

USO DE ONTOLOGIA NA RECUPERAÇÃO DA INFORMAÇÃO EM ACERVOS DIGITAIS DE JORNAIS

EL USO DE ONTOLOGÍAS EN LA RECUPERACIÓN DE INFORMACIÓN EN LAS COLECCIONES DIGITALES DE LOS DIÁRIOS

Luana Carla de Moura dos Santos*
Marisa Bräscher**

RESUMO:

Introdução: Visa a modelagem de uma ontologia do domínio do futebol para o tratamento de variações diacrônicas e sincrônicas da língua.

Objetivo: Apoiar a recuperação da informação em acervos digitais de jornais.

Metodologia: Trata-se de uma pesquisa aplicada, utilizando como base um acervo digital de jornal. Utiliza a metodologia *OntoForInfoScience*, de Mendonça (2015) para desenvolver a ontologia do domínio do futebol. Cronologicamente, o recorte estabelecido abrange terminologia utilizada entre 1900 a 2015, período que contempla a existência dos clubes de futebol no Brasil. A ontologia foi formalizada em linguagem lógica com o auxílio do editor Protegé. Como forma de avaliar a ontologia desenvolvida, elaborou-se questões de competência que foram executadas em linguagem SPARQL.

Resultados: 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.

Conclusão: 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.

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

*Mestre em Ciência da Informação pela Universidade Federal de Santa Catarina – UFSC. E-mail: luana-moura@hotmail.com.

**Doutora em Ciência da Informação. Professora do Departamento de Ciência da Informação da Universidade Federal de Santa Catarina - UFSC. E-mail: marisa.brascher@gmail.com.

1 INTRODUÇÃO

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.

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 Semântica¹ 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

¹O 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.

ser evitados (SOUZA; ALVARENGA, 2004). Ao possibilitar que os termos sejam ‘compreendidos’ em seus contextos de uso, é possível trazer significados próximos aos 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. Em acervos digitais que reúnem documentos de diferentes épocas e culturas, é também importante tratar, do ponto de vista da Recuperação da Informação (RI), 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 os contextos de uso, marcados muitas vezes por aspectos geográficos e socioculturais.

Assim, o futebol, objeto de estudo deste trabalho, tem o jornal como uma importante fonte de informação e apresenta variações linguísticas pertinentes ao estudo proposto. Diante disso, este artigo apresenta as etapas de elaboração de uma ontologia, aplicada ao domínio do futebol presente nos acervos digitais de jornais. O objetivo é oferecer 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, de forma a aprimorar a Recuperação da Informação (RI) em acervos digitais de jornais.

Para tal, foi utilizado como base o Acervo digital Estadão², 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.

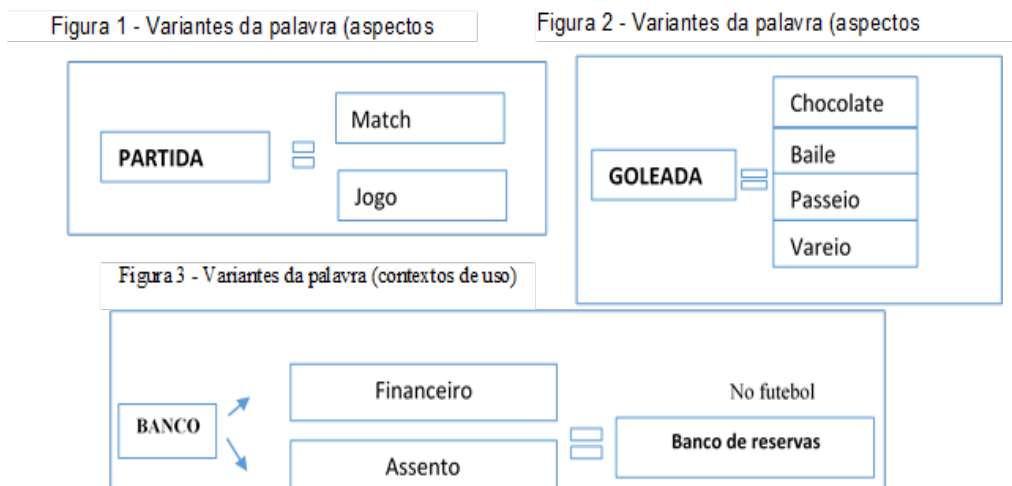
²Acervo Estadão: Disponível em: <http://acervo.estadao.com.br/>. Acesso em: 21 de jan. 2016.

2 RECUPERAÇÃO DA INFORMAÇÃO

Conforme Manning e Schütze (1999, p.1), a Recuperação da Informação (RI) consiste em “encontrar material (usualmente documentos) de natureza não estruturada (usualmente textos) que satisfaça 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”

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. Nos estudos de Bourne (1977 apud LOPES, 2002), estas “formas variantes de palavras” influenciam diretamente nas buscas. Considerando os exemplos apresentados nas figuras 1, 2 e 3, é possível entender de forma resumida algumas dessas questões:



Fonte: Elaborado pelas autoras

A figura 1 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 2 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 3, 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.

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. Krovetz e Croft (1992) avaliam essa problemática, destacando que 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.

Geralmente, os Sistemas de Recuperação da Informação (SRI) representam seus documentos apenas com palavras-chave e, nesse sentido, RI pode ficar limitada, pois aspectos linguísticos, tais como sinonímia, homografia e a polissemia podem interferir nos resultados de busca. Todavia, 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).

Conforme Moreiro González (2011), as ontologias permitem a definição de relações semânticas mais complexas, que as existentes em outros Sistemas de Organização do Conhecimento (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 ONTOLOGIAS DE DOMÍNIO

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 de como os conceitos estão inter-relacionados, o que resulta na estruturação do domínio e nas restrições de possíveis interpretações de seus termos”. Para 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”.

Um exemplo que ilustra o papel da ontologia é o uso do termo “banheira”, que de acordo com o Dicionário da Língua Portuguesa - Aurélio,³ 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.

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

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

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 de dez 2014].

tarefas e Ontologias de aplicação. A ontologia a ser utilizada nesta pesquisa é classificada como ontologia de domínio com formalidade leve, compreendida como aquela que descreve o vocabulário relativo a um domínio específico através da especialização de conceitos presentes na ontologia de alto nível.

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.

4 PROCEDIMENTOS METODOLÓGICOS

Como forma de traçar a tipologia da pesquisa, optou-se pelo uso dos critérios definidos por Lakatos e Marconi (1991). A pesquisa é caracterizada quanto aos objetivos como aplicada, haja vista a sua orientação para a resolução de problemas voltados à aplicação prática. Quanto aos procedimentos técnicos caracteriza-se como uma pesquisa exploratória e descritiva. Para o levantamento bibliográfico adotou-se como fonte principal o portal da Capes, por este conter inúmeras outras bases das mais variadas áreas do conhecimento que disponibilizam estudos que deram subsídios para o desenvolvimento da pesquisa. Devido aos objetivos da pesquisa, foram consultados dicionários e glossários especializados sobre o futebol para a formação do corpus de termos, além do uso do Acervo Estadão utilizado para coleta das informações e demonstração da recuperação da informação com a implementação da ontologia.

Para o alcance dos resultados desta pesquisa, a construção da ontologia do futebol para uso em acervos digitais de jornais, foi utilizada a metodologia *OntoForInfoScience*, proposta por Mendonça (2015). Dessa metodologia foram aplicadas as seguintes etapas: Avaliação da necessidade da ontologia; Especificação da ontologia; Aquisição de informações; Conceitualização;

Fundamentação da ontologia, Formalização da ontologia e Avaliação da ontologia. A ontologia desenvolvida na pesquisa foi nomeada de *OntoFootballForNewspapers*. Para construção da ontologia em nível formal foi utilizado o editor de ontologias *Protégé* 5.0. Após o desenvolvimento da ontologia, foi realizada a demonstração da recuperação da informação. As etapas são descritas de forma mais detalhada juntamente com os resultados.

5 RESULTADOS

Os resultados abrangem os caminhos utilizados para o desenvolvimento da *OntoFootballForNewspapers*, assim como sua avaliação.

Antes de iniciar o desenvolvimento da ontologia é preciso avaliar a necessidade de sua 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 adequada 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?”

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.

Nesse sentido, levando em consideração 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 que confirmam essa decisão são:

- a) possuir contexto dinâmico para descrição de recursos de um dado domínio do conhecimento – haja vista 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 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 o tipo de ontologia e o grau de formalidade que ela envolve. Essas informações estão detalhadas no quadro 1.

Quadro 1 - Template de Especificação da *OntoFootballForNewspapers*

Domínio/Esopo Geral
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.
Propósito Geral
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.
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 Unified Foundational Ontology (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. Seguidamente 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: <ol style="list-style-type: none">1. Impedimento e Escanteio é atribuição de quem? Resposta: Bandeirinha e seus sinônimos2. Quais times possuem jogadores que participaram da Copa do Mundo de Futebol?

Fonte: Elaborado pelas autoras.

5.2 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, 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.

O glossário de conceitos foi extraído de forma manual nas obras acima mencionadas. 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.

O glossário de verbos reutilizou os verbos *is_a* (é um) e *part_of* (parte de) empregados na ontologia sobre os componentes do sangue humano – Hemonto, desenvolvida como aplicação da metodologia *OntoForInfoScience*. Para coleta dos outros verbos da *OntoFootballForNewspapers* foram analisados 10 textos extraídos de documentos de referências contidos no acervo digital Estadão, ambiente de pesquisa deste trabalho. Esta etapa levantou o conhecimento a ser considerado na modelagem da ontologia. Muitas ontologias reutilizam instrumentos já construídos, como taxonomias e tesouros, para essa etapa. 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.

5.3 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. Nesse momento 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 dessa etapa é a elaboração do Dicionário de Conceitos, momento em que foram agregadas definições e sinônimos aos conceitos que fazem parte do Glossário de Conceitos, desenvolvido na etapa anterior de Aquisição das Informações. Para a formação do Dicionário de Conceitos foram utilizados os seguintes 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.

Os sinônimos incluídos nesta ontologia apresentam o mesmo propósito afirmado por Dodebei (2002, p.91) para as relações de equivalência 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 derivado 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 2.

Quadro 2 - Exemplo do Dicionário de Conceitos da OntoFootballForNewspapers

Dicionário de Conceitos			
ID	Conceito	Sinônimos	Definição
	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.
	Assistência	Passe; Shoot; 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.
	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.

Fonte: Elaborado pelas autoras com base em Mendonça (2015)

O próximo passo é o desenvolvimento da Tabela de Conceitos e Valores. Para realizar esta tarefa, foi utilizado o Dicionário de Conceitos e a partir de cada entrada, definidos 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. A Tabela de Conceitos e Valores apresentada no quadro 3 ilustra como ela se encontra estruturada.

Quadro 32 - Tabela de Conceitos e Valores

Tabela de Conceitos e Valores				
ID	Conceito	Sinônimos	Definição	Valores
	Armação	Marmelada	Jogo em que houve acerto de placar entre os dois times.	Time
	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.	-
	Avenida	-	Grande espaço que se abre para um jogador conduzir a bola sem encontrar adversário pela frente.	Adversário

Fonte: Elaborado pelas autoras com base em Mendonça (2015)

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 corresponde a 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. O quadro 4 apresenta parte do dicionário construído na pesquisa.

Quadro 4 - 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.

Fonte: Elaborado pelas autoras com base em Mendonça (2015)

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>, conforme exemplos do quadro 5. 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 5 – Exemplo do Glossário das Relações da *OntoFootballForNewspapers*

Glossário das Relações			
ID	Termo 1	Verbo	Termo 2
.	Assistência	É uma	Jogada
.	Banco	Contém	Jogador
.	Bandeirinha	Indica	Escanteio

Fonte: Elaborado pelas autoras com base em Mendonça (2015)

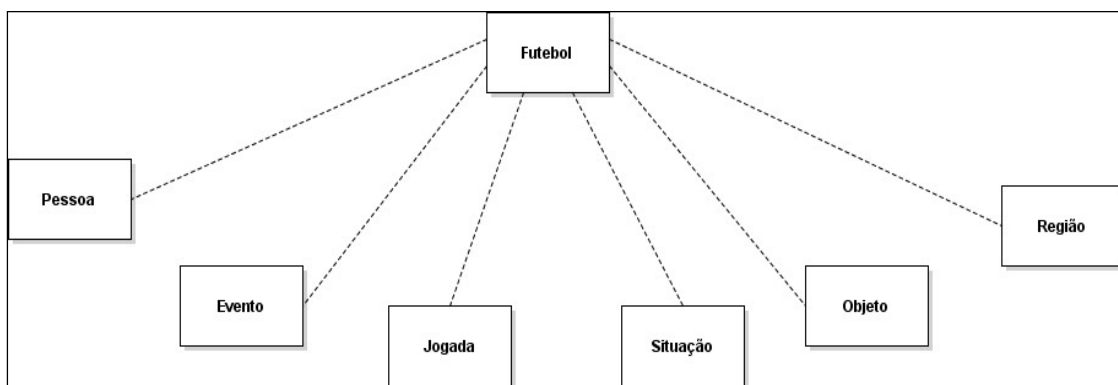
5.4 Conteúdo ontológico

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 em classes principais. Para desenvolvimento da

OntoFootballForNewspapers foi utilizada uma ontologia de fundamentação, a *Unified Foundational Ontology* (UFO) para a definição das classes principais da ontologia e inclusão dos conceitos da amostra do domínio.

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 4:

Figura 4 - Classes da *OntoFootballForNewspapers*



Fonte: Elaborado pelas autoras.

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 encaixa 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.

O quadro 6 exemplifica algumas das propriedades de relações realizadas na *OntoFootballForNewspapers*:

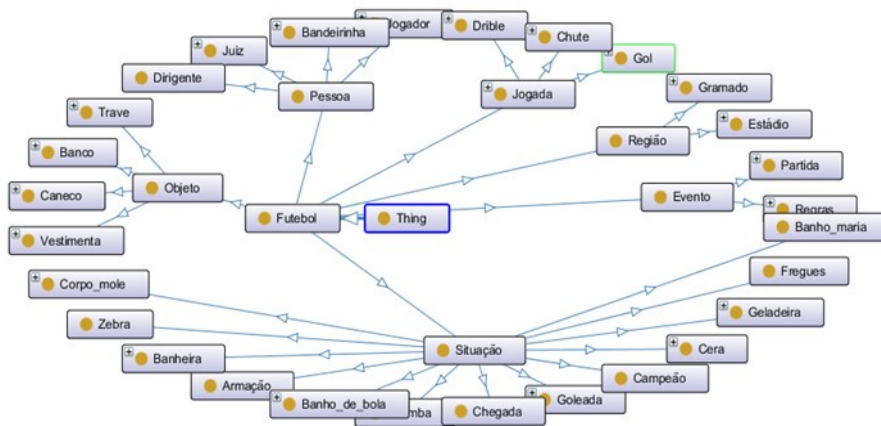
Quadro 6 - Preenchimento das relações ontológicas das classes no Protégé

<i>PROPERTY</i>	<i>DOMAINS</i>	<i>RANGE</i>
Bloqueia	Barreira Goleiro Zagueiro	Gol
Indicado_por	Regras	Bandeirinha Juiz
Parte_de	Trave Gramado Gaveta	Estádio

Fonte: Elaborado pelas autoras

E a figura 5 é possível visualiza-las de forma gráfica, por meio do suporte *OntoGraf*, disponibilizado no *Protégé*.

Figura 5 - Representação gráfica das principais classes OntoFootballForNewspapers



Fonte: Captura de tela Protégé da *OntoFootballForNewspapers*

Finalizada a ontologia, foram aplicadas questões de competência para validá-la, conforme exposto a seguir.

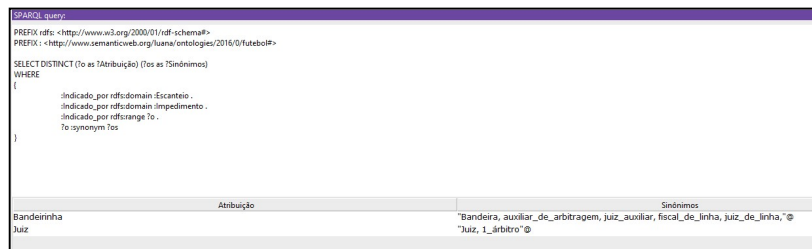
5.5 *Elucidação das questões de competência*

Para responder às questões de competência foi utilizado a linguagem *Protocol and RDF Query Language* (SPARQL), que permite realizar consultas e obter resultados em documentos semânticos. O SPARQL é dividido em três partes: linguagem de consulta, formato dos resultados e protocolo de acesso.

Com 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 6⁵.

⁵Dê zoom na imagem para visualizar melhor.

Figura 6 - Execução da questão de competência 1 no SPARQL



```
SPARQL query:
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX : <http://www.semanticweb.org/luana/ontologies/2016/0/futebol#>
SELECT DISTINCT (To as ?Atribuicao) (Tos as ?Sinonimos)
WHERE
{
  :indicado_por rdf:type domain :Escanteio .
  :indicado_por rdf:type domain :Impedimento .
  :indicado_por rdf:type range To .
  To synonym Tos
}
```

Atribuicao	Sinonimos
Bandeirinha	"Bandeira, auxiliar_de_arbitragem, juiz_auxiliar, fiscal_de_linha, juiz_de_linha,"@
Juiz	"Juiz, 1_arbitro"@

Fonte: Captura de tela do Protégé

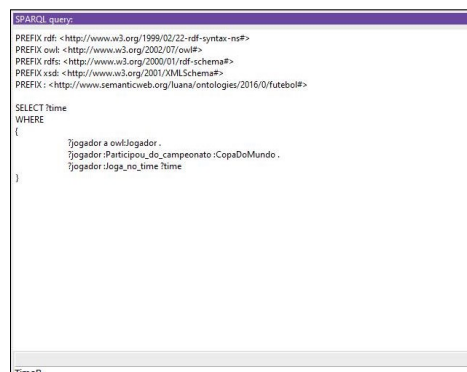
Para chegar nessas respostas foram incluídas as seguintes sentenças:

- Sujeito: Impedimento *and* Escanteio (*domain*);
- Predicado/relação: indicado_por (*range*);
- Objeto/Propriedade: é a resposta que buscamos;
- 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 - 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 7⁶.

Figura 7 - Execução da questão de competência 2 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 rdf: <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
}
```

TimeB

Fonte: Captura de tela do Protégé.

Na segunda questão as instâncias criadas para as subclasses jogadores (jogador 1, jogador 2, jogador 3...), time (time A time B, time C) e campeonato

⁶ Dê zoom na imagem para visualizar melhor.

(Copa do Mundo), foram ligadas as propriedades “Joga_no_time” e “Participa_do”Campeonato”, para inferir que o time B era o único que tinha jogadores participantes na copa do mundo, conforme mostrado na figura 10. Essa questão mostra como as ontologias podem relacionar objetos do mundo real e deduzir, por meio de inferência automatizada, respostas 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 semânticas 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 a 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.

5.6 Análise 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 desenvolvida. O acervo Estadão, ambiente de pesquisa deste trabalho, também foi utilizado nesta etapa. Para guiar este estágio uma pesquisa foi utilizada para apoiar na demonstração da recuperação da informação, sendo ela: Campeões no futebol no Brasil.

No desenvolver da 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 por palavra-chave no acervo os termos “futebol”, “*football*”, “caneco”, “taça”, “troféu”, “tropheu”, “tropheo”, conforme a figura 8:

Figura 8 - 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 7:

Quadro 7- Termos recuperados Acervo Estadão

Termo	Resultado	Período
Futebol	19.4248	1920 a 2015
<i>Football</i>	3.128	1900 a 2015
Taça	76.253	1.875 a 2015
Troféu	27.196	1875 a 2015
Tropheu	26.967	1875 a 2015
Tropheo	26.967	1875 a 2015
Caneco	1.458	1879 a 2015
Campeão	104.128	1875 a 2015

Fonte: Elaborado pelas autoras

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 os termos “Taça” “Caneco”, “Troféu”, “Tropheu” e “Trophéo” e atribui a propriedade “Conquista” para relacionar os termos “Jogador” e “Time”.

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”. 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 10.

Quadro 8 - Resultados obtidos com o uso de operadores booleanos

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

Fonte: Elaborado pelas autoras

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 9.

Figura 9 - Demonstração da recuperação da Informação no acervo digital



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”.

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.

6 CONSIDERAÇÕES FINAIS

A presente pesquisa mostrou como principal contribuição 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.

Como forma de apoiar a recuperação da informação, foi desenvolvida uma ontologia, que nesse caso conseguiu tratar os diferentes problemas que envolveram aspectos da 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, e também aspectos da sincronia, que auxiliou no tratamento de termos polissêmicos, como aqueles oriundos de diferentes regiões geográficas.

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 o letramento do usuário, pois com os relacionamentos formados, não é necessário realizar inúmeras buscas para recuperar conceitos e expressões equivalentes.

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.

REFERÊNCIAS

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. 2016.

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. 2016.

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 set. 2016.

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

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

BERNERS-LEE, T.; HENDER, J.; LASSILA, O.. **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.html#84A9809EC588EF21>>. Acesso em: 02 set. 2016.

BHOGAL, 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 set. 2016.

BOCCATO, V. R. C.; RAMALHO, R. A. S.; FUJITA, M. S. L. A contribuição dos tesauros 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.

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

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.

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

DODEBEI, V. **Tesauro: linguagem de representação da memória documentaria**. Rio de Janeiro: Intertexto, 2002.

FARIA, M. J. F. L. **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.

GUARINO, N. **Formal ontology in information systems**. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15, 1998.

HAAV, H.; LUBI, T. **A Survey of concept-based information retrieval tools on the web**. 2001. Disponível em: <<http://www.mii.vu.it/ADBIS/local2/haav.pdf>>. Acesso em: 24 set 2016.

KROVETZ, R.; CROFT, W. B. 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 set 2016.

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

LOPES, I. L. 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 set. 2016.

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 set 2016.

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

MELO, F. J. D.; BRASCHER, M. **Fundamentos da lingüística para a formação do profissional da informação**. Brasília: Thesaurus, 2011.
MENDONÇA, F. M. **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. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/BUBD-A35H3K>>. Acesso em: 17 set 2016.

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

MOURA, D. O. 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. 2016.

REIS, L. P.; LAU, N. **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 set 2016.

SALES, R.; CAFÉ, L. 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 set. 2016.

SCHIESSL, M. **Lexicalização de ontologias**: o relacionamento entre conteúdo e significado no contexto da Recuperação da Informação. 2015. 259 f. Tese (Doutorado) - Curso de Ciência da Informação, Programa de Pósgraduação em Ciência da Informação da Universidade, Universidade de Brasília, Brasília, 2015. Disponível em: <http://www.repositorio.unb.br/bitstream/10482/18663/1/2015_MarceloSchiessl.pdf>. Acesso em: 02 mar. 2017.

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: 06 set. 2016.

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. 2016.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. 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. 2016.

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 set de 2016.

USCHOLD, M.; JASPER, R. **A framework for understanding and classifying ontology applications**. 1999. Disponível em: <<http://ceur-ws.org/Vol-18/11-uschold.pdf>>. Acesso em: 10 ago. 2016.

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

Title

Use of ontology in the retrieval of information in digital collections of newspapers

Abstract:

Introduction: It aims at modeling an ontology of the soccer field for the treatment of diachronic and synchronic variations of the language;

Objective: To support the retrieval of information in digital collections of newspapers.

Methodology: This is an applied research, using as basis a digital journal collection. It uses the methodology OntoForInfoScience, de Mendonça (2015) to develop the ontology of the soccer field. Information collection was carried out on domain reference materials and newspaper news. Chronologically, the established cut covers terminology used between 1900 to 2015, a period that contemplates the existence of football clubs in Brazil. The ontology was formalized in logical language with the help of the editor Protegé. As a way of evaluating the developed ontology, competence issues were elaborated that were executed in SPARQL language. To verify the use of the ontology in environments composed by printed and digital newspapers, demonstrative searches were carried out in a real collection.

Results: The analysis of the results showed that without the use of the ontology in the digital collections of newspapers, the information retrieval is exhaustive and retrieves documents that are not relevant due to the absence of relationships between the terms that form the domain.

Conclusion: With the inclusion of the ontology, the search for information can dispense with both the user's literacy, because with the relationships formed, it is not necessary to perform numerous searches to retrieve equivalent concepts and expressions.

Keywords: Information Retrieval. Digital newspaper collection. Domain Ontology. Soccer.

Titulo

El uso de ontologías en la recuperación de información en las colecciones digitales de los diarios

Resumen:

Introducción: Visa modelado de una ontología de dominio de fútbol para el tratamiento de las variaciones diacrónico y sincrónico de la lengua;

Objetivo: apoyar la recuperación de la información en las colecciones digitales de los periódicos.

Metodología: Se trata de una investigación aplicada, utilizando como base una colección digital de periódicos. Utiliza la metodología OntoForInfoScience de Mendonça (2015) para desarrollar el dominio de fútbol ontología. La recolección de datos se llevó a cabo en el campo de los materiales de referencia y periódicos. Cronológicamente, el conjunto de recorte cubre la terminología utilizada entre 1900 a

2015, período que incluye la existencia de clubes de fútbol en Brasil. La ontología se formalizó en lenguaje lógico con la ayuda del editor de Protégé. Con el fin de evaluar la ontología desarrollada, se elaboró cuestiones jurisdiccionales que se han realizado en el lenguaje SPARQL. Para certificar el uso de la ontología en entornos compuestos para búsquedas de impresión y periódicos digitales demostrativa se llevaron a cabo en una colección real.

Resultados: El análisis de los resultados mostró que sin el uso de la ontología en colecciones digitales de los periódicos, la recuperación de la información es exhaustiva y recupera documentos no es relevante debido a la ausencia de relaciones entre términos que forman el dominio.

Conclusión: Con la adición de la ontología, la búsqueda de información puede hacer sin mucho alfabetización usuario, al igual que con las relaciones establecidas, no es necesario realizar numerosas búsquedas para recuperar conceptos y expresiones equivalentes

Palabras clave: Recuperación de la información. colección digital del periódico. ontología de dominio. Fútbol.

Recebido em: 02.04.2017

Aceito em: 05.10.2017