

UNIVERSIDADE ESTADUAL DE CAMPINAS - UNICAMP
FACULDADE DE TECNOLOGIA
SISTEMAS DE INFORMAÇÃO

**PROPOSTA DE SISTEMA PARA AVALIAÇÃO SEMIAUTOMÁTICA DE
QUESTÕES DISCURSIVAS BASEADA NA SIMILARIDADE DE CONCEITOS**

BRUNO DO VALLE SAES ADEGAS

LIMEIRA, 2020

BRUNO DO VALLE SAES ADEGAS

**PROPOSTA DE SISTEMA PARA AVALIAÇÃO SEMIAUTOMÁTICA DE
QUESTÕES DISCURSIVAS BASEADA NA SIMILARIDADE DE CONCEITOS**

Monografia apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como requisito para obtenção do título de Bacharel em Sistemas de Informação.

Orientadora: Profa. Dra. Gisele Busichia Baioco

LIMEIRA, 2020

AGRADECIMENTOS

Gostaria de agradecer minha namorada Juliana, que sempre me apoia e sempre está ao meu lado, independente da situação que for.

À minha família, que sempre me motivaram e apoiaram, ajudando no que fosse necessário para eu continuar a me desenvolver.

Aos amigos da graduação, em especial a “Turminha da Íris”, que tornaram toda a graduação mais divertida e agradável.

Ao meu amigo, Eduardo, que me auxiliou no desenvolvimento em Python.

Aos docentes que sempre lecionaram com carinho e com domínio, servindo de inspiração para cada vez mais querer aprender sobre a área. Principalmente pela minha orientadora Dra. Gisele Busichia Baioco, tanto das monitorias quanto do Trabalho de Conclusão de Curso, por me auxiliar quando precisava e sempre me guiando diante problemas.

RESUMO

Atividades no formato de questões discursivas são muito importantes para o aprendizado e desenvolvimento do aluno. Porém, geralmente, a quantidade de alunos e o fator tempo do professor torna muito exaustiva a correção de questões discursivas. Isso faz com que o professor utilize menos questões desse tipo em suas avaliações, substituindo por questões objetivas, que possibilitam correção mais simples e rápida, o que pode impactar na qualidade das atividades propostas aos alunos. Desse modo, este trabalho teve como objetivo propor um sistema de avaliação de questões discursivas de respostas curtas, de modo semiautomático por meio da comparação por similaridade de conceitos simples, presentes nas respostas dos alunos. Espera-se que o sistema proposto possa servir como uma ferramenta de auxílio ao professor, reduzindo o esforço para as correções de questões discursivas e, então, possibilitando a utilização de mais questões desse tipo nas avaliações. Foi realizado um estudo de caso que apresentou resultados satisfatórios para o objetivo e o escopo do trabalho, sendo que há muitos aspectos em que o sistema pode evoluir ainda no futuro.

Palavras-chave: Conceitos, Avaliação semiautomática, Comparação por similaridade de conceitos.

SYSTEM PROPOSAL FOR SEMIAUTOMATIC EVALUATION OF DISCURSIVE QUESTIONS BASED ON SIMILARITY OF CONCEPTS

ABSTRACT

Activities in the form of discourse questions are very important for the student's learning and development. However, in most cases, the number of students and the teacher's time factor makes correcting discursive questions very exhaustive. This causes the teacher to use fewer questions of this type in his assessments, replacing them with objective questions, which allow for simpler and faster correction, which can impact the quality of the activities proposed to students. So, this work aimed to propose a system for evaluating short answer discursive questions, in a semi-automatic way by comparing similarities in simple concepts, present in students' answers. It is hoped that the proposed system can serve as a tool to assist the teacher, reducing the effort to correct discursive questions and, thus, allowing the use of more questions of this type in evaluations. A case study was made and the results presented was satisfactory for the objective and scope of the work, and there are many aspects in which the system can be improved even in the future.

Keywords: Concepts, Semi automatic evaluation, Comparison by concept similarity.

SUMÁRIO

LISTA DE FIGURAS	7
LISTA DE ABREVIATURAS	8
1. INTRODUÇÃO	9
1.1. Objetivo	9
1.2. Metodologia	10
1.3. Estrutura do trabalho	10
2. CONCEITOS FUNDAMENTAIS	11
2.1. Pré-processamento Textual	11
2.2. Cálculo de Similaridade	12
2.3. Comparação por Similaridade de Conceitos	12
3. REVISÃO DA LITERATURA	17
4. ARQUITETURA DO PROTÓTIPO DO SISTEMA	20
5. ESTUDO DE CASO	23
5.1. Respostas utilizadas	23
5.2. Execução do protótipo	24
5.3. Análise dos Resultados	27
6. CONCLUSÃO E TRABALHOS FUTUROS	29
7. REFERÊNCIAS BIBLIOGRÁFICAS	31

LISTA DE FIGURAS

2.1 Definição da fórmula matemática da similaridade de conceitos.....	13
2.2 Mapeamento da similaridade entre dois conceitos.....	13
2.3 Equação da medida de distância entre os conceitos.....	13
2.4 Equação de coeficiente de Jaccard.....	14
2.5 Equação de similaridade do Cosseno.....	15
4.1 Arquitetura do protótipo do sistema.....	20
4.2 Estrutura obrigatória do arquivo de respostas.....	21
5.1 Enunciado da questão de história usada como exemplo.....	23
5.2 Exibição inicial da GUI do protótipo.....	24
5.3 Exemplo de inserção de Palavras-Chave no protótipo.....	25
5.4 Tela pop-up das respostas.....	26
5.5 Exibição das respostas por uma lista ordenada.....	27

LISTA DE ABREVIATURAS

AAC - Avaliação Assistida por Computador

PLN - Processamento de Linguagem Natural

PEG - *Project Essay Grader*

AVA - Ambiente Virtual de Aprendizagem

RA - Registro Acadêmico

KNN - *K-nearest-neighbor*

GUI - *Graphical User Interface*

CSV - *Comma-Separated Value*

NLTK - *Natural Language ToolKit*

1. INTRODUÇÃO

A avaliação de questões discursivas pode se tornar uma atividade muito massante e demorada para os professores, dependendo principalmente do número de alunos a serem avaliados. Isso faz com que tenham que reduzir a aplicação desses tipos de questões que são importantes para a formação do aluno. De acordo com Saxena e Gupta (2009), a construção de uma resposta discursiva permite que o aluno possa assimilar os conceitos e juntá-los de uma forma compreensiva em um texto. Entretanto, devido ao esforço para correção de questões discursivas, os professores utilizam mais as questões objetivas nas atividades propostas aos alunos, ou, fazem avaliações de questões discursivas de maneira mais simplificada, o que não permite explorar o máximo de potencial desse tipo de questão.

Conole e Warburton (2005) dizem que a aplicação de tecnologias educacionais para apoio à avaliação de aprendizagem, como as de AAC (Avaliação Assistida por Computador), adicionam eficiência e eficácia ao processo de ensino. Porém, para cada uso, as atividades e a interpretação sobre os dados diferirão, e isso é crucial para o uso correto desses sistemas. Quando se utiliza dessas ferramentas de maneira correta, a assistência do computador permite a redução do esforço necessário do professor para a obtenção de uma avaliação de qualidade e também da carga horária necessária para a correção.

1.1. Objetivo

O objetivo deste trabalho de conclusão de curso foi propor um sistema de avaliação semiautomática de questões discursivas de respostas curtas, através de técnicas que possibilitem ao professor verificar quais respostas chegaram mais perto do que ele deseja. O sistema irá receber um conjunto Q de conceitos e um conjunto $X = \{x_i\}_{i=1}^n$ de respostas dos alunos em formato textual. O conjunto Q representa os conceitos que o professor gostaria de pontuar em uma resposta correta. A resposta $\{x_i\}$ é o texto inserido pelo i -ésimo aluno. O sistema irá gerar para cada resposta um número real que indica a proximidade da resposta com o esperado pelo professor. Nesse caso, o sistema é uma função $\delta(Q, x_i) = d_{ij}$, com $d_{ij} \in \mathbb{R}$, para cada objeto $x_i \in X$. Terminado o processo, é possível gerar uma permutação de respostas do conjunto X com base nos valores d_{ij} obtidos por δ , o que fornece ao professor um ranqueamento com as melhores respostas no topo da lista. Como o objetivo é fazer uma proposta, o sistema desenvolvido neste trabalho está na etapa de protótipo inicial, permitindo fazer a avaliação de respostas a partir de conceitos limitados a uma palavra apenas.

1.2. Metodologia

A metodologia adotada neste trabalho teve início com a revisão do estado da arte nas áreas de avaliação semiautomática e automática de questões discursivas e na de comparação por similaridade de conceitos, visando entender o que foi pesquisado e desenvolvido na área. Em seguida foram analisados os algoritmos e técnicas que poderiam servir como base para o desenvolvimento de uma solução que atendesse ao objetivo do trabalho. Após verificar todas as técnicas relevantes para a solução, elas foram comparadas entre si, para verificar qual se encaixaria melhor na proposta. Após a seleção, foi feita a estruturação da arquitetura do protótipo do sistema, para então prosseguir para a etapa de desenvolvimento e assim ser aplicado a um estudo de caso de modo a testar a solução proposta pelo trabalho. Por fim, os resultados foram documentados e analisados, verificando se o teste da solução apresentou resultados satisfatórios para o objetivo deste trabalho.

1.3. Estrutura do trabalho

Além deste capítulo de Introdução, essa monografia está estruturada pelos seguintes capítulos:

- **Capítulo 2 - Conceitos Fundamentais:** apresentação dos conceitos utilizados na estruturação do trabalho, estudo das técnicas utilizadas para a comparação por similaridade e a avaliação delas.
- **Capítulo 3 - Revisão da Literatura:** breve revisão da literatura, apresentando os trabalhos já realizados na área para utilizá-los como base para a solução proposta.
- **Capítulo 4 - Arquitetura do Protótipo do Sistema:** descrição da arquitetura do protótipo do sistema, bem como as técnicas e as tecnologias utilizadas para desenvolvê-lo.
- **Capítulo 5 - Estudo de Caso:** apresentação da base de questões e respostas utilizadas como exemplo no estudo de caso, de modo a averiguar a solução desenvolvida nesta monografia, apresentação de como se utilizar o protótipo e a documentação dos resultados obtidos pelo estudo de caso.
- **Capítulo 6 - Conclusão e Trabalhos Futuros:** apresentação das conclusões obtidas com o desenvolvimento do trabalho e com a análise dos resultados, bem como fazer sugestões para melhorar o trabalho em desenvolvimentos futuros.

2. CONCEITOS FUNDAMENTAIS

Este capítulo apresenta os conceitos que servirão como base para a compreensão do trabalho, divididos nos seguintes tópicos: *Pré-processamento Textual*, *Cálculo de Similaridade*, *Técnicas para comparação por similaridade de conceitos*.

2.1. Pré-processamento Textual

O pré-processamento é uma etapa de atividades que serve para preparar os dados antes de serem processados, com o intuito de facilitar a aplicação de algoritmos e métricas neles. A partir de cinco atividades, os dados são limpos e organizados de forma com que permita um processamento mais rápido e até mesmo com resultados mais satisfatórios, pois dependendo dos dados, podem haver ruídos, atrapalhando o processamento deles, ou gerando resultados errôneos. As atividades são:

Formatação dos caracteres para forma minúscula (ou maiúscula): formatação de todos os caracteres presentes nos textos para uma única forma.

Remoção de acentos e pontuação: remoção de acentos das palavras e da pontuação em geral do texto, permitindo uma análise mais padronizada.

Remoção de *stopwords*: retirada das palavras consideradas irrelevantes para a análise do texto. A remoção varia de acordo com o contexto de uso.

Tokenização: divisão do texto em partes atômicas, que podem ser as palavras, pontuação e números. Por exemplo, a seguinte frase “Hoje o céu está muito bonito” é dividida em várias unidades atômicas, resultando em: “[Hoje], [o], [céu], [está], [muito], [bonito]”.

Stemming: redução da palavra ao seu radical, que é a forma mínima de uma palavra, onde está localizado o seu sentido básico. Por exemplo, a palavra “Estudar”, seu radical é “Estud”, pois a partir dele é possível derivar outras palavras como “Estudando”, “Estudante” e “Estudo”.

Utilizando o pré-processamento, a análise baseada na similaridade entre as palavras das respostas dos alunos e das palavras-chave do professor, será facilitada.

2.2. Cálculo de Similaridade

O cálculo de similaridade é uma das tarefas realizadas durante o processo de agrupamento (*clustering*) de objetos, que consiste na análise com a intenção de verificar quais são as características comuns entre eles. O grau de semelhança (similaridade) entre dois objetos é dado pelo resultado da aplicação de uma função de similaridade, que analisa as características semelhantes entre eles e retorna um valor numérico, sendo esses objetos um conjunto de dados, como exemplo, o texto de uma resposta, o qual contém várias palavras, sendo consideradas as características daquele texto. Pode-se comparar um texto com outro de modo a verificar a similaridade entre eles a partir das palavras em comum. Para fazer esse cálculo de similaridade, é possível utilizar medidas e métricas, de diferentes tipos, sendo elas:

Medidas Estatísticas: verificam a similaridade entre objetos de acordo com o número de atributos semelhantes entre os objetos e o número total de atributos. Exemplos de medidas estatísticas são: Coeficiente de Jaccard (PONTES, 2015) e a Similaridade do Cosseno (PONTES, 2015).

Medidas de Distância: distribuem os objetos em um espaço euclidiano, onde cada ordem representa uma característica do objeto. Assim, a similaridade entre dois objetos é dada por meio da proximidade deles em uma projeção do espaço. Um exemplo de medida de distância é: Distância Euclidiana (WIVES, 1999).

Métricas Fuzzy: por meio da lógica difusa, é possível usar do raciocínio aproximado para obter a similaridade entre dois objetos. Avaliando o quão presente é uma certa característica nos objetos comparados, por graus de inclusão. Exemplos de métricas *fuzzy* são: Grau de Igualdade de Pesos (WIVES, 1999) e a Média por Operadores *Fuzzy* (WIVES, 1999).

2.3. Comparação por Similaridade de Conceitos

A partir do estudo de formas de calcular a similaridade entre objetos, é possível verificar que há certas técnicas que podem ser utilizadas para calcular a similaridade entre dois conceitos.

Uma das abordagens utilizadas para verificar essa similaridade é o uso da taxonomia dentro de uma rede de conhecimento, como o WordNet (MILLER, 1995). O processo é baseado em um grafo, no qual cada nó é a representação de um conceito e o arco é a relação semântica entre os conceitos interligados e, através de um valor atribuído nessa relação, é possível verificar a similaridade entre os dois conceitos. Isso é feito por cálculos com o uso de uma métrica de similaridade semântica apresentada na Figura 2.1.

Figura 2.1 - Definição da fórmula matemática de distância taxonômica.

$$sim(c1, c2) = 2\gamma - (\alpha + \beta)$$

Fonte: FERREIRA e FILIPE (2020, p. 2).

Nessa equação, é possível verificar a distância entre dois conceitos (c1 e c2) dentro de uma taxonomia. A hierarquia do WordNet é construída a partir de relações "IS-A" dos conceitos, logo, temos alguns atributos que influenciam a força dessa relação: o atributo gama(γ) é o valor do nó pai comum mais próximo dos conceitos, alfa(α) é o valor do nó em que está localizado o c1 e beta(β), o valor do nó de c2. Outra forma de apresentar a similaridade entre dois conceitos é por uma definição de um domínio, apresentada Figura 2.2:

Figura 2.2 - Mapeamento da similaridade entre dois conceitos.

$$\sigma : C1 \times C2 \rightarrow [0, 1]$$

Fonte: FERREIRA e FILIPE (2020, p. 2).

Onde o valor sigma(σ) é o valor quantitativo da similaridade entre dois conceitos (c1 e c2) mapeado em um intervalo de 0 a 1, sendo 0, para situações totalmente dissimilares e 1 para aquelas totalmente similares (RESNIK, 1995). Para calcular o valor sigma(σ) devemos utilizar a fórmula de distância entre dois conceitos, apresentada na Figura 2.3:

Figura 2.3 - Equação da medida de distância entre os conceitos.

$$\sigma(c1, c2) = 1 - \delta(c1, c2)$$

Fonte: FERREIRA e FILIPE (2020, p.3).

Como visto na Figura 2.3, o valor de similaridade é obtido a partir da subtração do número 1 pelo valor delta(δ), uma medida de distância que se dá a partir do cálculo da soma dos resultados de uma diferença entre o peso da menor distância entre os conceitos comparados e o peso da distância entre esses conceitos até a raiz da taxonomia. Quanto maior a distância entre os nós em que os conceitos estão localizados e do nó raiz da taxonomia, maior será o valor de delta(δ) e consequentemente, menor será o valor da similaridade entre esses dois conceitos.

Embora, o uso de taxonomias em redes de conhecimento apresentarem bons resultados para o cálculo de similaridade através de uma análise semântica, por essas técnicas se basearem em uma rede de conhecimento previamente estabelecida, certas barreiras podem ser geradas na hora de utilizar elas para a avaliação de questões discursivas, pois seria necessário estabelecer uma rede de conhecimento para cada pergunta gerada pelo professor,

inviabilizando o uso da técnica, pois ao invés de auxiliar o professor, poderia consumir mais tempo no planejamento dessa estrutura.

Por outro lado, temos algumas outras técnicas que não necessitam de uma estruturação prévia para sua utilização e podem ser utilizada para comparação por similaridade, são elas: as medidas de similaridade apresentadas no Capítulo 2.2 (o Coeficiente de Jaccard, o Coeficiente de Jaccard baseado no uso de sinônimos e a Similaridade do Cosseno), as técnicas de modelo de linguagem por N-gramas e o uso do algoritmo de predição KNN(*K-Nearest Neighbors*). Como visto anteriormente, antes de utilizar essas técnicas, as respostas dos alunos devem ser tratadas por meio do pré-processamento, que se dá pelas tarefas apresentadas no Capítulo 2.1.

Apresentando primeiramente as medidas, a primeira a ser abordada é o coeficiente de Jaccard, uma medida baseada na razão entre os elementos comuns e os diferentes entre dois conjuntos, A e B, conforme apresentado na Figura 2.4.

Figura 2.4 - Equação do coeficiente de Jaccard.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Fonte: PONTES (2015. p. 34).

Onde:

$A \cap B$ = soma dos elementos comuns entre os conjuntos A e B .

$A \cup B$ = soma de todos os elementos dos dois conjuntos .

O coeficiente de Jaccard visa calcular a similaridade entre dois conjuntos, mapeada em um intervalo de [0,1], sendo que 0 é a representação de total dissimilaridade entre os dois conjuntos e 1 é a representação total similaridade (igualdade) entre os dois. Como visto no trabalho de Siqueira (2019), é possível utilizar uma versão alternativa do coeficiente de Jaccard, a qual aborda o uso de sinônimos para verificar as palavras de significado semelhante presentes no texto, dependendo de uma API (*Application Programming Interface*) para adquirir os sinônimos das palavras que estão sendo processadas.

Há também a similaridade do Cosseno, uma medida baseada no cálculo do cosseno do ângulo formado pela projeção vetorial no primeiro quadrante do plano cartesiano dos dois conjuntos a serem comparados. Assim, o valor de similaridade é apresentado no intervalo de [0,1] semelhante ao coeficiente de Jaccard. A equação é apresentada na Figura 2.5.

Figura 2.5 - Equação da similaridade do Cosseno.

$$Cos(P, Q) = \frac{P \cdot Q}{\|P\| \times \|Q\|}$$

Fonte: PONTES (2015, p. 34).

O valor da similaridade do Cosseno é obtido através dos vetores P e Q, os quais contém em cada coordenada, uma palavra. Então, pode-se calcular a razão entre o produto escalar desses vetores pela multiplicação da norma desses vetores, verificando a presença ou ausência das palavras em cada conjunto, na qual a presença de uma palavra gera o valor 1 para aquela posição e a ausência gera valor 0.

Por fim, Caldas e Favero (2009) apresentaram ser possível verificar a similaridade entre conceitos através de um método de modelagem de linguagem, o n-gramas, baseado na transformação de uma palavra em *tokens*, sendo a transformação de uma única palavra em um unigrama até um número n de palavras, um n-grama. De acordo com a pesquisa realizada, os formatos bigramas e trigramas são os mais eficientes para se utilizar comparação por similaridade de conceito. Com o uso do algoritmo KNN é possível gerar uma predição de nota ao Mapa Conceitual do aluno em relação ao do professor.

Visto que o presente trabalho busca o uso dessas técnicas para averiguar quais respostas estão mais perto do contexto que o professor pediu, de acordo com os conceitos escolhidos pelo professor. O uso de métodos baseados na hierarquia WordNet são úteis, porém, muito complexos por dependerem da formação de uma estrutura hierárquica para cada pergunta feita pelo professor, além de que o WordNet voltado para o Brasil ainda é muito precário, sendo pobre em dados morfológicos. As técnicas baseadas no modelo de linguagem n-gramas com o algoritmo de predição KNN permitem uma avaliação semântica ao analisar bigramas e trigramas contendo conceitos, porém, seu uso é mais voltado para a avaliação de mapas conceituais, devido aos arcos de conexão pelos quais os conceitos são interligados e pelo trabalho estar focado na avaliação de questões discursivas curtas, também torna-o inviável. Por fim, a similaridade do Cosseno, que trabalha a partir de mapeamento angular, não é tão interessante ao trabalho. Já o coeficiente de Jaccard, é uma medida que a partir da comparação dos radicais após as etapas de pré-processamento, é possível verificar quais são os conceitos presentes no texto, mesmo que seja de forma simples, sem uma verificação de significado, mas apenas de escrita, além do fato de que é possível incrementar

essa medida em trabalhos futuros, permitindo que possam ser utilizados sinônimos para uma avaliação mais acurada.

3. REVISÃO DA LITERATURA

Neste capítulo, é apresentada uma breve análise de trabalhos relacionados à avaliação semiautomática e automática de questões discursivas e à comparação por similaridade de conceitos.

As pesquisas na área de avaliações assistidas pelo computador tiveram início em 1967, quando Page (1967) propôs o PEG (*Project Essay Grader*), um sistema desenvolvido com o intuito de avaliar questões discursivas de respostas curtas. O sistema é baseado na análise sintática da resposta, ou seja, há apenas a verificação da qualidade da escrita da resposta, sem considerar a parte semântica, que se refere ao conteúdo da mesma. Assim, nos casos em que a resposta não esteja relacionada ao tema, entretanto, esteja com a escrita gramaticalmente correta, o PEG não irá apontar nenhum problema com a resposta.

Em 1998 o avanço na área já estava mais marcante, surgindo como um dos trabalhos o *E-rater*, proposto por Burstein et al. (1998) e, então, melhorado em uma nova versão em 2001. O *E-rater* é um sistema de avaliação de questões discursivas, capaz de avaliar várias características da qualidade de escrita das respostas, como a gramática, a complexidade léxica e o uso de palavras específicas.

Leacock e Chodorow (2003) apresentaram o sistema *C-rater*, que possibilita correção automática de respostas curtas usando PLN. O sistema pode identificar conceitos que aparecem nas respostas das questões discursivas e também consegue identificar aqueles que não deveriam estar nas respostas. Por meio da análise da estrutura de predicados e argumentos com a análise morfológica dos conceitos, é atribuída uma nota às respostas. O *C-rater* utiliza o banco de palavras WordNet criado por Miller (1995), que permite verificar as estruturas hierárquicas das palavras.

Gütl (2007) apresentou o sistema e-Examiner, capaz de avaliar respostas curtas a partir de estimativas de similaridade entre as respostas dos alunos e a resposta referência do professor, através de um modelo de espaço vetorial.

Um pouco mais adiante no tempo, Caldas e Favero (2009) apresentaram uma proposta de ferramenta de avaliação automática de questões discursivas abertas, com enfoque no uso da abordagem de Mapas Conceituais, o que permite uma visão mais detalhada sobre o aprendizado do aluno. Através de técnicas de N-gramas e KNN (*K-Nearest Neighbors*), foi possível avaliar os Mapas Conceituais tanto de modo qualitativo, com a geração de relatórios

para auxiliar os alunos no desenvolvimento dos Mapas Conceituais, como de quantitativo, com uma análise comparativa em relação a avaliadores humanos.

Mohler e Mihalcea (2009) abordaram técnicas não supervisionadas para a correção automática de questões discursivas de respostas curtas. Eles compararam diferentes métodos de similaridade, como os de similaridade de conceitos (*WordNet Shortest Path*) e os de similaridade de texto baseado em *corpus* de palavras (*LSA - Latent Semantic Analysis* e *ESA - Explicit Semantic Analysis*). Também analisaram os efeitos causados pelo domínio e do tamanho de um texto, quando utilizadas as técnicas baseadas em *corpus* de palavras. Por fim, incorporaram ao sistema um método de *feedback* automático das respostas dos alunos.

Ávila e Soares (2012) desenvolveram uma ferramenta de apoio à correção de questões discursivas voltadas para os AVAs (Ambientes Virtuais de Aprendizagem). Para o desenvolvimento do sistema, foram utilizadas técnicas de comparação de frases e métodos de pré-processamento de texto para aumentar o desempenho de índices de similaridade.

O trabalho de Pissinati (2014) apresentou uma proposta de um sistema de predições semiautomáticas de notas para questões discursivas. Foram utilizadas técnicas de pré-processamento de texto e de cálculo de similaridade entre documentos por um modelo vetorial. Além disso, o sistema é capaz de analisar a semântica das respostas. Para complementar a proposta, foram utilizadas técnicas de visualização de informação para apresentar os dados processados pelo algoritmo.

Pontes (2015) utilizou métodos de PLN e heurísticas baseadas em grafos e matrizes para resumir textos de forma automática, usando como princípio a verificação de certas sentenças do texto por meio da similaridade de conceitos, através de algumas métricas como Distância do Cosseno e Coeficiente de Jaccard e pela identificação dos conceitos mais importantes na sentença.

Spalenza (2018) deu continuação ao trabalho apresentado por Pissinati (2014), aprimorando certos aspectos do sistema de avaliação semiautomática, desenvolvido anteriormente. Uma das melhorias foi o uso do modelo de seleção de características, permitindo que haja uma redução no esforço necessário para a avaliação, além da possibilidade de identificar os termos relevantes para a compreensão da resposta. A parte de visualização de informação também foi melhorada com o uso da técnica de Mapa de Características.

O estudo de Siqueira (2019) trabalha com a comparação entre várias métricas de modo a verificar qual é o método mais interessante de se avaliar textos por meio da comparação por similaridade de conceitos, utilizando duas categorias de métricas: as

baseadas na sintaxe e as baseadas na semântica. Foi concluído que as semânticas são mais precisas, embora levem mais tempo de processamento, pois requerem mais etapas de pré-processamento, diferentemente das sintáticas, que abordam apenas a forma em que um texto é escrito, comparando elementos comuns entre elas.

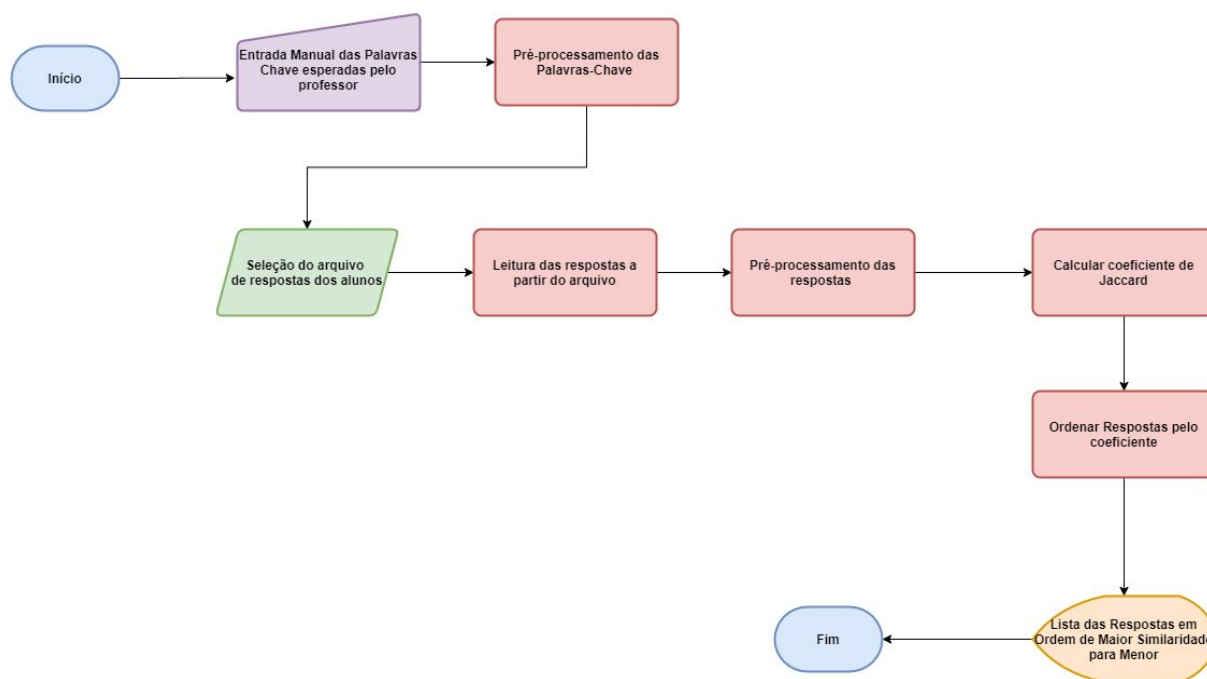
De acordo com o exposto, alguns trabalhos utilizam métodos para fazer a avaliação automática de questões discursivas. Esses métodos são mais complexos de serem implementados, considerando a necessidade de garantir a acurácia dos resultados, por se proporem a substituir a avaliação realizada pelo professor. Assim, a exposição deles se dá para apresentar técnicas que podem ser viáveis para calcular a similaridade entre conceitos mesmo de forma semiautomática, não substituindo o professor, mas sim o auxiliando no processo de avaliação das respostas.

4. ARQUITETURA DO PROTÓTIPO DO SISTEMA

Neste capítulo, são apresentados os detalhes da arquitetura do protótipo do sistema para avaliação semiautomática de questões discursivas proposto por este trabalho, desenvolvido utilizando a linguagem *Python* (ROSSUM, 1989).

O fluxo de processamento do protótipo é apresentado no fluxograma da Figura 4.1.

Figura 4.1 - Arquitetura do protótipo do sistema.



Fonte: O próprio autor.

Como pode ser observado na Figura 4.1, como entrada, o sistema receberá as palavras-chave que o professor deseja e espera que estejam contidas nas respostas dos alunos, bem como as respostas propriamente ditas. As palavras-chave são informadas diretamente no sistema, de forma manual. Em relação às respostas dos alunos, por se tratarem de várias, a inserção manual se torna inviável. Logo, o sistema permite carregá-las a partir de um arquivo do tipo CSV (*Comma-Separated Values*). Através do explorador de arquivos, é possível selecionar o arquivo com extensão .csv em que as respostas estão armazenadas, sendo que cada resposta deve estar associada à identificação do aluno. Desse modo, o arquivo CSV deve conter duas colunas, a primeira sendo os identificadores dos alunos e a segunda contendo as respostas.

Figura 4.2 - Estrutura obrigatória do arquivo de respostas.

	A	B	C
1	id	resposta	
2		1 resposta do aluno de RA 1	
3		2 resposta do aluno de RA 2	
4		3 resposta do aluno de RA 3	
5		...	...
6		N resposta do aluno de RA N	

Fonte: O próprio autor.

Ainda de acordo com a Figura 4.1, após a escolha do arquivo, o sistema faz a leitura de cada resposta, com a identificação do aluno (por exemplo, o Registro Acadêmico - RA). Finalizada a entrada de dados, as palavras-chave e as respostas dos alunos são pré-processadas usando o conjunto de tarefas de Pré-processamento de textos (Capítulo 2.1). Algumas atividades, como o processo de *stemming* e a remoção dos *stopwords*, foram desenvolvidas usando métodos disponíveis na biblioteca *NLTK* (*Natural Language ToolKit*) da linguagem *Python*, que realiza várias operações de pré-processamento. A formatação e a remoção dos caracteres especiais se deu pela biblioteca *unicodedata*. O processo de tokenização foi realizado com funções próprias do *Python*.

A próxima etapa do fluxo do protótipo, é realizar o cálculo da similaridade pelo coeficiente de Jaccard. Feito isto, o resultado será, então, apresentado de forma ordenada, começando pela resposta de maior similaridade para a de menor similaridade em relação às palavras-chave.

A *GUI* (*Graphical User Interface*) do protótipo foi modelada a partir da biblioteca *Tkinter* da linguagem *Python*.

O projeto completo do protótipo pode ser encontrado na plataforma de repositório *GitHub* no link: <https://github.com/brunoAdegas/TCCBrunoAdegas2020FT>

Por limitações de tempo e de desenvolvimento, o protótipo foi desenvolvido de forma a não ser capaz de lidar com a análise semântica de conceitos, para isso é necessária uma melhoria na forma em que a avaliação do sistema é feita para trabalhar com esse tipo de análise. Entretanto, para demonstrar a viabilidade da proposta desenvolvida neste trabalho, foi considerado para o protótipo realizar apenas análises de similaridade de escrita entre conceitos limitados a uma palavra, ou seja, palavras-chave, as quais o professor irá inserir no

protótipo, com o intuito de verificar se as respostas dos alunos possuem essas palavras-chave, por meio do radical das palavras.

5. ESTUDO DE CASO

Neste capítulo, é apresentado um estudo de caso de modo a apresentar o funcionamento do protótipo do sistema e também avaliar a solução proposta por este trabalho por meio da análise dos resultados.

5.1. Respostas utilizadas

A questão utilizada foi retirada da prova de História da segunda fase do Vestibular da Unicamp do ano de 2017, apresentada na Figura 5.1.

Figura 5.1 - Enunciado da questão de história utilizada como exemplo.

Questão 7

“Onde está aquela tua prudência? Onde está a sagacidade nas coisas que se devem discernir? Onde está a grandeza de alma? Já as pequenas coisas te afligem? (...) Nenhuma destas coisas é insólita, nenhuma inesperada. Ofender-te com estas coisas é tão ridículo quanto te queixares porque caíste em público ou porque te sujaste na lama. (...) O inverno faz vir o frio: é necessário gelar. O tempo traz de novo o calor: é necessário arder. A intempérie do céu provoca a saúde: é necessário adoecer. Uma fera em algum lugar se aproximará de nós, e um homem mais pernicioso que todas as feras. Algo a água, algo o fogo nos retirará. Esta condição das coisas não podemos mudar. Mas isto podemos: adotar um espírito elevado e digno do homem nobre para que corajosamente suportemos as coisas fortuitas e nos harmonizemos com a Natureza.”

(Sêneca, Carta de Sêneca a Lucílio, CVII. *Prometeus*, Maceió, ano 1 - n°1, p.121, jan.-jun. 2008. Disponível em <https://www.academia.edu/4204064>. Acessado em 19/12/2016.)

A partir da leitura do texto escrito pelo filósofo Sêneca,

- a) identifique e explique um princípio do estoicismo latino;
- b) cite dois legados culturais do mundo romano, além da filosofia, para a tradição ocidental.

Fonte: COMVEST 2017.

Para exemplificar o uso do sistema, será considerado o item b da questão apresentada na Figura 5.1, que tem a seguinte resposta esperada:

“As contribuições romanas em termos de legado cultural para o Ocidente incluem, entre outras, as artes (em especial a escultura e a arquitetura), o direito e o latim”.

Considerou-se, então, que as palavras-chave esperadas são:

Legado, Ocidente, Arte, Escultura, Arquitetura, Direito e Latim.

A partir da resposta esperada, foram desenvolvidas quatro respostas diferentes pelo próprio autor, com o intuito de verificar o comportamento do protótipo.

Resposta do Aluno 1 – “As contribuições do mundo romano foram a arte, a arquitetura, o direito e o latim”.

Resposta do Aluno 2 – “A roma desenvolveu a arte de uma forma muito diferente por causa do latim”.

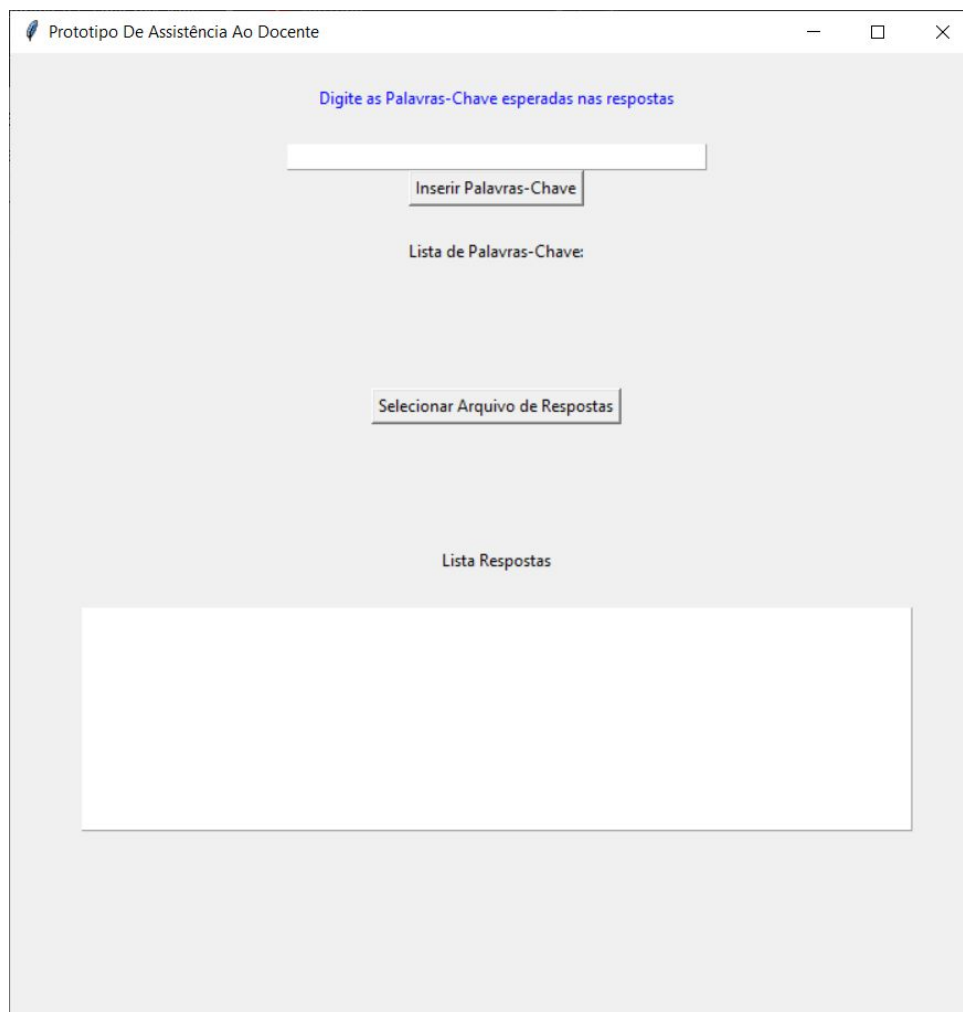
Resposta do Aluno 3 – “Estudos indicam que a maior contribuição romana para o ocidente foi a arte, a política e o latim”.

Resposta do Aluno 4 – “Roma foi uma das fontes importantes para o desenvolvimento filosófico e político que existe atualmente no Ocidente, graças ao direito e a arte do ócio”.

5.2. Execução do protótipo

Ao iniciar a execução do protótipo, o usuário irá se deparar com a GUI, como apresentada na Figura 5.2.

Figura 5.2 - Exibição inicial da GUI do protótipo.



Fonte: O próprio autor.

Primeiramente, o usuário deve digitar as palavras-chave que espera encontrar nas respostas dos alunos, separadas por um espaço e, então, clicar no botão “Inserir Palavras-Chave” para confirmar a ação. Desse modo, as palavras-chave serão pré-processadas e serão exibidas já com as modificações necessárias no frame “Lista de Palavras-Chave”, conforme mostrado na Figura 5.3.

Figura 5.3 - Exemplo de inserção de palavras-chave no protótipo.

Prototipo De Assistência Ao Docente

Digite as Palavras-Chave esperadas nas respostas

Legado Ocidente Arte Escultura Arquitetura Direito Latim

Inserir Palavras-Chave

Lista de Palavras-Chave:

art arquitet direit latim ocid escult leg

Selecionar Arquivo de Respostas

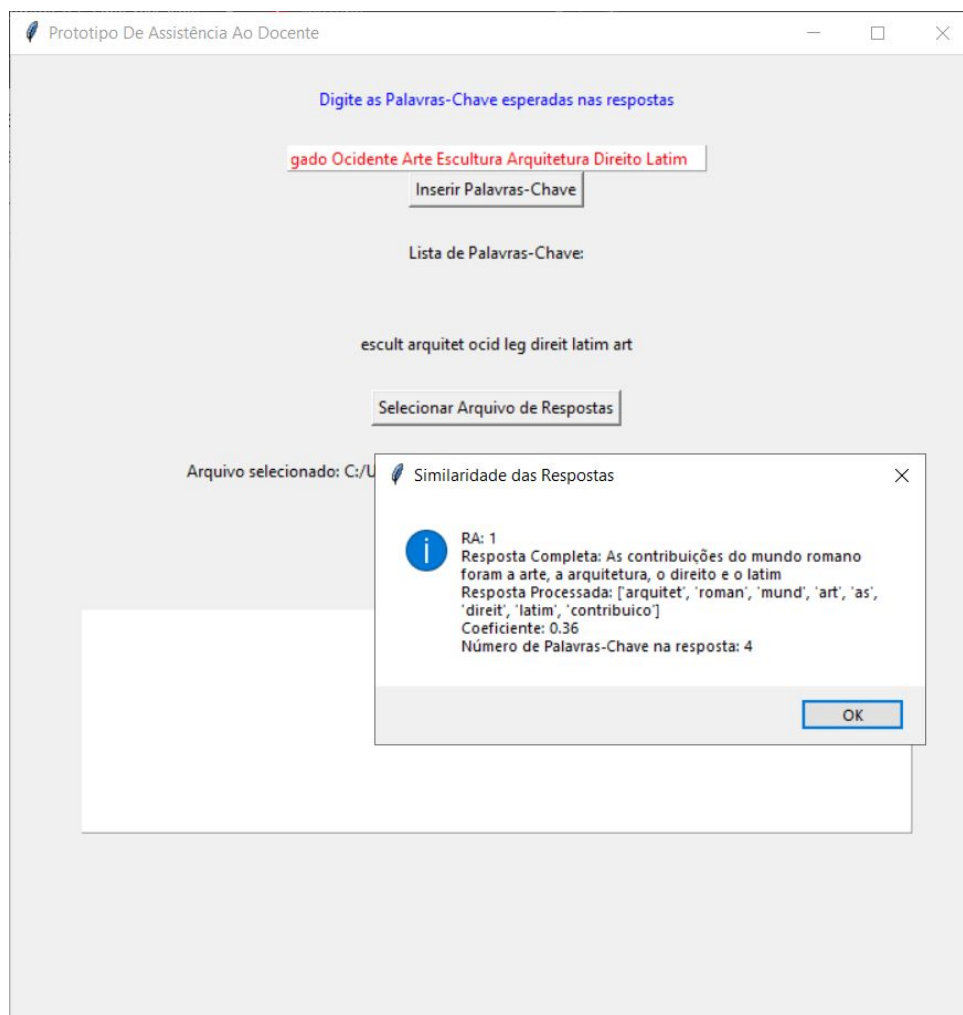
Arquivo selecionado:

Lista Respostas

Fonte: O próprio autor.

Para prosseguir, o usuário deve clicar no botão “Selecionar Arquivo de Respostas” de modo a abrir o explorador de arquivos do Windows, onde ele poderá procurar o arquivo CSV que contém as respostas dos alunos. Após ter selecionado o arquivo, o programa processa as respostas e apresenta uma tela *pop-up* para cada uma, apresentando seus dados relevantes, como o número identificador do aluno (RA), as respostas (completa e a processada), o valor do coeficiente de Jaccard e o número de palavras-chave presentes, conforme mostrado na Figura 5.4.

Figura 5.4 - Tela pop-up das respostas.



Fonte: O próprio autor.

Após apresentar a janela *pop-up* para todas as respostas, elas são ordenadas de acordo com o coeficiente de Jaccard de forma decrescente e então apresentadas através de uma lista no frame “Lista Respostas”, como apresentado na Figura 5.5.

Figura 5.5 - Exibição das respostas por uma lista ordenada.

Lista Respostas				
RA: 1	Resposta: As contribuições do mundo romano foram a arte, a arquitetura, o direito e o latim	Coefficiente: 0.36	Número Palavras-Chave: 4	<----->
RA: 3	Resposta: Estudos indicam que a maior contribuição romana para o ocidente foi a arte, a política e o latim	Coefficiente: 0.23	Número Palavras-Chave: 3	<----->
RA: 4	Resposta: Roma foi uma das fontes importantes para o desenvolvimento filosófico e político que existe atualmente	Coefficiente: 0.18	Número Palavras-Chave: 3	<----->
RA: 2	Resposta: A roma desenvolveu a arte de uma forma muito diferente por causa do latim	Coefficiente: 0.15	Número Palavras-Chave: 2	<----->

Fonte: O próprio autor

Como visto na Figura 5.5, a resposta processada foi retirada da apresentação para ficar uma apresentação mais resumida. Um ponto a se destacar é o fato das respostas muito longas não caberem totalmente na tela e também não sofrerem quebra de linha. Isso se deve ao frame utilizado no desenvolvimento para exibir a lista de respostas ordenada, para ler o que não está na tela, é possível fazer um rolamento, tanto horizontal como vertical para conseguir ler o restante dos dados.

5.3. Análise dos Resultados

O experimento de avaliação do método proposto pelo trabalho no estudo de caso se demonstrou satisfatório. O resultado do processamento das respostas pode ser visto na Tabela 5.1, ordenado pelo RA do aluno.

Tabela 5.1- Resultado do processamento do sistema para o estudo de caso.

RA	Resposta	Número de Palavras-chave $ A \cap B $	Número total de palavras* $ A \cup B $	Coefficiente de Jaccard
1	As contribuições do mundo romano foram a arte, a arquitetura, o direito e o latim.	4	11	0,36
2	A roma desenvolveu a arte de uma forma muito diferente por causa do latim.	2	13	0,15

Continua

RA	Resposta	Número de Palavras-chave $ A \cap B $	Número total de palavras* $ A \cup B $	Coefficiente de Jaccard
3	Estudos indicam que a maior contribuição romana para o ocidente foi a arte, a política e o latim.	3	13	0,23
4	Roma foi uma das fontes importantes para o desenvolvimento filosófico e político que existe atualmente no Ocidente, graças ao direito e a arte do ócio.	3	17	0,18

Fonte: Elaborada pelo próprio autor com base nos resultados da execução do estudo de caso.

* Número total de palavras ($|A \cup B|$) gerado pela união do conjunto Q de conceitos inseridos pelo professor e do conjunto X de palavras, após o pré-processamento das respostas.

A resposta do aluno 1 foi a que apresentou o maior índice de similaridade. Isso se deve ao fato de ser objetiva, apresentando as palavras-chave que o professor procura sem adicionar frases inúteis, que apenas aumentam o tamanho da resposta. As respostas dos alunos 3 e 4, apesar de apresentarem o mesmo número de palavras-chave, os valores de similaridade foram distintos. Pode-se considerar o mesmo princípio utilizado na análise da resposta do aluno 1, ou seja, a resposta do aluno 4 apresentou mais palavras (inúteis) em seu desenvolvimento do que a resposta do aluno 3. Como o coeficiente de Jaccard compara as palavras-chave (informadas pelo professor) com o conjunto formado pela união de todas as palavras presentes na resposta do aluno (após o pré-processamento) e as próprias palavras-chave. Logo, uma resposta que contenha menos palavras-chave terá um índice menor. Em relação à resposta do aluno 2, que teve o menor índice e apresentou o menor número de palavras-chave esperadas, pode ter seu resultado comprovado pelo professor ao fazer uma análise secundária, verificando que a resposta está quase totalmente fora de contexto.

Um ponto a ressaltar é que o protótipo não desenvolveu nenhum mecanismo para avaliar a ordem da escrita das palavras, nem do significado geral da resposta, focando apenas na similaridade de escrita entre as palavras contidas na resposta do aluno e as palavras-chave requeridas pelo professor. Assim, mesmo que respostas diferentes contenham um mesmo valor de similaridade, elas podem estar erradas devido à forma que foram escritas ou até mesmo o significado geral da resposta que não condiz com o pedido. Para avaliar esses componentes é necessário o desenvolvimento de métodos de análise sintática e semântica.

6. CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho teve como objetivo a proposta de um sistema que possa auxiliar professores na avaliação de questões discursivas de respostas curtas. Os resultados obtidos por meio do estudo de caso apresentado se mostraram satisfatórios. Eles mostraram que as respostas mais objetivas e que continham o que o professor esperava tiveram um índice de similaridade mais alto do que as respostas mais extensas e/ou que não continham o que o professor esperava. A análise de respostas mais longas apresenta um índice de similaridade mais baixo pelo fato de estarem fugindo do contexto da resposta do professor, utilizando palavras desnecessárias, sem objetividade. Entretanto, ainda seria possível observar se estão no contexto requerido de acordo com o valor do índice e número de palavras-chave retornados em comparação com as outras respostas. Mesmo assim, caso um aluno escreva uma resposta que não responda à pergunta do professor, mas contenha as palavras-chave que ele deseja, o sistema vai apontar um valor alto de similaridade, sendo assim, deve-se ter cautela por parte do professor ao fazer uma análise secundária das respostas.

Durante o desenvolvimento do trabalho, foi verificado que o pré-processamento é uma etapa fundamental para permitir uma análise de similaridade entre as palavras-chave e as respostas de forma eficiente, utilizando nessa comparação, apenas o radical das palavras e aquelas que não trazem utilidade para a análise são descartadas. Permitir com que o sistema utilize operações de leitura de arquivo permitiu com que as respostas de todos os alunos fossem processadas e avaliadas de uma vez, tornando o protótipo mais funcional para a rotina dos professores. Com isso, além da facilidade de processar tudo de uma vez, o sistema retorna ao professor uma lista ranqueada das respostas de acordo com a similaridade, permitindo com que ele possa fazer até análises entre as próprias respostas dos alunos para tirar uma conclusão.

Outros aspectos relacionados à avaliação das respostas de questões discursivas também foram observados durante o desenvolvimento deste trabalho, mas por limitações, ficando como recomendações para trabalhos futuros: tratamento efetivo de conceitos pelo sistema, uma vez que foram consideradas apenas palavras-chave, ou seja, conceitos compostos por mais de uma palavra não foram contemplados; análise sintática-semântica das respostas, permitindo verificar a influência que a ordem de escrita das palavras traz em uma sentença e ao significado geral por trás da frase e entender que um mesmo significado pode ser apresentado de diversas formas, tendo respostas feitas com palavras totalmente diferentes, mas que possuem a mesma semântica; para complementar a última sugestão, o uso de

sinônimos para a comparação, permitindo termos com significados idênticos ou muito semelhantes possam ser considerados para o cálculo de similaridade.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ATTALI, Yigal; BURSTEIN, Jill. **Automated essay scoring with e-rater® V. 2.** The Journal of Technology, Learning and Assessment, v. 4, n. 3, 2006.

CALDAS, Vanessa M.; FAVERO, Eloí L. **Uma ferramenta de avaliação automática para mapas conceituais como auxílio ao ensino em ambientes de educação a distância.** In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), 2009.

CONOLE, Graíne.; Warburton, Bill. **A review of computer-assisted assessment.** Research in Learning Technology, Co-Action Publishing, v. 13, n. 1, p. 17–31, 2005.

DE ÁVILA, Ricardo LF.; SOARES, José M. **Concepção de ferramenta de apoio à correção de questões dissertativas com base na adaptação de algoritmos de comparação e busca textual combinados com técnicas de pré-processamento de textos.** RENOTE-Revista Novas Tecnologias na Educação, v. 10, n. 3, 2012.

FERREIRA, João C.; FILIPE, Porfírio P. **Similaridade Semântica Baseada em Taxionomias.** Artigo Científico, Instituto Superior de Engenharia de Lisboa, 2020.

GÜTL, Christian. **e-Examiner: towards a fully-automatic knowledge assessment tool applicable in adaptive e-learning systems.** In: Proceedings of the Second International Conference on Interactive Mobile and Computer Aided Learning, p. 1-10, 2007.

LEACOCK, Claudia; CHODOROW, Martin. **C-rater: Automated scoring of short-answer questions.** Computers and the Humanities, v. 37, n. 4, p. 389-405, 2003.

MILLER, George A. **WordNet: a lexical database for English.** Communications of the ACM, v. 38, n. 11, p. 39-41, 1995.

MOHLER, Michael; MIHALCEA, Rada. **Text-to-text semantic similarity for automatic short answer grading.** In: Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009), p. 567-575, 2009.

PAGE, Ellis B. **Grading essays by computer: Progress report.** Proceedings of the Invitational Conference on Testing Problems, Princeton, p. 87–100, 1967.

PISSINATI, Eduardo M. **Uma Proposta de Correção Semi-Automática de Questões Discursivas e de Visualização de Atividades para Apoio a Atuação do Docente.** Dissertação de Mestrado, Universidade Federal do Espírito Santo, 2014.

PONTES, Elvys L. **Utilização de Grafos e Matriz de Similaridade na Sumarização Automática de Documentos Baseada em Extração de Frases**. Dissertação de Mestrado, Universidade Federal do Ceará, 2015.

RESNIK, Philip. **Using information content to evaluate semantic similarity in a taxonomy**. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1995.

ROSSUM, Guido V. **Python Language**. CWI - Centrum Wiskunde & Informatica, 1989.

SAXENA, Sachin; GUPTA, Poonam R. **Automatic assessment of short text answers from computer science domain through pattern based information extraction**. Proceeding of ASCNT, p. 109–118, 2009.

SIQUEIRA, Kaique Chiovetto. **Análise e Implementação de Algoritmos de Comparação por Similaridade de Conceitos em Mapas Conceituais**. Trabalho de Conclusão de Curso, Faculdade de Tecnologia da Unicamp, Limeira, 2019.

SPALENZA, Marcos A. **Aplicação de Seleção de Características em um Sistema Semiautomático de Avaliação de Questões Discursivas**. Dissertação de Mestrado, Universidade Federal do Espírito Santo, 2018.

WIVES, Leandro Krug; DE OLIVEIRA, José Palazzo M. **Agrupamento de Informações Textuais**. Dissertação de Mestrado, Porto Alegre: Instituto de Informática–UFRGS, 1997.