM Transactions on Information Systems**, v.10, n.2. p. 115-141, 1992. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.1928&rep=rep1&type=pdf>>_Acesso em: 01 fev 2015.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de metodologia científica.** 3ed. São Paulo: Atlas, 1991.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Metodologia científica: ciência e conhecimento científico, métodos científicos, teoria, hipóteses e variáveis, metodologia jurídica.** 6. ed. rev. e ampl. São Paulo: Atlas, 2012. 314 p.

LE GOFF, Jacques. **História e memória**. Campinas: UNICAMP, 1990, 553p

LE GOFF, Jacques. Documento/monumento. In: **Historia e memória**. 5. ed. Campinas, SP: UNICAMP, 2003. p. 525-539.

LÉVY, Pierre. **Cibercultura**. São Paulo: Ed. 34, 2000. 260p.

LOPES, Ilza Leite. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf**, Brasília, v. 1, n. 31, p.41-52, abr. 2002. Disponível em: <<http://www.scielo.br/pdf/ci/v31n1/a05v31n1.pdf>>. Acesso em: 24 fev. 2016.

LOPES, Carlos. A imagem e o sonho da arquivística. Rio de Janeiro: Arquivo Público, 1998.

MANINI, Miriam Paula; LIMA-MARQUES, Mamede; MIRANDA, Alex Santos Sandro. **Ontologias**: indexação e recuperação de fotografias baseadas na técnica fotográfica e no conteúdo da imagem. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. Anais... Salvador, BA: UFBA, 2007. Disponível em: <<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/246/GT2--151.pdf?sequence=1>>. Acesso em: 01 fev. 2015.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge, Massachusetts / London, England: The MIT Press, 1999. Disponível em: <http://ics.upjs.sk/~pero/web/documents/pillar/Manning_Schuetze_StatisticalNLP.pdf> Acesso em: 01 fev 2015.

MARTELOTTA, Mário Eduardo. **Mudança linguística: uma abordagem baseada no uso**. São Paulo: Cortez, 2011.

MÁXIMO, João. **Memórias do futebol brasileiro**. Estudos Avançados, [s.l.], v. 13, n. 37, p.179-188, set. 1999. FapUNIFESP (SciELO). Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-40141999000300009&script=sci_arttext>. Acesso em: 17 ago. 2015.

MELO, Adriano; CHAGAS, André; FREITAS, Fred. **Protégé-OWL**: Recife: Centro de Informática Ufpe, [20--]. 39 slides, color. Disponível em: <<http://www.cin.ufpe.br/~if684/EC/aulas/protege-owl.pdf>>. Acesso em: 02 mar. 2016

MELO, Fábio J. Dantas de; BRASCHER, Marisa. **Fundamentos da Linguística para a formação do profissional da informação**. Brasília: Thesaurus, 2011.

MENDONÇA, Fabrício Martins; ALMEIDA, Maurício Barcellos. Princípios metodológicos para desenvolvimento de ontologias: análise das práticas correntes e proposição de melhorias. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Além das nuvens, expandindo as fronteiras da Ciência da Informação**. Belo Horizonte: Eci/ufmg, 2014. v. 1, p. 430 - 447. Disponível em: <<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/2602/Mendonca;Almeida.pdf?sequence=1>>. Acesso em: 04 set. 2015.

MENDONÇA, Fabrício Martins. **Ontoforinfoscience: metodologia para construção de ontologias pelos cientistas da informação**: Uma aplicação prática no desenvolvimento da ontologia sobre componentes do sangue humano (HEMONTA). 2015. 311 f. Tese (Doutorado) - Curso de Ciência da Informação, Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2015.

MOOERS, C. Zatoncoding Applied to Mechanical Organization of Knowledge. **American Documentation**, 1951 v.2, n.1, p.20-32. Disponível em: <<https://courses.engr.illinois.edu/cs473/fa2013/misc/zatoncoding.pdf>>. Acesso em: 10 set. 2015.

MONARD, M.C.; BARANAUSKAS, J.A. **Conceitos de aprendizado de máquina**. In: REZENDE, S.O. Sistemas inteligentes: fundamentos e aplicações. Barueri, SP: Manole, 2003. Cap.4, p.89-114. Disponível em: <<http://dcm.ffclrp.usp.br/~augusto/publications/2003-sistemas-inteligentes-cap4.pdf>>. Acesso em: 09 ago. 2014.

MONTENEGRO, Rosilene Dias ; SILVA, Fábio Ronaldo da . **Por uma digitalização da memória jornalística**. BOCC. Biblioteca On-line de Ciências da Comunicação, Portugal, p. 1-10, 2005. Disponível em: <<http://bocc.ubi.pt/pag/rosilene-montenegro-fabio-silva-memoria-jornalistica.pdf>>. Acesso em: 13 ago. 2015.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A de P. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. **DataGramZero** - Revista de Ciência da Informação, v. 5 n. 6. 2004. Disponível em: <http://www.dgz.org.br/dez04/Art_01.htm>. Acesso em: 01 fev 2015.

MOREIRA, Alexandra. **Proposta de um framework apoiado em ontologias para detecção de frames**. 2012. 176 f. Tese (Doutorado) - Curso de PÓS-graduaÇÃO em Linguística, Universidade Federal de Juiz de Fora, Juiz de Fora, 2012. Disponível em: <<http://www.ufjf.br/ppglinguistica/files/2009/12/Alexandra-Moteira-Tese-2012.pdf>>. Acesso em: 24 fev. 2016.

MOREIRO GONZÁLEZ, José Antônio. **Linguagens documentárias e vocabulários semânticos para a web**. Salvador: Edufba, 2011.

MOURA, Dione Oliveira. O ‘sincrônico’, o ‘diacrônico’, o acontecimento e a errância de sentidos na análise do discurso jornalístico. **Comunicação & Informação**, Brasil, v. 2, n. 12, p.63-73, jul. 2009. Semestral. Disponível em: <<http://www.revistas.ufg.br/index.php/ci/article/view/12270/8132>>. Acesso em: 15 set. 2015

PÁDUA, Elisabete Matallo Marchesini de. **Metodologia da pesquisa: abordagem teórico-prática**. 13. ed. Campinas: Papirus, 2007. 124p.

PEDRAZA-JIMÉNEZ, R.; CODINA, L.; ROVIRA, C.. (2007). Web semántica y ontologías en el procesamiento de la información documental. // **El profesional de la información**, Barcelona, v.16, n.6, p. 569-578, nov. 2007. Disponível em: <<http://www.elprofesionaldeinformacion.com/contenidos/2007/noviembre/04.pdf>>. Acesso em: 15 set. 2015.

PALACIOS. Marcos. **Ruptura, Continuidade e Potencialização no Jornalismo Online**: o Lugar da Memória. In: MACHADO, Elias;

PALACIOS, Marcos (Orgs.), Modelos do Jornalismo Digital. Salvador: Editora Calandra, 2003. p.1-17.

PICKLER, Maria Elisa Valentim. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em ciência da informação**, Belo Horizonte, v. 12, n. 1, p.65-83, jan. 2007. Mensal. Disponível em:

<<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/251/468>>. Acesso em: 02 set. 2014.

QUEIROZ, João Machado de. **Vocabulário do futebol na mídia impressa: o glossário da bola**. 2005. 904 f. Tese (Doutorado) - Curso de Letras, Faculdade de Ciências e Letras de Assis, Universidade Estadual Paulista, Assis, 2005. Disponível em:

<http://repositorio.unesp.br/bitstream/handle/11449/102469/queiroz_jm_dr_assis.pdf?sequence=1>. Acesso em: 03 dez. 2015

REIS, Luiz Paulo; LAU, Nuno. **Coach Unilang: A Standard Language for Coaching a (Robo) Soccer Team**. Disponível em:

<[http://eaia07.di.fc.ul.pt/docs/reis/Reis%20Atelier/paginas.fe.up.pt/_lpreis/EAIA07/Docs/COACH_UNILANG_A_Standard_Language_for_Coaching_a_\(Robo\)_Soccer_Team.pdf](http://eaia07.di.fc.ul.pt/docs/reis/Reis%20Atelier/paginas.fe.up.pt/_lpreis/EAIA07/Docs/COACH_UNILANG_A_Standard_Language_for_Coaching_a_(Robo)_Soccer_Team.pdf)>. Acesso em: 24/02/2016

RIBOLDI, Ari. Cabeça de bagre: termos, expressões e gírias do futebol. São Paulo: AGE, 2008.

ROBREDO, Jaime. **Da ciência da informação revisitada aos sistemas humanos de informação**. Brasília, DF: Thesaurus Editora: SSRR Informações Consultoria e Projetos Ltda, 2003. xii, 245p

_____. **Do documento impresso à informação nas nuvens: reflexões**. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 11, 2010, Rio de Janeiro Brasília: Ibict, 2010 a. 26 p. Disponível em:

<http://repositorios.questoesemrede.uff.br/repositorios/bitstream/handle/123456789/982/Do_documento_-_Robredo.pdf?sequence=1>. Acesso em: 09 ago. 2014.

_____. **Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas?**. In: ROBREDO, Jaime;

BRÄSCHER, Marisa (Orgs.) **Passeios pelo bosque da informação:** estudos sobre a representação e organização da informação e do conhecimento – eróico . Brasília DF: IBICT, 2010b. 329 p. Disponível em:
<http://repositorio.unb.br/bitstream/10482/7952/6/CAPITULO_CiencialnformacaoWebSemantica.pdf>. Acesso em: 09 ago.2014.

RUEDA, Valéria Matias da Silva; FREITAS, Aline de; VALLS, Valéria Martin. Memória Institucional: uma revisão de literatura. **CrB-8 Digital**, São Paulo, v. 8, n. 1, p.78-89, abr. 2011. Disponível em:
<<http://revista.crb8.org.br/index.php/crb8digital/article/viewFile/62/64>>. Acesso em: 27 jun. 2016.

SALES, Rodrigo de; CAFÉ, Ligia. Semelhanças e Diferenças entre Tesouros e Ontologias. **Datagramazero**, Rio de Janeiro, v. 8, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_02.htm>. Acesso em: 23 fev. 2016.

SANTOS, Antonio Raimundo dos. **Metodologia científica:** a construção do conhecimento. 3.ed.Rio de Janeiro: DP&A Editora, 2000.

SANTOS, Luana Carla de Moura dos. **Clipping como fonte histórica no futebol:** o arquivo histórico do figueirense futebol clube (FFC). 2012. 90 f. TCC (Graduação) - Curso de Biblioteconomia, Centro de Ciências Humanas e da Educação,, Universidade do Estado de Santa Catarina, Florianópolis, 2012. Disponível em:
<<http://sistemabu.udesc.br/pergamumweb/vinculos/000000/000000000016/000016FA.pdf>>. Acesso em: 31 mar. 2016.

SANTOS, Luana Carla de Moura dos; VIERA, Angel Freddy Godoy. Avaliação da recuperação da informação em acervos digitais de jornais. **Em Questão**, Porto Alegre, v. 2, n. 21, p.49-73, maio 2015. Trimestral. Disponível em:
<<http://www.seer.ufrgs.br/EmQuestao/article/viewFile/52362/35082>>. Acesso em: 20 mar. 2016.

SANTOS, M T.; CORREA, R. F.; SILVEIRA, M. A. A. Estudos brasileiros sobre ontologia na Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 14, n. 1, fev. 2013. Disponível em:
<http://www.dgz.org.br/fev13/Art_05.htm>. Acesso em: 12 set. 2014.

SANTOS, Plácida Leopoldina Ventura Amorim da Costa; ALVES, Rachel Cristina Vesú. Metadados e Web Semântica para estruturação da Web 2.0 e Web 3.0. **Datagramazero**, Rio de Janeiro, v. 6, n. 10, dez. 2009. Disponível em: <http://www.dgz.org.br/dez09/Art_04.htm>. Acesso em: 28 fev. 2016

SARACEVIC, T. Ciência da informação: origem, evolução e relações. Perspec. **Ci.Inf.**, Belo Horizonte, v.1, n.1, p.41-62, jan./jun. 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 01 fev 2015.

_____. **Lexicalização de Ontologias: o relacionamento entre conteúdo e significado no contexto da Recuperação da Informação**. 2015. 261 f. Tese (Doutorado) - Curso de Ciência da Informação, Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2015

SCHIESSL, Marcelo; BRÄSCHER, Marisa. Do texto às ontologias: uma perspectiva para a ciência da informação. **Ciência da Informação**, [S.l.], v. 40, n. 2, set. 2011. ISSN 1518-8353. Disponível em: <<http://revista.ibict.br/cienciadainformacao/index.php/ciinf/article/view/1947/1427>>. Acesso em: 04 Set. 2015.

SCHWINGEL, Carla; ZANOTTI, Carlos A. **Produção e colaboração no jornalismo digital**. Florianópolis: Insular, 2010. 239 p.

SILVA, Augusto Soares da. **O cognitivo e o social nos estudos linguísticos: inimigos íntimos?** In: ENCONTRO NACIONAL DA ASSOCIAÇÃO PORTUGUESA DE LINGUÍSTIC, XXIV., 2009, Lisboa: 2009. p. 511 - 525. Disponível em: <<http://www.apl.org.pt/docs/24-textos-selecionados/35-Silva.pdf>>. Acesso em: 24 fev. 2016

SILVA, Edna Lúcia da; MENEZES, Estera Muszkat. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. rev. atual, Florianópolis: UFSC, 2005. Disponível em:

<http://200.17.83.38/portal/upload/com_arquivo/metodologia_da_pesquisa_e_elaboracao_de_dissertacao.pdf>. Acesso em: 17 ago. 2015

SILVA, D. L.; SOUZA, R. R.; e ALMEIDA, M. B. Ontologias e vocabulários controlados: comparação de metodologias para construção. **Ciência da Informação**, Brasília, v. 37, n. 3, p. 60-75, set./dez. 2008. Disponível em: <<http://www.scielo.br/pdf/ci/v37n3/.pdf>>. Acesso em: 04 set. 2015.

SILVA, George Batista da. Além das quatro linhas: dicionário futebolístico. Rio de Janeiro: Clube de Autores, 2014.

SIQUEIRA, André Henrique. Sobre a natureza da tecnologia da informação. **Ci. Inf.**, Brasília, v. 37, n. 1, p.85-94, jan. 2008. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-19652008000100008&script=sci_arttext>. Acesso em: 17 ago. 2015

SHAMSFARD, M.; BARFOROUSH, A. A. The state of the art in ontology learning: a framework for comparison. **The Knowledge Engineering Review**, Cambridge University Press, v. 18, n. 4, p. 293–316, 2003. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=CD6C53A35C907312637D475E2296952A?doi=10.1.1.99.5679&rep=rep1&type=pdf>>. Acesso em: 03 set. 2015.

SOERGEL, D. Functions of a Thesaurus / Classification / Ontological Knowledge Base. **College of Library and Information Services**, University of Maryland, 1997. Disponível em: <<http://www.dsoergel.com/cv/soergelfctclass.pdf>>. Acesso: 13 set. 2015.

SOUZA, Renato Rocha. Sistemas de Recuperação de Informações e Mecanismos de Busca na web: panorama atual e tendências. In.: **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 2, p.161-173, mai./ago.2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02>>. Acesso em: 01 fev 2015.

SOUZA, R. R.; ALVARENGA, L. A web semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v.

33, n. 1, p. 132-141, jan./ abr. 2004. Disponível em:

<http://www.scielo.br/scielo.php?pid=S0100-19652004000100016&script=sci_arttext>. Acesso em: 09 ago. 2014.

STUDER, Rudi; BENJAMINS, V. Richard; Fensel, Dieter. Knowledge Engineering: Principles and methods. **Data & Knowledge Engineering**, n.25, 1998, p.161-167.

Disponível em:

<<http://www.it.iitb.ac.in/~palwencha/ES/Knowledge%20engineering%20-%20Principles%20and%20methods.pdf>>. Acesso em: 10 ago. 2014.

SURE, Y.; STAAB, S.; e STUBER, R. **On-To-Knowledge Methodology (OTKM)**. 2003.

Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.14.798&rep=rep1&type=pdf>>.

Acesso em: 08 de maio de 2016.

TELLAROLI, Taís Marina; ALBINO, João Pedro. **Da sociedade da informação às novas tic's: questões sobre internet, jornalismo e comunicação de massa**.

Núcleo de Conforto

Ambiental – UNESP. 2007. Disponível em:

<<http://www.faac.unesp.br/publicacoes/anaiscomunicacao/textos/28.pdf>>.

Acesso em: 14 Ago 2014.

TRIVIÑOS, Augusto N. S. **Introdução à pesquisa em Ciências Sociais: a pesquisa qualitativa em Educação**. São Paulo: Atlas, 2010.

UNESCO. **Carta sobre a preservação do patrimônio digital**. 2003.

Disponível em:

< <http://unesdoc.unesco.org/images/0013/001311/131178e.pdf> > Acesso em : 10 ago. 2015

USCHOLD, Mike; JASPER, Robert. **A Framework for Understanding and Classifying**

Ontology Applications. 1999. Disponível em: <<http://ceur-ws.org/Vol-18/11-uschold.pdf>>. Acesso em: 10 ago. 2014.

VALENTIM, M. L. P. Processo de inteligência competitiva organizacional. In: VALENTIM, M. L.P. (Org.). **Informação, conhecimento e inteligência organizacional**. 2.ed. Marília: FUNDEPE Editora, 2006. 282p; p.9-24.

VARELA, Aida Varela; BARBOSA, Marilene Lobo Abreu. O caráter técnico, social e mediador da organização do conhecimento. In: DODEBEI, Vera; GUIMARÃES, José Augusto Chaves (Org.). **Complexidade e Organização do Conhecimento: Desafios do nosso século**. 2. ed. Rio de Janeiro: Fundepe, 2013. p. 277-282. Disponível em: <<http://isko-brasil.org.br/wp-content/uploads/2013/02/Estudos-avancados-2.pdf>>. Acesso em: 13 ago. 2015

VAZ, Paulo. Mediação e tecnologia. **Revista Famecos**, Porto Alegre, v. 1, n. 16, p.45-59, 2001. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/view/3137/2408>>. Acesso em: 06 maio 2014.

VIEIRA, Angel F. G. GARRIDO, Isadora dos S. Folksonomia como uma estratégia para recuperação colaborativa da informação. **DataGamaZero** – Revista de Ciência da Informação, v. 12, n. 2, abr. 2011. Disponível em: <http://www.dgz.org.br/abr11/Art_02.htm>. Acesso em: 01 fev 2015.

VICKERY, BC. Ontologies. **Journal of Information Science**, v.23, n.4, p. 227 – 286, Jan. 1997.

W3C. **SPARQL 1.1 Query Language**. 2013. Disponível em: <<https://www.w3.org/TR/sparql11-query/>>. Acesso em: 28 fev. 2016.

APÊNDICE A – GLOSSÁRIO DE CONCEITOS

	Glossário de Conceitos (coleta manual)
1.	Armação
2.	Assistência
3.	Avenida
4.	Banco
5.	Bandeirinha
6.	Banheira
7.	Banho de bola
8.	Banho- Maria
9.	Barreira
10.	Bicicleta
11.	Bico
12.	Bola
13.	Bola do jogo
14.	Boleiro
15.	Bomba
16.	Cama de gato
17.	Campeão
18.	Caneco
19.	Caneta
20.	Cartão amarelo
21.	Cartão vermelho
22.	Cartola
23.	Catimba
24.	Cera
25.	Chaleira
26.	Chegada
27.	Chiqueiro
28.	Chute de primeira
29.	Chuteira
30.	Chuveiro
31.	Clássico
32.	Conclusão

33.	Corpo Mole
34.	Corta-Luz
35.	Cozinhar
36.	Drible
37.	Elástico
38.	Escanteio
39.	Firula
40.	Freguês
41.	Furada
42.	Futebol
43.	Gaveta
44.	Geladeira
45.	Gol
46.	Goleada
47.	Goleador
48.	Gramado
49.	Impedimento
50.	Juiz
51.	Partida
52.	Passe
53.	Tiro de Meta
54.	Traves
55.	Zagueiro
56.	Zebra

Elaborado pela autora.

APÊNDICE B – DICIONÁRIO DE CONCEITOS

Dicionário de Conceitos			
ID	Conceito	Sinônimos	Definição
1.	Adversário	Rival;	Time contrário na partida. Time oposto. Time de combate.
2.	Advertência	Aviso	Situação em que o jogador é advertido verbalmente pelo árbitro sobre alguma infração cometido em capo. Situação em que o jogador recebe alguma punição (cartões).
3.	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.
4.	Assistência	Passe; <i>Shoot</i> ; Toque	Lançamento dado em boas condições para o companheiro prosseguir a jogada; Ação pela qual um jogador passa a bola para o outro; Jogador que recebe a bola do companheiro em boas condições.
5.	Atacante	Centroavante; Homem de frente; Dianteiro;	Jogador que se coloca na linha de frente do time com o objetivo de atacar e finalizar a jogador alcançando o gol.
6.	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.
7.	Banco	-	Lugar de permanência dos jogadores reservas ou aqueles que foram substituídos no decorrer do jogo e também da comissão técnica.
8.	Bandeirinha	Bandeira; Auxiliar de arbitragem; juiz auxiliar; fiscal de linha; juiz de linha;	Indivíduos que se encontram às margens do campo, carregando consigo uma pequena bandeira, para sinalizar as saídas da bola pela lateral ou linha de fundo, marcação de impedimentos e

		árbitro assistente.	escanteios, além de auxiliar o juiz nos lances do jogo.
9.	Banheira	-	Quando o jogador fica somente no ataque, esperando a bola chegar nele; ficar em situação de impedimento.
10.	Banho de bola	Baile	Perder de goleada; ganhar de goleada sem oferecer reação ao adversário.
11.	Banho-Maria	Jogo morno	Jogo em que os times não atacam, jogo morno e desinteressante.
12.	Barreira	-	Grupo de jogadores que se colocam em frente à bola para dificultar a cobrança de uma penalidade com direção ao gol.
13.	Bicicleta	-	O Jogador atira-se para o ar, de costas para a goleira, e acerta a bola como se estivesse pedalando uma bicicleta.
14.	Bico	-	Chute apenas com a ponta da chuteira, com a ponta do dedão.
15.	Bola	Pelota; balão de couro; gorduchinha.	Objeto redondo, que é chutado, rolado ou arremessado pelos jogadores com intuito de fazer gol.
16.	Bola do jogo	-	Lance para definir o jogo.
17.	Bomba	Balaço; bicuda.	Chute violento.
18.	Cama de gato	-	É uma falta cometida quando um jogador, em uma disputa de bola no alto, propositalmente não salta, fazendo com que o adversário se desequilibre e caia sobre as suas costas.
19.	Caneco	Taça, troféu, trofeu, trophéo	Troféu que simboliza a conquista de um campeonato.
20.	Campeão	“Levantar o caneco; “colocar a mão no caneco”;	Time que mais acumulou vitórias se tornando o vencedor do campeonato.

21	Campeonato	Copa; Torneio; Competição.	Evento em que os times são reunidos para disputar um título.
22	Caneta	Janelinha	Drible colocando a bola entre as pernas do adversário.
23	Cartão amarelo	Advertência	Mostrado ao jogador que cometeu alguma infração na partida. Serve como uma advertência, ao levar o segundo cartão amarelo o jogador recebe um vermelho e é expulso.
24	Cartão vermelho	Expulsão	Punição mais severa que o cartão amarelo. Mostrado a jogadores que tenham cometido infrações graves. Nesse caso o jogador é expulso do jogo.
25	Catimba	-	Uso de recursos, poucos esportivos para provocar a irritação do adversário e tumultuar o jogo.
26	Cera	Cozinhar	Recurso utilizado com o fim de reter a bola, retardar o jogo. Enrolar o jogo para ganhar tempo.
27	Chaleira	Charles	Jogada de rara habilidade em que a bola é batida com o lado externo do pé.
28	Chegada	-	Falta praticada no corpo do adversário. /Time de chegada é o que tem capacidade de reação e sempre está na disputa dos títulos.
29	Chiqueiro	-	Estádio em péssimas condições de jogo. Para o torcedor rival, chiqueiro é o estádio do time rival, independente das condições em que se encontra.
30	Chute	Pontapé	Ação que o jogador realiza um golpe/impulso com a ponta do pé.
31	Chute de primeira	-	Lance em que o jogador rebate imediatamente a bola, sem antes dominá-la.
32	Chuteira	Botina	Calçado usado para jogar futebol.
33	Chuveiro	Vestiário	Local para onde se diz que vai o jogador expulso de campo.

34.	Clássico	-	Confronto entre dois times de tradição ou de grande rivalidade da mesma cidade ou Estado.
35.	Conclusão	-	Lance que define a jogada.
36.	Corpo Mole	-	Fazer corpo mole é jogar sem disposição, facilitando a vitória do time adversário.
37.	Corta-Luz	-	Jogada em que o atleta vai com o corpo na bola, mas a deixa passar sem tocá-la, com o fim de enganar o adversário e deixa-la correr para um companheiro.
38.	Dirigente	Cartola	Responsável por dirigir os clubes de futebol.
39.	Drible	Dribling	Ato de gingar o corpo e ultrapassar o adversário com a bola dominada.
40.	Elástico	-	Drible em que a bola é jogada para frente, oferecida para o bote do adversário, e puxada rapidamente para trás, como se fosse um elástico.
41.	Escanteio	<i>Córner</i> ; Tiro de canto; Tiro de esquina;	Cobrança feita em um dos ângulos do campo; É ocasionado quando um jogador do time defensor toca a bola pela linha de fundo, então é determinado que um jogador do time atacante leve a bola até um dos ângulos do campo para realizar a cobrança.
42.	Estádio	Campo;	Local onde é realizado os jogos;
43.	Falta	-	Ato cometido contra um adversário, é uma irregularidade no futebol.
44.	Firula	-	Enfeitar a jogada.
45.	Freguês	-	Time que sempre perde para determinado adversário.
46.	Furada		Errar o pé na bola.
47.	Futebol	<i>Football</i>	Esporte disputado por duas equipes de 11 jogadores, cujo objetivo é levar a bola com os pés até a goleira adversária posicionada nas extremidades do

			campo. Não é permitido o uso das mãos, exceto para arremessar a bola em cobranças de lateral e para a defesa do gol pelos goleiros dentro da área permitida.
48	Gaveta	“Lugar onde a coruja dorme”	Ângulo superior da goleira.
49	Geladeira	-	Lugar em que se diz que fica o juiz, que por algum período, fica sem apitar jogos. É utilizado também para jogadores que ficam muitos jogos sem ser titular, ou jogadores que estão sem clube.
50	Gol	<i>Goal</i> ; caixa	Objetivo final do futebol. Espaço delimitado pelas duas traves laterais e o travessão superior, pelo qual deve entrar a bola.
51	Gol-contra	-	Gol do adversário;
52	Gol-pró	-	Gol a favor.
53	Goleada	Chocolate; Baile; Passeio; Vareio; levada.	Quando um time ganha de outro com grande diferença de gols; quando um time perde com uma grande diferença de gols.
54	Goleador	Brocador; Artilheiro	Jogador que faz muitos gols.
55	Goleiro	Arqueiro; Guarda redes;	Jogador que tem fica entre área delimitada em torno da trave com o objetivo de evitar o gol.
56	Gramado	Tapete (quando está em bom estado); campo.	Campo de jogo.
57	Impedimento		Situação irregular em que o jogador atacante fica antes da linha de fundo do campo, à frente de todos os jogadores adversários, exceto do goleiro adversário.
58	Jogador	<i>Footballer</i> ; Boleiro	Aquele que joga. Participante de uma partida de futebol.
59	Juiz	Árbitro; 1ª árbitro.	Profissional responsável por coordenar o cumprimento das regras do futebol.

60.	Partida	<i>Match</i> ; jogo;	Período de competição em que duas equipas formadas por jogadores disputam a vitória.
61.	Tiro de Meta	<i>Goal kick</i> ; <i>freekick</i>	Cobrança feita dentro da área do gol. Concedida quando a bola sai na linha de fundo do campo, depois de ter sido tocada por último em um jogador do time atacante.
62.	Time	Equipe; Grupo	É formado inicialmente por 11 jogadores para disputar partidas.
63.	Traves	Travessão; goleira	Partes verticais e superiores da goleira.
64.	Zagueiro	Beque, <i>beck</i>	Jogador que se posiciona na parte defensiva, entre a linha média e as goleiras do campo.
65.	Zebra	-	Resultado inesperado no futebol. Time considerado inferior que obtém vitória na partida. Time com pouca expressão no futebol que alcança as primeiras posições no campeonato.

Elaborado pela autora.

APÊNDICE C – TABELA DE CONCEITOS E VALORES

Tabela de Conceitos e Valores				
	Conceito	Sinônimos	Definição	Valores
1.	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.	Time
2.	Assistência	Passe; <i>Shoot</i> ; Toque	Lançamento dado em boas condições para o companheiro prosseguir a jogada; Ação pela qual um jogador passa a bola para o outro; Jogador que recebe a bola do companheiro em boas condições.	-
3.	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.	Adversário
4.	Banco	-	Lugar de permanência dos jogadores reservas ou aqueles que foram substituídos no decorrer do jogo e também da comissão técnica.	-
5.	Bandeirinha	Bandeira; Auxiliar de arbitragem; juiz auxiliar; fiscal de linha; juiz de linha; árbitro assistente.	Indivíduos que se encontram às margens do campo, carregando consigo uma pequena bandeira, para sinalizar as saídas da bola pela lateral ou linha de fundo, marcação de impedimentos e escanteios, além de auxiliar o juiz nos lances do jogo.	Bandeira
6.	Banheira	-	Quando o jogador fica somente no ataque, esperando a bola chegar	-

			nele; ficar em situação de impedimento.	
7.	Banho de bola	Baile	Perder de goleada; ganhar de goleada sem oferecer reação ao adversário.	-
8.	Banho-Maria	Jogo morno	Jogo em que os times não atacam, jogo morno e desinteressante.	Time; Ataque; Atacante
9.	Barreira	-	Grupo de jogadores que se colocam em frente à bola para dificultar a cobrança de uma penalidade com direção ao gol.	-
10.	Bicicleta	-	O Jogador atira-se para o ar, de costas para a goleira, e acerta a bola como se estivesse pedalando uma bicicleta.	-
11.	Bico	-	Chute apenas com a ponta da chuteira, com a ponta do dedão.	Chute
12.	Bola	Pelota; balão de couro;	Objeto redondo, que é chutado, rolado ou arremessado pelos jogadores com intuito de fazer gol.	-
13.	Bola do jogo	-	Lance para definir o jogo.	-
14.	Bomba	-	Chute violento.	Chute
15.	Cama de gato	-	É uma falta cometida quando um jogador, em uma disputa de bola no alto, propositalmente não salta, fazendo com que o adversário se desequilibre e caia sobre as suas costas.	Falta; Advertência

16.	Caneco	Taça, troféu, trophéu, trophéo	Troféu que simboliza a conquista de um campeonato.	Campeonato
17.	Campeão	“Levantar o caneco; “colocar a mão no caneco”;	Time que mais acumulou vitórias se tornando o vencedor do campeonato.	Campeonato
18.	Caneta	Janelinha	Drible colocando a bola entre as pernas do adversário.	-
19.	Cartão amarelo	Advertência	Mostrado ao jogador que cometeu alguma infração na partida. Serve como uma advertência, ao levar o segundo cartão amarelo o jogador recebe um vermelho e é expulso.	Advertência
20.	Cartão vermelho	Expulsão	Punição mais severa que o cartão amarelo. Mostrado a jogadores que tenham cometido infrações graves. Nesse caso o jogador é expulso do jogo.	-
21.	Catimba	-	Uso de recursos, poucos esportivos para provocar a irritação do adversário e tumultuar o jogo.	-
22.	Cera	Cozinhar	Recurso utilizado com o fim de reter a bola, retardar o jogo. Enrolar o jogo para ganhar tempo.	-
23.	Chaleira	Charles	Jogada de rara habilidade em que a bola é batida com o lado externo do pé.	-
24.	Chegada	-	Falta praticada no corpo do adversário. /Time de chegada é o	Falta; Título

			que tem capacidade de reação e sempre está na disputa dos títulos.	
25.	Chiqueiro	-	Estádio em péssimas condições de jogo. Para o torcedor rival, chiqueiro é o estádio do time rival, independente das condições em que se encontra.	Estádio; Torcedor rival; Time rival;
26.	Chute de primeira	-	Lance em que o jogador rebate imediatamente a bola, sem antes dominá-la.	Chute
27.	Chuteira	Botina	Calçado usado para jogar futebol.	-
28.	Chuveiro	Vestiário	Local para onde se diz que vai o jogador expulso de campo.	-
29.	Clássico	-	Confronto entre dois times de tradição ou de grande rivalidade da mesma cidade ou Estado.	Time de tradição; Time rival; Time da mesma cidade; Time do mesmo Estado.
30.	Conclusão	-	Lance que define a jogada.	-
31.	Corpo Mole	-	Fazer corpo mole é jogar sem disposição, facilitando a vitória do time adversário.	Time adversário
32.	Corta-Luz	-	Jogada em que o atleta vai com o corpo na bola, mas a deixa passar sem tocá-la, com o fim de enganar o adversário e deixa-la correr para um companheiro.	-
33.	Dirigente	Cartola	Responsável por dirigir os clubes de futebol.	Clube

34.	Drible	<i>Dribling</i>	Ato de gíngar o corpo e ultrapassar o adversário com a bola dominada.	-
35.	Elástico	-	Drible em que a bola é jogada para frente, oferecida para o bote do adversário, e puxada rapidamente para trás, como se fosse um elástico.	-
36.	<i>Escanteio</i>	<i>Córner;</i> Tiro de canto; Tiro de esquina;	Cobrança feita em um dos ângulos do campo; É ocasionado quando um jogador do time defensor toca a bola pela linha de fundo, então é determinado que um jogador do time atacante leve a bola até um dos ângulos do campo para realizar a cobrança.	Ângulos do campo.
37.	Firula	-	Enfeitar a jogada.	-
38.	Freguês	-	Time que sempre perde para determinado adversário.	Time
39.	Furada		Errar o pé na bola.	-
40.	Futebol	<i>Football</i>	Esporte disputado por duas equipes de 11 jogadores, cujo objetivo é levar a bola com os pés até a goleira adversária posicionada nas extremidades do campo. Não é permitido o uso das mãos, exceto para arremessar a bola em cobranças de lateral e para a defesa do gol pelos goleiros dentro da área permitida.	-
41.	Gaveta	“Lugar onde a	Ângulo superior da goleira.	Ângulos do campo.

		coruja dorme”		
42.	Geladeira	-	Lugar em que se diz que fica o juiz, que por algum período, fica sem apitar jogos. É utilizado também para jogadores que ficam muitos jogos sem ser titular, ou jogadores que estão sem clube.	-
43.	Gol	<i>Goal</i>	Objetivo final do futebol. Espaço delimitado pelas duas traves laterais e o travessão superior, pelo qual deve entrar a bola.	Gol Pró; Gol Contra.
44.	Goleada	Chocolate; Baile; Passeio; Vareio;.	Quando um time ganha de outro com grande diferença de gols; quando um time perde com uma grande diferença de gols.	-
45.	Goleador	Brocador; Artilheiro	Jogador que faz muitos gols.	-
46.	Gramado	Tapete (quando está em bom estado); campo.	Campo de jogo.	-
47.	Impedimento		Situação irregular em que o jogador atacante fica antes da linha de fundo do campo, à frente de todos os jogadores adversários, exceto do goleiro adversário.	Atacante; Goleiro
48.	Jogador	<i>Footballer</i> ; Boleiro	Aquele que joga. Participante de uma partida de futebol.	Jogadores
49.	Juiz	Árbitro; 1 ^a árbitro.	Profissional responsável por	-

			coordenar o cumprimento das regras do futebol.	
50.	Partida	Match; jogo	Período de competição em que duas equipes formadas por jogadores disputam a vitória.	-
51.	Tiro de Meta	<i>Goal kick; freekick</i>	Cobrança feita dentro da área do gol. Concedida quando a bola sai na linha de fundo do campo, depois de ter sido tocada por último em um jogador do time atacante.	Linha de fundo. Área do gol.
52.	Traves	Travessão; goleira	Partes verticais e superiores da goleira.	-
53.	Zagueiro	Beque, <i>beck</i>	Jogador que se posiciona na parte defensiva, entre a linha média e as goleiras do campo.	-
54.	Zebra	-	Resultado inesperado no futebol. Time considerado inferior que obtém vitória na partida. Time com pouca expressão no futebol que alcança as primeiras posições no campeonato.	Time

Elaborado pela autora.

APÊNDICE D – GLOSSÁRIO DAS RELAÇÕES

Glossário das Relações			
ID	Termo 1	Verbo	Termo 2
1.	Assistência	É um	Jogada
2.	Banco	Contém	Jogador
3.	Bandeirinha	Indica	Escanteio
4.	Bandeirinha	Indica	Impedimento
5.	Bandeirinha	Auxilia	Juiz
6.	Banheira	Causa	Impedimento
7.	Banho de bola	Permite	Gol
8.	Barreira	Protege	Gol
9.	Barreira	Bloqueia	Bola
10.	Bicicleta	É um	Jogada
11.	Bico	É um	Chute
12.	Bomba	É um	Chute
13.	Cama de gato	Causa	Cartão amarelo
14.	Campeão	Conquista	Caneco
15.	Caneta	É um	Drible
16.	Cartão vermelho	Causa	Expulsão
17.	Cera	Ocorre	Partida
18.	Chaleira	É um	Drible
19.	Chiqueiro	É um	Estádio
20.	Clássico	É uma	Partida
21.	Conclusão	É um	Chute
22.	Jogador	Faz	Corpo Mole
23.	Elástico	É um	Drible
24.	Firula	É uma	Jogada
25.	Gaveta	Parte de	Trave
26.	Goleador	É um	Jogador
27.	Gramado	Parte de	Estádio
28.	Traves	Parte de	Estádio

29.	Partida	Ocorre	Goleada
30.	Zagueiro	É um	Jogador
31.	Zebra	É um	Time

Elaborado pela autora.

ANEXO A – TEXTOS ANALISADOS PARA FORMAÇÃO DO GLOSSÁRIO DE VERBOS

Figura 30 – Texto 1 (Edição de 4 de dezembro de 1997, p.20)

O caneco é deles

MARSELHA — Lança-se hoje a sorte do Brasil na Copa do Mundo, com a escolha dos oito grupos em que se dividirão as 32 seleções na primeira fase, mas o caneco já não é nosso. Uma das mais colossais figuras femininas deste fim de século desembarcou, no começo da noite de ontem, nesta cidade portuária que se prepara para adotar a seleção brasileira no ano que vem — a Copa do Mundo está em Marselha.

Foi uma viagem sem pompas, embora o troféu conquistado pelo Brasil na Copa de 94 tenha vindo na classe executiva, discretamente escoltado por dois agentes da Polícia Federal, Nitor Vasconcelos e Jorge Guimarães, e acompanhado, de longe, pelo técnico Mário Jorge Lobo Zagallo. Coincidentemente, a Copa do Mundo viajou no bagageiro de número 13. Coincidentemente?

— Ela sabe escolher — brinca Zagallo, dando força à tenda de que é um dos mais supersticiosos profissionais do mundo da bola.

Na escala em Paris, um grupo de policiais foi receber seus colegas brasileiros na porta do avião, cuidando de levá-los diretamente para o setor em que a Copa do Mundo e Zagallo, seu mais demorado amor, embarcariam para Marselha. No desembarque na cidade em que hoje serão sorteados os grupos da Copa, mais policiais franceses a aguardavam para levá-la imediatamente ninguém sabe para onde. Certo é que o caneco já não é nosso.

— Está emprestado e vai voltar — faz questão de ressaltar o otimista Zagallo.

É o que esperam os brasileiros,

mas é bom saber, desde já: não será nada fácil conquistar o quinto título mundial em campos franceses. O próprio Zagallo, ao abordar o tema seriamente, avisa:

— Será uma Copa difícilíssima e a história mostra que ganhar o caneco em outro continente é uma raridade que só nós conseguimos, em 58, na Suécia.

Um rápido passsar de olhos pela história vai relembrar, por exemplo, que os alemães participaram de quatro das seis últimas decisões da Copa do Mundo em campos europeus e foram os vencedores em 54, 74 e em 90, fortíssimos dados ao **Caneco** França. Atualmente, são os fortes candidatos.



■ Roberto Benedites é editor de Esportes do Estado

Coincidentemente, a Copa do Mundo viajou no bagageiro de número 13. Coincidentemente?

os franceses são os donos da casa e estão com o caneco, ainda que por empréstimo, desde ontem.

Os europeus, um pouco por respeito e admiração pelo futebol da terra de Pelé e um pouco por esperteza, costumam dar o Brasil como franco favorito à conquista do título na França. Convém desconfiar e torcer, hoje, para que não caiam para a seleção brasileira adversários muito fortes logo na primeira fase da Copa. Zagallo terá menos de 20 dias para preparar o time antes da competição e ganhará precioso tempo para briliá-lo se, nos primeiros três jogos, não tiver de enfrentar equipes muito difíceis. Que a sorte nos ajude.

Termo: Caneco

Verbos coletados:
conquistar, ganhar,
enfrentar.

Fonte: Captura de tela acervo Estadão

Figura 31 – Texto 2 (Edição de 17 de julho de 2000, p.24)

Túlio brilha com gol de bicicleta

Em Jundiaí, artilheiro deu a vitória por 1 a 0 ao São Caetano e decretou o fim do título do título da A2

ANDRÉ AMARAL
Especial para o Estado

JUNDIAÍ – O São Caetano está bem próximo do título da Série A2 do Campeonato Paulista. Ontem, em Jundiaí, o time do ABC conteu o ataque de Túlio. Para chegar pela primeira vez a divisaõ de elite do futebol paulista, o São Caetano pode perder o segundo jogo por até um gol de diferença, domingo, em casa.

Empurrado por mais de 15 mil pessoas, que lotaram o Estádio Jayme Cintra, o Etti até foi melhor e teve a posse de bola por mais tempo na primeira etapa, mas não teve eficiência na hora das finalizações. Os atacantes Jean Carlo e Ailton trocavam bolas na entrada da área, mas ninguém ariscava o chute.

A equipe de Jundiaí, de propriedade da Parmalat, ariscou iniciais cruzamentos para a área, que paravam nas mãos do goleiro Sílvio Luis.

O São Caetano teve um completo domínio do meio-de-campo, com tabelas entre os experientes Leito, Claudcir e Zinho. Aos 6 minutos, o time do ABC perdeu a chance de abrir o marcador. Leto lançou para Esquerdinha, que chutou a bola com força na trave.

A equipe de Jundiaí me-



Túlio, de bicicleta, marca o gol que garante a vitória ao São Caetano, ontem, em Jundiaí, na primeira partida da final: "Todo mundo pensa que eu estou surtido, mas sou matador"

lhrou nos 15 minutos finais da primeira etapa. Aos 32, a defesa do São Caetano parou para pedir impedimento e Jean Carlo ficou livre, mas Sílvio Luis defendeu. No rebote, Juari errou o chute.

Um minuto depois, Ailton deixou Marcinho na cara do gol, mas ele perdeu o gol. Aos 40, após um cruzamento, Jean Elias dominou a bola na pequena área, mas ficou sem chances.

Túlio, do São Caetano, artilheiro da competição, jogou no seu estilo. Durante todo o primeiro tempo, tocou na bola apenas duas vezes, mas preocupou a zaga do Etti.

No segundo tempo, as

substituições mudaram o jogo. O técnico Luiz Carlos Ferreira, do Etti, colocou Mirandinha, ex-Corinthians, no lugar de Juari e errou ao tirar Ailton, que arriava a maioria das jogadas do time e optou por Caíli. Jair Picerni colocou Ademir no lugar de Zinho e deu nova cara ao ataque.

Dominato – Após os 10 minutos, o São Caetano começou a dominar completamente a partida. Ademir perdeu duas chances, aos 15 e 22 minutos. Aos 27, Túlio apareceu. Em um escanteio, após desvio do zagueiro Daniel, ele teve tranquilidade para tocar de bicicleta no canto esquerdo do goleiro César, fazendo seu 18º gol.

As chances de reação do Etti acabaram com a expulsão de Humberto, aos 35 mi-

nutos. Túlio, que tinha tomado o segundo cartão amarelo e ficaria fora do segundo jogo, provocou sua expulsão aos 40 minutos. Deve ser julgado pelo tribunal da federação no meio de semana e terá condições de jogo.

Etti/Jundiaí	0
São Caetano	1

Gol – Túlio aos 27 minutos do segundo tempo.
Artilheiro – César (Delegado), Jean Etti, Toga Júnior, José (Mirandinha), Humberto, Wagner Marinho e Humberto, Jean Carlo e Ailton (Caíli). **Técnicos** – Luiz Carlos Ferreira.
São Caetano – Sílvio Luis, Jorginho, Daniel, Douglas e César, Claudcir, Esquerdinha, Zinho (Ademir) e Leto (Esquerdinha), Túlio e Magno. **Feridas** – Lou Pôrto.
São Caetano – Sérgio, Sérgio Marinho, Ademir, César (E) e José Etti. **Cartões amarelos** – Humberto e César. **Bênção e expulsão** – Não divulgadas. **Local** – Estádio Jayme Cintra, em Jundiaí.

Termo: bicicleta

Verbos coletados:
conter, provocar.

Fonte: Captura de tela acervo Estadão

Alex, recordista em deixar companheiros na cara do gol

O meia participou de 16 jogadas de assistência que resultaram em gols, além de ter marcado 14 este ano

Pelas estatísticas da comissão técnica do Palmeiras, o meia Alex é o recordista da equipe nas jogadas de assistência que resultaram em gols. A sua volta contra o Coritiba, amanhã, no Palestra Itália, pode dar novamente ao Palmeiras um futebol objetivo e envolvente que faltou na derrota para o Botafogo, por 3 a 1, sábado.

Dos pés de Alex surgem os principais lances que deixam os atacantes Oseas e Paulo Nunes em condição de marcar. Além de ter feito 14 gols nesta temporada, o meia participou, em jogadas de assistência, de outros 16. O segundo jogador com melhor índice de aproveitamento nas jogadas de assistência é Júnior. O lateral-esquerdo participou de 15 gols na temporada.

Alex reconhece que vive um grande momento, mas divide seu

sucesso com os companheiros. “Estou aprimorando meu futebol, com mais movimentação em campo, dando combate e arriscando as jogadas ofensivas”, comentou. “Em pouco mais de um ano no Palmeiras tenho certeza que evolui bastante”, declarou o atleta.

O meia afirmou que se considera um jogador mais experiente e a sua participação nos dois amistosos da seleção brasileira sob o comando do técnico Wanderley Luxemburgo, contra a Iugoslávia e Equador, podem dar nova projeção ao seu futebol.

O capitão Zinho, companheiro de Alex no meio-de-campo, admitiu ontem que a ausência do companheiro em Niterói prejudicou o Palmeiras. “Sem querer desmerecer o Arílson, Alex realmente fez muita falta”, afirmou. “Ele vive uma grande fase, entrosado com a equipe e jogando um futebol para estar na seleção brasileira.” Paulo Nunes tem a mesma opinião. Disse que Alex é um dos destaques do Campeonato Brasileiro.

Termo: assistência

Verbo coletado:
resultar.

Fonte: Captura de tela acervo Estadão

Figura 33 – Texto 4 (Edição de 13 de junho de 1998, p.32)

A Copa mal começou e os árbitros já fazem história

Na França, a Itália foi beneficiada por um dos grandes erros de juiz deste Mundial

A tendência da Comissão de Arbitragens da Fifa de insistir com juizes de países pouco desenvolvidos no futebol para apitar partidas de potências da Copa do Mundo faz com que o Mundial da França já apareça como um campeonato de resultados distorcidos. Se pouca gente percebeu Dunga colocar a mão na bola no jogo Brasil e Escócia, houve praticamente unanimidade quanto ao erro do juiz Lucien Bouchardeau, do Níger, na partida Itália 2 x Chile 2, quinta-feira, em Bordeaux. Os italianos agradecem: só Bouchardeau viu o pênalti pelo qual a Azzurra chegou ao empate. Ontem, em Montpellier, foi a vez de Abdul Rahman Zeid, da Arábia Saudita, deixando de dar um pênalti para cada lado.

A Espanha, de José Maria García-Aranda, está entre os países de futebol profissional mais organizado do mundo, mas seus árbitros têm a tradição de errar com frequência. Não por serem venais: simplesmente são limitados tecnicamente e chegam a prejudicar jogos do Campeonato Espanhol e de outros torneios. Em 1993, um juiz espanhol tumultuou o jogo Palmeiras x São Paulo, que decidia o Torneio Ra-

món de Carranza, em Cádiz.

Pesquisa do InformEstado, ouvindo torcedores em São Paulo após a partida Brasil x Escócia, apontou elogios gerais para García-Aranda. Então, ainda não tinha sido destacada a imagem que mostra o pênalti de Dunga.

Em Bordeaux, foi gritante. O Chile poderia ter provocado a primeira grande surpresa da Copa, com uma merecida vitória de virada sobre a Itália, com dois bonitos gols de Salas, mas acabou cedendo o empate, a seis minutos do fim. Na verdade, o zagueiro Fuentes não cometeu o pênalti marcado pelo árbitro do Níger e convertido por Baggio, aquele que errou na final de 94.

Em Montpellier, a Bulgária esteve bem melhor no primeiro tempo e poderia ter chegado ao gol logo no início da partida. O juiz Abdul Rahman Zeid, da Arábia Saudita, não deu um pênalti cometido por um zagueiro paraguaio sobre o atacante Stoichkov. Irritado, o ídolo búlgaro caiu de rendimento e passou a fazer faltas. No segundo tempo, como que para compensar, o árbitro deixou de marcar um pênalti a favor do Paraguai e a partida terminou em 0 a 0. A Arábia Saudita, de Zeid, pelo menos está na Copa. O Níger, de Bouchardeau, só entraria no Mundial de 2006, se João Havelange continuasse no comando da Fifa, aumentando a cada Copa o número de vagas da África.

Termo: juiz/árbitro

Verbos coletados: provocar, prejudicar, fazer.

Fonte: Captura de tela acervo Estadão

Figura 34 – Texto 5 (Edição de 17 de janeiro de 2005, p.38)

Brasil pára na retranca do Paraguai

Time do técnico René Weber ficou no 1 a 1 em sua segunda partida no Sul-Americano Sub-20, na Colômbia

SUB-20
PEREIRA

O Brasil decepcionou e apenas empatou com o Paraguai, por 1 a 1, ontem, no Estádio Hernán Ramírez Villegas, em Pereira, na Colômbia, na segunda rodada do Sul-Americano Sub-20. A seleção brasileira soma quatro pontos em dois jogos no grupo B e mantém

boas chances de se classificar para a fase final. O próximo desafio do Brasil será quinta-feira, contra o Chile, em Armenia.

O primeiro tempo foi de poucas emoções. Lento no ataque, o time do técnico René Weber não conseguiu superar a forte retranca do adversário e quase não criou chances de gols.

Na etapa final, porém, as jogadas ofensivas da seleção brasilei-

ra começaram a surtir efeito, especialmente entre Rafael Sobis e Diego. E foi graças à dupla do Internacional que o Brasil abriu o placar. Aos 12 minutos, Rafael cruzou para Diego, que dividiu com os zagueiros e cabeceou, sem defesa para o goleiro Lopez.

O Paraguai não mudou seu esquema tático, mas foi feliz em uma das poucas descidas ao ataque. Aos 17, Cristian Bogado do-

minou na entrada da área e arriscou chute forte, que acertou o canto do goleiro Renan: 1 a 1.

Como as dificuldades para firmar o esquema defensivo do rival continuavam, o Brasil passou a insistir nos cruzamentos para a área e nas jogadas de bolas paradas. Aos 35, depois do escanteio, Qui-rino cabeceou no travessão. Os paraguaios se fecharam e seguraram o resultado.

ARGENTINA
Depois de duas brilhantes atuações, o jovem Lionel Messi, de 17 anos, deve ser titular da Argentina contra o Peru, em Muni-zales. Messi, atacante do Barcelona, da Espanha, foi o destaque das vitórias sobre Venezuela (3 a 0) e Bolívia (4 a 0), quando saiu do banco de reservas e marcou dois gols em cada confronto. Também hoje, jogam Bolívia e Venezuela. ■

Fonte: Captura de tela acervo Estadão

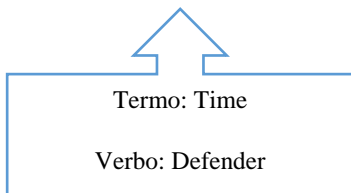


Figura 35 – Texto 6 (Edição de 8 de dezembro de 1998, p.38)

Para técnicos, só controle supera a catimba



Narciso: santista não se arrepende da falta cometida



Marcelinho Carioca: "Vou tirar Narciso de campo"

Os quatro comandantes semifinalistas alertam suas equipes para que não aceitem provocações

Os quatro técnicos semifinalistas do Campeonato Brasileiro têm um recado especial para seus jogadores nas partidas de amanhã, que definirão as duas equipes que vão disputar o título: não entrar na provocação de seus adversários. Eles foram unânimes em reconhecer que a catimba e a tentativa de desestabilizar o controle emocional do inimigo serão usados nos jogos do Pacembu e Candinho.

"É preciso que eles sejam lutadores, mas também inteligentes na mesma proporção, não adianta ganhar 10 mil quesito e ficar devendo em outro", afirmou o técnico do Santos, Emerson Leão. Acostumado a participar

de decisões, o ex-goleiro reconhece que é preciso haver um controle emocional.

O técnico Levir Culpi, do Cruzeiro, acha que o equilíbrio emocional vai definir quem vai para a decisão do Campeonato Brasileiro. Aqueles que se deixarem levar pela tensão da partida, entrando nas provocações dos adversários, vão acabar saindo prejudicado.

Os jogadores da equipe mineira são orientados a não se envolver em confusão. "Conto com um grupo formado por jogadores experientes, isso ajuda muito em um momento decisivo como este que estamos passando", destaca Levir, que possui jogadores como Valdo, Müller e Wilson Gottardo.

"Tudo isso faz parte do futebol", comenta o técnico do Corinthians, Wanderley Luxemburgo, sobre as provocações em campo entre os atletas. "O importante é ter controle emocional".

Candinho, técnico da Portuguesa e auxiliar de Luxemburgo na seleção brasileira, concorda. "O importante é não entrar em provocações, pois a catimba entre os jogadores durante as partidas é natural."

Segundo ele, nos dois primeiros jogos entre Portuguesa e Cruzeiro houve poucas faltas, pois "as duas equipes são muito técnicas e gostam de jogar com a bola no chão".

■ Reportagem de Anderson Costa, Esterson Costa, José Rodrigues e Paulo Guilherme

LEVIR CONFIANÇA
NA EXPERIÊNCIA
DOS SEUS
VETERANOS

Fonte: Captura de tela acervo Estadão

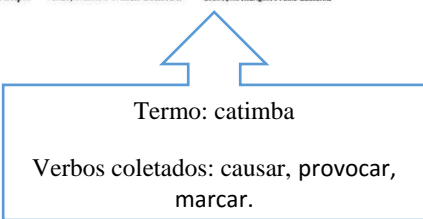


Figura 36 – Texto 7 (Edição de 21 de setembro de 1998, p.17)

Em goleada histórica, Lusa arrasa o São Paulo

Os 7 a 2 de ontem no Pacemba surpreendem a todos no Mirassol e Nelsinho pode cair

Portuguesa sempre, em meio a uma crise financeira, conseguiu vencer o São Paulo por 7 a 2 no Pacemba, em uma goleada histórica. O jogo, que aconteceu no domingo, foi marcado por gols de Lusa, de 10 minutos de jogo, e de outros jogadores da equipe lusitana. O São Paulo sofreu com a falta de concentração e erros defensivos. O técnico Nelsinho, que estava em sua primeira partida, pode sofrer consequências por esta goleada.



Atletas lusitanos comemoram gol, durante da partida entre o São Paulo e Portuguesa no Pacemba.

Candinho comemora o sucesso de sua estratégia

Candinho é o novo treinador do São Paulo. Desde o início de sua estratégia, o técnico português tem mostrado resultados positivos. A equipe lusitana conseguiu vencer o São Paulo por 7 a 2 no Pacemba. Candinho afirma que sua estratégia foi bem-sucedida e que ele está satisfeito com o desempenho da equipe. Ele também mencionou que o jogo foi muito emocionante e que ele espera continuar a trabalhar com a equipe para alcançar mais vitórias.

Redada de média de gol 80% melhor

A média de gols por jogo da Portuguesa é de 1,8, enquanto a do São Paulo é de 0,2. Isso indica uma melhoria significativa no desempenho ofensivo da equipe lusitana. O técnico Candinho atribui esse sucesso à melhor estratégia adotada e à maior concentração dos jogadores. Ele também mencionou que a defesa do São Paulo foi muito fraca durante o jogo.

Reuniao hoje pode afastar Nelsinho

Após a goleada sofrida pelo São Paulo, há especulações sobre o futuro do técnico Nelsinho. Uma reunião com o clube pode levar à sua demissão. A diretoria do São Paulo parece insatisfeita com o desempenho da equipe e com a atuação de Nelsinho. No entanto, Nelsinho afirma que ele não se desmotiva e que ele continuará trabalhando para melhorar a equipe.

O técnico diz que 'NASCER' FOI 'ACIDENTE'

O técnico Candinho afirmou que sua chegada ao São Paulo foi um acidente. Ele mencionou que ele não estava planejando vir para o Brasil e que ele acabou aceitando a oferta por uma série de circunstâncias. Candinho também mencionou que ele se sente bem no Brasil e que ele está aproveitando a oportunidade de trabalhar no futebol brasileiro.

CAMPEONATO BRASILEIRO DE FUTEBOL		CLASSIFICACAO		PROXIMOS JOGOS				
1	Corinthians	33	19	2	100	13	Flamengo	
2	Flamengo	30	18	4	29	14	Botafogo	
3	Paraná	27	18	8	21	7	14	Internacional
4	Botafogo	26	17	10	26	9	15	Grêmio
5	Fluminense	25	17	11	20	10	16	Portuguesa
6	Internacional	24	16	12	12	11	17	Corinthians
7	Fluminense	19	14	1	20	22	2	Flamengo
8	Corinthians	17	13	4	16	21	3	Fluminense
9	Flamengo	17	13	5	16	22	4	Corinthians
10	Paraná	16	12	6	17	18	5	Internacional
11	Internacional	15	12	7	18	19	6	Fluminense
12	Fluminense	14	11	9	17	17	8	Paraná
13	Botafogo	14	11	10	18	18	9	Flamengo
14	Grêmio	13	10	5	15	19	10	Corinthians
15	São Paulo	12	9	2	7	24	26	Fluminense
16	Flamengo	12	9	3	16	20	27	Paraná
17	São Paulo	11	8	1	14	19	0	Internacional
18	Internacional	11	8	1	15	17	1	Fluminense
19	Fluminense	10	7	1	12	17	2	Paraná
20	Paraná	10	7	1	13	21	4	Flamengo
21	Corinthians	9	6	1	11	16	5	Internacional
22	Flamengo	9	6	1	12	17	6	Fluminense
23	Fluminense	8	5	1	11	16	7	Paraná
24	Grêmio	8	5	1	12	17	8	Corinthians
25	Corinthians	7	4	1	11	16	9	Flamengo
26	Flamengo	7	4	1	12	17	10	Fluminense
27	Fluminense	6	3	1	11	16	11	Paraná
28	Paraná	6	3	1	12	17	12	Internacional
29	Internacional	5	2	1	11	16	13	Flamengo
30	Flamengo	5	2	1	12	17	14	Fluminense
31	São Paulo	4	1	1	11	16	15	Paraná
32	Corinthians	4	1	1	12	17	16	Internacional
33	Fluminense	3	0	1	11	16	17	Flamengo
34	Paraná	3	0	1	12	17	18	Corinthians
35	Internacional	2	0	1	11	16	19	Flamengo
36	Flamengo	2	0	1	12	17	20	Fluminense
37	Fluminense	1	0	1	11	16	21	Paraná
38	Paraná	1	0	1	12	17	22	Internacional
39	Internacional	0	0	1	11	16	23	Flamengo
40	Flamengo	0	0	1	12	17	24	Fluminense

Fonte: Captura de tela acervo Estádio

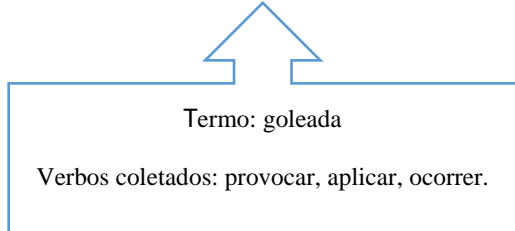


Figura 37 - Texto 8 (Edição de 14 de junho de 1981, p. 48)

FUTEBOL

FEDERAÇÃO BRASILEIRA DE FUTEBOL

A nova entidade nacional está fundada — Foi eleita a sua primeira directoria

Os esportistas brasileiros acabam de ser agradavelmente surpreendidos com a fundação da Federação Brasileira de Futebol, organizada com o apoio da Liga de Amadores de Futebol, de S. Paulo, e da Liga Metropolitana de Desportos Terrestres, do Rio de Janeiro.

Não poderia haver, por certo, mais promissora origem, pois contando, ao appropiar-se no cenário esportivo do Brasil, com a colaboração das duas entidades que dirigem o futebol dissidente de S. Paulo e Rio de Janeiro, pôde-se afirmar que essa nova Federação nasce victoriosa.

Nós aqui de S. Paulo sabemos perfeitamente qual a elevação de vistas que preside os actos da Laf, que negavelmente congrega em seu seio a elite esportiva da paulista. E os esportistas do Rio sabem, tambem, qual é o esforço que a Metropolitana dispõe em favor do futebol nacional.

Além disto, a primeira directoria da Federação é a melhor garantia de successo que poderia, porventura, obter, pois tendo na presidencia a figura sympathica de Manuel de Lacerda Franco, uma das expressões mais nitidas do verdadeiro esportista, ella surge fundada a representar o papel de salvadora do futebol de nossa terra, onde em virtude de incompreensíveis divergencias pessoais, nas quaes não se sabe se reprovar a teimosia ou o impatriotismo, o bom nome brasileiro nesse esporte se via ameaçado na sua moralidade e na sua technica.

Não reune, é verdade, a nova Federação, a totalidade

dos esportistas brasileiros. Os contingentes levados pela Metro e pela Laf, pouco representam em presenca da totalidade. Mas ella se caracteriza, ao nascer, não pela quantidade, e sim pela qualidade. Sabê-se, comtudo, que muitas entidades ou clubs, desgostosos com a orientação que o futebol brasileiro vem tendo nestes ultimos tempos, aguardavam impacientes a appareição de uma nova organização que o dirigisse com mais elevação de espirito e menor isenção de animo.

Para estes está, agora, satisffeito o desejo.

A Federação Brasileira de Futebol elegue a seguinte directoria:

Manuel de Lacerda Franco, presidente; Oswaldo Gomes, vice-presidente; Virgílio Guimarães, secretario; Janséio Gensérico Daimon, 2º secretario, e Carlos Reis Junior, thesoureiro. São nomes muito conhecidos no esporte do Brasil, e a operosa actividade que qualquer um delles tem desenvolvido em favor da definitiva implantação do esporte pelo esporte, são credencias valiosas que os fazem merecedores de todo acatamento e de toda consideração popular.

Outra medida inicial da Federação Brasileira de Futebol, que a fez immediatamente sympathica ao publico de S. Paulo, foi o facto de sua sede ser localisada em nossa cidade, justamente tida como a verdadeira capital do futebol brasileiro.

Embora nestes ultimos tempos os jogadores cariocas tenham provado, ás vezes, — vencendo club contra club, ou seleccionado contra selectio-

nado, — uma superioridade que apesar do incontestavel êxito, comtudo temporaria, e cuja causa sabemos resistir no fracçãoamento que se deu, aqui, no seio dos melhores valores futebolísticos, nem por isso S. Paulo deixou de ser o maior centro de futebol do Brasil.

A nossa actual inferioridade technica vém, como accentuamos acima, da repartição de forças entre a Apea e a Laf. Não houve esportista bem intencionado que não lamentasse a scisão que foi assumpto forçado de todos que temiam a queda — e com razão temiam — da supremacia de S. Paulo.

Mas agora, força é confessar, todos reconhecem que a scisão trouxe beneficios para o nosso futebol, sendo o maior delles a fundação da nova entidade nacional.

Tendo escolhido para suas cores esportivas as mesmas da bandeira brasileira, ella se torna a bandeira brasileira do futebol, e do seu seio é que ha de vir o verdadeiro futebol de nossa terra.

A Federação Brasileira está animada dos melhores propósitos, sendo o principal delles a realisação de partidas internacionaes. Para ser criado o seu primeiro fundo de reserva, vão ser jogadas duas partidas entre os seleccionados da Laf e da Metro, disputadas uma nesta capital e outra no Rio de Janeiro.

Mas, ainda que circunstancias inesperadas viessem impedir a de levar a bom termo os seus nobres propósitos, bastaria a intenção que tem de revigorar o doente futebol do Brasil, para merecer os applausos dos verdadeiros esportistas do Brasil.

Fonte: Captura de tela acervo Estadão

Termo: futebol;

Verbo coletado: tornar.

Figura 39 – Texto 9 (Edição de 27 de junho de 1929, p.12)

diB — O ESTADO DE S. PAULO

DOMINGO - 14 DE JUNHO DE 1981

Um grande clássico, hoje sem atrações

O grande clássico do futebol paulista — que beneficiou os times de 1978 a 1980 — agora está completamente desanimado: São Paulo e Santos não prometem atrações, tem um campeonato que está agora está provocando revulso interesse do público. O São Paulo, arrebolado por uma série de problemas, quer jogar até o último lugar, mas a reação é difícil. Já que o time está totalmente desmotivado e sem técnico. Talvez hoje a diretoria anuncie o provável substituto de Cláudio Oliveira. O Santos, que também passou por momentos de crise, ainda não apresenta um bom rendimento, porém a equipe está mais motivada e acreditando poder explorar a fragilidade do adversário.

Sergio Cláudio só tem uma opção, que é quanto ao aproveitamento do governo Marão, que retorna esta manhã de Toulon. João Leal ficou, ao contrário, sem muitas dúvidas e espera poder voltar com Everton, que também regressa da França. Na manhã, improvava uma equipe que irá abandonar tão logo a direção anuncie o novo contratado. Assim, tecnicamente, o Santos está em vantagem, seria a oportunidade para uma vitória em clássicos, como um risco para Sergio Cláudio, se ele não ocorrer. O jogo começa às 18 horas, com arbitragem de Raulino Marques G. Lacerda.

São Paulo — Tostão; Chiquinho, Oscar, Nel e Vladimir (Herberto); Elvio, Renato e Herberio (Everton); Paulo César, Tati e Everton (Valtinho).

Santos — Marão (Oss); Dummar, Jozélio, Marcelo e Washington; Marcos, Elvê e Pflar; Gilvan, Claudinho e João Paulo.

O velho estádio de São Jansuário estava esta manhã todo recheado, e o novo espaço que ocupa atualmente (que o time e o clube não conseguiram comprar) e o Parque Vitor-meigdy, pertencente ao clube, estava cheio de torcedores. O jogo começou às 18 horas, com arbitragem de Raulino Marques G. Lacerda. O Santos, que também passou por momentos de crise, ainda não apresenta um bom rendimento, porém a equipe está mais motivada e acreditando poder explorar a fragilidade do adversário.

Sergio Cláudio só tem uma opção, que é quanto ao aproveitamento do governo Marão, que retorna esta manhã de Toulon. João Leal ficou, ao contrário, sem muitas dúvidas e espera poder voltar com Everton, que também regressa da França. Na manhã, improvava uma equipe que irá abandonar tão logo a direção anuncie o novo contratado. Assim, tecnicamente, o Santos está em vantagem, seria a oportunidade para uma vitória em clássicos, como um risco para Sergio Cláudio, se ele não ocorrer. O jogo começa às 18 horas, com arbitragem de Raulino Marques G. Lacerda.

São Paulo — Tostão; Chiquinho, Oscar, Nel e Vladimir (Herberto); Elvio, Renato e Herberio (Everton); Paulo César, Tati e Everton (Valtinho).

Santos — Marão (Oss); Dummar, Jozélio, Marcelo e Washington; Marcos, Elvê e Pflar; Gilvan, Claudinho e João Paulo.

Ora pode ter sua chance no clássico de hoje, se Marão não tiver condições de jogar

Fonte: Captura de tela acervo Estadão

ALTAIR RAFFA

Vasco reabre estádio, em fase crítica

O velho estádio de São Jansuário estava esta manhã todo recheado, e o novo espaço que ocupa atualmente (que o time e o clube não conseguiram comprar) e o Parque Vitor-meigdy, pertencente ao clube, estava cheio de torcedores. O jogo começou às 18 horas, com arbitragem de Raulino Marques G. Lacerda. O Santos, que também passou por momentos de crise, ainda não apresenta um bom rendimento, porém a equipe está mais motivada e acreditando poder explorar a fragilidade do adversário.

Sergio Cláudio só tem uma opção, que é quanto ao aproveitamento do governo Marão, que retorna esta manhã de Toulon. João Leal ficou, ao contrário, sem muitas dúvidas e espera poder voltar com Everton, que também regressa da França. Na manhã, improvava uma equipe que irá abandonar tão logo a direção anuncie o novo contratado. Assim, tecnicamente, o Santos está em vantagem, seria a oportunidade para uma vitória em clássicos, como um risco para Sergio Cláudio, se ele não ocorrer. O jogo começa às 18 horas, com arbitragem de Raulino Marques G. Lacerda.

São Paulo — Tostão; Chiquinho, Oscar, Nel e Vladimir (Herberto); Elvio, Renato e Herberio (Everton); Paulo César, Tati e Everton (Valtinho).

Santos — Marão (Oss); Dummar, Jozélio, Marcelo e Washington; Marcos, Elvê e Pflar; Gilvan, Claudinho e João Paulo.

Termo: clássico

Verbo coletado: ocorrer.