



Programa Interdisciplinar de Pós-Graduação em

# **Computação Aplicada**

**Mestrado Acadêmico**

EMÍLIO LUIZ FARIA RODRIGUES

GERAÇÃO DE PERGUNTAS EM LINGUAGEM  
NATURAL A PARTIR DE BASES DE DADOS ABERTOS  
E CONECTADOS: UM ESTUDO EXPLORATÓRIO

São Leopoldo, 2017

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS  
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO  
PROGRAMA INTERDISCIPLINAR DE PÓS-GRADUAÇÃO  
EM COMPUTAÇÃO APLICADA  
NÍVEL MESTRADO

EMÍLIO LUIZ FARIA RODRIGUES

**GERAÇÃO DE PERGUNTAS EM LINGUAGEM NATURAL A PARTIR  
DE BASES DE DADOS ABERTOS E CONECTADOS: UM ESTUDO  
EXPLORATÓRIO**

São Leopoldo  
2017

EMÍLIO LUIZ FARIA RODRIGUES

**GERAÇÃO DE PERGUNTAS EM LINGUAGEM NATURAL A PARTIR  
DE BASES DE DADOS ABERTOS E CONECTADOS: UM ESTUDO  
EXPLORATÓRIO**

Dissertação apresentada à avaliação como requisito parcial para a obtenção do título de Mestre, pelo Programa Interdisciplinar de Pós-Graduação em Computação Aplicada da Universidade do Vale do Rio dos Sinos – UNISINOS.

Orientador: Prof. Dr. Sandro José Rigo

São Leopoldo  
2017

Dados Internacionais de Catalogação na Publicação (CIP)

R696g Rodrigues, Emílio Luiz Faria.

Geração de perguntas em linguagem natural a partir de bases de dados abertos e conectados: um estudo exploratório / Emílio Luiz Faria Rodrigues.

–. São Leopoldo, 2017.

93p. : il.; 30 cm.

Orientador: Prof. Dr. Sandro José Rigo.

Dissertação (mestrado em Computação Aplicada) – Universidade Vale do Rio dos Sinos - UNISINOS. Curso de Programa Interdisciplinar de Pós Graduação em Computação Aplicada, 2017.

1. Dados abertos e conectados. 2. Geração de linguagem natural. 3. Web semântica. I. Rigo, Sandro José (orientador). II. Título.

CDD- 004.678

Maria de Fátima Freire de Araújo – CRB/11-374

EMÍLIO LUIZ FARIA RODRIGUES

GERAÇÃO DE PERGUNTAS EM LINGUAGEM NATURAL A PARTIR DE BASES DE  
DADOS ABERTOS E CONECTADOS: UM ESTUDO EXPLORATÓRIO

Dissertação apresentada à Universidade do  
Vale do Rio dos Sinos – Unisinos, como  
requisito parcial para obtenção do título de  
Mestre em Computação Aplicada.

Aprovado em 04 de dezembro de 2017

BANCA EXAMINADORA

---

Prof. Dr. Sandro José Rigo – UNISINOS

---

Prof. Dr. Silvio Cazella – UFCSPA

---

Prof. Dr. Jorge Luís Victória Barbosa – UNISINOS

Prof. Dr. Sandro José Rigo (Orientador)

**Visto e permitida a impressão**

São Leopoldo, *(data deve ficar em branco)*

Prof. Dr. Sandro José Rigo  
Coordenador PPG em Computação Aplicada

## **AGRADECIMENTOS**

Primeiramente agradeço a minha esposa Cristina e meus filhos pela paciência e compreensão durante este período, a minha esposa, principalmente, sempre me incentivando e acreditando na conclusão, dando o suporte necessário na minha ausência, sem ela provavelmente seria mais difícil ficar longe da família e concluir essa caminhada.

Agradeço também ao meu Orientador professor Dr. Sandro José Rigo, um verdadeiro companheiro e amigo, onde não mediu esforços para me orientar e ajudar a traçar o meu caminho na direção certa, com sua experiência e paciência, o qual não teria chegado agora na reta final.

Aos meus pais, que sempre me incentivaram e me ajudaram na minha caminhada pessoal e profissional, sem medir esforços para que eu conseguisse trilhar o meu caminho, e ainda, aos meus irmãos André Luiz, Márcia Andréa, Valdir Júnior, Nelinielce e Sheyla, pelo apoio e incentivo.

Agradeço também aos colegas da UNISINOS, especialmente ao Augusto da Silva e a Jessica Braun, pelo suporte na implementação do protótipo e a avaliação de resultados e ainda do aluno Thiago Caauê, aluno do TADS do IFRR (Instituto Federal de Educação, Ciência e Tecnologia de Roraima).

Aos meus amigos e companheiros de trabalhos do IFRR, por trilharem junto comigo o sucesso dessa caminhada, sei que não foi fácil, mais quase todos concluíram.

E finalizando, não poderia esquecer de agradecer ao IFRR pela oportunidade de entrar no Mestrado pela parceria firmada com a UNISINOS e por acreditar e investir em seus servidores na sua qualificação profissional.

Obrigado a todos que torceram e torcem pelo meu sucesso!



## RESUMO

O crescimento acelerado das bases de dados abertas e conectadas vem sendo observado recentemente. Existem diversas motivações para tal, envolvendo desde a geração destas bases de forma automática a partir de textos, até a sua construção diretamente a partir de sistemas de informação. Este crescimento gerou um conjunto numeroso de bases de dados com grande volume de informações. Deste modo observa-se a possibilidade de sua utilização em larga escala, em sistemas de pergunta e resposta. Os sistemas de pergunta e resposta dependem da existência de uma estrutura de informações a ser usada como apoio na geração de frases e na conferência das respostas. O atual contexto das bases de dados abertas e conectadas proporciona este suporte necessário.

A partir de estudos da literatura, observou-se a oportunidade de maior utilização, em aplicações diversas, das possibilidades de geração de frases em linguagem natural a partir de bases de dados abertas conectadas. Além disso, foram identificados diversos desafios para a efetiva utilização destes recursos com esta finalidade. Desta forma, esse trabalho objetiva verificar quais os aspectos da estrutura de bases de dados abertas conectadas que podem ser utilizados como apoio na geração de perguntas em linguagem natural. Para tal foi desenvolvido um estudo exploratório e definida uma abordagem geral, testada em um protótipo que permitiu gerar frases de perguntas em linguagem natural com apoio em bases de dados abertas e conectadas. Os resultados foram avaliados por um especialista em linguística e foram considerados promissores.

**Palavras-Chave:** Dados Abertos e Conectados; Geração de Linguagem Natural; Web Semântica.



## **ABSTRACT**

The accelerated growth of open and connected databases has recently been observed. There are several motivations leading to it, some bases generate data from texts automatically, other bases are built straightaway from information systems. Therefore, a numerous set of data base with a huge volume of information is now available. Thus, the possibility of its use in large scale, in systems of question and answer is observed. The question and answer systems depend on the existence of an information structure to be used as support in generating sentences and checking the answers. The current background of open and connected data provides the essential support.

From literature studies, it was observed the opportunity to use more the possibilities of generating sentences in natural language from connected open databases, in different kinds of applications. In addition, several challenges have been identified to realize the effective use of this resource. So, this work aims to verify which aspects of the structure of connected open databases can be used as the support to generate questions in natural language. Since, an exploratory study was developed, and a general approach was established. Which was tested on a prototype that was able to generate natural language question sentences, supported by open and connected databases. The results were evaluated by a specialist in linguistics and were considered promising.

**Keywords:** Linked Open Data; Generation of Natural Language; Semantic Web.



## LISTA DE FIGURAS

|                                                                                                  |    |
|--------------------------------------------------------------------------------------------------|----|
| Figura 1 Camadas da Web Semântica.....                                                           | 25 |
| Figura 2 Integração de Dado DBpedia .....                                                        | 32 |
| Figura 3 Arquitetura típica de sistemas de perguntas e respostas .....                           | 34 |
| Figura 4 Arquitetura conceitual do sistema de perguntas-resposta.....                            | 38 |
| Figura 5 Arquitetura Pipeline do AskNow .....                                                    | 38 |
| Figura 6 Arquitetura Geral do VOX .....                                                          | 39 |
| Figura 7 Arquitetura do YAGO-QA.....                                                             | 40 |
| Figura 8 Tela do Demo SPARQL2NL.....                                                             | 42 |
| Figura 9 Visão geral do Sistema PortNLG.....                                                     | 43 |
| Figura 10 Representação do esquema do framework na lexicalização .....                           | 44 |
| Figura 11 Visão geral do processo de extração de Padrões Relacionais .....                       | 45 |
| Figura 12 Estágio dos processamentos e subtarefas do NaturalOWL .....                            | 46 |
| Figura 13 Exemplo de uso do método para a língua inglesa e alemã .....                           | 48 |
| Figura 14 Texto em língua natural (wikipedia) transformado em dados no formato RDF(dbpedia)..... | 54 |
| Figura 15 Dados da Wikipedia para a cidade de Porto Alegre .....                                 | 56 |
| Figura 16 Dados da DBpedia para Porto Alegre .....                                               | 57 |
| Figura 17 Relações da DBpedia para o conceito cidade.....                                        | 58 |
| Figura 18 Trecho em Jason para descrição parcial de resultado de consulta .....                  | 59 |
| Figura 19 Conceito sobre população total na DBpedia .....                                        | 61 |
| Figura 20 Exemplos de padrões de lexicalização .....                                             | 65 |
| Figura 21 Dinâmica geral do algoritmo definido .....                                             | 66 |
| Figura 22 Visão geral de componentes .....                                                       | 67 |
| Figura 23 Resumo de resultados .....                                                             | 72 |
| Figura 24 Tela Inicial .....                                                                     | 73 |
| Figura 25 Tela de seleção da Busca .....                                                         | 73 |
| Figura 26 Tela da Resposta da Consulta .....                                                     | 74 |



## LISTA DE TABELAS

|                                                                                                                                                                |    |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 Exemplos de Bases de Conhecimento .....                                                                                                               | 19 |
| Tabela 2 Tabela com os itens da proposta para Web Semântica .....                                                                                              | 25 |
| Tabela 3 Exemplo de padrões com triplas .....                                                                                                                  | 45 |
| Tabela 4 Comparação entre trabalhos sobre Perguntas e Respostas .....                                                                                          | 50 |
| Tabela 5 Comparação entre trabalhos sobre geração de linguagem .....                                                                                           | 51 |
| Tabela 6 Aspectos iniciais identificados .....                                                                                                                 | 62 |
| Tabela 7 Trecho parcial da tabela de análise, para o conceito<br><a href="http://dbpedia.org/page/Porto_Alegre">http://dbpedia.org/page/Porto_Alegre</a> ..... | 63 |



## LISTA DE SIGLAS

|        |                                                               |
|--------|---------------------------------------------------------------|
| SQL    | Structured Query Language (Linguagem de Consulta Estruturada) |
| SPARQL | Protocol and RDF Query Language                               |
| P&R    | Perguntas e Respostas                                         |
| GLN    | Geração de Linguagem Natural                                  |
| W3C    | World Wide Web Consortium                                     |
| OWL    | Web Ontology Language                                         |
| RDF    | Resource Description Framework                                |
| XML    | eXtensible Markup Language                                    |
| URI    | Uniform Resource Identifier                                   |
| URL    | Uniform Resource Locator                                      |
| HTTP   | Hypertext Transfer Protocol                                   |
| XLS    | Formato de arquivo do Microsoft Excel                         |
| CSV    | Comma-Separated Values (formato regular de arquivos de texto) |
| PLN    | Processamento de Linguagem Natural                            |
| IA     | Inteligência Artificial                                       |
| QA     | Question Answering (Perguntas e Respostas)                    |
| GLN    | Geração de Linguagem Natural                                  |
| BD     | Banco de Dados                                                |
| NQS    | Normalized Query Structure                                    |
| WEB    | Ou WWW – World Wide Web                                       |
| TREC   | Text REtrieval Conference                                     |



## SUMÁRIO

|                                                                                                                                                                |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>1 INTRODUÇÃO</b>                                                                                                                                            | <b>18</b> |
| 1.1 Motivação                                                                                                                                                  | 20        |
| 1.2 Questão de Pesquisa                                                                                                                                        | 21        |
| 1.3 Objetivos                                                                                                                                                  | 22        |
| 1.3.1 Objetivos Específicos                                                                                                                                    | 22        |
| 1.4 Metodologia                                                                                                                                                | 22        |
| 1.5 Organização do texto                                                                                                                                       | 23        |
| <b>2 EMBASAMENTO TEÓRICO</b>                                                                                                                                   | <b>24</b> |
| 2.1 Web Semântica                                                                                                                                              | 24        |
| 2.2 Ontologias                                                                                                                                                 | 26        |
| 2.2.1 Tipos Ontologias                                                                                                                                         | 27        |
| 2.2.2 Componentes da Ontologia                                                                                                                                 | 27        |
| 2.2.3 Linguagens e Ferramentas                                                                                                                                 | 27        |
| 2.3 Bases de Dados abertos e conectados                                                                                                                        | 28        |
| 2.3.1 Dados Abertos                                                                                                                                            | 28        |
| 2.3.2 Dados Conectados                                                                                                                                         | 29        |
| 2.3.3 DBpedia                                                                                                                                                  | 30        |
| 2.4 Processamento de Linguagem Natural (PLN)                                                                                                                   | 32        |
| 2.5 Sistemas de Pergunta e Resposta                                                                                                                            | 33        |
| 2.6 Geração de Linguagem Natural (GLN)                                                                                                                         | 35        |
| <b>3 TRABALHOS RELACIONADOS</b>                                                                                                                                | <b>37</b> |
| 3.1 Perguntas e Respostas                                                                                                                                      | 37        |
| 3.1.1 Sobre a aplicação de ontologias para orientar agentes a responder perguntas                                                                              | 37        |
| 3.1.2 AskNow: A Framework for Natural Language Query Formalization in SPARQL – Um Framework para formalização de consulta em Linguagem Natural no SPARQL       | 38        |
| 3.1.3 VOX System: A Semantic Embodied Conversational Agent exploiting Linked Data – Um agente conversacional incorporado explorando a semântica em Linked Data | 39        |
| 3.1.4 YAGO-QA: Answering Questions by Structured Knowledge Queries                                                                                             | 40        |
| 3.2 Geração de Frases                                                                                                                                          | 41        |
| 3.2.1 SPARQL2NL – <i>Verbalizing SPARQL queries</i>                                                                                                            | 41        |
| 3.2.2 PortNLG - Um Sistema de Realização Superficial para Geração de Textos em Português                                                                       | 42        |
| 3.2.3 RealText-lex: <i>A Lexicalization Framework for RDF Triples</i>                                                                                          | 43        |
| 3.2.4 NaturalOWL System - <i>Generating Natural Language Descriptions from OWL Ontologies</i>                                                                  | 46        |
| 3.2.5 <i>A language-independent method for the extraction of RDF verbalization templates</i>                                                                   | 47        |
| 3.2.6 <i>Generating Natural Language from Linked Data: Unsupervised template extraction</i>                                                                    | 48        |
| 3.3 Análise de trabalhos estudados                                                                                                                             | 49        |
| 3.3.1 Sobre Perguntas e Respostas                                                                                                                              | 49        |
| 3.3.2 Sobre Geração de Frases                                                                                                                                  | 50        |
| 3.3.3 Pontos de atenção observados                                                                                                                             | 51        |
| <b>4 MATERIAIS E MÉTODO</b>                                                                                                                                    | <b>53</b> |
| 4.1 Visão geral                                                                                                                                                | 53        |
| 4.2 Estudo de bases de dados abertas e conectadas                                                                                                              | 61        |
| 4.3 Algoritmo geral proposto para geração de frases e arquitetura de apoio                                                                                     | 64        |
| 4.4 Aspectos de Implementação e avaliação                                                                                                                      | 68        |
| <b>5 CONCLUSÃO</b>                                                                                                                                             | <b>75</b> |
| 5.1 Limitações da pesquisa                                                                                                                                     | 75        |
| 5.2 Trabalhos futuros                                                                                                                                          | 76        |

## 1 INTRODUÇÃO

Alguns avanços tecnológicos recentes, tais como o aumento na capacidade de comunicação, processamento e armazenamento de dados definiram uma situação na qual a quantidade de dados gerados e compartilhados é muito grande e seu potencial de uso muito promissor. Várias aplicações foram criadas a partir do uso destes recursos, tais como as aplicações de redes sociais, ferramentas de publicação de notícias, ambientes virtuais de aprendizagem, entre outras. A partir deste contexto é possível observar uma situação em que o número de dados disponíveis para os mais diversos setores da sociedade é muito grande, cresce em ritmo acelerado e o seu uso se torna cada vez mais difícil, devido ao volume e formato. Esta situação vem estimulando iniciativas para melhor utilização destes dados, sendo que uma delas é a iniciativa conhecida como Dados Abertos Conectados (HEATH, BIZER, 2011).

A definição de Dados Abertos Conectados dada por (Isotani e Bittencourt, 2015) descreve em primeiro lugar o aspecto também conhecido como Dados Abertos, que está ligado com iniciativas governamentais para publicação e disponibilização de dados gerados no âmbito de atividades institucionais. Algumas iniciativas como a Lei de Acesso à informação (Lei 12.527, 18 novembro de 2011<sup>1</sup>) promovem o suporte jurídico que estimula as instituições a também divulgarem constantemente os dados de suas atividades. Outro aspecto dos Dados Abertos está relacionado com iniciativas populares de divulgação de informação de forma espontânea e colaborativa, como no caso da Wikipedia (Paletta, Mucheroni, 2017).

As iniciativas conhecidas como Dados Abertos, entretanto, possuem uma fragilidade no aspecto de utilização, dependendo do tipo de formato empregado para sua descrição. Em geral os formatos tradicionalmente utilizados (documentos textuais, planilhas de cálculo, relatórios de bases de dados, por exemplo), carecem de padronização, podem depender de softwares específicos e proprietários, além de apresentarem uma descrição formal restrita, ou simplesmente não possuírem esta descrição. A Web Semântica proporciona recursos para descrever dados de modo a possibilitar a automatização do seu uso e de sua integração. Recursos como

---

<sup>1</sup> Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm)>

ontologias e padrões de descrição de recursos vêm sendo amplamente utilizados com esta finalidade. Estas iniciativas são suportadas por setores em diversas áreas (HASLHOFER, ISAAC, 2011) Um exemplo de melhorias proporcionadas com a Web Semântica é a iniciativa da DBPedia (LEHMANN, 2015), projeto que objetiva automaticamente anotar as informações disponibilizadas na Wikipedia, em formato estruturado.

Com base neste cenário, a iniciativa conhecida como Dados Abertos Conectados envolve não apenas o aspecto de acesso aos dados, que é proporcionado pelas iniciativas de Dados Abertos. Esta iniciativa destaca justamente o potencial para aplicações quando os dados também estiverem anotados segundo os padrões da Web Semântica. Além de estarem acessíveis, os dados estão representados de modo a serem usados por aplicações, diretamente (Isotani e Bittencourt, 2015).

O crescimento acelerado das bases de dados abertas e conectadas vem sendo observado ultimamente. Existem diversas motivações para tal. Algumas situações são a geração destas bases a partir de textos, outras a partir de sistemas de informação. Este crescimento gerou um conjunto número de bases de dados com grande volume de informações. Alguns exemplos destas bases são a DBpedia<sup>2</sup>, YAGO3<sup>3</sup>, Wikidata<sup>4</sup>, entre outros. Informações preliminares sobre estas iniciativas estão descritas na Tabela 1.

**Tabela 1 Exemplos de Bases de Conhecimento**

| Nome           | Descrição                                                                                                                                                                         | Tamanho dos Dados                  |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------|
| <b>DBpedia</b> | Base de conhecimento que representa um esforço colaborativo para extrair informações estruturadas da Wikipedia, tornando estas informações disponíveis na Web em formato anotado. | Mais de 16,9 milhões de entidades. |
| <b>YAGO3</b>   | É uma enorme base de conhecimento semântica, derivada da Wikipedia, WordNet                                                                                                       | Mais de 10 milhões de entidades.   |

<sup>2</sup> <http://www.dbpedia.pt/>

<sup>3</sup> <http://yago-knowledge.org/>

<sup>4</sup> [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page)

---

e GeoNames.

---

|                 |                                                                                                                                                                                                                                    |                                       |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|
| <b>Wikidata</b> | Projeto da Fundação Wikimedia: um banco de dados secundário, multilíngue, colaborativo e livre, que coleta dados estruturados para fornecer suporte à Wikipédia, ao Wikimedia Commons, aos demais projetos Wikimedia e muito mais. | Mais de 26 milhões de itens de dados. |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------|

---

**Fonte:** dbpedia.pt/, yago-knowledge.org/, wikidata.org

Este contexto apresenta uma possibilidade de utilização em larga escala, como fonte de informações para sistemas de pergunta e resposta (Shekarpour et al., 2016). Os sistemas de pergunta e resposta possuem um funcionamento geral que está baseado em uma estrutura de informações a ser usada na geração de frases e na conferência das respostas. Alguns dos trabalhos nesta área são destinados a realizar uma transcrição entre as perguntas feitas em Linguagem natural pelos usuários e um formato de consulta, como SQL ou SPARQL (Dubey et al., 2016). Outros sistemas atuam também na geração das frases contendo as perguntas a serem realizadas (Ell e Harth, 2016). Para estas atividades observa-se uma tendência no uso ontologias como fonte destas informações, tanto para a geração de perguntas como para consulta de respostas (Androutsopoulos et al., 2013).

### 1.1 Motivação

A partir do contexto descrito, observa-se uma situação em que existe um conjunto de recursos (as bases de dados abertos conectados) que está em franco crescimento e que pode ser utilizado para sistemas de perguntas e respostas, seja com finalidades gerais, seja para finalidades educativas.

O desafio que se coloca para a sua utilização automática reside na definição de técnicas que permitam gerar automaticamente frases em linguagem natural, a partir dos dados abertos conectados, de modo a permitir que estas sejam usadas pelos sistemas de perguntas e respostas.

Os métodos para geração de linguagem natural necessitam alguns subsídios para sua execução (Reiter e Dale, 1997), sendo que um destes subsídios consiste em informações para o planejamento do texto a ser gerado. Este problema é

observado em diversos trabalhos estudados e em geral é resolvido com intervenção manual, o que é custoso e pode não ser eficiente em diversos contextos. Ocorre que nas bases de dados abertos conectados, parte dos recursos consiste exatamente neste tipo de subsídio, ou pode ser usado com esta finalidade. Ou seja, a estrutura interna de anotação dos dados nestas bases descreve exatamente um conjunto de relações que pode ser utilizada neste sentido. As bases de dados abertos conectados possuem uma estrutura e semântica definidas, com conceitos descritos e relacionados entre si. Desta forma, oferecem o suporte para os métodos de GLN, de modo que diversos trabalhos recentes indicam esta possibilidade [Staykova 2016; Hoffner et al. 2016] como promissora e analisam alguns de seus desafios.

Este recurso pode ser utilizado como base para sistemas de pergunta e resposta, bem como para jogos educacionais. Alguns trabalhos já foram desenvolvidos com propósitos restritos a um determinado contexto, como é o caso do trabalho de Dannélls et al. (2012), que apresenta uma aplicação multilíngue baseada em ontologias para informações de museus na Web. A abordagem baseia-se no uso da Web de dados ligados aplicados ao domínio do património cultural, tendo sido desenvolvido um aplicativo Web que usa ontologias da Web Semântica com técnicas de geração de linguagem natural para gerar descrições multilíngues coerentes sobre objetos de museus. Trabalho com objetivo similar pode ser observado em Molina et al (2016), com foco no uso de dados geográficos abertos para gerar automaticamente descrições para redes de sensores Hidrológicos.

Este trabalho visa desenvolver um estudo sobre as possibilidades de uso da estrutura de bases de dados abertos conectados para apoiar a descrição de um modelo que utilize esses dados disponíveis na web para montar frases contendo perguntas em linguagem natural.

## **1.2 Questão de Pesquisa**

Durante as pesquisas realizadas, observou-se a oportunidade de maior utilização, em aplicações diversas, das possibilidades de geração de frases em linguagem natural a partir de uma base de dados abertos conectados. Além disso, foram identificados diversos desafios para a efetiva utilização dos recursos de dados abertos conectados com esta finalidade.

Desta forma, esse trabalho visa responder a seguinte questão de pesquisa: Quais os aspectos da estrutura de bases de dados abertas conectadas que podem ser utilizados como apoio na geração de perguntas em linguagem natural?

### **1.3 Objetivos**

O objetivo geral deste trabalho é realizar um estudo exploratório que utiliza bases de dados abertos e conectados para gerar frases com perguntas em linguagem natural.

#### **1.3.1 Objetivos Específicos**

- a) Avaliar as características e potencialidades dos recursos de web semântica, em especial no caso de dados abertos e conectados, para seu uso na geração de perguntas em linguagem natural.
- b) Modelar um mecanismo de geração automática de perguntas usando dados abertos e conectados.
- c) Descrever e implementar uma aplicação para testar a geração de frases curtas com perguntas em linguagem natural.

### **1.4 Metodologia**

Os aspectos metodológicos considerados no desenvolvimento desta pesquisa iniciam com a realização de pesquisa bibliográfica para colher informações sobre o tema em análise. O Segundo passo envolveu uma análise de bases de dados abertos e conectados, para colher detalhes de interesse do trabalho. No terceiro passo foi especificada a solução a ser desenvolvida, junto com a definição de tecnologias para o desenvolvimento de um protótipo e de aspectos de avaliação.

Foi considerada a abordagem descrita por (KÖCHE, 2011), segundo a qual o presente trabalho pode ser caracterizado como uma pesquisa de caráter exploratório, pois a mesma busca inicialmente descrever em detalhes um determinado cenário no qual se identifica o problema tratado. Com base neste ponto inicial, busca-se demonstrar possibilidades para o encaminhamento de uma solução, que neste sentido foi descrita e validada com base no estudo de um caso específico.

Adotamos (WAZLAWICK, 2009) para definir o seguinte método de trabalho:

a) realização de revisão bibliográfica, estudo documental e de bases de dados abertas conectadas; b) realização de estudo e análise de trabalhos relacionados com as áreas tratadas; c) Modelagem e uma solução com os dados obtidos nas etapas anteriores; d) Descrição de um protótipo e de uma abordagem para avaliação; e) realização de estudo de caso para avaliar os resultados possíveis; f) descrição e documentação da pesquisa.

### **1.5 Organização do texto**

O texto deste trabalho está organizado como segue. No capítulo 2 é apresentado o referencial teórico. A seguir, no capítulo 3, são apresentados alguns trabalhos relevantes na área de geração de linguagem natural, perguntas e respostas e dados abertos conectados, que foram estudados neste trabalho. O capítulo 4 apresenta o estudo preliminar realizado e o modelo proposto. Aspectos de implementação são descritos no capítulo 4, juntos com elementos de avaliação. Por fim, no capítulo 5 são expostas as conclusões, limitações do trabalho e indicações de trabalhos futuros.

## 2 EMBASAMENTO TEÓRICO

Neste capítulo serão apresentados conceitos utilizados como embasamento para o trabalho desenvolvido. O referencial teórico apresentado tem como temas: Web Semântica, Ontologias, Bases de Dados abertos e conectados, Processamento de Linguagem Natural, Sistemas de Perguntas e Respostas e Geração de Linguagem Natural.

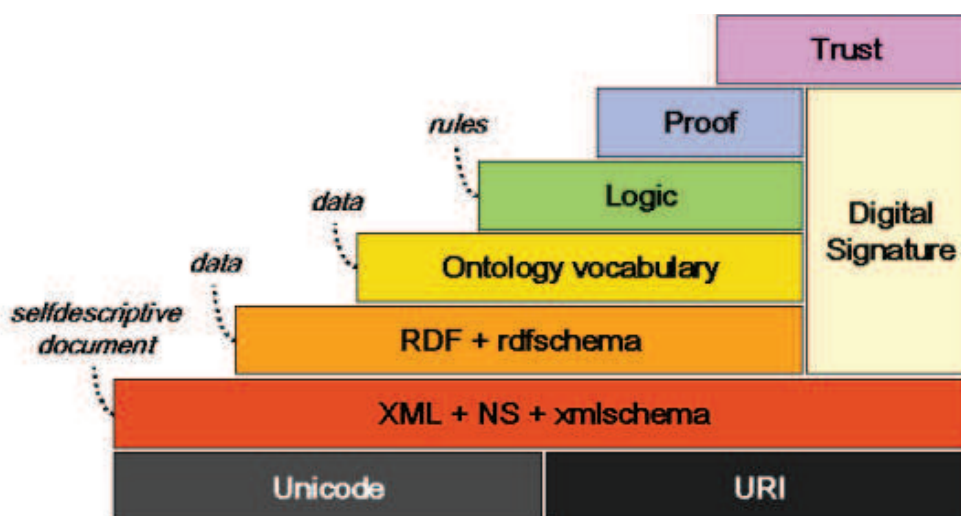
### 2.1 Web Semântica

Devido ao grande crescimento da internet nos últimos anos e com o aumento das possibilidades dos usuários em publicar e compartilhar suas próprias informações observa-se a disponibilidade de uma enorme quantidade de dados, o que torna o seu uso uma atividade bastante complexa (BEX et al., 2014). A partir da iniciativa de instituições como o consórcio W3C e seus grupos de trabalhos, a Web Semântica trata de iniciativas para facilitar a interoperabilidade dos dados e assim apoiar a criação de repositórios, vocabulários e regras para que se possam auxiliar nesta tarefa.

Alguns dos recursos utilizados consistem de taxonomias, que definem classes de objetos e as ligações entre elas. Para descrever uma ontologia na web, é possível utilizar-se da Linguagem de Ontologia para Web (OWL – *Web Ontology Language*). O uso de recursos para inferência permite melhorar a análise e utilização dos dados. A linguagem OWL é dotada de capacidade superior de propriedades e descrição de classes em relação ao RDF, incluindo a expressão de relações entre classes, restrição de cardinalidade, características de propriedades e classes enumeradas (MCGUINNESS; HARMELEN, 2003).

A Figura 1 descreve a estrutura para web semântica descrita por Tim Berners-Lee (BERNERS-LEE, 2015).

Figura 1 Camadas da Web Semântica



Fonte: ("W3C Semantic Web Activity", 2015)

Os elementos da Figura 1 são descritos na Tabela 2, a partir de seu significado e função. O vocabulário RDF é uma recomendação do W3C, sendo que ele será utilizado no modelo proposto neste trabalho e por isso foi reservado um tópico para descrevê-lo.

Tabela 2 Tabela com os itens da proposta para Web Semântica

| Item                     | Descrição                                                                                                                                                                                               |
|--------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| URI                      | Identificador Único de Recursos que possibilita a definição e adoção, de maneira precisa, de nomes aos recursos e seus respectivos endereços na Internet.                                               |
| UNICODE                  | Esquema padronizado de codificação dos caracteres, que diminui consideravelmente a possibilidade de redundâncias dos dados, pois funciona independentemente da plataforma utilizada.                    |
| <i>Digital Signature</i> | Conjunto de tecnologias desenvolvidas com o intuito de substituir em ambiente computacional a função exercida pela assinatura formal de uma pessoa em um suporte físico.                                |
| <i>Encryption</i>        | Consiste de um processo em que as informações são cifradas de modo que não possam ser interpretadas por qualquer pessoa ou sistema computacional, garantindo assim a confidencialidade das informações. |
| XML                      | É uma linguagem que possibilita a estruturação dos dados por meio da definição de elementos e atributos.                                                                                                |
| <i>Namespace</i>         | Coleção de nomes, identificados por um URI, que são utilizados em documentos XML para validar elementos e atributos.                                                                                    |
| RDF Core                 | Núcleo que compreende as especificações do modelo e a sintaxe da <i>Resource</i>                                                                                                                        |

|                        |                                                                                                                                                                                                                                                                       |
|------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                        | <i>Description Framework</i> (Estrutura de Descrição de Recursos), possibilitando a descrição dos recursos por meio de suas propriedades e valores.                                                                                                                   |
| RDF<br><i>Schema</i>   | Utilizada para a descrição do vocabulário RDF, possibilitando a definição de taxonomias de recursos em termos de uma hierarquia de classes.                                                                                                                           |
| SparQL                 | Linguagem computacional utilizada para realizar consultas a partir de estruturas RDF, favorecendo a recuperação de informações de maneira mais eficaz.                                                                                                                |
| OWL                    | Linguagem computacional recomendada pelo W3C para o desenvolvimento de ontologias.                                                                                                                                                                                    |
| <i>Rules</i>           | Permite a definição de regras lógicas relacionadas aos recursos informacionais.                                                                                                                                                                                       |
| <i>Logic Framework</i> | Camada para a definição de regras mais abrangentes, utilizadas no tratamento de informações, possibilitando que agentes computacionais possam realizar inferências a partir das relações existentes entre os recursos informacionais.                                 |
| <i>Proof</i>           | Espera-se que esta camada possibilite a verificação/comprovação da coerência lógica dos recursos, de modo que os aspectos semânticos das informações estejam descritos de maneira consideravelmente adequada, atendendo a todos os requisitos das camadas inferiores. |
| Trust                  | Camada de Confiança, a partir da qual se espera garantir que as informações estejam representadas de modo correto, possibilitando certo grau de confiabilidade.                                                                                                       |

**Fonte:** (RAMALHO, VIDOTTI, FUJITA, 2008)

## 2.2 Ontologias

O uso de ontologias no ambiente da Computação ocorre há vários anos, visando criar representações que fossem além da descrição de simples instâncias do domínio considerado (GAVA e MENEZES, 2003). GRUBER (1992) indica que uma ontologia define um conjunto de representações primitivas com o qual se modela um domínio de conhecimento.

Para GÓMES-PÉREZ (1999), ontologia é um conjunto de termos ordenados hierarquicamente para descrever um domínio que pode ser usado como um esqueleto para uma base de conhecimentos.

O SPARQL (*SPARQL Protocol and RDF Query Language*) fornece uma linguagem padronizada de consultas para grafos RDF (SEGARAN et al, 2009). A linguagem SPARQL permite a criação de consultas utilizando, entre outros, padrões de triplas, conjunções, disjunções e padrões opcionais.

### 2.2.1 Tipos Ontologias

Para GUARINO (1998) são cinco as classificações básicas para ontologias:

- ✓ Ontologias Genéricas: descrevem conceitos mais amplos. Não dependem de um problema específico (domínio).
- ✓ Ontologias de Domínio: representam conceitos de um domínio (área) específico.
- ✓ Ontologias de Aplicação: descrevem conceitos que estão relacionados com um domínio específico (área) e com uma tarefa específica.
- ✓ Ontologias de Representação: descrevem os conceitos que são usados para a representação do conhecimento.
- ✓ Ontologias de Tarefa: descrevem conceitos que são usados por processos (tarefas e atividades) de uma maneira geral. Não dependem de um problema específico (domínio).

### 2.2.2 Componentes da Ontologia

De acordo com GRUBER (1992) e NOY & MCGUINNESS (2001) os elementos que compõem uma ontologia são:

- ✓ Classes: ontologias organizadas em taxonomias (classificação e identificação dos termos presentes na ontologia).
- ✓ Relações: representam um tipo de interação entre conceitos e o domínio.
- ✓ Axiomas: são usadas para modelagem de sentenças que são sempre verdadeiras.
- ✓ Instâncias: utilizadas para representar elementos específicos, ou seja, os próprios dados.

### 2.2.3 Linguagens e Ferramentas

Existem dois grupos principais de linguagens para desenvolvimento de ontologias dividido em linguagens que se baseiam em uma lógica de primeira ordem e as que se baseiam em XML (eXtensible Markup Language). A OWL é a linguagem

de definição e criação de ontologias. Contempla um conjunto de axiomas e pode incluir descrições de classes e seus relacionamentos.

Existe um grande número de ferramentas de desenvolvimento disponíveis para se trabalhar com ontologias, que são: Apollo, OLEd, Servidor Ontolingua, OntoSaurus, Protégé, Jena e WEbOnto.

Uma das mais usadas é o Protegé que é uma ferramenta distribuída sob a licença *Open Source* Mozilla Public License (MPL) para desenvolvimento de ontologias para WEB, construída em JAVA e possui uma grande variedade de *plugins* para consultas e exportar a ontologia em diversas linguagens (OWL, RDF, RDFS, XML SCHEMA). Foi criada pela Universidade de Stanford em parceria com a Universidade de Manchester, estando disponível em: <http://protege.stanford.edu>.

## 2.3 Bases de Dados abertos e conectados

Nesta seção serão descritos alguns dos principais conceitos relacionados com Dados Abertos e Conectados e um dos exemplos mais relevantes de sua disponibilização em domínio aberto.

### 2.3.1 Dados Abertos

De acordo com a *Open Definition* (Open Definition, 2014), dados abertos são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa – sujeitos, no máximo, à exigência de atribuição à *fonte original* e ao *compartilhamento pelas mesmas* licenças em que as informações foram apresentadas. Diante disto, os dados quando publicados e disponibilizados não deverão conter restrições de acesso e a sua utilização seja distribuída de forma livre para qualquer pessoa que assim desejar utilizar.

Por esta perspectiva, a definição do termo dados abertos carrega três normas fundamentais (Open Knowledge Foundation, 2010):

- **Disponibilidade e acesso:** os dados devem estar disponíveis como um todo e sob custo não maior que um custo razoável de reprodução, e

preferencialmente devem ser possíveis de ser baixados pela Internet. Os dados devem também estar disponíveis de uma forma conveniente e modificável.

- **Reuso e redistribuição:** os dados devem ser fornecidos sob termos que permitam a reutilização e a redistribuição, inclusive a combinação com outros conjuntos de dados.

- **Participação universal:** todos devem ser capazes de usar, reutilizar e redistribuir. Não deve haver discriminação contra áreas de atuação ou contra pessoas ou grupos. Por exemplo, restrições de uso “não comercial” que impediriam o uso “comercial”, ou restrições de uso para certos fins (ex.: somente educativos) excluem determinados dados do conceito de “abertos”.

De acordo com Chignard (Chignard, 2013) o termo “Dados Abertos”, surgiu em 1995 e atualmente é crescente o movimento sobre dados abertos e esse crescimento é considerável. Em 2010 Floridi destaca em um estudo sobre a quantidade de dados gerados, que entre 2006 a 2010 o número cresceu de 166 exabytes para 988 exabytes. O trabalho de Klein (2017) destaca o crescimento da disponibilidade de bases de dados abertas, inclusive enfatizando o crescimento de dados governamentais abertos (Klein, 2017).

### 2.3.2 Dados Conectados

Os dados conectados ou dados vinculados, ou em inglês *Linked Data* (dados ligados entre si) é um conjunto de práticas introduzidas por Tim Berners-Lee (2006) em suas notas sobre arquitetura Web. A iniciativa de Linked Data usa a Web para conectar diversos dados publicados em diferentes fontes e baseia-se no uso de documentos contendo dados em RDF (Resource Description Framework) formato (Klyne e Carroll, 2004).

Berners-Lee (2006)<sup>5</sup> delineou um conjunto de "regras" para a publicação de dados na Web de os dados publicados se tornam parte de um único espaço de dados global:

1. Usar URIs como nomes para as coisas.
2. Usar URIs HTTP para que as pessoas possam procurar esses nomes.

---

<sup>5</sup> <https://www.w3.org/DesignIssues/LinkedData.html>

3. Quando alguém procurar um URI, fornecer informações úteis, usando os padrões (RDF, SPARQL).

4. Incluir links para outros URIs, para que eles possam descobrir mais coisas.

Para ISOTANI e BITTENCOURT o termo “Dados Conectados” representa práticas de boas publicações na Web usando padrões internacionais recomendados pelo W3C. (ISOTANI; BITTENCOURT; pág. 35, 2014).

A conexão e a publicação de dados de forma aberta vêm preocupando muitas entidades e governos, pensando nessa visão, Tim Berners-Lee<sup>6</sup> propôs um “Sistema de 5 Estrelas” para classificar o grau de abertura dos dados. Segundo este trabalho, as 5 estrelas para Dados Abertos são:

(1) Disponível na Internet (em qualquer formato; por exemplo, PDF), desde que com licença aberta, para que seja considerado Dado Aberto.

(2) Disponível na Internet de maneira estruturada (em um arquivo Excel com extensão XLS).

(3) Disponível na Internet, de maneira estruturada e em formato não proprietário (CSV em vez de Excel).

(4) Seguindo todas as regras anteriores, mas dentro dos padrões estabelecidos pelo W3C (RDF e SPARQL): usar URL para identificar coisas e propriedades, de forma que as pessoas possam direcionar para suas publicações.

(5) Todas as regras anteriores, mais: conectar seus dados a outros dados, de forma a fornecer um contexto.

### 2.3.3 DBpedia

A DBpedia representa um esforço colaborativo para extrair informações estruturadas da Wikipedia, tornando estas informações disponíveis na Web em formato anotado. A DBpedia permite realizar perguntas sofisticadas sobre informações na Wikipedia, opcionalmente incluindo ligações com outras fontes de dados ligados na Web. Um dos objetivos deste projeto é facilitar o uso da quantidade e variedade de informações na Wikipedia de maneiras novas e

---

<sup>6</sup> <http://5stardata.info/pt-BR/>

interessantes, e que inspirem novos mecanismos de navegação, interligação e melhoramento da própria enciclopédia<sup>7</sup>.

A versão inglesa da base de conhecimento da DBpedia descreve atualmente 6.0M de entidades, dos quais 4.6M possuem resumos (abstracts), 1.53M possuem coordenadas geográficas, além 1.6M em imagens / fotos. No total, existem 5.2M de recursos classificados em uma ontologia consistente, contendo 1.5M descrições de pessoas, 810K descrições de lugares (incluindo 505K descrições de capitais / regiões populosas), 490K em obras (incluindo 135K de álbuns musicais, 106K em filmes e 20K de vídeo games), 275K organizações (incluindo 67K em empresas e 53K instituições educacionais), 301K espécies e 5K em doenças.

O número total de recursos da DBpedia Inglesa totaliza 16.9M, além de 6.0M em recursos, incluindo 1.7M conceitos skos (categorias), 7.3M páginas de redirecionamento, 260K páginas de desambiguação e 1.7M em nós intermediários. No total, a DBpedia consiste em 9.5 bilhões de triplas dos quais 1.3 bilhões foram extraídas da versão inglesa da Wikipedia, 5.0 bilhões foram extraídos de outras edições do idioma e 3.2 bilhões da DBpedia Commons e da Wikidata<sup>8</sup>.

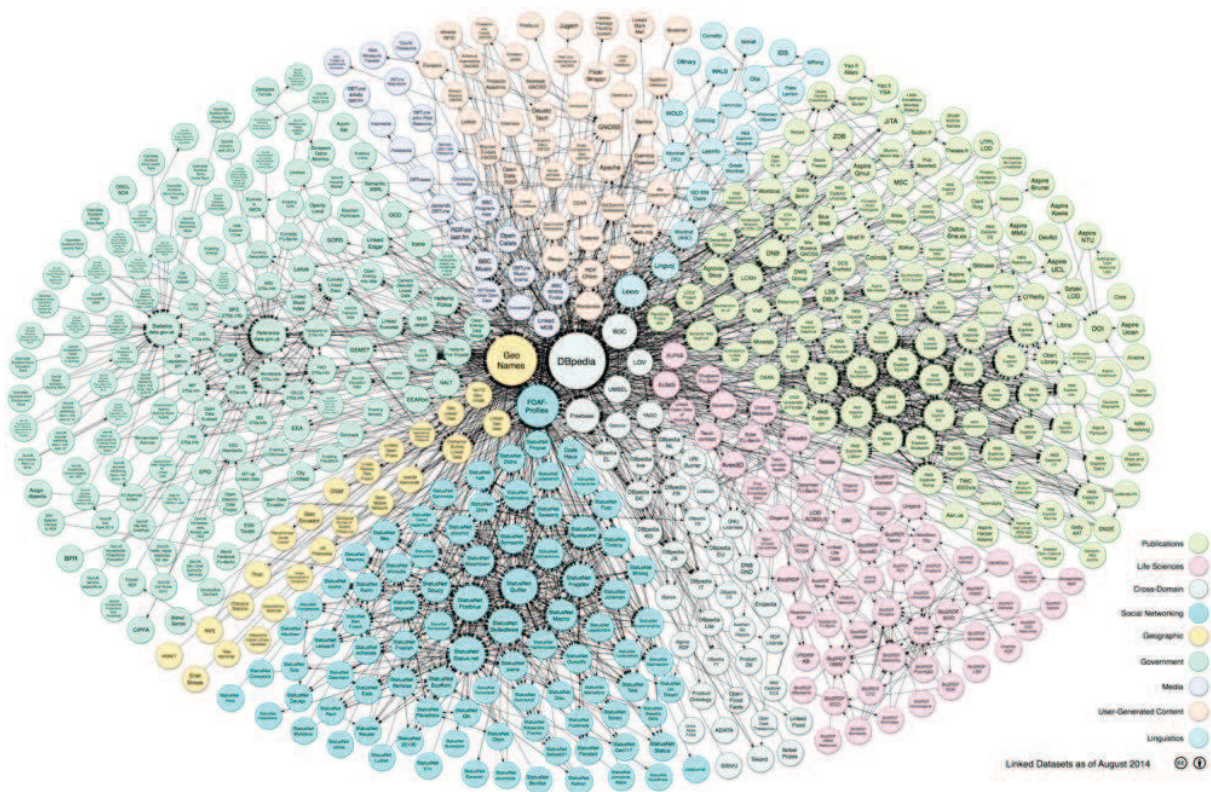
Na DBpedia a integração de Dados consiste em combinar dados provenientes de fontes de dados distintas de forma a prover uma visão unificada destes dados. Na Figura 2 (criada por Richard Cyganiak e Anja Jentzsch) mostra os objetivos de integrar a DPpedia com o maior número possível de fontes de dados, de forma a servir como um núcleo para a Web de Dados.

---

<sup>7</sup> [http://pt.dbpedia.org/o\\_que\\_e/](http://pt.dbpedia.org/o_que_e/)

<sup>8</sup> <http://www.dbpedia.pt/releases/2016/10/17/dbpedia-2016-04.html>

**Figura 2 Integração de Dado DBpedia**



**Fonte:** [http://pt.dbpedia.org/como\\_participar/integracao](http://pt.dbpedia.org/como_participar/integracao)

Acesso: 07nov2017

## 2.4 Processamento de Linguagem Natural (PLN)

O Processamento de Linguagem Natural (PLN) é uma subárea da Inteligência Artificial (IA), crucial para sistemas que necessitam tratar de modo mais amigável a entrada de dados do usuário. Utilizando-se de técnicas e padrões linguísticos, o PLN procura automatizar a compreensão da linguagem humana.

O PLN pode analisar dados em nível fonético (relacionando palavras através de seus sons), em nível morfológico (analisando a formação da palavra e seu radical), em nível sintático (considerando o papel de cada palavra na frase), em nível semântico (buscando o significado da palavra através de combinações com outras) e em nível pragmático (analisando o contexto para inferir ou mudar o significado) (GONZALES; LIMA, 2003).

A partir destes aspectos, é possível definir que a aplicação de um sistema PLN deverá passar por cinco etapas principais para realizar o que ela propõe são:

- ✓ **Análise Morfológica:** cada palavra é analisada e classificada isoladamente, levando em consideração a sua própria categoria gramatical da língua natural. Comparando com o estudo da língua portuguesa, equivaleria a classificar cada palavra segundo sua classificação morfológica (adjetivo, substantivo, verbo, etc.).
- ✓ **Análise Sintática:** analisa-se uma sequência linear de palavras a fim de analisar o seu relacionamento e emprego na frase. Comparando com o estudo da gramática portuguesa, equivaleria a classificar cada expressão segundo seu emprego (análise sintática) na frase, como sujeito, predicado verbal, etc. É esta a etapa correspondente à transformação da frase plena em estrutura hierárquica das unidades de significado da frase, também conhecido como parsing.
- ✓ **Análise Semântica:** nesta etapa as estruturas criadas pelo analisador sintático são associadas ao seu significado, mapeando assim cada estrutura sintática a um objeto no domínio na base de conhecimento. Esta é a fase onde é feito o processamento léxico, identificando o significado de cada palavra, além da resolução de problemas de ambiguidade.
- ✓ **Integração de Discurso:** esta parte pode ser comparada à interpretação do texto, onde a compreensão do todo é importante para que se possa compreender cada parte. Geralmente, o significado de uma frase é melhor compreendido a partir da interpretação da mesma junto com todas as frases antecedentes e sucessoras a ela.
- ✓ **Análise Pragmática:** o objetivo dessa etapa é de fazer a reinterpretação da estrutura frasal, para analisar se o que foi dito é realmente o que se quis dizer (uma solicitação, uma pergunta, uma afirmação, etc.) a fim de determinar melhor qual deverá ser a reação correta do programa.

## **2.5 Sistemas de Pergunta e Resposta**

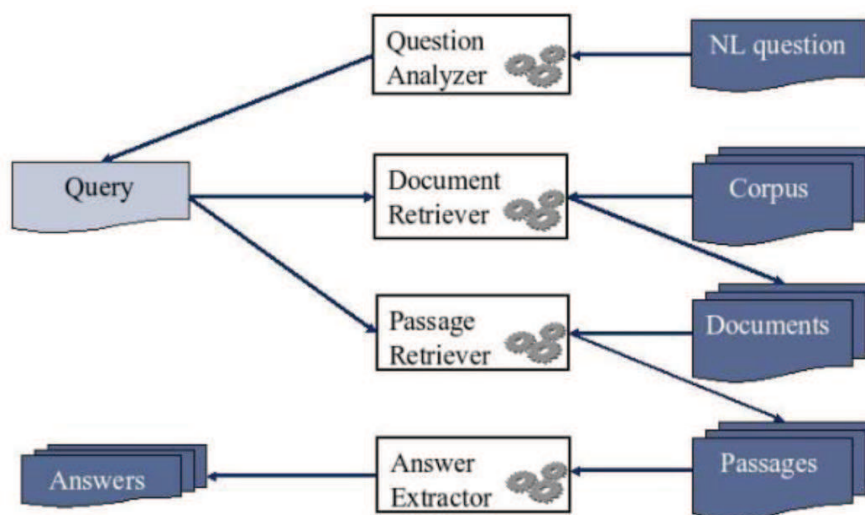
*Question Answering (QA)*, em português Pergunta e Resposta, é uma área de pesquisa que envolve outros campos, tais como recuperação da informação, extração da informação e processamento de linguagem natural (ALLAM; HAGGAG, 2012).

Sistema de pergunta e resposta tem como objetivo responder a uma pergunta formulada em Linguagem Natural, analisando diversas fontes de dados e documentos, e tentar obter a resposta mais exata para a questão (MISHRA; BANGALORE, 2010).

Os sistemas de pergunta e resposta podem ser classificados como *closed-domain* ou *open-domain* (LIN, 2002). Sistemas *closed-domain* têm o objetivo de encontrar respostas para questões sobre um determinado domínio. Estes sistemas podem fazer uso de ontologias sobre aquele assunto a fim de obter as respostas de forma precisa. Já os sistemas de pergunta e resposta *open-domain* tentam responder as questões sobre vários domínios. Esses sistemas podem usar diferentes abordagens para encontrar uma resposta satisfatória.

A arquitetura mais comum (Figura 3) encontrada entre sistemas de pergunta e resposta é constituída por três etapas fundamentais executadas por componentes distintos: a classificação da questão (*question processing*), a recuperação de informação (*information retrieval*) e o processamento e extração da resposta (*answer processing*).

**Figura 3 Arquitetura típica de sistemas de perguntas e respostas**



**Fonte:** Comparing Improved Language Models for Sentence Retrieval in Question Answering (MERKEL; KLAKOW, 2007)

O componente de classificação das questões é fundamental para o sistema de pergunta e resposta. Ele é responsável por encontrar o tópico da pergunta. Já o componente responsável pela recuperação de informação obtém os documentos ou registros das fontes de dados do sistema que se encaixam com o tópico

reconhecido. Por fim, a fase de extração de respostas (*answer extractor*) classifica as possíveis respostas e as validam (GUPTA; GUPTA, 2012).

## 2.6 Geração de Linguagem Natural (GLN)

A Geração de Linguagem Natural tem como principal tarefa converter dados não estruturados em linguagem humana (LINDBERG, 2013). A Geração de Linguagem Natural (GLN) é uma tarefa na área do Processamento de Linguagem Natural que utiliza dados em uso por sistemas de computação para transformar estes em textos com uma representação de linguagem natural. Esses dados podem ser originados em uma base de conhecimento ou base de dados abertos e conectados disponíveis na web. Dito de outra forma, a GLN é um subcampo da inteligência artificial e da linguística computacional que se preocupa com a construção de sistemas computacionais que podem produzir textos compreensíveis em diversas línguas, dentro das dificuldades mais encontradas é a existência de sistemas na língua portuguesa.

O sistema de GLN necessita de conhecimento sobre a linguagem e o domínio da aplicação para produzir automaticamente documentos e/ou relatórios compreensíveis e entendíveis por pessoas ou usuários não especializados em computação.

De acordo com REITER e DALE (1997) há seis tipos básicos de atividades que precisam ser executadas, indo de dados de entrada para um texto de saída final, e isso não significa que um Sistema de NLG precisa ter obrigatoriamente os seis módulos.

Estas seis atividades básicas são as seguintes:

- a) Determinação de conteúdo: é o processo de decidir quais as informações dever ser comunicadas no texto;
- b) Planejamento do discurso: é o processo de impor ordenação e estrutura sobre o conjunto a serem transmitidas;
- c) Agregação de sentenças: é o processo de agrupar as mensagens em frases;

d) Lexicalização: é o processo de decidir quais palavras e frases específicas devem ser escolhidas para expressar os conceitos e relações de domínio que aparecem nas mensagens;

e) Geração de expressões de referência: é a tarefa de selecionar palavras ou frases para identificar entidades de domínio;

f) Realização linguística: é o processo de aplicação das regras da gramática para produzir um texto sintático, morfológico e ortográfico correto.

### **3 TRABALHOS RELACIONADOS**

Neste tópico serão apresentados trabalhos relacionados com o tema proposto, que foram analisados para a identificação de tendências e lacunas. A abordagem usada para selecionar os trabalhos apresentados envolveu a pesquisa de referências em periódicos acadêmicos e anais de eventos buscando identificar avanços e experiências nesta área. Foram estudados trabalhos voltados para a área de Perguntas e Respostas, pelo interesse geral no seu tipo de aplicação e na identificação de suas necessidades. Também foram estudados trabalhos sobre geração de linguagem natural a partir de recursos de dados abertos e conectados.

A seguir são apresentados os principais aspectos relacionados de cada trabalho pesquisado e após uma análise comparativa com os elementos de contexto considerados mais relevantes para o estudo desenvolvido. Estes trabalhos foram apresentados de modo a destacar os dois agrupamentos principais observados durante o processo de pesquisa. Em primeiro lugar são apresentados os trabalhos descrevendo iniciativas relacionadas com o tema Perguntas e Respostas. Em segundo lugar são descritos trabalhos dedicados à Geração de Frases em Linguagem Natural.

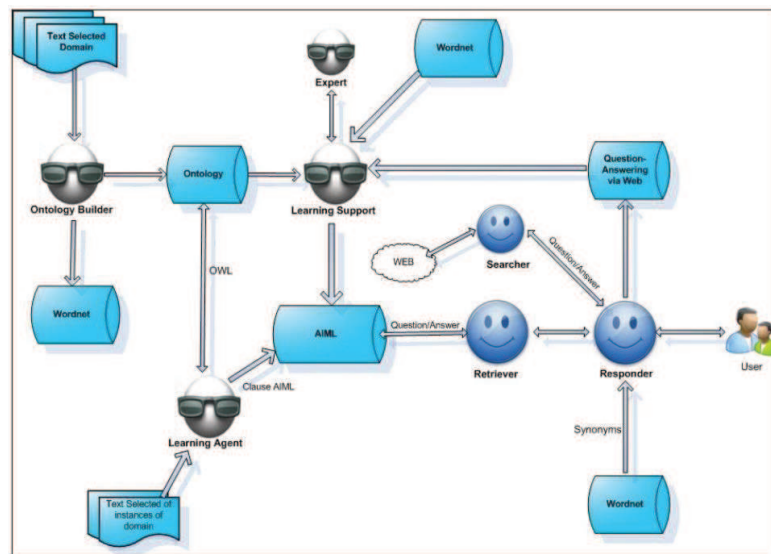
#### **3.1 Perguntas e Respostas**

Neste tópico são descritos artigos que trabalham com o uso de ontologias para responder perguntas e o uso do BDA para perguntas e repostas.

##### **3.1.1 Sobre a aplicação de ontologias para orientar agentes a responder perguntas**

Neste trabalho (MENEZES et al., 2014) apresenta um sistema de perguntas e respostas que utiliza ontologias e técnicas de recuperação de informação na análise da pergunta e dessa forma busca aperfeiçoar a extração de respostas além das possibilidades das buscas baseadas em palavras chaves.

**Figura 4 Arquitetura conceitual do sistema de perguntas-resposta**



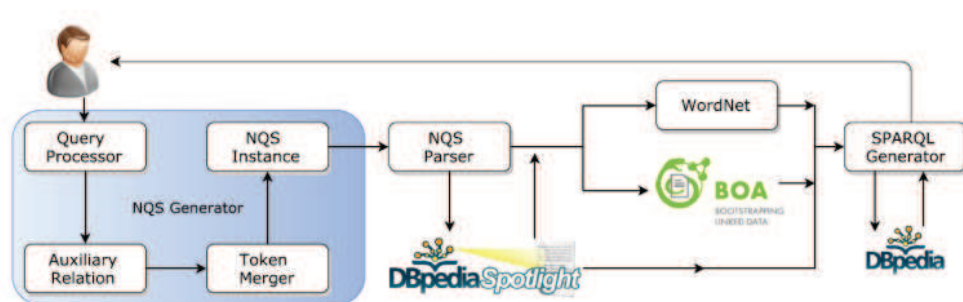
Fonte: MENEZES et al., 2014

A arquitetura proposta por (MENEZES et al., 2014), ilustrada na Figura 4, combina ontologias com a família de agentes de softwares usando agentes para flexibilizar a arquitetura, favorecendo assim o uso de mecanismos inteligentes específicos, sintonizados com as necessidades de cada ação do sistema.

### 3.1.2 AskNow: A Framework for Natural Language Query Formalization in SPARQL – Um Framework para formalização de consulta em Linguagem Natural no SPARQL

Neste trabalho (DUBEY et al., 2016) propõe a criação de um framework chamado AskNow, onde os usuários podem realizar consultas em inglês em uma base de conhecimento em RDF. Estas consultas são inicialmente normalizadas em uma forma sintática canônica intermediária, chamada de *Normalized Query Structure* (NQS) e traduzidas para consultas SPARQL.

**Figura 5 Arquitetura Pipeline do AskNow**



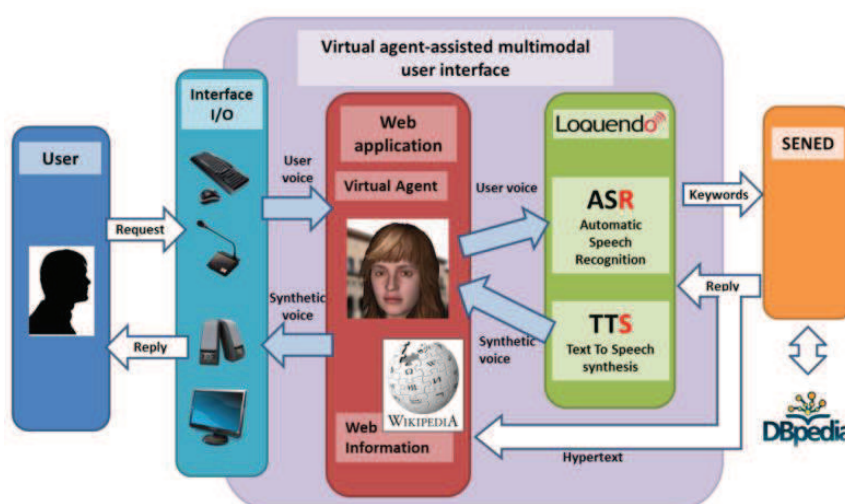
Fonte: DUBEY et al., 2016

Segundo DUBEY, o algoritmo de ajuste NQS normaliza as consultas parafraseadas em uma estrutura comum. Outro objetivo da abordagem é ajudar o tradutor SPARQL a identificar facilmente o desejo da consulta e a entrada da consulta, ou seja, as informações adicionais fornecidas pelo usuário, bem como a sua relação semântica mútua.

### 3.1.3 VOX System: A Semantic Embodied Conversational Agent exploiting Linked Data – Um agente conversacional incorporado explorando a semântica em Linked Data

Neste trabalho, SERÓN e BOBED (2016) apresentam o sistema VOX, uma plataforma de agente conversacional semântica que explora os dados semânticos disponíveis em qualquer nó de extremidade SPARQL público. Em particular, apresenta uma plataforma de agentes como uma estratégia híbrida de busca baseada em palavras-chaves que explora o conhecimento armazenado em ontologias para formar e enriquecer o processo de pesquisa em dados estruturados.

Figura 6 Arquitetura Geral do VOX



Fonte: SERÓN e BOBED, 2016.

A Figura 6 mostra a arquitetura geral do VOX que é composta de dois módulos principais:

- ✓ **Interface Multimodal:** Este módulo fornece uma interface multimodal baseada em um ECA. É responsável por capturar todos os inputs (detectar e

perceber os diferentes Canais de entrada), e processamento e encaminhando-os para os serviços subjacentes adequados. Além disso, este módulo é também responsável pelas reações do sistema e as animações dos agentes virtuais 3D, fornecendo a dimensão visual do agente.

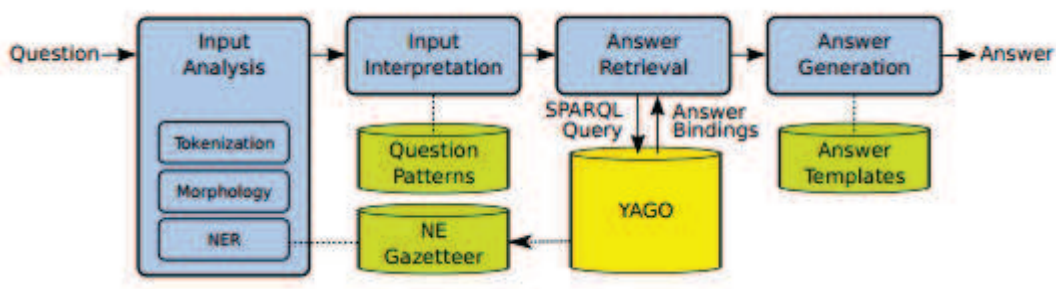
- ✓ **Módulo SENED:** Ele fornece acesso ao conhecimento semântico armazenado em repositórios de dados vinculados, que são anexados ao sistema, definindo seu domínio de pesquisa usando ontologias. Este módulo implementa técnicas para processar as consultas baseadas em palavras-chave que são recebidas a partir da interface multimodal durante a interação com o usuário.

### 3.1.4 YAGO-QA: Answering Questions by Structured Knowledge Queries

Neste trabalho Adolphs et al. (2011) apresentam um sistema de perguntas e respostas em linguagem natural para dar acesso ao conhecimento na Web e Wikipédia, através de uma base de conhecimento estrutura automaticamente. Neste sistema, o objetivo principal é estabelecer mapeamentos de expressões de linguagem natural para representações semânticas.

A linguagem SPARQL fornece recursos de consultas em bases de dados de conhecimento e fontes de dados conectadas, mas para os usuários sem conhecimento técnico a linguagem natural é muito mais simples. Este artigo propõe uma abordagem de perguntas e respostas sobre bases de conhecimentos com técnicas e métodos para mapear questões de linguagem natural em consultas SPARQL estruturadas.

**Figura 7 Arquitetura do YAGO-QA**



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

O YAGO-QA consiste em três componentes principais. O primeiro usa a base de conhecimento YAGO com uma interface de consulta SPARQL para fornecer o conteúdo para responder as perguntas formuladas usando uma tripla *RDF* extraída da Wikipédia. O segundo componente usa um repositório de padrões de textos e paráfrases de perguntas compiladas por duas abordagens complementares: processar automaticamente um *corpus* Web muito grande (ClueWeb) para reunir padrões que podem denotar relações em YAGO e coletar paráfrases para perguntas por *crowdsourcing* para humanos. O terceiro componente dedica-se ao processamento de perguntas reais e decompõe questões de linguagem natural, mapeando os constituintes em padrões do SPARQL.

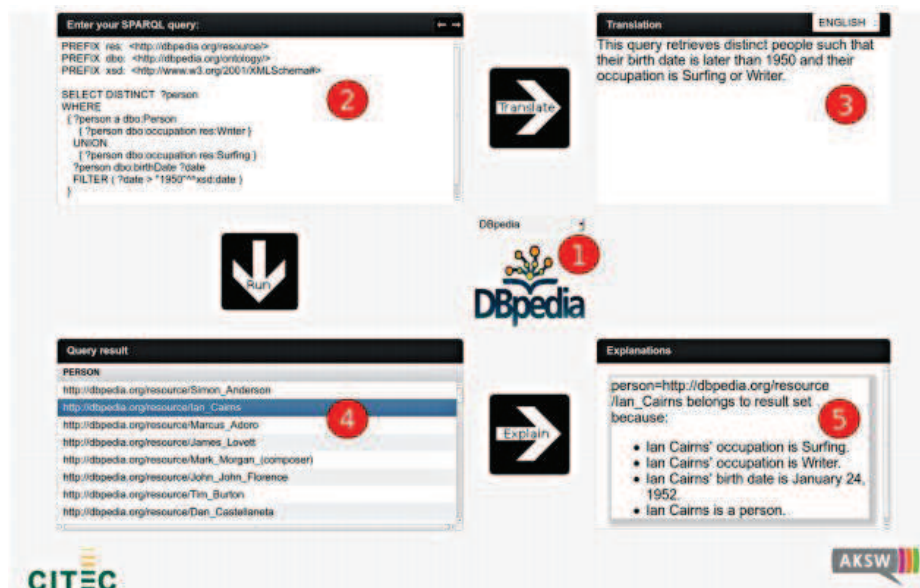
### 3.2 Geração de Frases

Neste tópico são descritos os trabalhos estudados com o tema de geração de frases em linguagem natural.

#### 3.2.1 SPARQL2NL – *Verbalizing SPARQL queries*

Neste trabalho, Nogmo et al. (2013) apresentam o SPARQL2NL um framework que permite verbalizar consultas em uma base de conhecimento como a DBpedia utilizando consultas em SPARQL e convertendo-as em linguagem natural. Para gerar verbalização, o SPARQL2NL possui em seu processamento básico quatro etapas principais: pré-processamento, processamento, pós-processamento e verbalização. O passo de pré-processamento normaliza a consulta e extrai informações de tipo para as variáveis que ocorrem; no passo de processamento durante uma representação genérica é gerada de uma consulta, no passo de pós-processamento são aplicadas regras de redução e substituição para melhorar a legibilidade da verbalização. Em uma etapa final de realização léxica toda a informação gerada nas etapas anteriores é integrada na representação final da linguagem natural (NOGMO et al. 2013).

Figura 8 Tela do Demo SPARQL2NL



Fonte: Nogmo et al. (2013)

Parte do protótipo do SPARQL2NL consiste na interface mostrada na Figura 8. O usuário começa selecionando o conjunto de dados contra o qual ele quer que a consulta SPARQL seja executada (item 1). No item 2, o usuário deve selecionar uma consulta SPARQL para a qual uma verbalização é necessária. A seleção do botão "Traduzir" executa a consulta a ser verbalizada. A verbalização da consulta é mostrada no item 3. Além de visualizar a verbalização da consulta, o usuário pode optar por executar a consulta clicando no botão "Executar". Isso leva os resultados da consulta a serem exibidos em formato tabular em 4. Após a seleção de um resultado e um clique no botão "Explicar", as instruções RDF utilizadas são recuperadas e exibidas no painel marcado como item 5.

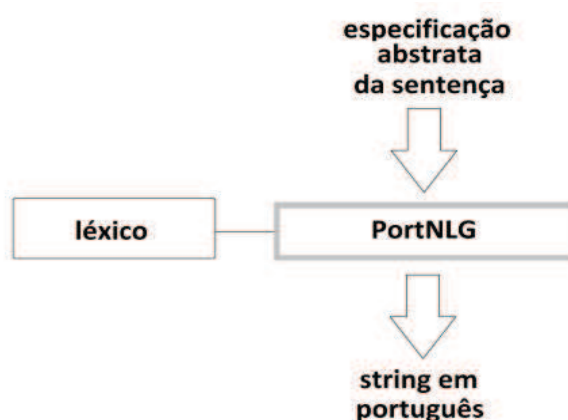
### 3.2.2 PortNLG - Um Sistema de Realização Superficial para Geração de Textos em Português

Neste trabalho Douglas et al. (2013) propõe a implementação de um módulo de realização textual baseado em regras para o português brasileiro, chamado *PortNLG*, que trata da tarefa de linearização sentencial para aplicações computacionais que necessitem apresentar dados de saída em formato textual. Segundo DOUGLAS, o *PortNLG* é disponibilizado em forma de uma biblioteca de métodos JAVA para construção de sentenças. Esta biblioteca foi projetada para ser

inserida no código da aplicação subjacente e, em ambos os sistemas, exige certo grau de conhecimento de programação e desenvolvimento orientado a objetos para sua utilização efetiva. A Figura 9 ilustra a organização geral do sistema *PortNLG*.

A entrada do sistema é uma especificação abstrata da sentença a ser gerada. Fazendo uso da base lexical implementada e regras próprias, o sistema constrói a cadeia correspondente em português, a qual constitui o resultado final do processo.

**Figura 9** Visão geral do Sistema *PortNLG*



**Fonte:** DOUGLAS et al. (2013)

Como forma de tornar o presente sistema mais facilmente adaptável a possíveis aplicações, foram implementadas duas modalidades de chamada ao sistema: uma modalidade em que a aplicação subjacente realiza uma série de chamadas incrementais aos métodos de construção da sentença, e uma modalidade de chamada única, na qual é fornecida uma árvore representando a sentença a ser gerada. As duas modalidades de chamada ao sistema fazem uso dos mesmos métodos básicos de construção da sentença, porém implementando funcionalidades ligeiramente distintas. A construção de uma frase mediante chamadas incrementais é a maneira tradicional de uso do sistema, na qual a aplicação possui total controle sobre a forma superficial produzida. Na construção da sentença mediante chamada única, por outro lado, não há nenhum esforço de programação. Neste caso, o sistema espera receber como entrada uma representação simbólica - em formato de árvore - da sentença a ser gerada.

### 3.2.3 RealText-lex: *A Lexicalization Framework for RDF Triples*

Neste trabalho Perera e Nand (2016), apresentam o framework RealText-lex que tem como objetivo construir uma estrutura de padrões para lexicalização de triplas de RDF extraídos da DBpedia. Neste trabalho considera-se um padrão de lexicalização como um padrão sintático que transformará um conjunto de triplas em uma frase ou parágrafo de linguagem natural sintaticamente correta.

O framework é baseado em quatro módulos de mineração de padrões que são executados na seguinte prioridade: Padrões de metonímias ocupacionais, padrões de gramática livre de contexto, padrões relacionais e padrões de propriedade (PERERA e NAND 2016).

**Figura 10 Representação do esquema do framework na lexicalização**



**Fonte:** Perera e Nand (2016).

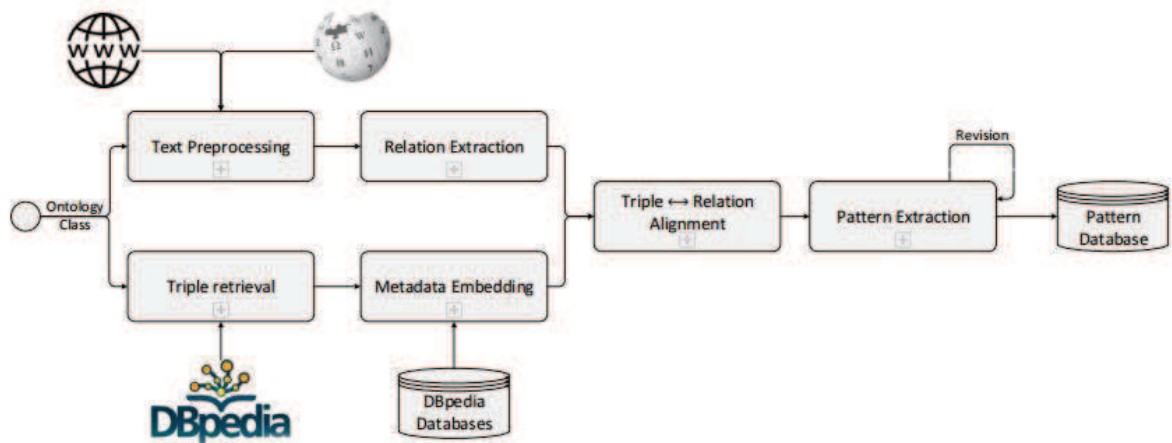
De acordo com a Figura 10, Perera e Nand (2016) definem que os quatros módulos são orientados para a geração de padrões de lexicalização. Embora eles sejam colocados em um pipeline, se um dos módulos encontrar um padrão de lexicalização para uma dada tripla, então os módulos restantes não serão executados.

Padrões de metonímias ocupacionais são compostos por uma única palavra ou frase não definida por seu próprio nome, mas por um nome associado ao seu significado (Kövecses e Radden, 1998) que são usadas para identificar uma pessoa baseada na ocupação. Portanto, pode-se conceber que a metonímia ocupacional seja uma representação semântica da pessoa, sendo que existem várias metonímias que são criadas com base nas ocupações. Algumas delas são "comandante", "proprietário" e "produtor", que são usadas, respectivamente, para denotar alguém que dá ordens a uma ou mais pessoas, alguém que possui alguma coisa e alguém que produz alguma coisa (PERERA e NAND, 2016).

O módulo de padrões de gramática livre de contexto é considerado de duplo propósito na PNL. Isso significa que ele pode ser usado para entender a linguagem,

bem como para gerar linguagem, com base em determinadas regras gramaticais (PERERA e NAND, 2016). O módulo de processamento de padrões relacionais concentra-se em derivar o padrão de lexicalização que são derivados do texto não estruturado usando a extração de relação. Em resumo, processa um grande número de recursos de texto não estruturados relacionados às relações triples e extraídas usando *Open Information Extraction* (OpenIE) (Etzioni et al., 2008). Estas relações são então alinhadas com os triplos para extrair padrões de lexicalização. A Figura 11 mostra a visão geral do processo de extração de padrões relacionais (PERERA e NAND, 2016).

**Figura 11 Visão geral do processo de extração de Padrões Relacionais**



**Fonte:** Perera e Nand (2016).

O módulo de padrões de propriedade especifica um léxico limitado onde certos predicados estão associados a uma lista pré-especificada de modelos como padrões de lexicalização. São especificados cinco padrões, que serão aplicados apenas à lista predeterminada de predicados. A Tabela 3 ilustra os 5 padrões com alguns exemplos de lexicalização quando aplicados a triplas com predicados predeterminados (PERERA e NAND, 2016).

**Tabela 3 Exemplo de padrões com triplas**

| Pattern | Predicate | Triple | Resulting lexicalization triple |
|---------|-----------|--------|---------------------------------|
|---------|-----------|--------|---------------------------------|

|                                     |               |                                                           |                                                                  |
|-------------------------------------|---------------|-----------------------------------------------------------|------------------------------------------------------------------|
| $\langle S? sP?, is, O? \rangle_L$  | height        | $\langle Michael\ Jordan, height, 1.98 \rangle_T$         | $\langle Michael\ Jordan's, height, is, 1.98 \rangle_{LT}$       |
| $\langle S?, has, O? P? \rangle_L$  | championships | $\langle Rubens\ Barrichello, championships, 0 \rangle_T$ | $\langle Rubens\ Barrichello, has, 0\ championship \rangle_{LT}$ |
| $\langle S?, is, O? \rangle_L$      | occupation    | $\langle Natalie\ Portman, occupation, actress \rangle_T$ | $\langle Natalie\ Portman, is, an\ actress \rangle_{LT}$         |
| $\langle P? inS?, is, O? \rangle_L$ | largesCity    | $\langle Australia, largesCity, Sydney \rangle_T$         | $\langle Largest\ city\ in\ Australia, is, Sydney \rangle_{LT}$  |
| $\langle S?, P?, O? \rangle_L$      | isPartOf      | $\langle Delfi, isPartOf, South\ Holland \rangle_T$       | $\langle Delfi, is, part\ of, South\ Holland \rangle_{LT}$       |

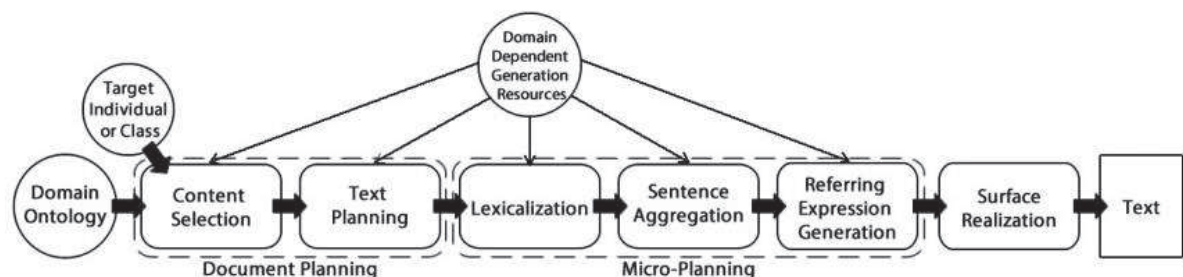
Fonte: Perera e Nand (2016).

### 3.2.4 NaturalOWL System - Generating Natural Language Descriptions from OWL Ontologies

Neste trabalho Androutsopoulos et al. (2013) apresentam um sistema de geração de linguagem natural que produz textos que descrevem indivíduos ou classes de ontologias OWL. No sistema, podem-se publicar informações em OWL na Web, juntamente com textos fluentes e coerentes, produzidos automaticamente.

O NaturalOWL define o seu processamento em etapas e recursos com base em uma arquitetura de pipeline. O sistema gera textos em três etapas, planejamento de documentos, micro-planejamento e realização de superfície, conforme o processo ilustrado na Figura 12 ANDROUTSOPOULOS et al. (2013).

Figura 12 Estágio dos processamentos e subtarefas do NaturalOWL



Fonte: Androutsopoulos et al. (2013).

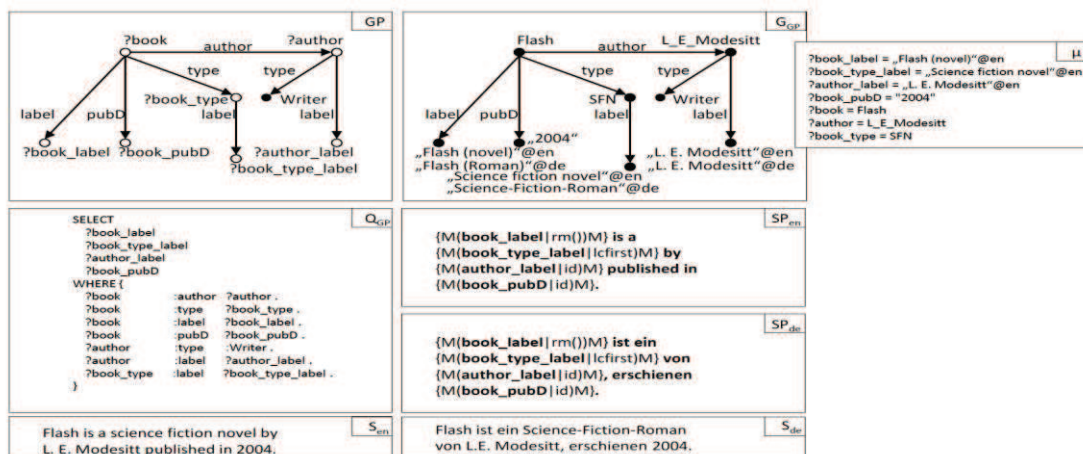
No planejamento de documentos o NaturalOWL faz a seleção dos conteúdos e o sistema seleciona as informações a serem transmitidas junto com o planejamento do texto com a estrutura a ser gerada. Dentro do planejamento, a seleção de conteúdo primeiro recupera da ontologia toda a informação OWL que será relevante para a classe ou indivíduo a ser descrito e converte as instruções OWL em triplas de mensagens e frases fáceis de serem expressas. Ainda na etapa de planejamento, o planejamento de texto para cada alvo utiliza mecanismos que produzem as triplas de mensagens a serem expressas e cada tripla é destinada a ser facilmente expressível como uma única sentença (ANDROUTSOPoulos et al. 2013).

A etapa de Micro Planejamento ele é dividido em três sub-etapas: i) lexicalização: para cada propriedade da ontologia e cada linguagem natural suportada, podem ser especificadas uma ou mais sentenças do modelo para indicar as triplas de mensagens envolvendo essa propriedade. ii) agregação de sentenças: nos planos das sentenças anteriores, considera-se uma sentença separada para cada tripla de mensagem. Essa agregação de sentenças é definida por um conjunto de regras criadas manualmente, destinadas a serem independentes do domínio. iii) geração de expressões referencias: um plano de sentença pode exigir uma expressão de referência a ser gerada a partir de uma mensagem tripla.

### 3.2.5 *A language-independent method for the extraction of RDF verbalization templates*

O trabalho de Ell e Harth (2016) atua na utilização de recursos de dados abertos e conectados para a geração de frases em linguagem natural, tendo como diferencial o fato de utilizar *templates* com aspectos que relacionam tanto os elementos dos dados conectados, considerados neste trabalho como grafos, além de elementos de lexicalização.

Figura 13 Exemplo de uso do método para a língua inglesa e alemã



Fonte: Ell e Harth, 2016.

A Figura 13 acima ilustra um exemplo de uso da abordagem proposta neste trabalho, segundo a qual um grafo (representado como GGP na figura) contendo as informações a serem utilizadas pode ser consultado com padrões como o SPARQL (consulta representada como QGP na figura). Os resultados das consultas serão utilizados juntamente com os padrões existentes (elementos SP<sub>en</sub> e SP<sub>de</sub> na figura).

### 3.2.6 Generating Natural Language from Linked Data: Unsupervised template extraction

O trabalho de Dumas e Klein (2013) atua em uma área bastante necessária para que os sistemas de geração de linguagem natural sejam amplamente utilizados e adotados. Neste caso são exploradas técnicas para que os padrões de lexicalização a serem utilizados sejam gerados automaticamente, a partir de análise de textos disponíveis.

Desta forma, o trabalho propõe uma arquitetura para geração de linguagem natural a partir de dados vinculados que automaticamente aprende modelos de sentenças e planejamento de documentos estatísticos a partir de conjuntos de dados RDF e texto paralelos. O trabalho também atuou na construção de uma metodologia para um sistema de prova de conceito treinado em texto não-anotado. Foram utilizados no trabalho a Wikipédia em inglês e a DBpedia, focando-se exclusivamente em dados factuais e não temporais. O objetivo do sistema é gerar

descrições curtas, equivalentes a entradas da Wikipédia, de entidades encontradas em conjuntos de dados vinculados.

### 3.3 Análise de trabalhos estudados

Após estudos e leituras nos resultados da pesquisa, verificou-se que existe uma lacuna quando se considera sistemas para geração de perguntas e respostas automáticas no que se refere ao estudo das relações definidas em bases de dados abertas e conectadas, de modo a facilitar a utilização destas relações como parte determinante dos algoritmos de geração das perguntas em linguagem natural. A seguir são descritos resumos dos aspectos estudados.

#### 3.3.1 Sobre Perguntas e Respostas

O trabalho de MENEZES et al. 2014, é um sistema de perguntas e respostas que utiliza ontologias e técnicas de recuperação de informações na análise da pergunta, além de usar palavras chaves com o uso da combinação de ontologias com agentes de softwares.

O AskNow, DUBEY et al. 2016, é um framework através do qual o usuário pode colocar consultas em Inglês em uma base de conhecimento RDF e traduzir estas para consultas SPARQL.

O VOX System, SERÓN e BOBED, 2016, é uma plataforma ECAs semântica que explora os dados semânticos disponíveis em qualquer nó de extremidade SPARQL público. O sistema usa *agentes* como estratégia híbrida de busca baseada em palavras-chaves que explora o conhecimento armazenado em ontologias.

O YAGO-QA, ADOLPHS et al. 2011, é um sistema de perguntas e respostas em linguagem natural para favorecer o acesso ao conhecimento da Web e Wikipédia. Neste sistema, a chave principal é estabelecer mapeamentos de expressões de linguagem natural para representação semântica.

**Tabela 4 Comparação entre trabalhos sobre Perguntas e Respostas**

| Artigos                | Ontologias | BDA | Dados Abertos | Dados Conectados |
|------------------------|------------|-----|---------------|------------------|
| (ADOLPHS et al., 2011) | NÃO        | NÃO | SIM           | SIM              |
| (MENEZES et al., 2014) | SIM        | SIM | NÃO           | NÃO              |
| (DUBEY et al., 2016)   | NÃO        | SIM | SIM           | SIM              |
| (SÉRON e BOBEB, 2016)  | SIM        | SIM | SIM           | SIM              |

**Fonte:** Autor

A tabela 4 resume e organiza visualmente os atributos analisados nos sistemas de pergunta e resposta utilizados neste estudo.

### 3.3.2 Sobre Geração de Frases

SPARQL2NL (NOGMO et al. 2013) é um framework que permite verbalizar consultas utilizando o SPARQL em uma base de conhecimento como a DBpedia e convertendo-as em linguagem natural. O framework normaliza a consulta e extrai informações, processando a representação genérica gerada através de uma consulta, aplicando regras de redução e substituição para melhorar a legitimidade da verbalização e transformando toda a informação gerada em representação final da linguagem natural.

PortNLG (DOUGLAS et al. 2013) é um sistema de realização superficial para geração de textos em português, baseado em regras para o português brasileiro, que trata da tarefa de linearização sentencial para aplicações computacionais que necessitem apresentar dados de saída em formato textual.

O RealText-lex (PERERA e NAND 2016) é um framework que tem como objetivo construir uma estrutura de padrões para lexicalização de *triplas* de RDF extraídos da DBpedia. O framework tem quatro módulos orientados para a geração de padrões de lexicalização.

NaturalOWL (ANDROUTSOPOULOS et al. 2013) é um sistema de geração de linguagem natural que produz textos que descrevem os indivíduos ou classes de ontologias OWL. No sistema, podem-se publicar informações em OWL na Web, juntamente com textos produzidos automaticamente correspondentes em vários idiomas.

No trabalho de Ell e Harth (2016) o sistema atua na utilização de recursos de dados abertos e conectados para geração de frases em linguagem natural utilizando templates com aspectos que relacionam os elementos de dados conectados além de elementos de lexicalização.

No trabalho de Dumas e Klein (2013), o sistema propõe uma arquitetura de geração de linguagem natural a partir de dados vinculados que automaticamente aprende com modelos de sentenças e planejamento de documentos estatísticos a partir de conjuntos RDF e textos paralelos. O objetivo do sistema é gerar descrições curtas, equivalentes a entradas da Wikipédia, de entidades encontradas em conjuntos de dados vinculados.

**Tabela 5 Comparação entre trabalhos sobre geração de linguagem**

| <b>Artigos</b>                | <b>SPARQL</b> | <b>Lexicalização</b> | <b>Dados Abertos</b> | <b>Dados Conectados</b> |
|-------------------------------|---------------|----------------------|----------------------|-------------------------|
| (NOGMO et al., 2013)          | SIM           | NÃO                  | SIM                  | SIM                     |
| (DOUGLAS et al., 2013)        | NÃO           | SIM                  | NÃO                  | NÃO                     |
| (ADROUTSOPOULOS et al., 2013) | NÃO           | SIM                  | NÃO                  | NÃO                     |
| DUMAS e KLEIN, 2013)          | NÃO           | NÃO                  | SIM                  | SIM                     |
| (PERERA e NAND, 2016)         | NÃO           | SIM                  | SIM                  | SIM                     |
| ELL e HARTH, 2016)            | SIM           | NÃO                  | SIM                  | SIM                     |

**Fonte:** Autor

Na tabela 5 estão agrupados os atributos analisados nos sistemas de geração de linguagem que foram alvo deste estudo. As principais considerações desta análise estão apresentadas na seção seguinte.

### 3.3.3 Pontos de atenção observados

Por meio do estudo dos trabalhos relacionados apresentados anteriormente, realizou-se um processo de identificação dos principais desafios da área e também análise da experimentação de soluções constatadas. Foram identificadas, nestes trabalhos, as principais técnicas atuais para geração de frases, sendo que neste

aspecto pode-se observar um conjunto de abordagens que vai desde o uso de padrões textuais até recursos linguísticos e aprendizagem de máquina.

Ainda não são observados trabalhos consolidados com base na existência de um estudo aprofundado sobre os dados abertos conectados para identificar relacionamentos que possam ser de utilidade para a geração de frases curtas automáticas. Sendo assim, de acordo com os objetivos deste trabalho, estas lacunas serão abordadas a partir do desenvolvimento de um estudo futuro sobre padrões de dados abertos e conectados e características que podem ser de interesse para a geração de frases contendo perguntas em linguagem natural.

## 4 MATERIAIS E MÉTODO

Neste capítulo são apresentadas as etapas desenvolvidas neste trabalho, a partir do estudo realizado nas áreas de conhecimento relacionadas, nos trabalhos pesquisados e na exploração de possibilidades de implementação e avaliação.

Uma das etapas desenvolvidas está destinada à análise detalhada das bases de dados abertas conectadas no que se refere aos elementos que podem ser usados na geração de linguagem natural. Esta etapa descreve exemplos dos dados encontrados e formas de sua utilização. Complementando esta etapa são apresentados estudos e considerações sobre a estrutura de relações que podem ser encontradas em bases de dados abertas e conectadas de modo a apoiar a geração de frases em linguagem natural.

Outra etapa está relacionada com a descrição de elementos de um modelo para apoiar a prototipação de uma ferramenta que permita realizar a geração automática de frases a partir das bases de dados abertas e conectadas. Este modelo incorpora as principais necessidades observadas em trabalhos relacionados e inclui especificamente aspectos de contribuição originárias do trabalho aqui descrito.

Por fim, foram analisadas abordagens para a avaliação de resultados e realizado um experimento que consistiu na implementação de aspectos do modelo para a geração de frases com perguntas, tomando como base um conjunto de dados da BDPedia. Estes resultados foram analisados por um especialista em linguística, procedimento que permitiu verificar aspectos positivos e necessidades de melhoria.

### 4.1 Visão geral

Este trabalho aborda a geração de perguntas em linguagem natural, tendo como ponto de apoio os dados disponibilizados em bases de dados abertos conectados. Nesta seção serão analisadas as relações disponíveis nestas bases de dados e indicadas as suas possibilidades de uso na geração de linguagem natural. Como é uma das bases mais conhecidas e uma das mais numerosas em termos de elementos, tomou-se como ponto inicial de observação a DBPedia.

O potencial para geração de perguntas que a DBpedia possui pode ser avaliado a partir do estudo de seu contexto. O projeto da DBPedia possui como



rótulos (*labels*) usados para descrever textualmente aspectos do conceito representado, neste caso em duas línguas, inglês e italiano (*rdfs:label "Berlin"@en; "Berlino"@it;*). Um terceiro exemplo são os conceitos que relacionam Berlin (conceito geral da página) com os conceitos de cidade e outros conceitos complementares a este, como o conceito de local e de local muito populoso (*dbpedia:Berlin rdfs:type dbpedia-owl:City , dbpedia:Berlin dbpedia-owl:PopulatedPlace, dbpedia:Berlin dbpedia-owl:Place*).

A partir destes dados anotados nas bases de dados abertas e conectadas como o exemplo da DbPedia, são vislumbradas possibilidades de geração automática ou semiautomática de frases curtas com perguntas em linguagem natural. As relações e sua semântica serão empregadas como elemento para auxiliar e guiar o planejamento da geração das frases.

Para destacar com maiores detalhes os aspectos gerais com os quais este trabalho está relacionado, a seguir é descrito um exemplo de um possível processo de extração automática dos dados da DBpedia para geração de frases em linguagem natural.

As etapas deste processo são sumarizadas da seguinte forma:

- Inicialmente são apresentados exemplos de páginas da Wikipedia (na figura 15) e da Dbpedia (na figura 16) descrevendo o conjunto de informações disponíveis sobre um mesmo assunto. Neste caso, o assunto é a cidade de Porto Alegre e a primeira página descreve informações não estruturadas enquanto que a segunda página apresenta as mesmas informações com base em uma formalização definida como relações com base na notação RDF.
- A semântica destas relações anotadas em RDF é disponibilizada para o uso público, logo este fato permite que sejam utilizadas de modo complementar à diversas aplicações de integração de dados, tal como, no nosso caso, para a geração de frases em linguagem natural. A figura 17 descreve um exemplo parcial desta semântica, para o conceito de cidade, na qual são exemplificados os tipos de dados que podem ser obtidos em consultas à conceitos e suas relações.
- A etapa seguinte é dedicada à análise do processo de consulta dos dados na DBPedia, com uso da linguagem SPARQL. A figura 18 ilustra

um exemplo de conjunto de dados obtido com este processo de consultas.

- A etapa final consiste na utilização dos resultados obtidos, em conjunto com a semântica das relações disponíveis, para a finalidade de geração de frases em linguagem natural. O quadro 1 ilustra um exemplo desta possibilidade.

A seguir são comentadas estas etapas e ao final são indicados alguns desafios para este processo, bem como os encaminhamentos gerais adotados neste trabalho.

A figura 15, a seguir, descreve um exemplo de página da wikipedia com textos em linguagem natural. Neste caso, o tema da página é uma cidade, a cidade de Porto Alegre. Pode ser verificada na figura a disponibilidade de informações sobre a cidade em termos de material gráfico (imagens) e também em termos de material textual. Apesar do material textual estar associado com hyperlinks, ele não é um material anotado semanticamente e desta forma não é um material que possa ser utilizado diretamente por outros sistemas em atividades de composição destas informações.

**Figura 15 Dados da Wikipedia para a cidade de Porto Alegre**

## Porto Alegre

Origem: Wikipédia, a enciclopédia livre.

**Nota:** Para outros significados, veja *Porto Alegre (desambiguação)*.

**Porto Alegre** é um município brasileiro e a capital do estado mais meridional do Brasil, o Rio Grande do Sul.<sup>10</sup> Pertence à *mesorregião metropolitana de Porto Alegre* e à *microrregião de Porto Alegre*.<sup>1</sup> Com uma área de quase 500 km², possui uma *geografia* diversificada, com morros, baixadas e um grande lago, o Guaíba. Dista 2 027 quilômetros de Brasília, a capital nacional.<sup>2 10 11</sup>

A cidade constituiu-se a partir da chegada de casais *açorianos* em meados do século XVIII. No século XIX contou com o influxo de muitos *imigrantes alemães* e *italianos*, recebendo também *espanhóis*, *africanos*, *poloneses* e *libaneses*. Desenvolveu-se com rapidez e hoje abriga mais de 1,4 milhão de habitantes.<sup>10</sup> A cidade enfrenta muitos desafios, entre eles a grande população ainda vivendo em condições de *pobreza* e *sub-habitação*,<sup>12</sup> alto *custo de vida*,<sup>13</sup> alta incidência de *obesidade* e *tabagismo*,<sup>14 15</sup> deficiências sérias no tratamento de *esgotos*,<sup>16</sup> muita *poluição*<sup>17</sup> e degradação de *ecossistemas* originais,<sup>18</sup> índices de *crime* elevados<sup>19 20</sup> e crescentes problemas de *trânsito*.<sup>21</sup>

Por outro lado, ostenta mais de 80 prêmios e títulos que a distinguem como uma das melhores capitais brasileiras para morar, trabalhar, fazer negócios, estudar e se divertir. Foi destacada em anos recentes também pela *ONU* como a *Metrópole nº1 em qualidade de vida* do Brasil por três vezes; como possuindo um dos 40 melhores modelos de gestão pública democrática pelo seu *Orçamento Participativo* e por ter o melhor *Índice de Desenvolvimento Humano* (IDH) entre as metrópoles nacionais.<sup>22</sup> Dados do IBGE a apontaram em 2009 como a capital brasileira com a menor taxa de *desemprego*,<sup>23</sup> a empresa de consultoria britânica *Jones Lang LaSalle* a incluiu em 2004 entre as 24 cidades com maior potencial para atrair investimentos no mundo<sup>24</sup> e figura na lista da *Pricewaterhouse Coopers* entre as cem cidades mais ricas do mundo.<sup>25</sup> Porto Alegre é uma cidade influente no cenário global, recebendo a classificação de *cidade global "nema -"* por parte do *Globalization and World Cities Study Group & Network* (GaWC).<sup>26</sup>

### Município de Porto Alegre

"Capital dos Gaúchos"  
"PoA"

Da esquerda para a direita: Usina do Gasômetro, Estátua do Lapaador, Cais Mauá, Mercado Público e Monumento aos Açorianos diante do prédio do Centro Administrativo do Estado

**Fonte:** Wikipedia

Um resultado do processo de análise e representação formal destes dados textuais disponíveis na figura 15 podem ser observados na página equivalente

disponível na DBPedia. O resultado da exportação dos dados para a DBPedia pode ser observado na figura 16, que inclusive pode ser consultado no seguinte endereço: [http://dbpedia.org/page/Porto\\_Alegre](http://dbpedia.org/page/Porto_Alegre). Nesta figura podem ser observadas diversas relações específicas e delimitadas, tais como área, país de localização, data de fundação, entre diversas outras.

Figura 16 Dados da DBpedia para Porto Alegre

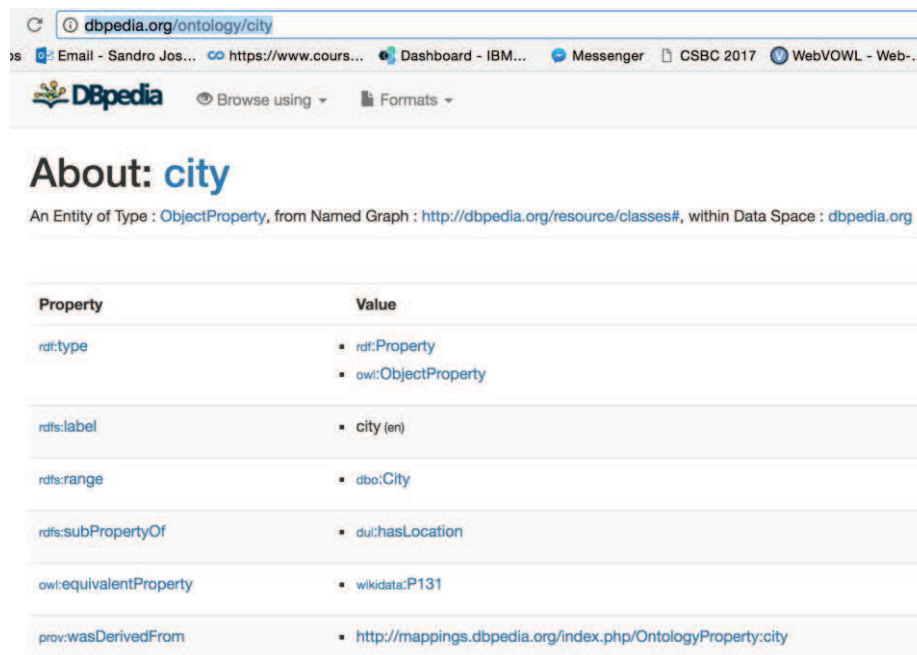
| About: Porto Alegre                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| An Entity of Type : <a href="#">municipality</a> , from Named Graph : <a href="http://dbpedia.org">http://dbpedia.org</a> , within Data Space : <a href="#">dbpedia.org</a>                                                                                                                                                                                                                                                                                                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <p>Porto Alegre (local Portuguese pronunciation: [ˈpɔɾtu aˈlɛɡɾi] - Happy Harbour) is the capital and largest city in the Brazilian state of Rio Grande do Sul. Founded in 1769 by Manuel Sepúlveda, its population of 1,509,939 inhabitants (2010) makes it the tenth most populous city in the country and the centre of Brazil's fourth largest metropolitan area, with 4,405,769 inhabitants (2010). The city is the southernmost capital city of a Brazilian state.</p> |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| Property                                                                                                                                                                                                                                                                                                                                                                                                                                                                     | Value                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| <a href="#">dbpedia-owl:PopulatedPlace/area</a>                                                                                                                                                                                                                                                                                                                                                                                                                              | <ul style="list-style-type: none"> <li>1.0E-6</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <a href="#">dbpedia-owl:PopulatedPlace/areaTotal</a>                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>496.627</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <a href="#">dbpedia-owl:PopulatedPlace/populationDensity</a>                                                                                                                                                                                                                                                                                                                                                                                                                 | <ul style="list-style-type: none"> <li>394.0</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:abstract</a>                                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>Porto Alegre (local Portuguese pronunciation: [ˈpɔɾtu aˈlɛɡɾi] - Happy Harbour) is the capital and largest city in the Brazilian state of Rio Grande do Sul. Founded in 1769 by Manuel Sepúlveda, its population of 1,509,939 inhabitants (2010) makes it the tenth most populous city in the country and the centre of Brazil's fourth largest metropolitan area, with 4,405,769 inhabitants (2010). The city is the southernmost capital city of a Brazilian state. Porto Alegre is one of the top cultural, political and economic centers of Brazil. Porto Alegre was founded in 1769 by Manuel Jorge Gomes de Sepúlveda, who used the fictitious name or pseudonym José Marcelino de Figueiredo (fictitious name), to hide their identity, but the official date is 1772 with the act signed by immigrants from the Azores, Portugal. In the late 19th century the city received many immigrants from other parts of the world, particularly from Germany, Italy, and Poland. The vast majority of the population is of European descent. The city lies on the eastern bank of the Rio Guaíba (Guaíba Lake), where five rivers converge to form the Lagoa dos Patos (Lagoon of the Ducks), a giant freshwater lagoon navigable by even the largest of ships. This five-river junction has become an important alluvial port as well as a chief industrial and commercial center of Brazil. The port of Porto Alegre is important for transporting local produce. The "capital gaúcha" has a broad-based economy that lays particular emphasis on agriculture and industry. Agricultural production includes produce such as plums, peaches, rice and cassava grown on rural smallholdings. The shoe and leather industries are also important, especially in Novo Hamburgo, in the Metropolitan Region of Porto Alegre. In recent years, Porto Alegre hosted the World Social Forum, an initiative of several non-government organizations. The city became famous for being the first city that implemented participatory budgeting. The 9th Assembly of the World Council of Churches was held in Porto Alegre in 2006. Since 2000, Porto Alegre also hosts one of the world's largest free software events, called FISL. The city will be one of the host cities of the 2014 FIFA World Cup. Additionally, Porto Alegre hosted the 1950 FIFA World Cup.</li> <li>Porto Alegre é um município brasileiro e a capital do estado mais meridional do Brasil, o Rio Grande do Sul. Pertence à mesorregião metropolitana de Porto Alegre e à microrregião de Porto Alegre. Com uma área de quase 500 km², possui uma geografia diversificada, com morros, baixadas e um grande</li> </ul> |
| <a href="#">dbpedia-owl:area</a>                                                                                                                                                                                                                                                                                                                                                                                                                                             | <ul style="list-style-type: none"> <li>1.000000 (xsd:double)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:areaCode</a>                                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li>+55 51</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <a href="#">dbpedia-owl:areaTotal</a>                                                                                                                                                                                                                                                                                                                                                                                                                                        | <ul style="list-style-type: none"> <li>496827000.000000 (xsd:double)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <a href="#">dbpedia-owl:country</a>                                                                                                                                                                                                                                                                                                                                                                                                                                          | <ul style="list-style-type: none"> <li><a href="#">dbpedia:Brazil</a></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| <a href="#">dbpedia-owl:elevation</a>                                                                                                                                                                                                                                                                                                                                                                                                                                        | <ul style="list-style-type: none"> <li>10.000000 (xsd:double)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <a href="#">dbpedia-owl:foundingDate</a>                                                                                                                                                                                                                                                                                                                                                                                                                                     | <ul style="list-style-type: none"> <li>1772-03-26 (xsd:date)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:isPartOf</a>                                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li><a href="#">dbpedia:Porto_Alegre_Metro</a></li> <li><a href="#">dbpedia:Rio_Grande_do_Sul</a></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <a href="#">dbpedia-owl:leaderName</a>                                                                                                                                                                                                                                                                                                                                                                                                                                       | <ul style="list-style-type: none"> <li><a href="#">dbpedia:Democratic_Labour_Party_(Brazil)</a></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| <a href="#">dbpedia-owl:leaderTitle</a>                                                                                                                                                                                                                                                                                                                                                                                                                                      | <ul style="list-style-type: none"> <li>Mayor</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:motto</a>                                                                                                                                                                                                                                                                                                                                                                                                                                            | <ul style="list-style-type: none"> <li>Loyal and Valiant city of Porto Alegre</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
| <a href="#">dbpedia-owl:populationDensity</a>                                                                                                                                                                                                                                                                                                                                                                                                                                | <ul style="list-style-type: none"> <li>394.000000 (xsd:double)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| <a href="#">dbpedia-owl:populationMetro</a>                                                                                                                                                                                                                                                                                                                                                                                                                                  | <ul style="list-style-type: none"> <li>4405769 (xsd:integer)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:populationTotal</a>                                                                                                                                                                                                                                                                                                                                                                                                                                  | <ul style="list-style-type: none"> <li>1509939 (xsd:integer)</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:postalCode</a>                                                                                                                                                                                                                                                                                                                                                                                                                                       | <ul style="list-style-type: none"> <li>90000-000</li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
| <a href="#">dbpedia-owl:thumbnail</a>                                                                                                                                                                                                                                                                                                                                                                                                                                        | <ul style="list-style-type: none"> <li><a href="http://commons.wikimedia.org/wiki/Special:FilePath/Montagem_de_porto_alegre_2.jpg?width=300">http://commons.wikimedia.org/wiki/Special:FilePath/Montagem_de_porto_alegre_2.jpg?width=300</a></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |
| <a href="#">dbpedia-owl:timeZone</a>                                                                                                                                                                                                                                                                                                                                                                                                                                         | <ul style="list-style-type: none"> <li><a href="#">dbpedia:Coordinated_Universal_Time</a></li> </ul>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |

Fonte: Dbpedia

Usando o exemplo do conceito cidade, pode-se observar na Figura 17 abaixo, uma lista preliminar de relações disponíveis para este conceito, existentes na ontologia da DBPEDIA para um conceito tal como “cidade” (“city”, em inglês, como na figura). A partir do exemplo da Figura 17, uma análise preliminar permite identificar aspectos detalhados das relações formalizadas nestes conjuntos de dados. Esta análise foi realizada neste trabalho para buscar subsídios para as atividades de gerações de frases em linguagem natural. Para fins de um exemplo simples, pode-se identificar o elemento “*label*”, que é definido pela padronização “*RDF-Schema*” e que permite identificar um termo descrito para a identificação ou a menção do conceito em língua natural. Este tipo de identificação sobre os recursos e

sua semântica possibilita que sejam utilizadas estas conclusões para a geração de regras e de heurísticas para a produção automática de frases em linguagem natural.

**Figura 17 Relações da DBpedia para o conceito cidade**



| Property                            | Value                                                                                                                      |
|-------------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <code>rdfs:type</code>              | <ul style="list-style-type: none"> <li><code>rdfs:Property</code></li> <li><code>owl:ObjectProperty</code></li> </ul>      |
| <code>rdfs:label</code>             | <ul style="list-style-type: none"> <li><code>city (en)</code></li> </ul>                                                   |
| <code>rdfs:range</code>             | <ul style="list-style-type: none"> <li><code>dbo:City</code></li> </ul>                                                    |
| <code>rdfs:subPropertyOf</code>     | <ul style="list-style-type: none"> <li><code>dui:hasLocation</code></li> </ul>                                             |
| <code>owl:equivalentProperty</code> | <ul style="list-style-type: none"> <li><code>wikidata:P131</code></li> </ul>                                               |
| <code>prov:wasDerivedFrom</code>    | <ul style="list-style-type: none"> <li><code>http://mappings.dbpedia.org/index.php/OntologyProperty:city</code></li> </ul> |

**Fonte:** dbpedia

No caso do exemplo da Figura 18, abaixo, pode-se identificar o elemento *label* associado com os termos “cidade”, em português, bem como o mesmo termo em diversas outras línguas. O trecho em destaque (com negrito) ilustra esta situação. Neste caso, a formalização utilizada permite tanto identificar termos que podem ser utilizados como base para a geração de linguagem natural, bem como possibilita a descrição de informações adicionais. Neste exemplo, como citado, o formalismo descreve também a grafia do termo em diversas outras línguas, identificadas com os sufixos específicos (“en” para inglês, “fr” para francês e assim por diante). Os dados da figura 18 são obtidos com consultas SPARQL a partir de DBpedia. Isso pode ser realizado a partir de locais de consulta com formatos SPARQL, tais como o seguinte endereço (<https://dbpedia.org/sparql>), a partir dos quais é possível o acesso e a realização de consultas nesta linguagem.

A seguir estão relacionados alguns exemplos de resultados obtidos nestas consultas SPARQL, tendo como foco o conceito “Porto Alegre” e duas relações

associadas com este conceito na ontologia utilizada na DBpedia (*populationTotal* e *country*). Para o conceito “Dbpedia: Porto Alegre” podem ser obtidos exemplos de relações e valores como: “Dbpedia-owl: populationTotal 15099399” ou “Dbpedia-owl: country dbpedia: Brazil”. Em uma interpretação livre das relações indicadas, seria possível considerar a geração de exemplos de perguntas em linguagem natural que podem ser geradas, tais como “Qual é a população total de Porto Alegre?” para a primeira relação, ou então “Em que país fica Porto Alegre?” para a segunda relação.

**Figura 18 Trecho em Jason para descrição parcial de resultado de consulta**

```
{
  "http://dbpedia.org/ontology/city" :
  [ { "type" : "uri", "value" : "http://www.w3.org/2002/07/owl#ObjectProperty" } ,
    { "type" : "uri", "value" : "http://www.w3.org/1999/02/22-rdf-syntax-ns#Property" }
  ] ,
  "http://www.w3.org/2000/01/rdf-schema#subPropertyOf" :
  [ { "type" : "uri", "value" :
    "http://www.ontologydesignpatterns.org/ont/dul/DUL.owl#hasLocation" }
  ] ,
  "http://www.w3.org/2002/07/owl#equivalentProperty" :
  [ { "type" : "uri", "value" : "http://www.wikidata.org/entity/P131" }
  ] ,
  "http://www.w3.org/2000/01/rdf-schema#label" :
  [ { "type" : "literal", "value" : "\u03C0\u03CC\u03BB\u03B7" , "lang" : "el" } ,
    { "type" : "literal", "value" : "ville" , "lang" : "fr" } ,
    { "type" : "literal", "value" : "coathair" , "lang" : "ga" } ,
    { "type" : "literal", "value" : "Stadt" , "lang" : "de" } ,
    { "type" : "literal", "value" : "city" , "lang" : "en" } ,
    { "type" : "literal", "value" : "miasto" , "lang" : "pl" } ,
    { "type" : "literal", "value" : "stad" , "lang" : "nl" }
  ] ,
  "http://www.w3.org/2000/01/rdf-schema#range" :
  [ { "type" : "uri", "value" : "http://dbpedia.org/ontology/City" }
  ] ,
  "http://www.w3.org/2002/07/owl#sameAs" :
  [ { "type" : "uri", "value" : "http://dbpedia.org/ontology/city" }
  ] ,
  "http://www.w3.org/ns/prov#wasDerivedFrom" :
  [ { "type" : "uri", "value" :
    "http://mappings.dbpedia.org/index.php/OntologyProperty:city" } ] }
}
```

**Fonte:** dbpedia.org, w3.org

Portanto um desafio identificado neste trabalho é a avaliação sobre como associar as relações da DBpedia ou de outras ontologias com as frases em linguagem natural. No sentido de atuar na busca de uma resposta para este desafio, um primeiro passo a ser realizado deve ser o estudo de partes da ontologia da DBpedia, para identificar as relações existentes e também como estas relações podem auxiliar a gerar as frases.

Neste sentido, pode-se considerar um exemplo ilustrativo, como descrito no quadro 1, abaixo. A partir de um conceito “populationTotal” pode ser identificado o

uso do valor disponível na relação “label”, bem como sinônimos destes valores, para a geração de frases significativas para geração de perguntas sobre este conceito.

#### Quadro 1 – exemplo ilustrativo

- a) Conceito da DBPedia: “Dbpedia-owl: populationTotal”
- b) Estimativa de geração de perguntas relacionados:
  - 1. “Qual é a população total de ...”;
  - 2. “Me diga qual é a população total da cidade de ...”;
  - 3. “Diga qual é a quantidade de pessoas que vivem em ...”.
- c) Elementos disponíveis diretamente: semântica do elemento `rdfs:label`, que contém uma descrição textual do conceito associado. O termo “população total” está representado como valor do elemento `rdfs:label`.
- d) Elementos disponíveis indiretamente: sinônimos dos termos do elemento `rdfs:label` podem ser usados. Por exemplo, o termo “população total” pode ser considerado um termo sinônimo de “quantidade de pessoas que vivem em”.

Conforme observado nos trabalhos relacionados estudados e descritos no capítulo anterior, existem diversas abordagens utilizadas para a tarefa de uso dos recursos de dados abertos e conectado para a geração de frases em linguagem natural. Uma parte considerável dos trabalhos utiliza padrões de lexicalização, que podem variar quanto à sua origem, que pode ser desde a geração manual até a utilização de recursos de aprendizagem de máquina. Também são observados recursos oriundos de estudos linguísticos, que possibilitam o apoio na geração das frases.

Neste trabalho existe uma delimitação do escopo para a geração de frases com perguntas, em sentenças curtas. Esta delimitação será usada no estudo como um elemento importante, tanto para definir requisitos como para apoiar no estudo de aspectos das bases de dados abertos e conectados.

Associado a este aspecto, um dos diferenciais almejados com este trabalho está na análise de relações descritas pelas ontologias e na identificação de possibilidades para relacionamento automático entre estas relações e os elementos de lexicalização. Outro diferencial do trabalho está na busca de identificação de aspectos das relações que possam eventualmente ser usados para apoiar tarefas de correções de respostas, a serem usados em trabalhos futuros.

Sendo assim, a próxima seção apresenta alguns detalhes dos estudos realizados neste sentido. Em seguida são expressas as primeiras definições no

sentido de um algoritmo para a implementação desta tarefa de geração de frases curtas com perguntas, a partir de dados abertos e conectados.

## 4.2 Estudo de bases de dados abertas e conectadas

Este item descreve o estudo realizado para identificar informações que podem apoiar o processo de geração de linguagem, para perguntas curtas, em fontes de dados abertas e conectadas.

A primeira fonte de informação escolhida foi a DBpedia. A metodologia a seguir foi definida para esta primeira etapa: estudar as descrições da ontologia da DBpedia para identificar quais os itens possuem elementos com viabilidade para auxiliar na automatização da geração de frases com perguntas a partir de dados abertos e conectados.

A análise inicial partiu do pressuposto de que as relações definidas para os elementos da ontologia estudada possuem aspectos que descrevem uma semântica bem definida e que podem ser utilizados para o processo de geração de linguagem. Além da descrição das relações, estima-se ser possível utilizar sua qualificação, como tipo de dados, intervalos de valores, ou outras restrições.

**Figura 19** Conceito sobre população total na DBpedia

This website uses cookies

**About: [population total](#)**

An Entity of Type : [Property](#), from Named Graph : <http://wikidata.dbpedia.org/resource/classes#>, within Data Space : <http://wikidata.dbpedia.org/>

| Property                                                 | Value                                                                                                                                                                                                         |
|----------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <a href="#">rdf:type</a>                                 | <ul style="list-style-type: none"> <li>owl:FunctionalProperty</li> <li>rdf:Property</li> <li>owl:DatatypeProperty</li> </ul>                                                                                  |
| <a href="#">rdfs:label</a>                               | <ul style="list-style-type: none"> <li>Einwohnerzahl</li> <li>population totale</li> <li>inwonersaantal</li> <li>population total</li> <li>população total</li> <li>συνολικός_πληθυσμός</li> </ul>            |
| <a href="#">rdfs:range</a>                               | <ul style="list-style-type: none"> <li>xsd:nonNegativeInteger</li> </ul>                                                                                                                                      |
| <a href="#">owl:equivalentProperty</a>                   | <ul style="list-style-type: none"> <li><a href="http://www.wikidata.org/entity/P1082">http://www.wikidata.org/entity/P1082</a></li> </ul>                                                                     |
| <a href="#">http://www.w3.org/ns/prov#wasDerivedFrom</a> | <ul style="list-style-type: none"> <li><a href="http://mappings.dbpedia.org/index.php/OntologyProperty:populationTotal">http://mappings.dbpedia.org/index.php/OntologyProperty:populationTotal</a></li> </ul> |
| <a href="#">owl:sameAs</a>                               | <ul style="list-style-type: none"> <li>dbpedia-owl:populationTotal</li> </ul>                                                                                                                                 |

Fonte: Dbpedia

A partir da Figura 19 acima, alguns aspectos preliminares são analisados e comentados. Estes foram descritos na Tabela 6 abaixo.

**Tabela 6 Aspectos iniciais identificados**

| <b>Elemento</b>                | <b>Significado</b>                                                                                                                                                                                |
|--------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| rdfs: label                    | Contém o rótulo da relação. Pode ser usado para gerar a frase. Sinônimos deste “label” também podem ser usados.                                                                                   |
| rdfs: range                    | Indica o tipo de dados que será usado para completar a relação. Pode ser usado para apoiar a correção da resposta dada à pergunta?                                                                |
| owl: sameAs                    | Relação que indica uma existência de outra relação que tem o mesmo significado. Permite considerar o uso dos termos da relação indicada como apoio na geração de frases.                          |
| rdfs: type: datatype property  | Relação que descreve literais como valores e, portanto pode também indicar tipos específicos de dados e variações destes valores. Todos estes aspectos podem ser utilizados na geração de frases. |
| rdfs: type: object property... | Relação que descreve uma conexão com outros conceitos. Permite identificação de conjuntos restritos de valores.                                                                                   |

**Fonte:** Autor

Com base na metodologia definida, foi ampliada a quantidade de elementos estudados, sendo que uma parte destes está resumida na Tabela 7 abaixo. O total de elementos analisados até o momento ainda não permite considerar uma ampla cobertura sobre as relações e conceitos existentes, mas já possibilita identificar algumas tendências e sinalizar possibilidades de aplicação de algoritmos para aproveitar estas situações na geração de frases curtas para perguntas. O anexo A disponível neste trabalho apresenta em maiores detalhes outras consultas e análises realizadas.

**Tabela 7 Trecho parcial da tabela de análise, para o conceito [http://dbpedia.org/page/Porto\\_Alegre](http://dbpedia.org/page/Porto_Alegre)**

| Propriedade / Recurso                | Property     | Value                                                                                                                             |
|--------------------------------------|--------------|-----------------------------------------------------------------------------------------------------------------------------------|
| dbo:PopulatedPlace/area              | rdf:type     | owl:DatatypeProperty                                                                                                              |
|                                      | rdfs:domain  | dbo:PopulatedPlace                                                                                                                |
|                                      | rdfs:label   | área (km <sup>2</sup> )                                                                                                           |
|                                      | rdfs:range   | <a href="http://dbpedia.org/datatype/squareKilometre">http://dbpedia.org/datatype/squareKilometre</a>                             |
|                                      | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                                           |
| dbo:PopulatedPlace/areaTotal         | rdf:type     | owl:DatatypeProperty                                                                                                              |
|                                      | rdfs:domain  | dbo:PopulatedPlace                                                                                                                |
|                                      | rdfs:label   | área total (km <sup>2</sup> )                                                                                                     |
|                                      | rdfs:range   | <a href="http://dbpedia.org/datatype/squareKilometre">http://dbpedia.org/datatype/squareKilometre</a>                             |
|                                      | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                                           |
| dbo:PopulatedPlace/populationDensity | rdf:type     | owl:DatatypeProperty                                                                                                              |
|                                      | rdfs:domain  | dbo:PopulatedPlace                                                                                                                |
|                                      | rdfs:label   | population density (/sqkm) (en)                                                                                                   |
|                                      | rdfs:range   | <a href="http://dbpedia.org/datatype/inhabitantsPerSquareKilometre">http://dbpedia.org/datatype/inhabitantsPerSquareKilometre</a> |
|                                      | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                                           |
| dbo:ontology/country                 | rdf:type     | rdf:Property; owl:ObjectProperty                                                                                                  |
|                                      | rdfs:comment | The country where the thing is located. (en)                                                                                      |
|                                      | rdfs:label   | country (en)                                                                                                                      |
|                                      | rdfs:range   | dbo:Country                                                                                                                       |
| dbo:ontology/isPartOf                | rdf:type     | rdf:Property; owl:ObjectProperty                                                                                                  |
|                                      | rdfs:label   | is part of (en)                                                                                                                   |
| dbo:ontology/populationTotal         | rdf:type     | owl:FunctionalProperty; rdf:Property;<br>owl:DatatypeProperty                                                                     |
|                                      | rdfs:label   | population total (en); população total (pt)                                                                                       |
|                                      | rdfs:range   | xsd:nonNegativeInteger                                                                                                            |

**Fonte:** Autor

A análise prevista nesta etapa parte, portanto, do uso de recursos de lexicalização já definidos nas bases de dados abertos e conectados, sendo que o primeiro elemento investigado é o elemento “*label*”. Para este elemento foram realizados testes de consultas para verificar o seu conteúdo, além de realizar a sua validação, quanto à possibilidade de utilização na lexicalização das frases, com apoio de profissionais especializados.

Como segundo elemento investigado, foram catalogados e analisados os exemplos de relações que permitem a geração automática de lexicalização, seja de forma automática, com recursos de processamento de linguagem natural, seja com anotação manual. Nestes casos, a validação destes elementos como possibilidade

permite a inclusão e constante atualização de uma etapa ou opção no algoritmo de geração de perguntas em linguagem natural.

Um terceiro elemento a ser considerado são as definições conjuntas de tipo (rdf:type) e de faixa de valores (rdf:range), que foram considerados, nesta análise, como potenciais elementos para apoiar a definição de alguns dos aspectos de perguntas, tais como o tipo de pergunta e os valores possíveis. Por exemplo, uma relação cuja faixa de valores está definida como valores inteiros permite considerar este aspecto na geração da pergunta e também em sua posterior avaliação.

### **4.3 Algoritmo geral proposto para geração de frases e arquitetura de apoio**

O algoritmo de geração de frases com perguntas curtas proposto neste trabalho consiste de uma derivação do estudo realizado com trabalhos relacionados e da qualificação das relações estudadas com as bases de dados conectados que foram analisadas.

Seguindo a linha geral de trabalhos como (Ell e Harth, 2016; Duma e Klein, 2015), diversos recursos foram descritos para apoiar a geração das frases. Alguns dos recursos são padrões definidos manualmente, com base no estudo das relações das bases de dados analisadas, contendo aspectos de lexicalização e de substituição de valores. Outros padrões são mais elaborados e utilizam as relações semânticas detectadas, para então definir o tipo de pergunta a ser gerada e os elementos de lexicalização a serem usados. Além disso, são previstos recursos adicionais para compor um cenário de flexibilidade de geração de frases, com uso de variações linguísticas, uso de sinônimos e termos compostos.

Os elementos inicialmente definidos iniciam pelos padrões gerais de pergunta com aspectos de lexicalização, que associam de modo manual uma relação e conceito da base de dados aberta e conectada com um termo a ser usado para a geração das frases. Estes padrões possuem o formato descrito na Figura 20 abaixo:

**Figura 20 Exemplos de padrões de lexicalização**

Exemplo de padrão:

[tipo de pergunta] [tipo de objeto] [ tipo de qualificação]

Exemplo de frases:

[Qual a] [quantidade] [ ...]

[Qual a] [data] [...]

[Qual o] [nome] [...]

**Fonte:** Autor

O segundo grupo de elementos é descrito a partir de uma associação de relações com tipos de pergunta, sendo que as perguntas, nestes casos, são geradas de forma mais flexível. Nestes casos são mapeadas preliminarmente associações entre relações e as perguntas respectivas, levando em conta a semântica da relação e atributos, como os seus tipos de dados, por exemplo. Nestes casos, alguns exemplos são, para a relação do tipo “Total”, as perguntas do tipo “Qual”, ou “Quanto”. Ou então, para as relações do tipo “Date”, as perguntas do tipo “Quando”.

A diferença fundamental deste segundo grupo para com o primeiro grupo de padrões é que este padrão não é gerado manualmente, mas gerado a partir de um conjunto de relacionamentos entre padrões de lexicalização para as perguntas e um conjunto de relações associado.

O terceiro grupo de elementos está definido a partir da experimentação de uso dos elementos analisados nas bases de dados abertas e conectadas. Por exemplo, um estudo inicial realizado permitiu identificar que o elemento “label” possui padrões de lexicalização com potencial de uso. Deste modo, este tipo de estudo será continuado e serão catalogados continuamente novos elementos e sua possibilidade de utilização. De modo a fomentar a automatização sempre que possível estes padrões são definidos com base na listagem das relações, dos itens a serem utilizados e também em operações a serem realizadas, apoiando assim o ajuste dos termos sempre que necessário.

A dinâmica geral do algoritmo proposto está descrita na Figura 21 a seguir:

**Figura 21 Dinâmica geral do algoritmo definido**

Seleção do conjunto de bases de dados abertos a utilizar.  
Seleção do tipo de conceito desejado pelo usuário.  
Realização de consulta SPARQL para recuperar as relações e valores associados com o conceito buscado.  
Análise do conjunto de relações obtido com a consulta  
Utilização do conjunto de padrões de lexicalização para a geração das frases curtas.

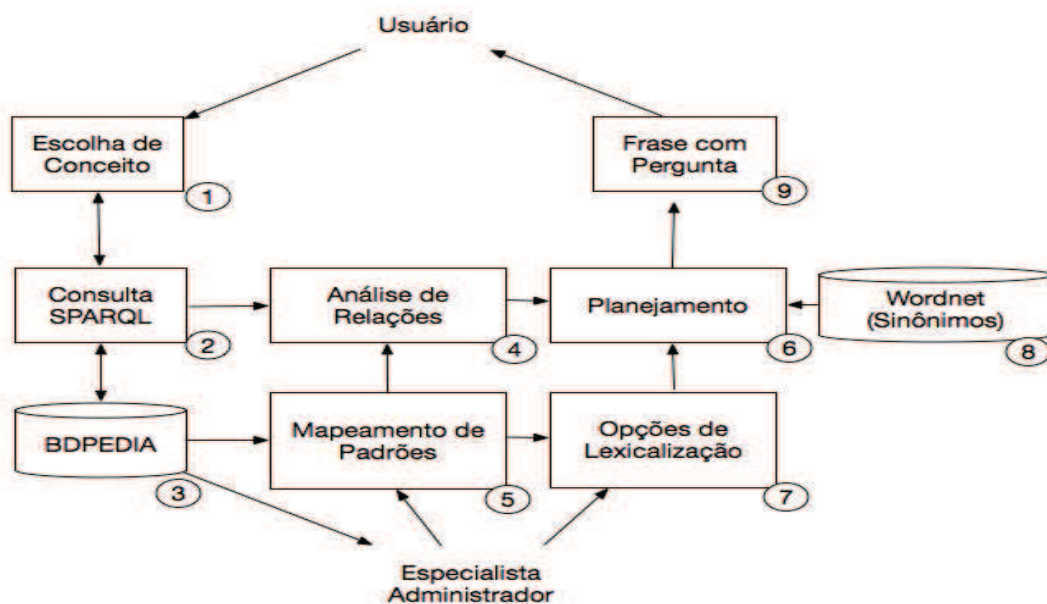
**Fonte:** Autor

Para apoiar a execução deste algoritmo proposto, os componentes necessários estão relacionados na arquitetura geral, resumida na Figura 22 abaixo. Na figura 22, que contém os principais componentes da arquitetura proposta, podem ser observados os elementos e processos descritos a seguir.

Existem duas formas de interação com o sistema, sendo uma de parte dos usuários, que irão utilizar o sistema para consultar conceitos e receber as frases com as perguntas em linguagem natural. A segunda forma de interação é por parte de especialistas, responsáveis pela identificação e documentação dos padrões de lexicalização a serem empregados.

Sobre a forma de operação prevista, existem também duas etapas diferenciadas. Uma das etapas é dedicada ao atendimento dos acessos do usuário, processo em que são consultados os conceitos e geradas as frases com perguntas em linguagem natural. Nesta etapa de operação, são seguidas as etapas numeradas na figura. A segunda etapa prevista consiste de uma etapa dedicada ao atendimento do usuário Especialista ou administrador, sendo que nesta etapa apenas os itens numerados como 3, 5 e 7 estão envolvidos. As atividades desta etapa envolvem o estudo das relações da base de dados abertos e conectados estudada, da geração de padrões de interesse e de geração de padrões de lexicalização.

Figura 22 Visão geral de componentes



Fonte: Autor

Durante a operação prevista para o atendimento do usuário desta arquitetura, os módulos de número 1 e 9 consistem apenas de elementos dedicados à interface do sistema com os usuários para permitir a escolha de conceitos pelo mesmo e a exibição das frases geradas como resultado do processo.

O módulo identificado como item 2 possui como objetivo a tradução da seleção de um conceito em uma consulta SPARQL para ser realizada em uma base de dados aberta e conectada. No caso deste exemplo da Figura 22, a base indicada é a DBPEDIA (item 3) da figura. Estes dois itens (2 e 3), portanto, são os itens envolvidos na consulta às bases de dados abertas e conectadas, sendo que neste caso foi indicado como a DBpedia.

Os componentes de Mapeamento de Padrões e de Análise de relações utilizam os aspectos do algoritmo definido para permitir que sejam relacionadas diferentes relações com o seu correspondente item lexical, compondo assim um banco de informações necessárias para a geração das frases. O Item de Mapeamento de padrões (item 5) é o responsável por armazenar um conjunto de informações sobre as relações analisadas na base de dados abertos e conectados e que foram consideradas como viáveis para a utilização na geração de linguagem natural. O item de Análise de relações (item 4) é o item responsável por realizar a

verificação das relações obtidas na consulta realizada pelo usuário sobre um determinado conceito e identificar se o conjunto de padrões já selecionados possui alguma destas relações. No caso de terem sido encontrados elementos nestes padrões, isso significa que estes padrões podem ser utilizados na etapa seguinte, de lexicalização.

Por fim os elementos de Opções de Lexicalização (item 7) e de Planejamento (item 6) são os responsáveis pela integração final das informações. O elemento de Opções de lexicalização mantém uma base de padrões para serem usados na geração das frases. Já o item de Planejamento possui como tarefa a integração dos aspectos lexicais com os dados obtidos com a consulta SPARQL feita sobre o conceito escolhido, junto com recursos adicionais, tais como o Wordnet, que permitem a expansão das etapas de lexicalização, com uso de recursos linguísticos.

#### **4.4 Aspectos de Implementação e avaliação**

Os aspectos considerados como centrais neste trabalho, em termos de contribuição, estão associados com o estudo preliminar do atual contexto ligado às bases de dados abertas e conectadas, bem como o estudo das relações descritas nestas bases e na sua semântica. Desta forma, buscou-se avançar principalmente nestes aspectos do estudo.

Entretanto, como forma de avaliação sobre os possíveis resultados destes subsídios identificados, foi realizado no escopo do trabalho uma implementação parcial da arquitetura proposta, o que proporcionou os elementos necessários para uma etapa de geração de frases curtas com perguntas em linguagem natural e também permitiu a avaliação destes resultados.

Foram desenvolvidas experimentações com o uso de NodeJs e com o uso de Java, para realizar etapas do processo geral descrito. Estes experimentos permitiram validar os procedimentos de acesso às bases de dados abertas e conectadas, tendo sido usada como caso de estudo a DBPedia, tanto para busca de dados disponíveis em língua portuguesa como para dados em língua inglesa. O principal objetivo foi a implementação do acesso e recuperação de dados, seguido pelo posterior tratamento destes para a geração de frases com perguntas em

linguagem natural e, por fim, a sua avaliação crítica por parte de uma especialista em linguística.

Foi utilizado no experimento 1 uma parte do conjunto de relações estudadas. Definiu-se, por conveniência de acesso e facilidade de obtenção de maior quantidade de dados, a busca de conceitos em língua inglesa. Os mesmos procedimentos podem ser utilizados em outras línguas, dentro de um escopo de restrições de alguns de seus aspectos. Deste modo, definiu-se o uso de perguntas do tipo “Who”, “How much” e “What”. Estes tipos iniciais de perguntas definidos estão associados com aspectos da semântica de relações estudadas. Desta forma, o primeiro caso está associado com a identificação de uma pessoa, o segundo está associado com relações descrevendo quantidades e o último está associado com outras situações gerais, como forma de avaliar resultados não mapeados previamente.

No caso da execução de uma consulta com uso da linguagem SPARQL, o resultado típico permite que sejam consultadas diversas relações para o conceito que foi o motivo da consulta. A figura 18 ilustra, por exemplo, as diversas relações associadas com um conceito como “cidade”, dentre as quais é possível identificar as relações como “label”, ou “range”, entre diversas outras. Quando a consulta é realizada para a busca do nome de uma pessoa, por exemplo, o mesmo mecanismo de identificação de relações possibilita a descrição de relações que identificam o tipo e domínio, indicando o pertencimento à classe pessoa. Da mesma forma, em diversas outras categorias de conceitos, existem relações apropriadas para a sua identificação, que podem ser então utilizadas na geração de subsídios para a geração de perguntas em linguagem natural.

Na geração deste teste prevendo a avaliação, foram usadas algumas destas relações especificamente. Deste modo, quando é realizada uma consulta e o elemento que representa o sujeito da tupla recebida na consulta pode ser verificado como possuindo a relação do tipo “dbo:domain”, então esta informação pode ser utilizada com confiança como o delimitador para o tipo de questão a ser associada. Por exemplo, se a relação de domínio (dbo:domain) apontar para uma pessoa como domínio, a questão pode ser realizada com segurança a respeito deste aspecto. O mesmo ocorre com as demais relações mapeadas.

Para a avaliação do protótipo procedeu-se à realização de experimentos com a finalidade de identificar os aspectos positivos gerados pelo estudo das relações estudadas. Um dos aspectos de interesse foi a funcionalidade do protótipo na busca de tratamento dos dados. O outro aspecto foi a sua capacidade de geração de frases com perguntas em linguagem natural. Para a avaliação dos resultados obtidos foi executada uma análise por especialistas em linguística, buscando identificar problemas de lexicalização e de uso da linguagem.

Como forma de diversificar as possibilidades de obter subsídios para a análise do procedimento desenvolvido, foram definidos elementos de consulta que representam categorias diversificadas. Por exemplo, foram identificadas categorias como pessoas (“Bill Gates”, “Barack Obama”, “Neymar”), cidades (“Chicago”, “Porto Alegre”) e outras entidades como animais (“lion”, “tiger”), doenças (“Influenza”) e eventos (“World War II”, “French Revolution”). Uma listagem completa das consultas realizadas e das frases geradas, junto com a sua avaliação pela especialista em linguística está disponível no Anexo B deste trabalho.

Com base neste conjunto de tópicos escolhidos e nos três tipos de perguntas elencados inicialmente, foram geradas as frases curtas com perguntas e depois as mesmas foram analisadas por uma especialista em linguística. Um ponto de vista inicial da análise foi identificar se a frase está correta gramaticalmente. Outro ponto de vista de análise foi buscar identificar a correção no uso da linguagem da forma como foi gerada a pergunta. Estes dois aspectos estão relacionados intimamente com o mapeamento das relações disponíveis nas bases de dados abertas e conectadas e a sua semântica. Melhorias nestes estudos levam a melhorias nos resultados obtidos. Por exemplo, uma mesma pergunta pode ser formulada corretamente com diferentes estilos de escrita, o que precisa ser identificado como possibilidade a partir do estudo das relações e de informação gerada pelos especialistas em linguística.

Foram geradas cerca de 205 questões, com base nos tópicos definidos. A análise dos resultados por especialista em linguística permitiu identificar situações que serão tratadas em trabalhos futuros. Além da identificação de correção das sentenças, esta análise foi realizada para indicar também outras formas de gerar as frases com as perguntas, o que poderá servir de subsídio futuro para a

diversificação dos resultados. No anexo C deste trabalho está disponível este conjunto de informações.

Uma das situações a destacar na avaliação é a existência de relações contendo descrições que não são adequadas para a geração de perguntas, pois indicam elementos adicionais, por vezes técnicos, dos conceitos tratados. Por exemplo, uma destas é a situação em que as relações são usadas para identificar termos adicionais sobre material a respeito do conceito. Por exemplo, é comum encontrar-se o termo “*has abstract*” associado com diversos conceitos indicando um texto adicional sobre aquele conceito. No caso de um conceito associado com uma pessoa, não terá sentido a realização de uma pergunta sobre esta relação. Ou então, para alguns conceitos, especialmente animais ou fenômenos atmosféricos, é comum a existência de relações contendo imagens sobre os mesmos, nas quais os termos usados são tais como “image”, “image width”, “image caption”. Ou então, em mais um exemplo de interesse, conceitos descrevendo doenças podem conter relações apontando para recursos como bases de dados nas quais existem mais informações sobre as mesmas, como “DiseasesDb”.

Além desta identificação, os resultados possibilitaram identificar situações de correção gramatical das frases. Na figura 23 estão resumidas algumas das análises realizadas e seus resultados. Todas as situações de interesse identificadas podem ser observadas nesta figura. A figura 23 representa alguns dos termos de consulta acessados na Base de dados aberta, na sua primeira coluna. Neste caso, estão representados os termos “French Revolution”, “World War II”, “Porto Alegre”, “Barack Obama”. Para cada um destes termos, a primeira coluna mostra as frases geradas. As frases corretas estão representadas em letra preta e as frases incorretas estão representadas em letra vermelha.

A segunda coluna, na figura 23, exhibe exemplos de resultados das correções e avaliações realizadas, sendo que podem ser identificadas tanto as correções das frases, quando necessário, como exemplos das mesmas frases contendo formulação diferente, como exemplo de diversidade na geração dos resultados.

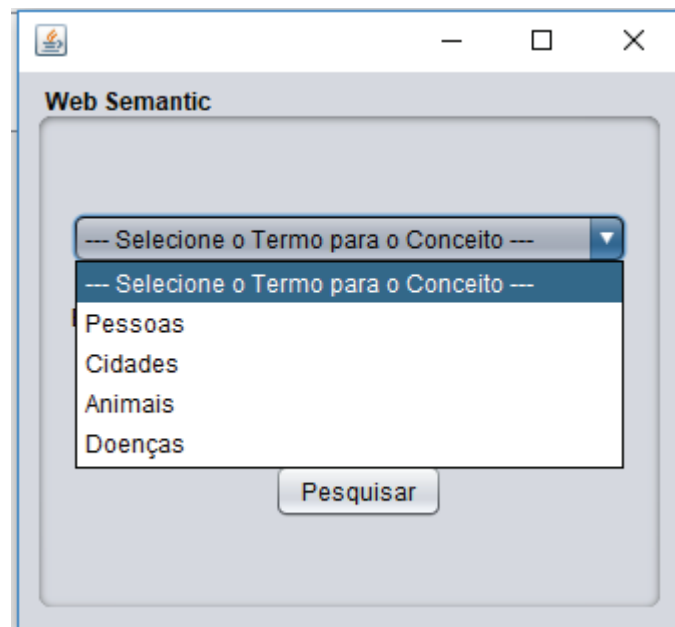
Figura 23 Resumo de resultados

|                                                   |  |                                       |
|---------------------------------------------------|--|---------------------------------------|
| <b>French Revolution</b>                          |  | CORRETA                               |
| What's French Revolution's has abstract?          |  | What's the French Revolution history? |
| What's French Revolution's date?                  |  |                                       |
| What's French Revolution's Event Name?            |  |                                       |
| What's French Revolution's image caption?         |  |                                       |
| <b>World War II</b>                               |  | CORRETA                               |
| What's World War II's has abstract?               |  | What's the World War II history?      |
| What's World War II's casualties?                 |  | What were the WWII effects?           |
| What's World War II's combatant?                  |  | Who were the WWII soldiers?           |
| What's World War II's place of military conflict? |  | Where did the WWI take place?         |
| What's World War II's result?                     |  |                                       |
| <b>Porto Alegre</b>                               |  | CORRETA                               |
| What's Porto Alegre's area (km2)?                 |  | What's Porto Alegre's area?           |
| What's Porto Alegre's area total (km2)?           |  |                                       |
| What's Porto Alegre's population density (/sqkm)? |  |                                       |
| What's Porto Alegre's has abstract?               |  |                                       |
| How much is Porto Alegre's area (m2)?             |  |                                       |
| What's Porto Alegre's area code?                  |  |                                       |
| How much is Porto Alegre's area total (m2)?       |  |                                       |
| <b>Barack Obama</b>                               |  | CORRETA                               |
| What's Barack Obama's birth place?                |  | What's Barack Obama's birthplace?     |
| What's Barack Obama's birth date?                 |  | What's Barack Obama's birth date?     |
| What's Barack Obama's name?                       |  | What's Barack Obama's name?           |
| What's Barack Obama's has abstract?               |  | What's Barack Obama's history?        |

Além destes aspectos citados, que priorizam a avaliação da correção das frases, do ponto de vista gramatical ou do ponto de vista de utilização da linguagem, deve ser destacado que o resultado desta avaliação foi considerado positivo do ponto de vista da funcionalidade. A abordagem avaliada permite que sejam acessados e tratados os dados oriundos de bases de dados abertas e conectadas, de forma eficiente, levando em conta tanto o acesso remoto, como a possibilidade de acesso local, tendo em vista que diversas destas permitem a sua cópia para utilização em contextos com necessidades maiores de performance.

Foi utilizado no experimento 2 o desenvolvido de um protótipo para executar a consulta SPARQL com a geração de frases curtas, tendo com fonte de dados DBPEDIA. Na figura 24, o usuário começa a “Selecionar o Termo para o Conceito” que pode ser: “Pessoas, Cidades, Animais e Doenças”.

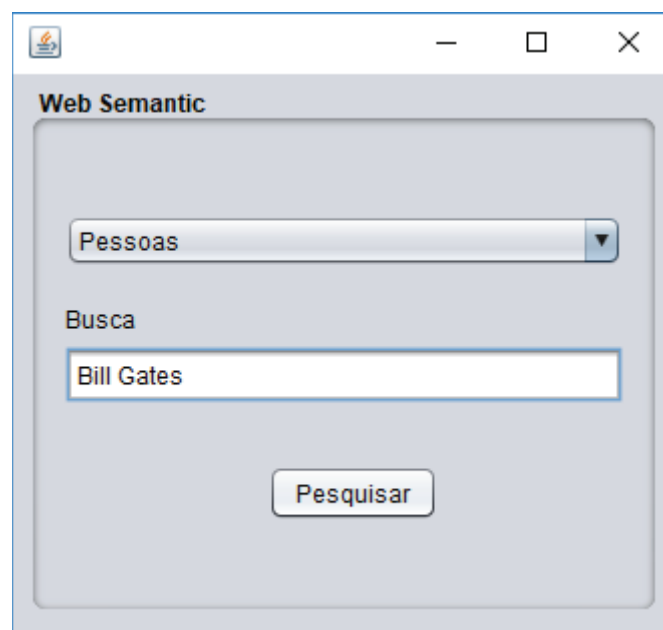
Figura 24 Tela Inicial



Fonte: Autor

Na figura 25, o usuário deve digitar o nome para a aplicação realizar a consulta com as perguntas geradas, de acordo com a figura 26 onde será gerado o questionário.

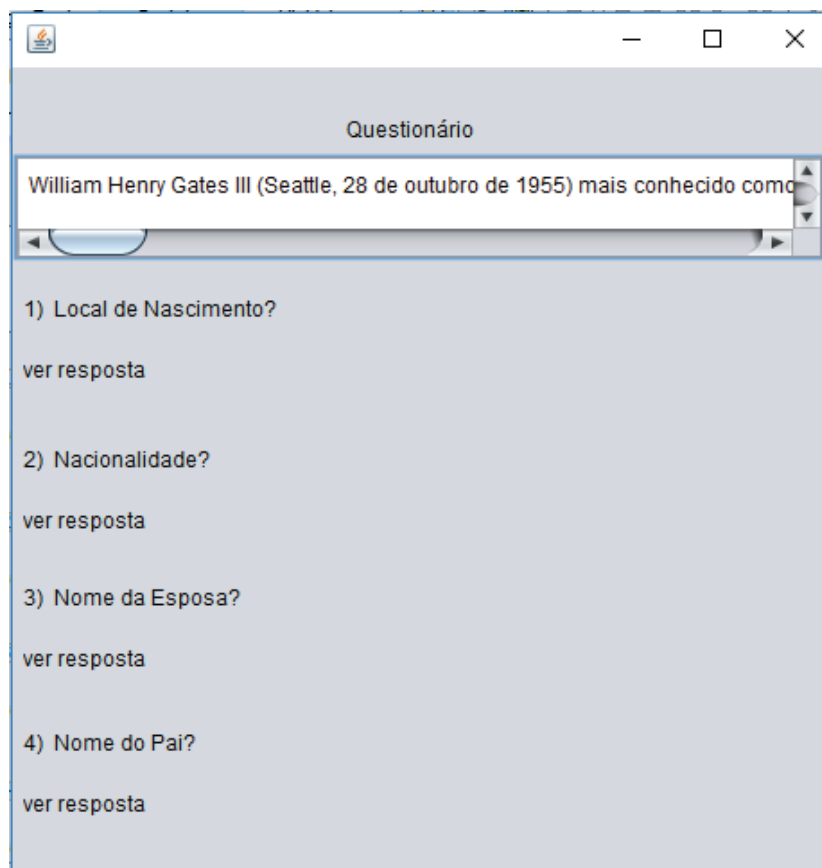
Figura 25 Tela de seleção da Busca



Fonte: Autor

Este segundo experimento teve como objetivo o acesso aos dados e implementação do protótipo para testar a geração das perguntas explorando o local de origem da resposta.

**Figura 26 Tela da Resposta da Consulta**



The screenshot shows a web application window with a title bar containing a small icon and standard window controls (minimize, maximize, close). The main content area has a header labeled "Questionário". Below the header is a text input field containing the text "William Henry Gates III (Seattle, 28 de outubro de 1955) mais conhecido como". Below the input field is a horizontal scrollbar. The main content area lists four questions, each followed by a "ver resposta" link:

- 1) Local de Nascimento?  
ver resposta
- 2) Nacionalidade?  
ver resposta
- 3) Nome da Esposa?  
ver resposta
- 4) Nome do Pai?  
ver resposta

**Fonte:** Autor

## 5 CONCLUSÃO

O trabalho apresentado descreve a pesquisa realizada para explorar a possibilidade de geração de linguagem natural a partir de bases de dados abertas e conectadas. Foram identificadas potencialidades nesta área, de modo a ampliar o aproveitamento dos atuais conjuntos de dados disponíveis. Estes conjuntos possuem uma tendência de crescimento continuado. Além disso apresentam aspectos que favorecem o seu uso nesta tarefa de geração de linguagem natural, tais como a sua estruturação formal, a partir de ontologias ou vocabulários bem definidos. Estes conjuntos também apresentam como aspecto relevante a sua tendência de atualização constante, em especial aqueles conjuntos associados com iniciativas de governo aberto ou então conjuntos associados com redes sociais e de colaboração, tais como a Wikipedia, por exemplo.

Foram analisadas bases de dados abertas e conectadas no que diz respeito às relações descritas no formalismo RDF, o que possibilitou a indicação de procedimentos para o uso destas relações na geração automática de frases em linguagem natural, contendo perguntas que utilizam estas relações. Também foi descrito um modelo geral para a implementação de procedimentos de utilização destas relações para a geração de perguntas. Aspectos deste modelo foram validados e avaliados, com um protótipo que permitiu identificar aspectos positivos e melhorias a serem realizadas em atividades futuras.

A colaboração apresentada neste trabalho está associada com o estudo amplo dos atuais conjuntos de dados abertos e conectados e de suas relações. Espera-se, a partir desta abordagem, ter sido definido passos na direção de um procedimento que auxilie a identificar elementos para apoiar as tarefas de geração de linguagem natural. A partir deste estudo, espera-se, portanto, qualificar os métodos de geração de frases em linguagem natural, de forma a ampliar o coeficiente de possibilidades de automatização desta atividade. Atualmente os trabalhos estudados possuem um grau elevado de utilização de padrões lexicais definidos de modo manual.

### 5.1 Limitações da pesquisa

O trabalho desenvolvido apresenta algumas limitações que são destacadas neste item. Uma delas está relacionada com o aspecto de análise de relações existentes em bases de dados abertas e conectadas, que permite gerar subsídios para a geração de frases, na medida em que é ampliada. Em função de restrições de tempo e de restrições técnicas a pesquisa neste aspecto ficou limitada em um conjunto preliminar de dados. Além disso, não foi possível ampliar o conjunto de bases de dados utilizadas, o que certamente geraria uma maior perspectiva de resultados. Por exemplo, algumas bases de dados abertas e conectadas, por estarem ligadas à algumas áreas específicas, tais como a saúde, ou educação, constituem um conjunto de relações com uma semântica bastante peculiar, que pode ser utilizado no contexto de geração de frases.

Neste trabalho não foi realizada a avaliação de aspectos do público alvo das questões geradas, o que poderia trazer informações importantes sobre os resultados obtidos e sobre o processo de mapeamento das informações.

A quantidade de frases geradas e o conjunto de avaliadores envolvidos ficou reduzido neste momento, com base na disponibilidade de recursos. Desta forma, considera-se que uma maior quantidade de frases e um conjunto maior de avaliadores poderiam ter enriquecido a análise apresentada.

A implementação do protótipo não foi um item privilegiado durante o desenvolvimento do trabalho, tendo sido escolhido como foco o estudo das relações disponíveis em bases de dados abertas e conectadas. Entretanto, considera-se que este ponto é também uma limitação do trabalho atual, pois a estrutura completa do protótipo, segundo as diversas etapas descritas, permitiria validar de forma mais consistente o modelo proposto.

## **5.2 Trabalhos futuros**

Como trabalhos futuros, está prevista a ampliação do conjunto de estudos sobre as relações disponibilizadas em bases de dados abertas e conectadas. Além disso, existe em andamento um processo de desenvolvimento de um protótipo mais completo, envolvendo colaboração com outros pesquisadores, o que vai

proporcionar o avanço em etapas importantes, como as avaliações de frases geradas.

Outro aspecto a ampliar com trabalhos futuros é a avaliação realizada. Está prevista a ampliação de testes de avaliação para um contexto mais amplo, não apenas no conjunto de temas tratados, mas também no conjunto de línguas consideradas. Por fim, espera-se realizar esta nova etapa de avaliações contando com um grupo de especialistas, o que provavelmente ampliará a qualidade das observações destacadas atualmente.

O trabalho atual não se preocupou com o tratamento das informações obtidas nas bases de dados abertas e conectadas com o objetivo de realizar uma validação de respostas para as perguntas geradas. Deste modo, pretende-se ampliar o escopo do trabalho atual para que sejam mantidas informações que permitam a verificação automática da correção de respostas para as perguntas geradas.

## REFERÊNCIAS

AMORIM, Marta Talitha Carvalho Freire; CURY, Davidson; MENEZES, Crediné Silva de. **Sobre a aplicação de ontologias para orientar agentes a responder perguntas**. Revista brasileira de informática na educação. Florianópolis. Vol. 22, n. 3 (2014), p. 1-12, 2014.

ANDROUTSOPOULOS, Ion; LAMPOURAS, Gerasimos; GALANIS, Dimitrios. **Generating natural language descriptions from OWL ontologies: the NaturalOWL system**. Journal of Artificial Intelligence Research, v. 48, p. 671-715, 2013.

BAUER, Florian; KALTENBÖCK, Martin. **Linked open data: The essentials**. Edition mono/monochrom, Vienna, 2011.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python**: "O'Reilly Media Inc.". 2009.

BOBED, Carlos; ESTEBAN, Guillermo; MENA, Eduardo. **Enabling keyword search on Linked Data repositories: An ontology-based approach**. International Journal of Knowledge-Based and Intelligent Engineering Systems, v. 17, n. 1, p. 67-77, 2013.

BOUTTAZ, Thomas et al. **A policy-based approach to context dependent natural language generation**. In: Proceedings of the 13th European Workshop on Natural Language Generation. Association for Computational Linguistics, 2011. p. 151-157.

DA SILVA JUNIOR, Douglas Fernandes Pereira; DE NOVAIS, Eder Miranda; PARABONI, Ivandrê. **Um Sistema de Realização Superficial para Geração de Textos em Português**. Revista de Informática Teórica e Aplicada, v. 20, n. 3, p. 31-48, 2013.

DANNÉLLS, Dana et al. **Multilingual online generation from semantic web ontologies**. In: Proceedings of the 21st International Conference on World Wide Web. ACM, 2012. p. 239-242.

DUBEY, Mohnish et al. **AskNow: A Framework for Natural Language Query Formalization in SPARQL**. In: International Semantic Web Conference. Springer International Publishing, 2016. p. 300-316.

DUMA, Daniel; KLEIN, Ewan. **Generating natural language from linked data: Unsupervised template extraction**. Association for Computational Linguistics, Potsdam, Germany, p. 83-94, 2013.

ELL, Basil; HARTH, Andreas. **A language-independent method for the extraction of RDF verbalization templates**. INLG-2014, 2014.

ELL, Basil; HARTH, Andreas. **A language-independent method for the extraction of RDF verbalization templates**. INLG-2014, 2014.

FABIAN, M. S. et al. **Yago: A core of semantic knowledge unifying wordnet and wikipedia**. In: 16th International World Wide Web Conference, WWW. 2007. p. 697-706.

GALANIS, Dimitrios; ANDROUTSOPOULOS, Ion. **Generating multilingual descriptions from linguistically annotated OWL ontologies: the NaturalOWL system**. In: Proceedings of the Eleventh European Workshop on Natural Language Generation. Association for Computational Linguistics, 2007. p. 143-146.

HARRIS, Ian G. **Extracting design information from natural language specifications**. In: Proceedings of the 49th Annual Design Automation Conference. ACM, 2012. p. 1256-1257.

HASLHOFER, Bernhard; ISAAC, Antoine. **data. europeana. eu: The europeana linked open data pilot**. In: International Conference on Dublin Core and Metadata Applications. 2011. p. 94-104.

HEATH, Tom; BIZER, Christian. **Linked data: Evolving the web into a global data space**. Synthesis lectures on the semantic web: theory and technology, v. 1, n. 1, p. 1-136, 2011.

ISOTANI, Seiji; BITTENCOURT, Ig Ibert. **Dados Abertos Conectados: Em busca da Web do Conhecimento**. Novatec Editora, 2015.

Klein, R. H. **MECANISMOS DE AMPLIAÇÃO DA TRANSPARÊNCIA EM PORTAIS DE DADOS ABERTOS GOVERNAMENTAIS BRASILEIROS À LUZ DA ACCOUNTABILITY THEORY**. Tese de doutorado PUCRS. Porto Alegre, 2017.

LEHMANN, Jens et al. **DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia**. Semantic Web, v. 6, n. 2, p. 167-195, 2015.

Lindberg, D. Popowixh, F. **Generating Natural Language Questions to Support Learning On-Line**. Proceedings of the 14th European Workshop on Natural Language Generation, pages 105–114, Sofia, Bulgaria, 2013.

MOLINA, Martin; SANCHEZ-SORIANO, Javier; CORCHO, Oscar. **Using open geographic data to generate natural language descriptions for hydrological sensor networks**. Sensors, v. 15, n. 7, p. 16009-16026, 2015.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de textos. Relatório Técnico–Instituto de Informática (UFG)**, 2007.

MORAIS, EDISON ANDRADE MARTINS. **Contextualização de documentos em domínios representados por ontologias utilizando mineração de textos**. 2007. Tese de Doutorado. Master's thesis, Institute of Informatics of the Federal University of Goiás, Brazil.

NGONGA NGOMO, Axel-Cyrille et al. **Sorry, i don't speak SPARQL: translating SPARQL queries into natural language**. In: Proceedings of the 22nd international conference on World Wide Web. ACM, 2013. p. 977-988.

NGONGA NGOMO, Axel-Cyrille et al. **SPARQL2NL: verbalizing sparql queries**. In: Proceedings of the 22nd International Conference on World Wide Web. ACM, 2013. p. 329-332.

PERERA, Rivindu; NAND, Parma; KLETTE, Gisela. **RealText-lex: A Lexicalization Framework for RDF Triples**. The Prague Bulletin of Mathematical Linguistics, v. 106, n. 1, p. 45-68, 2016.

PERERA, Rivindu; NAND, Parma. **Lexicalizing Linked Data towards a Human Friendly Web of Data**. In: Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on. IEEE, 2016. p. 873-878.

POWER, Richard. **Complexity assumptions in ontology verbalisation**. In: Proceedings of the ACL 2010 Conference Short Papers. Association for Computational Linguistics, 2010. p. 132-136.

SERÓN, Francisco J.; BOBED, Carlos. **VOX system: a semantic embodied conversational agent exploiting linked data**. Multimedia Tools and Applications, v. 75, n. 1, p. 381-404, 2016.

SHEKARPOUR, Saeedeh et al. **Question answering on linked data: Challenges and future directions**. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, 2016. p. 693-698.

WAZLAWICK, Raul. **Metodologia de pesquisa para ciência da computação, 2a edição**. Elsevier Brasil, 2014.

WEAL, Mark J. et al. **Ontologies as facilitators for repurposing web documents**. International Journal of Human-Computer Studies, v. 65, n. 6, p. 537-562, 2007.

Paletta, Francisco Carlos, and Marcos Luiz Mucheroni. **"O desenvolvimento da WEB 3.0: Linked Data e DBPEDIA."** Prisma. (2017).

## Anexo A

Listagem de bases de dados abertas e conectadas e principais relações estudadas.

[http://dbpedia.org/page/Elis\\_Regina](http://dbpedia.org/page/Elis_Regina)

| Propriedade / Recurso | Property                                                     | Value                                                                                                                                                                                                 |
|-----------------------|--------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>dbo:birthDate</b>  | rdf:type<br>rdfs:domain<br>rdfs:label<br>rdfs:range<br>Graph | owl:FunctionalProperty; rdf:Property;<br>owl:DatatypeProperty<br>dbo:Person<br>birth date (en)<br>xsd:date<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
| <b>dbo:birthPlace</b> | rdf:type<br>rdfs:domain<br>rdfs:label<br>rdfs:range<br>Graph | rdf:Property; owl:ObjectProperty<br>dbo:Person<br>birth place (en)<br>dbo:Place<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                            |

[http://dbpedia.org/page/Barack\\_Obama](http://dbpedia.org/page/Barack_Obama)

| Propriedade / Recurso | Property                                                     | Value                                                                                                                                                                                                 |
|-----------------------|--------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>dbo:birthDate</b>  | rdf:type<br>rdfs:domain<br>rdfs:label<br>rdfs:range<br>Graph | owl:FunctionalProperty; rdf:Property;<br>owl:DatatypeProperty<br>dbo:Person<br>birth date (en)<br>xsd:date<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
| <b>dbo:birthPlace</b> | rdf:type<br>rdfs:domain<br>rdfs:label<br>rdfs:range<br>Graph | rdf:Property; owl:ObjectProperty<br>dbo:Person<br>birth place (en)<br>dbo:Place<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                            |
| <b>dbo:religion</b>   | rdf:type<br>rdfs:label<br>Graph                              | rdf:Property; owl:ObjectProperty<br>religion (en); religião (pt)<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                                           |
| <b>dbo:spouse</b>     | rdf:type<br>rdfs:domain<br>rdfs:label<br>rdfs:range<br>Graph | rdf:Property; owl:ObjectProperty<br>dbo:Person<br>spouse (en)<br>dbo:Person<br><a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                                |
| <b>dbp:birthName</b>  | rdf:type<br>rdfs:label<br>Graph                              | rdf:Property<br>birth name (en)<br><a href="http://dbpedia.org">http://dbpedia.org</a>                                                                                                                |

| <a href="http://dbpedia.org/page/Porto_Alegre">http://dbpedia.org/page/Porto_Alegre</a> |              |                                                                                                                 |
|-----------------------------------------------------------------------------------------|--------------|-----------------------------------------------------------------------------------------------------------------|
| Propriedade / Recurso                                                                   | Property     | Value                                                                                                           |
| <b>dbo:PopulatedPlace/area</b>                                                          | rdf:type     | owl:DatatypeProperty                                                                                            |
|                                                                                         | rdfs:domain  | dbo:PopulatedPlace                                                                                              |
|                                                                                         | rdfs:label   | área (km <sup>2</sup> )                                                                                         |
|                                                                                         | rdfs:range   | <a href="http://dbpedia.org/datatype/squareKilometre">http://dbpedia.org/datatype/squareKilometre</a>           |
|                                                                                         | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                         |
| <b>dbo:PopulatedPlace/areaTotal</b>                                                     | rdf:type     | owl:DatatypeProperty                                                                                            |
|                                                                                         | rdfs:domain  | dbo:PopulatedPlace                                                                                              |
|                                                                                         | rdfs:label   | área total (km <sup>2</sup> )                                                                                   |
|                                                                                         | rdfs:range   | <a href="http://dbpedia.org/datatype/squareKilometre">http://dbpedia.org/datatype/squareKilometre</a>           |
|                                                                                         | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                         |
| <b>dbo:PopulatedPlace/populationDensity</b>                                             | rdf:type     | owl:DatatypeProperty                                                                                            |
|                                                                                         | rdfs:domain  | dbo:PopulatedPlace                                                                                              |
|                                                                                         | rdfs:label   | population density (/sqkm) (en)                                                                                 |
|                                                                                         | rdfs:range   | <a href="http://dbpedia.org/datatype/inhabitantsPerSquare">http://dbpedia.org/datatype/inhabitantsPerSquare</a> |
|                                                                                         | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                         |
| <b>dbo:areaCode</b>                                                                     | rdf:type     | rdf:Property; owl:DatatypeProperty                                                                              |
|                                                                                         | rdfs:comment | Area code for telephone numbers. Use this not ph                                                                |
|                                                                                         | rdfs:label   | area code (en)                                                                                                  |
|                                                                                         | rdfs:range   | xsd:string                                                                                                      |
|                                                                                         | Graph        | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a>                         |
| <b>dbo:ontology/country</b>                                                             | rdf:type     | rdf:Property; owl:ObjectProperty                                                                                |
|                                                                                         | rdfs:comment | The country where the thing is located. (en)                                                                    |
|                                                                                         | rdfs:label   | country (en)                                                                                                    |
|                                                                                         | rdfs:range   | dbo:Country                                                                                                     |
|                                                                                         | rdf:type     | rdf:Property; owl:DatatypeProperty                                                                              |
| <b>dbo:elevation</b>                                                                    |              | altitude média acima do nível do mar (pt); averag                                                               |
|                                                                                         | rdfs:comment | above the sea level (en)                                                                                        |
|                                                                                         | rdfs:label   | altitude (μ) (pt); elevation (μ) (en)                                                                           |
|                                                                                         | rdfs:range   | xsd:double                                                                                                      |
|                                                                                         | rdf:type     | rdf:Property; owl:DatatypeProperty                                                                              |
| <b>dbo:foundingDate</b>                                                                 | rdfs:label   | founding date (en)                                                                                              |
|                                                                                         | rdfs:range   | xsd:date                                                                                                        |
|                                                                                         | rdf:type     | owl:FunctionalProperty; rdf:Property; owl:Datatyp                                                               |
| <b>dbo:ontology/populationTotal</b>                                                     | rdfs:label   | population total (en); população total (pt)                                                                     |
|                                                                                         | rdfs:range   | xsd:nonNegativeInteger                                                                                          |

<http://dbpedia.org/page/Tiger>

| Propriedade / Recurso    | Property    | Value                                                                                   |
|--------------------------|-------------|-----------------------------------------------------------------------------------------|
| <b>rdf:type</b>          | rdf:type    | owl:Thing; dbo:Eukaryote; umbel-rc:Mammal                                               |
| <b>rdfs:label</b>        | rdfs:label  | Tiger (en); Tigre (pt)                                                                  |
| <b>is dbo:species of</b> | rdf:type    | rdf:Property; owl:ObjectProperty                                                        |
|                          | rdfs:domain | dbo:Species                                                                             |
|                          | rdfs:label  | species (en)                                                                            |
|                          | Graph       | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
| <b>is dbp:animal of</b>  | rdf:type    | rdf:Property                                                                            |
|                          | rdfs:label  | Animal (en)                                                                             |
|                          | Graph       | <a href="http://dbpedia.org">http://dbpedia.org</a>                                     |

<http://dbpedia.org/page/Influenza>

| Propriedade / Recurso | Property   | Value                                                              |
|-----------------------|------------|--------------------------------------------------------------------|
| <b>rdf:type</b>       | rdf:type   | owl:Thing; wikidata:Q12136; dbo:Disease; umbel-rc:AilmentCondition |
|                       | rdfs:label | Influenza (en); Gripe (pt)                                         |

[http://dbpedia.org/page/World\\_War\\_II](http://dbpedia.org/page/World_War_II)

| Propriedade / Recurso | Property    | Value                                                                                   |
|-----------------------|-------------|-----------------------------------------------------------------------------------------|
| <b>dbo:causalties</b> | rdf:type    | rdf:Property; owl:DatatypeProperty                                                      |
|                       | rdfs:domain | dbo:MilitaryConflict                                                                    |
|                       | rdfs:label  | causalties (en)                                                                         |
|                       | rdfs:range  | xsd:string                                                                              |
| <b>dbo:place</b>      | Graph       | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
|                       | rdf:type    | rdf:Property; owl:ObjectProperty                                                        |
|                       | rdfs:domain | dbo:MilitaryConflict                                                                    |
|                       | rdfs:label  | place of military conflict (en)                                                         |
| <b>dbo:result</b>     | rdfs:range  | dbo:PopulatedPlace                                                                      |
|                       | Graph       | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
|                       | rdf:type    | rdf:Property; owl:DatatypeProperty                                                      |
|                       | rdfs:domain | dbo:MilitaryConflict                                                                    |
| <b>rdfs:label</b>     | rdfs:label  | result (en)                                                                             |
|                       | rdfs:range  | xsd:string                                                                              |
|                       | Graph       | <a href="http://dbpedia.org/resource/classes#">http://dbpedia.org/resource/classes#</a> |
|                       | rdf:type    | World War II (en); Segunda Guerra Mundial (pt)                                          |
|                       | Graph       | <a href="http://dbpedia.org">http://dbpedia.org</a>                                     |

## Anexo B

Listagem geral de resultados obtidos.

Os resultados abaixo foram gerados com uso de consultas para pessoas (“Bill Gates”, “Barack Obama”, “Neymar”), cidades (“Chicago”, “Porto alegre”) e outras entidades como animais (“lion”, “tiger”), doenças (“Influenza”) e eventos (“World War II”, “French Revolution”).

Para cada um dos tópicos identificados, na coluna da esquerda, a frase em vermelho representa um erro na geração da pergunta. A coluna à direita (com o label “Correta”) apresenta a frase correta.

| Bill Gates                                     | CORRETA                                   |
|------------------------------------------------|-------------------------------------------|
| What's Bill Gates' birth place?                | What's Bill Gates' birthplace?            |
| What's Bill Gates' birth date?                 | What's Bill Gates' birth date?            |
| What's Bill Gates' has abstract?               | What's Bill Gates' history?               |
| What's Bill Gates' active years start year?    | What are the Bill Gates' active years?    |
| What's Bill Gates' alma mater?                 | What's Bill Gates' alma mater?            |
| What's Bill Gates' birth name?                 | What's Bill Gates' birth name?            |
| What's Bill Gates' birth year?                 | What's Bill Gates' birth year?            |
| What's Bill Gates' board?                      | Is Bill Gates on the board?               |
| How much is Bill Gates' networth (\$)?         | How much is Bill Gates' net worth?        |
| Who is Bill Gates' parent?                     | Who are Bill Gates' parents?              |
| What's Bill Gates' person function?            | What's Bill Gates' function?              |
| What's Bill Gates' residence?                  | What's Bill Gates' place of residence?    |
| Who is Bill Gates' spouse?                     | Who is Bill Gates' spouse?                |
| What's Bill Gates' title?                      | What's Bill Gates' title?                 |
| What's Bill Gates' 1a?                         |                                           |
| What's Bill Gates' 1p?                         |                                           |
| What's Bill Gates' 1y?                         |                                           |
| What's Bill Gates' 2a?                         |                                           |
| What's Bill Gates' 2p?                         |                                           |
| What's Bill Gates' 2y?                         |                                           |
| Barack Obama                                   | CORRETA                                   |
| What's Barack Obama's birth place?             | What's Barack Obama's birthplace?         |
| What's Barack Obama's birth date?              | What's Barack Obama's birth date?         |
| What's Barack Obama's name?                    | What's Barack Obama's name?               |
| What's Barack Obama's has abstract?            | What's Barack Obama's history?            |
| What's Barack Obama's active years end date?   | What are the Barack Obama's active years? |
| What's Barack Obama's active years start date? | When did Barack Obama start working?      |
| What's Barack Obama's alma mater?              | What's Barack Obama's alma mater?         |
| What's Barack Obama's order in office?         | Is Barack Obama in the office?            |
| What's Barack Obama's residence?               | What's Barack Obama's place of            |

What's Barack Obama's seniority?  
 Who is Barack Obama's spouse?  
 What's Barack Obama's birth name?  
 What's Barack Obama's book?  
 What's Barack Obama's by?  
 What's Barack Obama's children?  
 What's Barack Obama's commons?  
 What's Barack Obama's congbio?  
 What's Barack Obama's d?  
 What's Barack Obama's district?  
 What's Barack Obama's fec?

residence?

Who is Barack Obama's spouse?

What's Barack Obama's book?

Who are Barack Obama's children?

What's Barack Obama's history?

What's Barack Obama's district?

### Neymar

What's Neymar's birth place?  
 What's Neymar's birth date?  
 What's Neymar's height (cm)?  
 What's Neymar's has abstract?  
 What's Neymar's career station?  
 How much is Neymar's height (μ)?  
 What's Neymar's number?  
 What's Neymar's align?  
 What's Neymar's bg?  
 What's Neymar's bordercolor?  
 What's Neymar's Caption?  
 What's Neymar's fg?  
 What's Neymar's image size?  
 What's Neymar's nationalcaps?  
 What's Neymar's nationalgoals?  
 What's Neymar's nationalteam?  
 What's Neymar's nationalyears?  
 What's Neymar's quote?  
 What's Neymar's source ?  
 What's Neymar's title?

CORRETA

What's Neymar's birthplace?

What's Neymar's birth date?

What's Neymar's height?

What's Neymar's history?

What's Neymar's number?

What are Neymar's goals?

What's Neymar's national team?

What's Neymar's quote?

What titles does Neymar's have?

### Chicago

What's Chicago's area metro (km2)?  
 What's Chicago's area total (km2)?  
 What's Chicago's area urban (km2)?  
 What's Chicago's population density (/sqkm)?  
 What's Chicago's has abstract?  
 What's Chicago's area code?  
 How much is Chicago's area land (m2)?  
 How much is Chicago's area metro (m2)?  
 How much is Chicago's area total (m2)?  
 How much is Chicago's area urban (m2)?

CORRETA

What's Chicago's area?

What's Chicago's area code?

How much is Chicago's area water (m2)?

How much is Chicago's elevation (μ)?

What's Chicago's founding date?

What's Chicago's governing body?

Who is Chicago's leader name?

What's Chicago's leader title?

How much is Chicago's maximum elevation (μ)?

How much is Chicago's minimum elevation (μ)?

What's Chicago's motto?

What's Chicago's percentage of area water?

What's Chicago's population as of?

How much is Chicago's population density (/sqkm)?

What's Chicago's population metro?

What's Chicago's population total?

What's Chicago's total population ranking?

What's Chicago's postal code?

What's Chicago's time zone?

What's Chicago's UTC offset?

What's Chicago's align?

What's Chicago's area code type?

### Porto Alegre

CORRETA

What's Porto Alegre's area (km2)?

What's Porto Alegre's area?

What's Porto Alegre's area total (km2)?

What's Porto Alegre's population density (/sqkm)?

What's Porto Alegre's has abstract?

How much is Porto Alegre's area (m2)?

What's Porto Alegre's area code?

How much is Porto Alegre's area total (m2)?

How much is Porto Alegre's elevation (μ)?

What's Porto Alegre's founding date?

Who is Porto Alegre's leader name?

What's Porto Alegre's leader title?

What's Porto Alegre's motto?

How much is Porto Alegre's population density (/sqkm)?

What's Porto Alegre's population metro?

What's Porto Alegre's population total?

What's Porto Alegre's postal code?

What's Porto Alegre's time zone?

What's Porto Alegre's UTC offset?

What's Porto Alegre's Apr high C?

What's Porto Alegre's Apr humidity?

What's Porto Alegre's Apr low C?

What's Porto Alegre's Apr mean C?

What's Porto Alegre's Apr precipitation days?

What's Porto Alegre's Apr precipitation mm?

What's Porto Alegre's Apr record high C?  
 What's Porto Alegre's Apr record low C?  
 What's Porto Alegre's Apr sun?  
 What's Porto Alegre's Aug high C?  
 What's Porto Alegre's Aug humidity?  
 What's Porto Alegre's Aug low C?

### Lion

CORRETA

Who is Lion?

What's Lion's name?  
 What's Lion's has abstract?  
 What's Lion's image?  
 What's Lion's image2 caption?  
 What's Lion's image2 width?  
 What's Lion's image caption?  
 What's Lion's image width?  
 What's Lion's range map?  
 What's Lion's range map2 caption?  
 What's Lion's range map2 width?  
 What's Lion's range map width?  
 What's Lion's status?  
 What's Lion's status system?  
 What's Lion's subdivision ranks?  
 What's Lion's synonyms?  
 What's Lion's taxon?  
 What's Lion's trend?

### Tiger

CORRETA

What's Tiger's history?

What's Tiger's has abstract?  
 What's Tiger's align?  
 What's Tiger's c?  
 What's Tiger's Caption?  
 What's Tiger's date?  
 What's Tiger's direction?  
 What's Tiger's image?  
 What's Tiger's image caption?  
 What's Tiger's p?  
 What's Tiger's range map?  
 What's Tiger's range map caption?  
 What's Tiger's status?  
 What's Tiger's status system?  
 What's Tiger's subdivision?  
 What's Tiger's subdivision ranks?  
 What's Tiger's synonyms?  
 What's Tiger's taxon?  
 What's Tiger's title?  
 What's Tiger's trend?

What's Tiger's url?  
 What's Tiger's width?

### Influenza

CORRETA

What's Influenza's has abstract?  
 What's Influenza's ICD10?  
 What's Influenza's ICD9?  
 What's Influenza's MeSH ID?  
 What's Influenza's OMIM id?  
 What's Influenza's Caption?  
 What's Influenza's DiseasesDB?  
 What's Influenza's eMedicineSubj?  
 What's Influenza's eMedicineTopic?  
 What's Influenza's field?  
 What's Influenza's MedlinePlus?  
 What's Influenza's width?

What's Influenza's history?  
 What is ICD10?

### World War II

CORRETA

What's World War II's has abstract?  
 What's World War II's causalities?  
 What's World War II's combatant?  
 What's World War II's place of military conflict?  
 What's World War II's result?  
 What's World War II's combatants header?  
 What's World War II's b?  
 What's World War II's book?  
 What's World War II's Caption?  
 What's World War II's casualties?  
 What's World War II's commons?  
 What's World War II's d?  
 What's World War II's date?  
 What's World War II's n?  
 What's World War II's portal?  
 What's World War II's q?  
 What's World War II's s?  
 What's World War II's v?  
 What's World War II's wikt?

What's the World War II history?  
 What were the WWII effects?  
 Who were the WWII soldiers?  
 Where did the WWI take place?

### French Revolution

CORRETA

What's French Revolution's has abstract?  
 What's French Revolution's date?  
 What's French Revolution's Event Name?  
 What's French Revolution's image caption?  
 What's French Revolution's image name?  
 What's French Revolution's label?  
 What's French Revolution's location?

What's the French Revolution history?

What's French Revolution's onlinebooks?

What's French Revolution's participants?

What's French Revolution's result?

What's French Revolution's title?

## Anexo C

Listagem geral de resultados e avaliação, contendo a indicação de alternativas para geração das frases.

| Bill Gates                                  | CORRETA                                | ALTERNATIVA 1                      | ALTERNATIVA 2                        | ALTERNATIVA 3                     | ALTERNATIVA 4                     |
|---------------------------------------------|----------------------------------------|------------------------------------|--------------------------------------|-----------------------------------|-----------------------------------|
| What's Bill Gates' birth place?             | What's Bill Gates' birthplace?         | What's Bill Gates' place of birth? | Where was Bill Gates born?           | Where is Bill Gates from?         |                                   |
| What's Bill Gates' birth date?              | What's Bill Gates' birth date?         | When is Bill Gates' birthday?      | When was Bill Gates born?            |                                   |                                   |
| What's Bill Gates' has abstract?            | What's Bill Gates' history?            | What's Bill Gates' biography?      |                                      |                                   |                                   |
| What's Bill Gates' active years start year? | What are the Bill Gates' active years? | When did Bill Gates start working? | When did Bill Gates stop working?    |                                   |                                   |
| What's Bill Gates' alma mater?              | What's Bill Gates' alma mater?         | Where did Bill Gates go to school? | Which school did Bill Gates attend?  |                                   |                                   |
| What's Bill Gates' birth name?              | What's Bill Gates' birth name?         | What's Bill Gates' real name?      | What's Bill Gates' middle name?      | What's Bill Gates' surname?       | What's Bill Gates' lastname?      |
| What's Bill Gates' birth year?              | What's Bill Gates' birth year?         | What's Bill Gates' year of birth?  |                                      |                                   |                                   |
| What's Bill Gates' board?                   | Is Bill Gates on the board?            | Is Bill Gates off the board?       |                                      |                                   |                                   |
| How much is Bill Gates' networth (\$)?      | How much is Bill Gates' net worth?     | How much is Bill Gates worth?      | How much does Bill Gates worth?      |                                   |                                   |
|                                             |                                        |                                    |                                      | What's Bill Gates' mother's name? | What's Bill Gates' father's name? |
| Who is Bill Gates' parent?                  | Who are Bill Gates' parents?           | Who is Bill Gates' mother?         | Who is Bill Gates' father?           | What's Bill Gates' task?          | What's Bill Gates' work?          |
| What's Bill Gates' person function?         | What's Bill Gates' function?           | What's Bill Gates' spot?           | What's Bill Gates' job?              |                                   |                                   |
| What's Bill Gates' residence?               | What's Bill Gates' place of residence? | Where does Bill Gates live?        | Where is Bill Gates' house located?  |                                   |                                   |
| Who is Bill Gates' spouse?                  | Who is Bill Gates' spouse?             | Who is Bill Gates' wife?           | What's the name of Bill Gates' wife? |                                   |                                   |
| What's Bill Gates' title?                   | What's Bill Gates' title?              | Does Bill Gates hold               | Does Bill Gates                      |                                   |                                   |

a title?

have a title?

| Barack Obama                                   | CORRETA                                   | ALTERNATIVA 1                           | ALTERNATIVA 2                           | ALTERNATIVA 3                  | ALTERNATIVA 4                   |
|------------------------------------------------|-------------------------------------------|-----------------------------------------|-----------------------------------------|--------------------------------|---------------------------------|
| What's Barack Obama's birth place?             | What's Barack Obama's birthplace?         | What's Barack Obama's place of birth?   | Where was Barack Obama born?            | Where is Barack Obama from?    |                                 |
| What's Barack Obama's birth date?              | What's Barack Obama's birth date?         | When is Barack Obama's birthday?        | When was Barack Obama born?             |                                |                                 |
| What's Barack Obama's name?                    | What's Barack Obama's name?               | What's Barack Obama's real name?        | What's Barack Obama's middle name?      | What's Barack Obama's surname? | What's Barack Obama's lastname? |
| What's Barack Obama's has abstract?            | What's Barack Obama's history?            | What's Barack Obama's biography?        |                                         |                                |                                 |
| What's Barack Obama's active years end date?   | What are the Barack Obama's active years? | When did Barack Obama stop working?     |                                         |                                |                                 |
| What's Barack Obama's active years start date? | When did Barack Obama start working?      |                                         |                                         |                                |                                 |
| What's Barack Obama's alma mater?              | What's Barack Obama's alma mater?         | Where did Barack Obama go to school?    | Which school did Barack Obama attend?   |                                |                                 |
| What's Barack Obama's order in office?         | Is Barack Obama in the office?            | Is Barack Obama off the office?         |                                         |                                |                                 |
| What's Barack Obama's residence?               | What's Barack Obama's place of residence? | Where does Barack Obama live?           | Where is Barack Obama's house located?  |                                |                                 |
| What's Barack Obama's seniority?               |                                           |                                         |                                         |                                |                                 |
| Who is Barack Obama's spouse?                  | Who is Barack Obama's spouse?             | Who is Barack Obama's wife?             | What's the name of Barack Obama's wife? |                                |                                 |
| What's Barack Obama's birth name?              |                                           |                                         |                                         |                                |                                 |
| What's Barack Obama's book?                    | What's Barack Obama's book?               | What's the name of Barack Obama's book? | Which book is Barack Obama's?           |                                |                                 |
| What's Barack Obama's by?                      |                                           |                                         |                                         |                                |                                 |

|                                 |                                  |                                             |                                                 |                                   |                                  |
|---------------------------------|----------------------------------|---------------------------------------------|-------------------------------------------------|-----------------------------------|----------------------------------|
| What's Barack Obama's children? | Who are Barack Obama's children? | What's the name of Barack Obama's daughter? | What are the names of Barack Obama's daughters? | Who are Barack Obama's kids?      | Does Barack Obama have children? |
| What's Barack Obama's commons?  |                                  |                                             |                                                 |                                   |                                  |
| What's Barack Obama's congbio?  | What's Barack Obama's history?   | What's Barack Obama's biography?            |                                                 |                                   |                                  |
| What's Barack Obama's d?        |                                  |                                             |                                                 |                                   |                                  |
| What's Barack Obama's district? | What's Barack Obama's district?  | Where does Barack Obama govern?             | Where does Barack Obama work?                   | Which is Barack Obama's district? |                                  |

## Neymar

|                                  |                                        |                                                  |                                         |                                        |
|----------------------------------|----------------------------------------|--------------------------------------------------|-----------------------------------------|----------------------------------------|
| What's Neymar's birth place?     | CORRETA<br>What's Neymar's birthplace? | ALTERNATIVA 1<br>What's Neymar's place of birth? | ALTERNATIVA 2<br>Where was Neymar born? | ALTERNATIVA 3<br>Where is Neymar from? |
| What's Neymar's birth date?      | What's Neymar's birth date?            | When is Neymar's birthday?                       | When was Neymar born?                   |                                        |
| What's Neymar's height (cm)?     | What's Neymar's height?                | How tall is Neymar?                              | Do you know Neymar's height?            |                                        |
| What's Neymar's has abstract?    | What's Neymar's history?               | What's Neymar's biography?                       |                                         |                                        |
| What's Neymar's career station?  |                                        |                                                  |                                         |                                        |
| How much is Neymar's height (μ)? |                                        |                                                  |                                         |                                        |
| What's Neymar's number?          | What's Neymar's number?                | What's Neymar position?                          | What's Neymar's jersey number?          | What's Neymar's shirt number?          |