



Associação Propagadora Esdeva
Centro de Ensino Superior de Juiz de Fora – CES/JF
Curso de Bacharelado em Sistemas de Informação
Trabalho de Conclusão de Curso – Artigo

CONSULTA EM BASE DE CONHECIMENTO UTILIZANDO LINGUAGEM NATURAL

Cristiano Paschoalim de Almeida¹

Centro de Ensino Superior de Juiz de Fora, Juiz de Fora, MG

Daves Márcio Silva Martins²

Centro de Ensino Superior de Juiz de Fora, Juiz de Fora, MG

RESUMO

Com o crescimento do volume de dados disponíveis na *web* e estando cada vez mais automáticos e ágeis os meios que os criam, é fundamental a organização desses dados e a sua interpretação para obtenção de respostas rápidas e confiáveis, tornando as pesquisas mais precisas e completas, para que as pessoas tomem melhores decisões e criem novos conceitos. Dando foco a essa necessidade, este trabalho tem como objetivo explorar, com base teórica, a evolução das pesquisas em computação cognitiva, em que procuramos uma maior interação entre homem e máquina, com técnicas de reconhecimento da linguagem natural e recuperação da informação, visto que a comunicação e a recuperação das informações são as maiores dificuldades encontradas. Com esse propósito, daremos início ao processo de desenvolvimento de um modelo de aplicação pensando na próxima geração de sistemas inteligentes, capazes de compreender a fala e responder perguntas que fazem referência a conhecimentos previamente definidos em um banco de dados orientado a grafos. Ao longo da pesquisa, viu-se que a reestruturação da *web* não se mostrou ser uma tarefa fácil, tampouco rápida.

Palavras-chave: Computação cognitiva. Processamento de linguagem natural. Recuperação de informação. Banco de dados orientado a grafo.

¹ Discente do Curso de Bacharelado em Sistemas de Informação do Centro de Ensino Superior de Juiz de Fora – CES/JF. Endereço: Rua Halfeld 1.179 – 36.016-000 – Juiz de Fora – MG - Brasil. E-mail: cristiano_pascholaim@hotmail.com

² Docente do Curso de Bacharelado em Sistemas de Informação do Centro de Ensino Superior de Juiz de Fora. Orientador.

1 INTRODUÇÃO

O anseio do ser humano em criar uma inteligência artificial vem crescendo muito ao longo dos anos. Desde Turing, em 1940, e sua primeira máquina criada por sua equipe, com o propósito de decifrar mensagens alemãs durante a Segunda Guerra Mundial até os dias de hoje, quando tentamos simular o funcionamento do cérebro humano, houve um salto tecnológico enorme, passando por válvulas eletrônicas e cartões perfurados até microchips e as linguagens de programação de alto nível. Graças a esse avanço, ensaiamos hoje como seria a automação de processos que, até o momento, só era realizada pelos seres humanos (RUSSELL, 2013).

Segundo Russell (2013), a linguística moderna e a Inteligência Artificial (IA) nasceram “juntas” e a linguística provou ser mais complexa quando *Skinner* publicou, em *Verbal Behavior* (1957), um artigo falando sobre sua formulação, dando foco à gramática e desprezando o conceito para a formulação, o qual veio a ganhar destaque após uma crítica feita por Noam Chomsky, que atentou para o fato de o referido artigo não tratar da noção de criatividade na linguagem. O entendimento da linguagem não está apenas vinculado à compreensão da estrutura da frase, mas sim à compreensão do assunto e do contexto.

Hoje, possuímos um vasto acervo de informações disponíveis na *web*. O primeiro passo foi dado com os buscadores *Excite*, *Yahoo* e *Altavista* e seus modelos de indexação de páginas da internet por palavras-chave (BAEZA-YATES; NETO, 1999), devolvendo rapidamente uma lista de *links* relacionados à pesquisa efetuada. Mesmo com esses mecanismos, algumas vezes o usuário recebe uma lista muito extensa ou que não tem vínculo com o que está buscando, justamente pelo fato de uma máquina não entender o contexto que se procura (GONÇALVES, 2003).

São exploradas duas vertentes para superar esse problema: a primeira visa criar padrões semânticos para disponibilizar os dados na *web*, de forma que estes sejam compreensíveis pelos computadores; a segunda busca criar sistemas dotados de inteligência e autonomia, que possam “varrer” a internet, aprendendo informações úteis (OATES; PRASAD; LESSER, 1994).

Com o intuito de auxiliar na criação de modelos de agentes inteligentes que sejam capazes de tornar as estruturas já existentes compreensíveis para o computador, visto que sua reestruturação dos conteúdos disponíveis na *web* não se mostrou uma tarefa fácil, este trabalho terá como objetivo estudar e desenvolver um sistema que vincule a análise sintática a um banco de dados orientado a grafos, que possa responder questionamentos efetuados em linguagem natural controlada.

A aplicação será desenvolvida utilizando técnicas de reconhecimento de voz hoje nativos nos sistemas operacionais (*Android* e *Windows*), ferramenta disponível na biblioteca Apache Open NLP para identificação e classificação das palavras. Para armazenamento, utilizaremos o Neo4J como banco de dados.

No capítulo 2 falaremos brevemente sobre a comunicação, suas fases e como são classificadas (neste trabalho, abordaremos somente a consulta). No capítulo 3, trataremos da análise sintática para identificação dos sujeitos e ações presentes em uma frase. No capítulo 4, será apresentado o conceito de ligação entre os dados com os modelos empregados atualmente como o *Resource Description Framework [RDF]*, assim como a teoria dos grafos e o uso para orientar banco de dados. No capítulo 5 serão apresentados os passos adotados para o desenvolvimento do projeto, usando os conceitos adquiridos ao longo do artigo.

2 COMUNICAÇÃO

O que nos difere dos demais animais é o ato de conseguirmos nos comunicar e armazenar o conhecimento atemporal em linguagens tanto audiovisuais, quanto escritas. A comunicação é o cerne do desenvolvimento de nossa espécie e o que nos difere de outros mamíferos, como golfinhos e chimpanzés, que também possuem um vocabulário com milhares de símbolos. Apesar de outras espécies mostrarem habilidades comunicacionais semelhantes, somos os únicos a realizá-las e construir, com extrema maestria, novos conceitos e tecnologias (Russell, Stuart J., 2013).

As fases que compõem a comunicação são:

Intenção: a vontade de expressar uma ideia para um ou mais receptores; e

Geração: formulação de como será transmitida essa informação ou essa vontade.

Segundo Chomsky (1957), os diálogos podem se dividir em quatro classes de formalização gramatical, que se diferem apenas na forma de se escrever. São elas:

- Consultar: quando efetuamos uma pergunta, a fim de obtermos uma resposta ou informação que precisamos saber.
- Informar: quando se expressa uma ideia ou conhecimento para outras pessoas ou seres.
- Solicitar: quando pedimos ou solicitamos algo ou alguma informação.
- Confirmar: perante uma fala ou execução de um comando, onde se apresenta um sinal de que a informação está certa ou errada, ou se é verdadeira ou falsa.

2.1 Linguagem Natural Controlada

Quando falamos em linguagem natural, consideramos qualquer forma de comunicação entre um emissor e um receptor em que eles possam manter um diálogo por sons, símbolos e gestos, entre outras maneiras. A Linguagem Natural Controlada diz respeito a uma forma de comunicação pré-definida, como som ou escrita e com termos que serão usados para evitar gírias e ambigüidades (FERREIRA, 2013).

3 SINTAGMAS DA GRAMÁTICA

De acordo com Russell e Norvig (2013), a maioria dos formalismos das regras gramaticais baseia-se na ideia de estrutura sintagmática, de que as cadeias são compostas de subcadeias chamadas sintagmas.

- Sentença(S): frase que expressa um pensamento de ordem geral e de valor moral.

- *Sintagma Nominal (SN)*: é um grupo de palavras que tem como base ou núcleo um substantivo ou termo equivalente.

Exemplo: As verdadeiras amizades nunca morrem. SN = As verdadeiras amizades, núcleo amizades.

- *Sintagma Verbal (SV)*: é um grupo de palavras que tem como base ou como núcleo um verbo.

Exemplo: As margaridas murcharam. SV=murcharam, núcleo = murcharam.

Sintaxe:

[S -> SN SV]

(S: [SN: {Artigo: As}{Nome: margaridas}] [SV:{Verbo: murcharam}]).

Onde dentro de S: (Sentença) possuímos SN: [*Sintagmas* Nominais] e SV: [*Sintagmas* Verbais]

3.1 Análise sintática (*Parsing*)

Segundo Russell e Norvig (2013), “análise sintática é o processo que analisa uma cadeia de palavras para definir uma estrutura frasal, obedecendo as regras gramaticais”.

A mesma pode ser representada de duas formas: de cima para baixo, começando da raiz (S) até as folhas (palavras); ou o inverso, de baixo para cima, começando das folhas (palavras) e terminando na raiz (S). Contudo, vale destacar que as duas abordagens talvez venham a ser ineficientes, pois podem nos levar a becos sem saídas.

O quadro a seguir ilustra como efetuar a análise sintática da raiz (S) até as folhas (palavras).

Frase a ser analisada: *O wumpus está morto.*

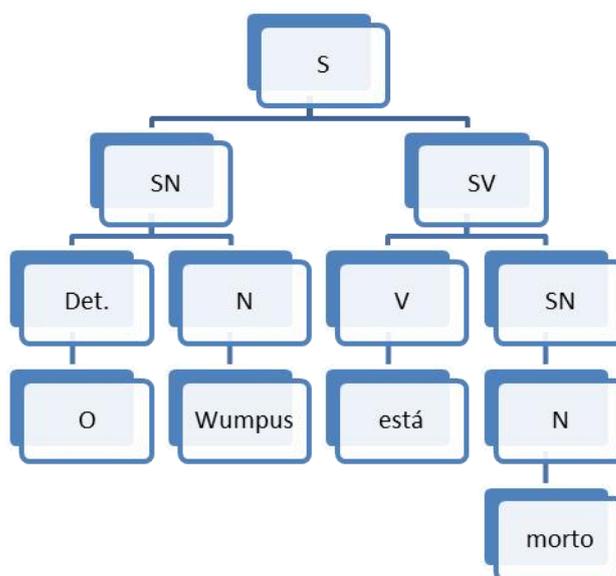
QUADRO 1 - Representação detalhada da análise sintática

<i>Lista de itens</i>	<i>Regra</i>
S	
SN SV	$S \rightarrow SN SV$
SN SV Adjetivo	$SV \rightarrow SV Adjetivo$
SN Verbo Adjetivo	$SV \rightarrow Verbo$
SN Verbo morto	$Adjetivo \rightarrow \mathbf{morto}$
SN está morto	$verbo \rightarrow \mathbf{está}$
Artigo Substantivo está morto	$SN \rightarrow Artigo Substantivo$
Artigo wumpus está morto	$Substantivo \rightarrow \mathbf{wumpus}$
owumpus está morto	$Artigo \rightarrow \mathbf{o}$

Fonte: RUSSELL (2013)

Segue a representação em forma de árvore da análise sintática realizada no quadro 1.

Figura 1: Representação da árvore da análise sintática



Fonte: próprio autor

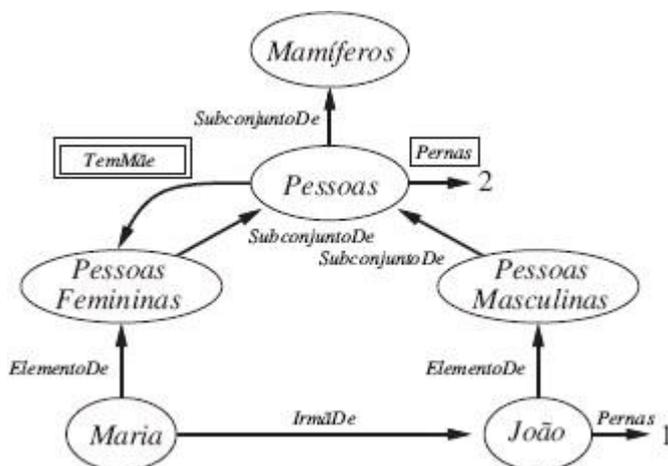
3.2 Ligações entre os dados

A web como conhecemos hoje visa disponibilizar acesso a vários documentos de pesquisa, independentemente de sua plataforma de criação. A ideia de ligar dados teve início com o surgimento do *Hiperlink*, cujo objetivo era poder ligar parte de um texto com seus respectivos conceitos. Como bem explica Tim Berners-Lee em sua palestra, “(...) um dado isolado sem contexto não traz qualquer informação, mais um grupo de dados ligados podem nos dizer muito mais que uma só ‘coisa’ (...)” (BERNERS-LEE, 2009).

Charles S. Peirce propôs, em 1909, uma notação em forma gráfica para representar o conhecimento onde estabeleceu-se arcos que rotulou como “Lógica do futuro”, sendo capaz de representar objetos individuais, categorias e relação entre eles. Uma notação gráfica exhibe nomes de objetos ou categorias em elipses ou retângulos e os conecta por meio de arcos rotulados (RUSSELL, 2013).

Na figura 2 a seguir, temos um arco *ElementoDe* entre Maria e Pessoas Femininas, onde há a asserção lógica $\text{Maria} \in \text{Pessoas Femininas}$, assim como o arco *IrmãDe* entre Maria e João que corresponde à asserção $\text{IrmãDe}(\text{Maria}, \text{João})$, mas não podemos traçar uma relação *TemMãe* entre Pessoas e Pessoas Femininas, pois essa relação seria entre pessoa e sua mãe, e não possuímos mães como categoria.

Figura 2: Representação gráfica da relação entre Maria e João



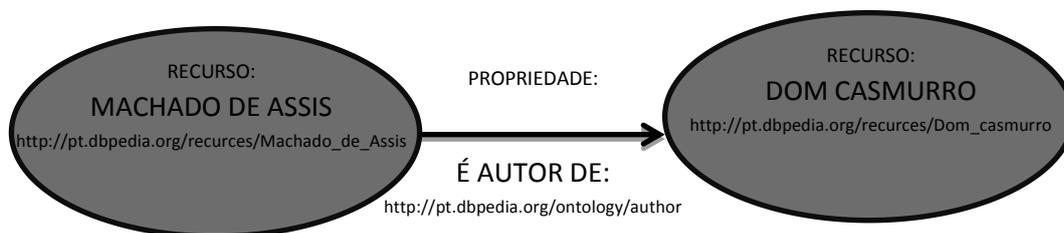
Fonte: RUSSEL (2013)

No intuito de se criar modelos representativos que associem os nomes aos objetos e que representem um universo definido, foram criadas as ontologias, que são dicionários de vocabulários, como o OWL (*Ontology Web Language*), uma lógica descritiva que pode ser aplicada nos mecanismos RDF (BRICKLEY, D.; GUHA, R. V., 2014).

Segundo HEATH; BIZER (2011), o mecanismo *Resource Description Framework (RDF)* é um modelo de dados em grafo que permite estabelecer relação entre quaisquer coisas publicadas na web que possam ser descritas, como locais, pessoas, objetos etc. Em essência, são *links* que conectam as coisas ao mundo. Esse mecanismo baseia-se em triplas Recurso-Propriedade-Valor, onde cada um desses itens é representado por um *Uniform Resource Identifier (URI)*, que nada mais é que um *link* que liga a web à “coisa” procurada com suas respectivas características, tornando-a, assim, única e de acesso direto por pessoas ou máquinas (FREITAS; AZEVEDO, 2016).

Na figura 3, vemos a representação no modelo *Resource Description Framework (RDF)*, onde Machado de Assis e Dom Casmurro são recursos e Author é a propriedade que os relaciona.

Figura 3: Representação do modelo *Resource Description Framework (RDF)*



Fonte: FREITAS; AZEVEDO (2016)

Russell (2013) cita o autor Khare (2006) ao afirmar que o uso dos *Resource Description Framework (RDF)* com as ontologias está inversamente proporcional a sua complexidade, visto que 99% de todo conteúdo disponível na web está em formatos tradicionais, como *Hypertext Markup Language (HTML)* e *Cascading Style*

Sheets (CSS), possuindo textos livres e seguindo esquemas mais simples de representação.

3.3 Teoria e aplicação dos grafos

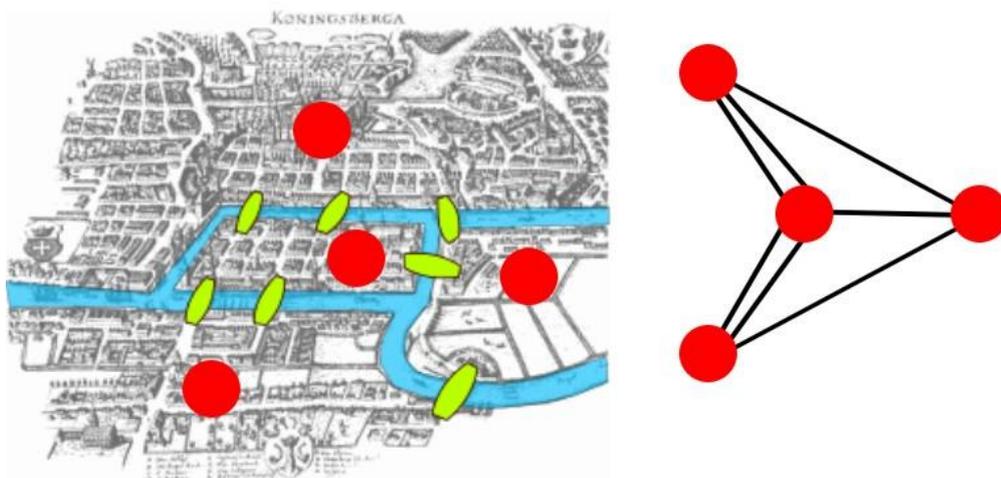
Conforme Penteado (2014), apesar da ampla utilização dos modelos dos bancos de dados relacionais pelos meios empresariais, os mesmos podem ser ineficientes para grandes volumes de dados e representação de aplicações complexas e específicas.

Uma das alternativas adotadas pelas redes sociais como o *Twitter* foi a utilização de base de dados orientados a grafo, onde os usuários são vértices e as ligações entre eles são arestas (PENTEADO, 2014).

Leonhard Euler (1707-1783) escreveu seu primeiro artigo relacionado a grafos, o famoso enigma hoje conhecido como o “Problema das sete pontes de Königsberg”, que consistia em determinar um caminho passando pelas sete pontes sem atravessar duas vezes a mesma ponte, o que ele resolveu, e criou um modelo geral para resolver problemas deste tipo (POSSANI, 2011).

Na figura 4 temos a representação do problema das pontes de Königsberg e a solução gráfica encontrada para o problema.

Figura 4: Representação simplificada das pontes de Königsberg



Fonte: <https://universoracionalista.org/origem-da-teoria-dos-grafos-as-7-pontes-de-konigsberg/>

Existem inúmeras formas de representação de um grafo. Este artigo se aterá à formação básica, em que um grafo é composto por $G(V,E)$, onde:

- V é conjunto finito e não vazio de vértices; e
- E de Arestas, formadas por pares não ordenados de elementos de V .

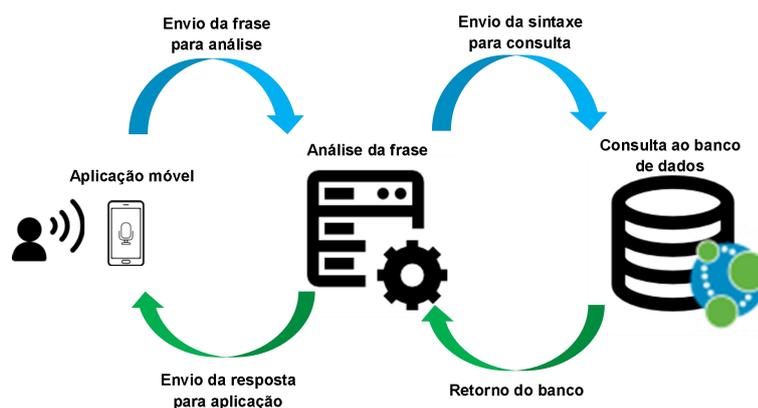
Graficamente, os vértices são representados por nós (pontos), que são unidos ou não por arestas (traços), indicando, assim, a ligação entre eles.

Segundo OLIVEIRA, F. R. (2015); apud SADALAGE, P. J.; FOWLER, M. (2012), o uso de banco de dados NoSQL (que não utilizam modelos de relacionamento com a linguagem *Structured Query Language* – SQL - empregados nos bancos de dados relacionais) tem crescido muito com os grandes volumes de dados e o aumento da complexidade das aplicações, que representam uma dificuldade ainda maior para os desenvolvedores por se tratar de bancos que possuem uma lógica específica para cada modelo de dados (PRIGOL, Eliézer, 2016).

4 DESENVOLVIMENTO DA APLICAÇÃO

Extraíndo conhecimentos obtidos ao longo das pesquisas, desenvolvemos um modelo de aplicação pensando na próxima geração de sistemas inteligentes, capazes de compreender a fala utilizando uma biblioteca nativa do *Android* e *Windows*, que transcreverá a fala em texto. Posteriormente, será efetuada uma análise sintática vinculada a uma base de conhecimento por um servidor remoto, que dará uma resposta inerente a perguntas realizadas pelo usuário.

Figura 5: Representação do fluxo de execução



Fonte: Próprio autor

Procuramos uma maneira concisa e simples de representar e desenvolver, abrindo, assim, portas para futuras pesquisas no âmbito da interpretação da linguagem natural, da extração do conhecimento e armazenamento do mesmo para consultas posteriores em linguagem natural.

4.1 Reconhecimento da fala e resposta áudio visual

Trata-se de bibliotecas nativas do *Windows* e *Android* (principais sistemas operacionais utilizados) que permitem a percepção da fala em diversos idiomas, transformando-a em texto que pode ser usado como comandos ou até mesmo para transcrever diálogos. Usaremos essas ferramentas para reconhecer as perguntas pela fala e fornecer respostas audiovisuais, o que é essencial para casos de deficiência visual.

4.2 Classificação gramatical

Utilizamos a biblioteca *Apache Open NLP*, que possui diversas ferramentas, desde separação de sentenças até destaque dos *tokens* e classificação gramatical das palavras. Em alguns idiomas, como o inglês, a biblioteca dispõe de algoritmos de análise sintática, conforme mencionado no início do artigo. Como em português não dispomos de tal funcionalidade, neste trabalho usaremos somente a ferramenta *pt-pos-maxent.bin*, que nos informa a classificação gramatical das palavras na frase, como podemos ver no exemplo a seguir:

Exemplo 01:

Frase: Quem e filho de Maria?

Resultado: Quem[pron-indp] e[conj-c] filho[n] de[prp]

Maria[prop] [?:punc]

4.3 Classificação das perguntas

A classificação das perguntas é fundamental para se determinar a resposta, uma vez que por meio dessa classificação é definido o tipo de resposta esperada, como valor, local e data, entre outros. O mesmo pode ser feito por algoritmos de satisfação ou somente por consulta a uma base de relacionamento (TAHRI; TIBERMACHINE, 2013), como mostra o exemplo abaixo:

QUADRO 2 - Classificação das perguntas

Termo	Resposta	Definição
Quem	Nome	Pessoa, Ator, Autor, Usuário
Onde	nome, substantivo	Local
Como	adjetivo, verbo,	Estado, Definição

Fonte: próprio autor

Exemplo:

Frase: Quem e filho de Maria?

Resultado: **Pessoa.nome** e[conj-c] filho[n] de[prp]

Maria[prop] [?:punc]

Vale ressaltar que em alguns casos, como na frase “qual a idade do filho de Maria?”, a mesma já informa a resposta esperada, que é a propriedade ou característica idade.

4.4 Identificação de recursos e propriedades

Podemos dizer que seria uma consulta a um dicionário, no qual usaremos uma tabela com palavras e suas classificações e o que elas representam no contexto, como Pessoa, Carro, Cachorro, entre outra classificação que podemos adotar. Isso é fundamental para posteriormente montarmos a sintaxe de consulta ao banco de dados.

No quadro a seguir, exemplo para consulta das palavras para rotular entidades e relacionamentos.

QUADRO 3 - Dicionário de palavras

Palavra	Classificação Gramatical	Rótulo
---------	--------------------------	--------

João	Prop	Pessoa.nome
Maria	Prop	Pessoa.nome
CESJF	Prop	Faculdade.nome, Local.nome
filho	N	Filho

Fonte: próprio autor

Exemplo:

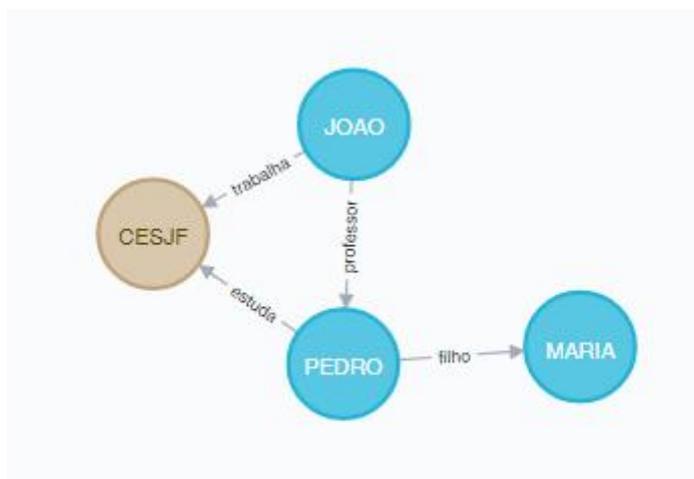
Frase: Quem é filho de Maria?

Resultado: Pessoa.nome e[conj-c] filho de[prp]Pessoa.nome:“Maria” [?:punc]

4.5 Consulta ao banco de dados

Usando um banco de dados orientado por grafos (Neo4J), simulamos o conceito de trípliques RDF (BRICKLEY; GUHA, 2014) para armazenar os conhecimentos, sendo os sujeitos ou objetos nossos vértices e as nossas ações ou propriedades nossas arestas, como podemos ver na imagem a seguir.

Figura 6: Representação dos dados no SGBDG



Fonte: Próprio autor

O neo4j utiliza, para consulta, a linguagem *Cypher* para manipulação dos dados. Neste artigo, como trataremos da consulta no banco de dados, exploraremos a sintaxe *MATCH*, como podemos ver no exemplo a seguir.

Exemplo:

`MATCH (B.Faculdade)<-[:estuda]-(F.Pessoa{nome:'PEDRO'}) return B.nome`

4.6 Adaptando a frase à sintaxe do banco

Como pode ser percebido, o resultado do tópico 4.4 supracitado, devem ser adaptado para tornar-se a sintaxe de consulta ao banco, removendo, assim, todos os artigos, preposições e pontuações e adicionando letras para representar os recursos e propriedade. Logo depois, devem ser reorganizados, colocando o que queremos saber no *return*.

Exemplo:

Pessoa.nome e[conj-c] filho de[prp] Pessoa.nome: "Maria" [?:punc]

Pessoa.nome - filho – Pessoa.nome: "Maria"

MATCH (a:Pessoa)-[b:filho]-(c:Pessoa{nome: "Maria"}) retorna.nome

A imagem a seguir mostra a interface, onde pode se efetuar consulta por voz ou escrita, com o respectivo resultado apresentado pela aplicação.

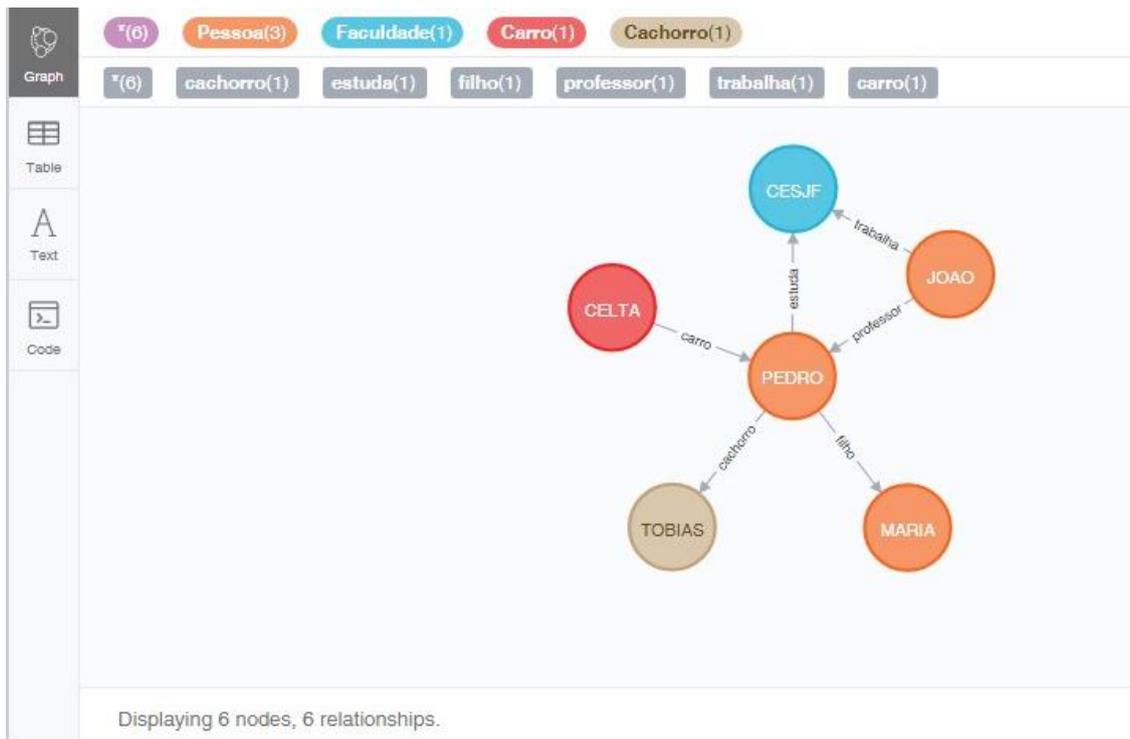
Figura 7: Tela da aplicação desenvolvida com os conhecimentos construídos ao longo da pesquisa



Fonte: próprio autor

A seguir, segue o conhecimento pré-existente no banco de dados (Neo4J).

Figura 8: Ilustração da base de conhecimento representada em forma de grafo



Fonte: próprio autor

5 CONSIDERAÇÕES FINAIS E SUGESTÕES PARA FUTUROS TRABALHOS

O trabalho de desenvolver um sistema que possa interagir com o ser humano em linguagem natural, armazenando os conhecimentos obtidos para auxiliá-lo no dia-dia, tem se mostrado não ser uma tarefa fácil, pois a forma como nos comunicamos e guardamos nossos conhecimentos não obedecem aos padrões lógicos utilizados pelas máquinas. Podemos observar esse problema na *web*, que ainda não segue um padrão semântico compreensível pelas máquinas, sendo que a maior parte se encontra disponível em estruturas simples, como *Hypertext Markup Language* (HTML) e *Cascading Style Sheets* (CSS), e sua reestruturação, como explorado pela outra vertente apresentada na pesquisa, não é uma tarefa fácil, tampouco rápida.

Entretanto, nos últimos anos isso tem se mostrado possível e, como vimos ao longo da pesquisa, há o surgimento de novos estudos e tecnologias cognitivas que tentam simular as ações humanas, como o uso de banco de dados orientado a grafos para expressar e armazenar os conhecimentos, onde acreditamos que uma das maiores dificuldades seja a compreensão da formulação linguística de cada idioma (como o entendimento de expressões, gírias e dialetos), que é complexo até mesmo para as pessoas.

Este trabalho mostrou, com base teórica, que é possível o desenvolvimento de um sistema que interaja com o ser humano, como a criação de uma aplicação que consiga recuperar informação por meio de linguagem natural controlada, utilizando ferramentas existentes como o Apache Open NLP e o banco de dados Neo4J. Um próximo passo é a construção de modelos que sejam capazes de entender as especificidades da linguagem para a correta compreensão, por partes das máquinas, dos atuais dados disponíveis na *web*.

RÉSUMÉ

Face à l'augmentation du volume des données disponibles sur le web et étant chaque fois plus automatiques et ágiles les moyens qui les créent, est fondamentale l'organisation de ces données et son interprétation pour l'obtention de réponses rapides et fiables, en faisant les recherches plus précises et complètes pour que les gens prennent des décisions plus appropriées et créent des nouveaux concepts. En mettant en lumière cette nécessité, ce travail a le but d'explorer, sur une base théorique, l'évolution des recherches dans la computation cognitive, dans laquelle on recherche une plus grande interaction homme-machine, avec des techniques de reconnaissance du langage naturel et de la récupération de l'information, en voyant que la communication et la récupération des informations sont les plus grandes difficultés rencontrées. Avec cet objectif, on va commencer le procès de développement d'un modèle d'application orienté vers la prochaine génération de systèmes intelligents, capables de comprendre la parole et de répondre aux questions qui font référence aux connaissances pré définis dans une base de données orientée vers les graphes. Pendant la recherche, on a voit que la réstruturation de web n'est pas un travail facile ni rapide.

Mots-clés: Informatique cognitive. Traitement du langage naturel. Recherche d'information. Base de données orientée vers les graphes.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. Addison Wesley. New York, 1999.

BERNERS-LEE, T. **On the next Web**. TED 2009. Filmed Feb, 2009. Disponível em: <http://www.ted.com/talks/tim_berners_lee_on_the_next_Web?language=pt-br>. Acesso em: 04 de junho de 2019.

BRICKLEY, D.; GUHA, R. V. (2014) **RDF Schema 1.1. 2014**. Disponível em: <<http://www.w3.org/TR/rdf-schema/>>. Acesso em: 04 de junho de 2019.

CHOMSKY, Noam. (1957) **Syntactic Structures**. Mouton, 1957.

FERREIRA, Fabiano. (2013) **Consulta a Ontologias utilizando Linguagem Natural Controlada**. 2013. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/45/45134/tde-02012014-224412/publico/fabianoDissertacao.pdf>> Acesso em: 04 de junho de 2019.

FREITAS, N; AZEVEDO, Mark. (2016) **Um protótipo Linked data para catalogação semântica de publicações**. 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362016000400048#B16> Acesso em: 04 de junho de 2019.

GONÇALVES, F. (2015) **Ontologias e a Web Semântica**. 2015. Disponível em: <http://www.inf.ufsc.br/~fernando.gauthier/EGC6006/material/Aula%203/Ontologia_Web_semantica%20Freitas.pdf>. Acesso em: 05 de junho 2019.

HEATH, Tom; BIZER, Christian. (2011) **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. E-book. Disponível em: <<http://linkeddatatool.com/editions/1.0/>>. Acesso em: 04 de junho de 2019.

KHARE, R. **Microformats: The next (small) thing on the semantic web**. IEEE Internet Computing, 2006.

OATES, T.; PRASAD, M.; Lesser, V. (1994). **Cooperative Information Gathering: A Distributed Problem Solving Approach**. Computer Science Technical Report 94-66-version 2. University of Massachusetts, Amherst, USA.

OLIVEIRA, F. R. CURA, L. M. del V. (2015). **Avaliação do desempenho de gerenciadores de bancos de dados multimodelo em aplicações com persistência poliglota**. Disponível em:

<<http://www.cc.faccamp.br/Dissertacoes/FabioRobertoOliveira.pdf> > Acesso em: 04 de junho de 2019.

PEIRCE, C. S. (1909). **Existential graphs**. Unpublished manuscript; reprinted in BUCHLER, 1955.

PENTEADO, R. R. et al. (2014) **Um estudo sobre bancos de dados em grafos nativos**. 2014. Disponível em: < <http://www.inf.ufpr.br/carmem/pub/erbd2014-artigo.pdf> > Acesso em: 08 de junho de 2019.

POSSANI, P. (2011) **Teoria de grafos e suas aplicações**. 2011. Disponível em: < <http://www.rc.unesp.br/tmelo/diss-polyanna.pdf> > Acesso em: 10 de junho de 2019.

RUSSELL, Stuart J.; NORVIG, Peter. (2013) **Inteligência artificial**. Tradução: Regina Célia Simille. Rio de Janeiro: Elsevier, 2013.

SADALAGE, P. J.; FOWLER, M. (2012) **NoSQL distilled: a brief guide to the emerging world of polyglot persistence**. [S.l.]: Pearson Education, 2012.

TAHRI, Adel; TIBERMACHINE, Okba. (2013) **DBPEDIA base factoid question answering system**. 2013. Disponível em: < <https://pdfs.semanticscholar.org/dcb3/04003dde5a9fbbb2d42772405e3c8ba4ca75.pdf> > Acesso em: 04 de junho de 2019.

Neo4j. Disponível em: <<https://neo4j.com/>>. Acesso em: 05 de junho de 2019.

OpenNLP. Disponível em: <<https://opennlp.apache.org/>>. Acesso em: 05 de junho de 2019.