Exploração do uso de bases de conhecimento e processamento de linguagem natural em um simulador de casos clínicos

Diego Pinheiro[1], Blanda Mello[1], Paulo Ricardo Barros[12], Sandro Rigo[1], Marta Bez[2], Luana D. S. Rockenback[2]

[1] Universidade do Vale do Rio dos Sinos

[2] Universidade Feevale

diegopinheiro@feevale.br, blandamellus@gmail.com, pbarros1979@gmail.com, rigo@unisinos.br, martabez@feevale.br, luanarockenback@gmail.com

Abstract. *This research aims to explore the possibilities of integration of resources about Natural Languages Processing and ontology in the context of a Virtual Patient Simulator called Health Simulator, whose purpose is to assist the development of clinical reasoning and diagnosis of the students on health area, following his behavior and evolution during the simulation. A translation module was build, that uses ontology based on the questions and answers of the simulator about the disease Headache, and consequently, the final answer is formatted based in the context of the clinical case.*

Resumo. *Esta pesquisa visa explorar as possibilidades de integração de recursos sobre Processamento de Idiomas naturais e ontologia no contexto de um Simulador de Pacientes Virtuais chamado Health Simulator. O objetivo é auxiliar no desenvolvimento de raciocínio clínico e diagnóstico dos alunos na área de saúde, seguindo seu comportamento e evolução durante a simulação. Foi construído um módulo de tradução, que usa a ontologia com base nas perguntas e respostas do simulador sobre a doença Cefaleia e, consequentemente, a resposta final é formatada com base no contexto do caso clínico.*

1. Introdução

Para facilitar a interação entre o ser humano e as máquinas, é preciso que a diferença da representação simbólica seja superada, removendo esta lacuna que envolve tanto capturar as informações, quanto o seu uso e disponibilização. Um agente que deseja adquirir conhecimento precisa entender a linguagem que os seres humanos usam [Russell; Norvig, 2013]. A Geração de Linguagem Natural (GLN) é uma técnica, na área de Processamento de Linguagem Natural (PLN), utilizada para gerar automaticamente frases em linguagem natural (LN), com base em modelos diversos.

Dentro do campo de PLN, observa-se crescente interesse no tratamento de textos na área da saúde, com o objetivo de produzir resumos concisos de documentos e relacionamentos de dados, usando técnicas de diferentes graus de sofisticação, como apresentado por [Banaee; Amy, 2015], [Hsu et al. 2012], [Puppala et al. 2015] e [dos Santos et al. 2016]. Entretanto, com base na revisão da literatura é possível observar uma

lacuna na integração de PLN em simuladores do tipo Paciente Virtual (PV) para o apoio ao ensino à prática em saúde [Russell; Norvig, 2013].

As ontologias vêm sendo utilizadas na computação como uma estrutura para integrar o conhecimento compartilhado e informações heterogêneas [Souza et al. 2014]. Dentre as vantagens de seu uso, destaca-se a possibilidade dada aos programadores de reaproveitar o conhecimento, podendo realizar alterações e implementar extensões para sua utilização. A criação de estruturas de conhecimento é uma das atividades mais longas, caras, e complexas de um sistema especialista[1]. Há um alto ganho de tempo, trabalho e investimento quando reutilizada uma ontologia, visto que existe uma ampla quantidade de ontologias disponíveis para uso, reutilização e integração, podendo ser complementadas de acordo com os conceitos dos domínios [Zanetti et al. 2014].

Em trabalhos relacionados, verificou-se que Kataria & Juric (2010) criaram uma ontologia para reutilização em ambientes de saúde, onde é usada para gerenciamento semântico de aplicações. Samuel et al. (2017) utilizou PLN para comparar perguntas semelhantes em fóruns de saúde, afim de que não cheguem perguntas repetidas aos especialistas. Santos et al. (2016) criaram um modelo para integração automática de conteúdos através de PLN, buscando melhorar a qualidade dos mesmos em conversação dos usuários. Não foram encontrados artigos que tratam da temática do uso Ontologias e PLN para utilização em simuladores de casos clínicos ou ensino na saúde.

O propósito geral deste trabalho é explorar as possibilidades de integração de recursos de PLN para o contexto de um simulador do tipo PV denominado Health Simulator, cuja finalidade é auxiliar no desenvolvimento do raciocínio clínico e diagnóstico do aluno da área da saúde, acompanhando sua conduta e evolução durante a simulação. Para isso, foi construída uma ontologia em OWL (*Web Ontology Language*) baseada nas perguntas e respostas do simulador sobre a doença Cefaleia, aplicando consultas de palavras-chave (adquiridas com técnicas de PLN e validadas na ferramenta web LXparser) para descobrir e associar as mesmas. Os dados foram retirados de [Bez 2013]. Para a construção da ontologia, foi utilizado o software Protégé, sendo aplicadas as consultas com o auxílio da linguagem de programação Python e dos padrões SPARQL (*Simple Protocol and RDF Query Language*).

Este artigo está organizado do seguinte modo: a seção 2 apresenta alguns trabalhos relacionados. Já a seção 3 descreve o desenvolvimento da Ontologia, a proposta de implementação é apresentada na seção quatro. Na 5 pode ser acompanhado o experimento realizado com os resultados encontrados. A seção 6 apresenta a validação do modelo proposto e as conclusões e sugestões de trabalhos futuros logo em seguida.

2. Trabalhos correlatos

Kataria & Juric (2010) desenvolveram uma aplicação no software SA for Go-CID, que auxilia na recuperação de informações heterogêneas sobre saúde. Utilizou-se camadas ontológicas que realizam o alinhamento, integração e fusão de ontologias através de regras de raciocínio, resolvendo heterogeneidades semânticas da formação em saúde.

A camada ontológica criada no software Go-CID é reutilizável em ambientes de saúde, pois as aplicações não são específicas do domínio. O sistema foi utilizado para o gerenciamento semântico de requisitos para a criação de aplicações de software

[1] Todos os especialistas consultados são relacionados à área da saúde.

conscientes da situação, que dependem de uma infinidade de dispositivos, preferências do usuário e da prestação de serviços semânticos na computação [Kataria; Juric, 2010].

Através da camada, é possível resolver conflitos semânticos através do raciocínio e sem afetar os repositórios originais, isto é, sem impor qualquer mudança neles. Ela também é capaz de abordar heterogeneidades sem usar uma variedade de algoritmos e palavras-chave correspondentes, com base em uma "ponderação" para inferir conhecimento sobre esses ambientes [Kataria; Juric, 2010].

Já Samuel et al. (2017) propôs um método para comparar perguntas semelhantes em fóruns de saúde utilizando TF-IDF ponderada, heurística de relevância e expansões de termos baseadas em relações semânticas e abreviaturas. Os experimentos usaram um conjunto de dados do desafio HDAC 2015, e também coletaram dados padrões para avaliação de sistemas de recuperação de perguntas, como fóruns de saúde.

A abordagem de busca de informações dos usuários inclui um especialista em domínio que, dada uma pergunta recebida, procura por questões semelhantes nos fóruns, determinando palavras-chave relevantes. Há um sentido de palavras semanticamente relacionadas, como sinônimos e abreviaturas, que poderiam ser correspondidas no corpus. Finalmente, o especialista tem conhecimento do contexto da pergunta que chega e o que está sendo perguntado. Sua metodologia é dividida em: pré-processamento, pontuação, relevância heurística e expansão de prazo [Samuel et al. 2017].

Os autores dos Santos et al. (2016) criaram um modelo para a integração automática de conteúdo com um agente inteligente de comunicação, com o propósito de realizar publicações de conteúdo apontados semanticamente para páginas da web e absorver este conteúdo através do agente de comunicação. O protótipo possui como domínio um caso de propagação de dados de um instituto de ensino.

As respostas dos conteúdos mostraram a oportunidade de uma conexão entre o agente de comunicação e a estrutura de usabilidade do AIML (*Artificial Intelligence Markup Language*), permitindo que se ultrapassasse o obstáculo de uma reprodução de conteúdo que sirva de estrutura para as conversas. Na validação, os usuários notaram uma melhora quanto a utilização do protótipo em plataformas de conversação por LN, do qual sujeita-se a qualidade dos conteúdos disponíveis pelas aplicações de diálogo, visto que esse é um ponto a melhorar na interface de conversação dos usuários. Conforme os resultados, os autores puderam autenticar os conceitos existentes no modelo, buscando analisar o mapeamento das informações conforme a língua portuguesa [dos Santos et al. 2016].

3. Desenvolvimento Da Ontologia

A utilização de ontologias é um importante passo para a construção de um domínio de conhecimento estruturado onde será possível, no futuro, realizar inferências sobre o modelo proposto, tornando o conteúdo acessível e compreensível por um humano, além de habilitar o seu processamento adequado por uma máquina. Com base nas perguntas e respostas pré-definidas dos casos clínicos, pode-se futuramente realizar buscas e inferências em ontologias existentes na área médica, tanto manualmente como de forma automática. O uso das ontologias também facilitará o processo de análise semântica e de contexto, possibilitando o trabalho com inferências e similaridade, consequentemente, torna-se viável trabalhar com consultas SPARQL para a composição das frases de um modo mais flexível.

Os dados para o desenvolvimento da estrutura da ontologia foram retirados de Bez (2013), que apresenta um quadro de perguntas e respostas relacionados a doença Cefaleia, para a utilização em um simulador de casos clínicos. A rede foi construída por Fonseca (2013) e teve como base as Diretrizes Clínicas da Sociedade Brasileira de Medicina. As perguntas, respostas e condutas são separadas por nodos (definições de um determinado tipo de sintoma) que, quando unidos de acordo com a resposta do paciente, levam a um diagnóstico. Nas respostas, existem variáveis pré-definidas que serão substituídas, aleatoriamente, de acordo com a doença. Em uma visão final, o processo que antecede a substituição das variáveis ocorre com o retorno da inferência no Modelo Bayesiano (MB) através de um retorno em valor percentual. O mesmo é convertido à uma notação aplicada à um filtro de intervalos, onde identifica-se a variável a ser utilizada na substituição [Bez 2013]. A estrutura e o conhecimento inseridos são representados da seguinte forma:

Classes: a classe Casos Clínicos apresenta, de forma geral, todas as doenças a serem simuladas. Através da classe Investigação, conecta-se às demais subclasses de conversação (Condutas, Diagnósticos, Perguntas e Respostas), bem como, a classe Tipos, que possui as subclasses com todos os nodos existentes.

Propriedades: foram desenvolvidos dezenove tipos de propriedades para armazenar o conhecimento. associado_ao_caso_clinico (identificação da doença), ID (chave de identificação), diagnóstico (resposta de acordo com o conjunto de nodos), nodo (classificação do sintoma), min (peso em termo de tempo), r$ (peso em termo de valor), tipo (tipo de pergunta), pergunta (pergunta pré-definida para o nodo), ligado_a_pergunta (liga a resposta a uma pergunta pré-definida), palavra_chave (palavras-chave extraídas da pergunta), resposta (modelo de resposta), variável (valores que serão inseridos para substituição no modelo de resposta), ligado_ao_nodo (ligação entre os nodos), tradução (tradução do nome do nodo) e descrição (descrição do nodo).

Instâncias: nas instâncias são inseridos os dados reais da base de conhecimento. As informações são cadastradas conforme as propriedades inseridas em cada classe. A Figura 1 ilustra um exemplo de instâncias da ontologia.

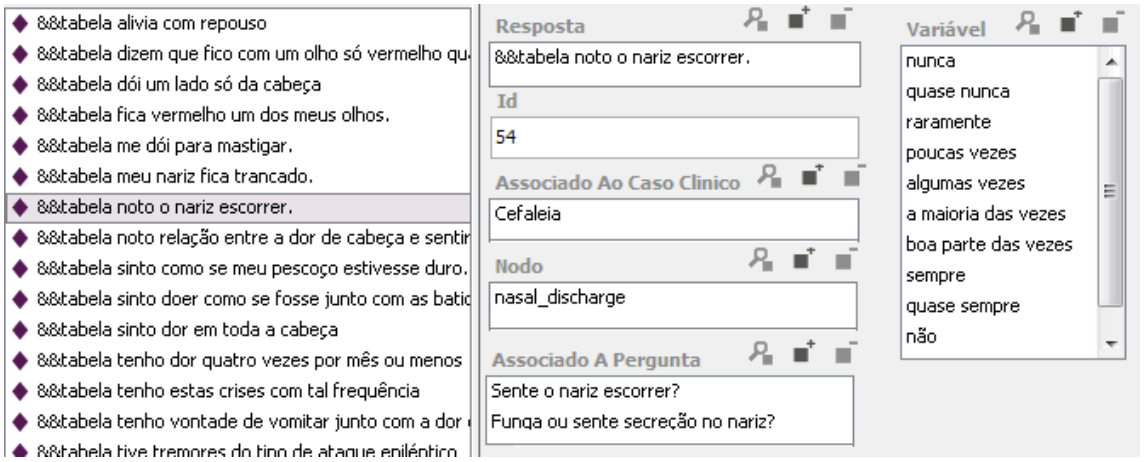

Figura 1 – Exemplo de instâncias da ontologia (dos autores, 2017).

Para a construção do modelo proposto, foi exportada a ontologia no formato OWL para ser utilizada como base de conhecimento. Mais detalhes sobre o desenvolvimento serão exibidos na próxima seção.

4. Construção do Modelo

A abordagem deste trabalho baseia-se na busca de informações com base em uma ontologia, conforme apresentado na seção anterior. Dada uma pergunta recebida, o sistema procura por questões semelhantes no corpus, determinando palavras-chave relevantes a partir da pergunta recebida e o domínio do conhecimento dessas palavras-chave. Esta seção descreve a arquitetura do módulo de tradução utilizado no processo de investigação do Health Simulator, conforme apresentada na Figura 2.

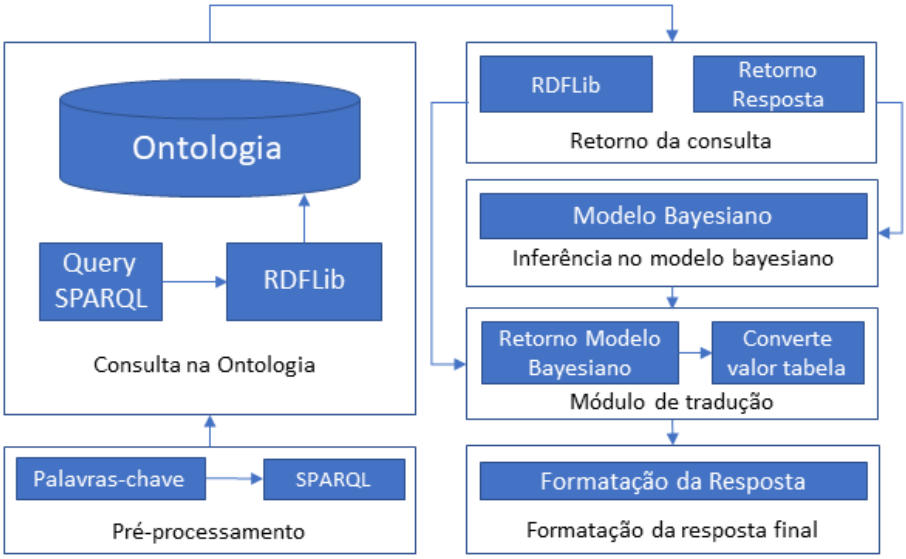

Figura 2 – Estrutura da construção do modelo (dos autores, 2017).

O módulo de tradução representa o núcleo da aplicação, responsável por receber os dados da pergunta, tratamento e extração das palavras chave, consulta a ontologia e seu tratamento do retorno, servindo como base para inferência no MB, resultando em uma probabilidade que sofre uma conversão com base na escala de representação. Desta forma, é possível obter-se a essência, para gerar as respostas curtas em texto coerente, produzidas com base no modelo estatístico.

Para a realização de tal protótipo, utilizou-se a linguagem Python versão 2.7 em apoio às consultas e manuseio com o modelo de domínio – ontologia nesta linguagem, utilizou-se a biblioteca RDFLib. O modelo ilustrado na Figura 3 pode ser resumido nas etapas a seguir.

Pré-processamento: neste primeiro momento, é necessário realizar a extração das palavras-chave com base na pergunta realizada. Este processo pode ser exemplificado ao analisar uma simples pergunta "A dor alivia com repouso?". Ao retirar as *stop-words*, tais como artigos, preposições e pontuações, obtêm-se as palavras de relevância 'dor', 'alivia' e 'repouso'. Para a análise das palavras-chave, utilizou-se a ferramenta web LXparser [Silva et al. 2010]. Após realizadas as primeiras experimentações, observou-se que, utilizando-se três palavras-chave, o resultado já se apresentava de maneira satisfatória, portanto, definiu-se para os casos em que o número de palavras relevantes sejam superiores à 3, deve-se limitar para as 3 primeiras, isso para cada pergunta processada. Posteriormente, o módulo de processamento é responsável por gerar uma consulta em linguagem SPARQL com estas palavras-chave extraídas. As consultas são geradas com o intuito de localizar as perguntas e, por consequência, as respostas vinculadas às mesmas.

Consulta na ontologia: a consulta na ontologia é realizada com o auxílio da biblioteca RDFLib, que permite trabalhar com arquivos em formato OWL, onde pode-se realizar uma consulta com SPARQL e, ao receber o retorno desta consulta, trabalhar com a mesma como um objeto na linguagem Python, realizando interações a fim de formatar os resultados obtidos.

Retorno da consulta: o retorno da consulta é diretamente relacionado à qualidade das palavras-chave extraídas, desta forma, espera-se sempre o retorno de uma e somente uma resposta a ser tratada. Para os casos em que a definição das palavras-chave foi insatisfatória e o retorno da inferência apresentou-se com mais de uma resposta, é indispensável tratá-los em separado. Para tanto, visando uma solução futura, evidenciou-se a necessidade e preocupação não apenas com as palavras-chave destacadas na extração, mas também um contexto de similaridade semântica ao analisar as respostas resultantes da inferência.

Formulação da resposta: para as respostas que forem encontradas a referência "&&tabela", significa que esse texto será substituído, em função da consulta, pelas opções de respostas do paciente, conforme segue: "nunca", "quase nunca", "raramente", "poucas vezes", "algumas vezes", "a maioria das vezes", "boa parte das vezes", "sempre", "quase sempre", "não" e "sim". Esta substituição é norteada pelo retorno da inferência realizada no MB, que traz um valor em percentual, que indicará qual das palavras deve ser utilizada para a substituição e formulação da resposta final. É importante ressaltar que a relação do percentual retornado do MB e o valor é ditado pelo profissional especialista [Bez 2013].

Flexibilização da Resposta: essa etapa foi dividida em quatro partes, sendo elas Remoção de Stopwords, Análise e etiquetagem de palavras, Remoção de Afixos e Busca de Sinônimos, descritas a seguir:

a) Remoção de *stopwords*: a remoção de *stopwords* é uma prática **frequentemente** adotada durante o processo de análise de orações. São palavras irrelevantes, as quais não contribuem para o significado geral da frase (Hardeniya et al., 2016). Este processo de remoção é realizado na resposta obtida após a consulta na ontologia, onde o objetivo é permanecer apenas com as palavras relevantes ao contexto, de forma a conseguir trabalhar com sua flexibilização por meio de sinônimos. Para o protótipo utiliza-se a ferramenta disponibilizada na própria biblioteca NLTK.

b) Análise e etiquetagem de palavras: após a remoção das *stopwords*, restam as palavras de maior relevância para o contexto da resposta. Neste momento, realiza-se a etiquetagem das mesmas, para obtenção de sua função morfossintática. É somente após a etiquetagem que pode-se trabalhar com as palavras de modo diferenciado, pois sabe-se sua função na resposta. Para este processo de rotulagem, utilizou-se o projeto NlpNet, o qual pode ser acessado como uma biblioteca para a linguagem Python. A biblioteca tem como objetivo realizar tarefas de PLN com apoio de alguns modelos treinados de redes neurais. Atualmente, ele realiza rotulagem parcial e algumas rotulagens de função semântica. Foi escolhido para este trabalho por possuir algumas funções especialmente adaptadas para trabalhar com a língua portuguesa (Nlpnet 2017), (Fonseca & Rosa 2013).

c) Remoção de Afixos: a redução é utilizada para buscar o cerne da palavra relevante em análise, desta forma pretende-se identificar sinônimos com base em sua forma original e trabalhar na flexibilização de respostas. Para esta tarefa, utilizou-se a biblioteca NLTK, a qual disponibiliza material para a tarefa, também denominada *stemmer*. Na morfologia linguística, a remoção de afixos trata-se do processo de redução

de palavras inflexivas (ou derivadas) em sua forma de caule, base ou raiz - geralmente uma forma de palavra escrita. Importante ressaltar que o caule não precisa ser idêntico à raiz morfológica da palavra, muitas vezes é suficiente que as palavras relacionadas correspondam ao mesmo tronco, mesmo que esse tronco não seja em si uma raiz válida.

d) Busca de Sinônimos: com as etapas anteriores finalizadas, é possível realizar a busca de sinônimos por meio de um dicionário de *synsets,* este disponibilizado via API pelo projeto *wordnetPT*, o qual viabiliza acesso a alguns métodos de consulta (Wordnetpt 2017).

Formatação da Resposta Final: após realizar a consulta dos sinônimos, é realizada a substituição para a formulação da resposta final. Durante os experimentos, foram identificados problemas na construção da sentença, como a utilização de sinônimos que possam alterar o sentido e/ou resultar em uma má formação da mesma, isso ocorre devido a semântica da sentença, a qual não foi considerada neste primeiro momento.

5. Resultados Preliminares

Para avaliação do modelo, foi realizado um experimento com base na estrutura gerada pelo especialista, que elencou um total de 36 perguntas e respostas denominadas corretas e, a partir desta informação, foram realizadas as interações no modelo, onde obteve-se um resultado satisfatório, mas intimamente ligado a correta extração das palavras-chave da pergunta inferida no modelo. Conforme a breve caracterização para extração das palavras-chave na seção anterior, percebeu-se alguns aspectos distintos no retorno das respostas após a consulta, onde verificou-se 3 situações isoladas, conforme apresentado na sequência:

a. Consulta com número satisfatório de palavras-chave e com múltiplas respostas: a busca possui o número mínimo de palavras-chave e retorna múltiplas respostas, contudo, o retorno é aceitável quando observado que tais perguntas são pertencentes ao mesmo nodo da rede representada no domínio, portanto suas respostas apresentam-se em torno de conhecimentos análogos.

b. Consulta com número de palavras-chave inferior à regra com múltiplas respostas: uma busca sem um número mínimo de palavras-chave, eventualmente, trará múltiplas respostas. Neste cenário, ao analisar o caso separadamente, pôde-se concluir que a pergunta foi mal formulada, com informações insuficientes para uma resposta conclusiva. Desta forma, a mesma foi considerada irrelevante para alteração nos resultados gerais.

c. Consulta com número de palavras-chave inferior e retorno satisfatório (expansão da regra): buscas com apenas uma resposta de retorno, apresentaram, em primeiro momento, uma situação semelhante ao item (b), onde o número mínimo de palavras-chave não é atingido. Entretanto, ao analisar o caso detalhadamente, percebeu-se que a pergunta apresentou melhor filtro para consulta, onde o critério aplicado à oração consistiu em remover todas as *stop-words*, e posteriormente optando-se por uma expansão à regra, ao tratar-se de verbos no infinitivo. Este caso pode ser melhor compreendido no exemplo demonstrado na Tabela 1, onde o critério de extração destas *stop-words* (artigos, advérbios, pronomes, preposições, conjunções, flexões para o gerúndio e particípio, pontuações e verbos – exceto no infinitivo) constantes nas perguntas, ocasiona um retorno positivo, mesmo com o número mínimo de palavras-chave não atingido. A Tabela 1 apresenta a análise morfossintática da pergunta “A dor alivia no repouso?”.

Tabela 1. **Análise da pergunta "**A dor alivia no repouso**?"**.

Token	**Análise morfológica**
A	Artigo definido.
dor	Substantivo.
alivia	Verbo conjugado na 2ª pessoa do singular.
no	Contração e pronome pessoal.
repouso	Substantivo.
?	Pontuação.

A Tabela 2 apresenta o resultado da aplicação do exemplo ilustrado na Tabela 1 junto a outros experimentos das consultas no modelo e as respectivas respostas.

Tabela 2. Resultados da consulta no modelo.

Key Words	**Resposta**	**Resposta Formatada**
['dor', 'repouso']	&&tabela alivia com repouso	poucas vezes alivia com repouso
['dor', 'espalhada', 'cabeca']	&&tabela sinto dor em toda a cabeca	poucas vezes sinto dor em toda a cabeca
['luz', 'doi', 'piscante']	A luz &&tabela piora minha dor	A luz a maioria das vezes piora minha dor
['dor', 'cabeca', 'vomitar']	&&tabela tenho vontade de vomitar junto com a dor de cabeca.	raramente tenho vontade de vomitar junto com a dor de cabeca.
['dor', 'olhos', 'vermelho']	&&tabela fica vermelho um dos meus olhos.	algumas vezes fica vermelho um dos meus olhos.

O experimento foi realizado com um número reduzido de perguntas, servindo como um indicador de validade. É necessário aplicar o modelo a uma quantidade maior de perguntas, buscando confrontar com o modelo gerado pelo especialista e a aplicação de métodos estatísticos, para assim validá-lo.

6. Experimentação com o modelo

Para o processo de validação do modelo, optou-se por uma abordagem baseada em cenário. Para isto, serão descritas situações onde o módulo será utilizado e o seu contexto, observando o processo de interação do aluno com o jogo e analisando as perguntas e respostas fornecidas durante a simulação, possibilitando, com isto, validação do modelo proposto e elucidando as principais características e possíveis falhas.

A Modelagem do Conhecimento do Health Simulator é realizada por um especialista de domínio, que deve ter claro quais limites usará na representação do problema, com apoio de uma diretriz clínica [Fonseca 2013]. Esta representação é feita através de uma RB previamente construídas pelo especialista, que será importada no simulador e servirá como fonte de conhecimento durante o processo de desenvolvimento dos casos clínicos e da simulação.

A Montagem do Caso Clínico no Health Simulator é realizada pelo professor. São informados alguns dados básicos do caso, como um histórico clínico anterior, uma história inicial e a seleção das variáveis Sinais e Sintomas, que devem estar presentes no caso. Ao incluir livremente sintomas e sinais disponíveis na rede, o professor propaga as probabilidades, fazendo emergir um ou mais diagnósticos e suas respectivas condutas, modelando, assim, o caso que será simulado pelos alunos. Estas probabilidades são

apresentadas a todo o momento ao professor, para que este possa acompanhar como será o desenvolvimento do caso.

No ambiente do simulador, o estudante seleciona o caso clínico a ser executado. É apresentado ao aluno um resumo do caso e a ficha do paciente. Posteriormente, é iniciada a fase da simulação (Investigação, Diagnóstico e Conduta). Na fase de diagnóstico, o aluno deve fazer perguntas ao PV, digitadas em texto livre, conforme segue, ***"Sente alguma alteração para enxergar durante a dor de cabeça?"***, dando início no processo de raciocínio diagnóstico. O simulador reconhece as palavras-chave da pergunta e realiza a consulta na RB, propagando esta evidência. Deste modo, o resultado da inferência no MB é utilizado para formular a resposta com base na ontologia. Um exemplo de resultado ou resposta seria **"*Enquanto tenho dor [raramente] sinto minha vista diferente*."**, expressando de forma coloquial a probabilidade do nodo a que se refere a pergunta. De posse da resposta, o aluno continua o processo, tentando confirmar ou refutar a hipótese inicial.

Nessa forma dialogada, o aluno tem a possibilidade de formular suas hipóteses diagnósticas a partir da seleção das perguntas e respostas realizadas ao PV, que poderão reforçá-las ou refutá-las no decorrer do processo de investigação.

7. Conclusão

Este artigo apresentou uma proposta de uso de ontologias e PLN para a recuperação de respostas curtas no Health Simulator. O processo compreende a construção da ontologia, a consulta através de protótipo, resultados das consultas efetuadas, retorno do modelo de resposta, bem como, a simulação de variáveis como retorno de um domínio de conhecimento incerto (modelo bayesiano), e sua substituição com base na escala de representação. Por fim, realiza-se a formatação final da resposta, e esta é apresentada ao aluno como resposta do processo de investigação para resolução do caso clínico. Após a implementação completa, será possível realizar as primeiras avaliações e possível validação do mesmo junto ao Simulador de Casos Clínicos em estudo.

O sistema apresentou resultados satisfatórios. Já é possível perceber que existem fortes indícios que o módulo proposto terá aderência ao projeto Health Simulator, cumprindo sua função no simulador. Após a implementação completa, será possível realizar as primeiras avaliações e validação do mesmo junto a alunos da área da saúde.

Com isso, resolve-se um problema ocorrido na validação dos resultados do simulador sem o uso de PLN, onde especialistas em educação alertaram que, o modo que o simulador apresenta perguntas pré-definidas é ineficiente, pois já induz os alunos a uma resposta, o que não acontece em situações reais. O uso de ontologia adequou-se bem ao modelo de estruturação de perguntas e respostas e, existem milhares de ontologias de saúde que poderão ser reaproveitadas e utilizadas para o contexto do simulador. Como trabalhos futuros, estão previstas explorações do potencial dos dados organizados na ontologia para flexibilizar a geração de frases automaticamente, assim como incorporar ao protótipo, um módulo de análise semântica para flexibilização correta das sentenças.

Referências

Banaee, H.; Amy, L. (2015) "Data-Driven Rule Mining and Representation of Temporal Patterns in Physiological Sensor Data." IEEE journal of biomedical and health informatics 19.5: 1557-1566.

Bez, M. R. (2013) Construção de um Modelo para o Uso de Simuladores na Implementação de Métodos Ativos de Aprendizagem das Escolas de Medicina. Porto Alegre. 314 f. Tese (Doutorado em Informática na Educação) – PGIE/UFRGS, Porto Alegre.

dos Santos, F. R., Rigo, S. J.; Barbosa, J. L. V.; Rodrigues, C. (2016, November). EDUARDO: A Semantic Model for Automatic Content Integration with an Conversational Intelligent Agent. In Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web (pp. 255-262). ACM.

Fonseca, J. M. L. (2013). Descrição de um simulador baseado em redes bayesianas como método de avaliação do aprendizado de diretrizes clínicas em ensino a distância para medicina de família e comunidade (Doctoral dissertation).

Fonseca, E. R.; Rosa, J. L. G. (2013). Mac-morpho revisited: Towards robust part-of-speech tagging. In Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology (pp. 98-107). sn.

Hardeniya, N., Perkins, J., Chopra, D., Joshi, N., Mathur, I. (2016, November): Natural Language Processing: Python and NLTK. Publisher: Packt Publishing.

Hsu, W., Taira R. K., El-Saden S., Kangarloo H., Bui A. A. (2012) "Context-based electronic health record: toward patient specific healthcare." IEEE Transactions on information technology in biomedicine 16.2: 228-234.

Kataria, P. & Juric, R. (2010) "Sharing E-Health Information through Ontological Layering", Hawaii International Conference on System Sciences (HICSS), Proceedings of the 43rdAnnual, IEEE Computer society.

NLPNET, "Python library for Natural Language Processing", acessado em 30 de junho de 2017, disponível em: http://nilc.icmc.usp.br/nlpnet/

Puppala, M., He, T., Chen, S., Ogunti, R., Yu, X., Li, F., ... & Wong, S. T. (2015). METEOR: an enterprise health informatics environment to support evidence-based medicine. IEEE Transactions on Biomedical Engineering, 62(12), 2776-2786.

Russell, S. & Norvig, P. (2013) Inteligência Artificial. 3. ed. Rio de Janeiro: Elsevier.

Samuel, H., Kim, M. Y., Prabhakar, S., Jabbar, M. S. M., & Zalane, O. (2017, February). Community question retrieval in health forums. In Biomedical & Health Informatics (BHI), 2017 IEEE EMBS International Conference on (pp. 77-80). IEEE.

Silva, J. R., Branco, A., Castro, S., & Reis, R. (2010, April). Out-of-the-Box Robust Parsing of Portuguese. In PROPOR (Vol. 6001, pp. 75-85).

Souza, A. S., Duran, A., & Vieira, V. (2014). Uma Ontologia de Domínio para a Metodologia de Aprendizagem Baseada em Problemas. In Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE) (Vol. 25, No. 1, p. 1253).

WordnetPT, acessado em 29 de junho de 2017, disponível em: http://wordnet.pt/api.

Zanetti, H. A., & Bonacin, R. (2014). Uma Metodologia Baseada em Semiótica para Elaboração e Análise de Práticas de Ensino de Programação com Robótica Pedagógica. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 25, No. 1, p. 1233).