

Processamento de Linguagem Natural em Artefatos Textuais Educacionais: Um Mapeamento Sistemático no Contexto Brasileiro

Anderson Felinto Barbosa¹, Cláudio E. C. Campelo¹

¹Departamento de Ciência da Computação – Universidade Federal de Campina Grande (UFCG) – Campina Grande – PB – Brasil

anderson.fbarbosa@gmail.com, campelo@dsc.ufcg.edu.br

Abstract. *Independent of the learning approach, normally, students create textual artifacts or interact in conversational environments that require some kind of text production. However, the analysis of produced artifacts is time consuming and demands significant effort from teachers and tutors. On the other hand, there has been considerable recent advances in the field of Natural Language Processing (NLP) that might help address different issues in the area of Education. This paper presents a secondary research with the objective of characterizing how NLP has been applied to textual artifacts produced by students, with a focus on the Brazilian education. We conducted searches in 7 digital libraries that resulted in 47 candidate studies, from which we selected 12 for a deeper analysis.*

Resumo. *Independente da abordagem de ensino, comumente, os discentes criam artefatos textuais ou interagem em ambientes conversacionais que exigem a produção de textos. Por sua vez, a análise desses artefatos/atividades pode demandar tempo e esforço para professores e tutores. Diante disso, esse artigo apresenta a condução de um estudo secundário com a finalidade de caracterizar como o Processamento de Linguagem Natural é aplicado a artefatos textuais produzidos por alunos no contexto da Informática na Educação no Brasil. Foram realizadas buscas em 7 bibliotecas digitais, que resultou em 47 publicações candidatas, sendo selecionadas 12 para análises.*

1. Introdução

No contexto da educação presencial ou mediada por computador, é comum que o aluno participe de atividades que promovam, diretamente ou indiretamente, a construção de artefatos textuais que serão utilizados no processo avaliativo. Na educação mediada por computador, as interações são realizadas através de ambientes que facilitem a comunicação entre os participantes – normalmente fóruns e ambientes de bate-papo (*Chats*) [de Lima et al. 2011]. Por suas vezes, as atividades, quase na sua totalidade, são realizadas através da escrita de artefatos textuais. O modelo presencial não difere desse, porém, é mais comum que as interações sejam realizadas através de fala (e.g., debates, discussões ou seminários). Contudo, no modelo presencial, também há produção de atividades textuais.

Em ambos os modelos, costuma ser de responsabilidade do professor a avaliação do que é discutido nas interações e artefatos produzidos que, em número considerável,

pode demandar tempo e esforço por parte do professor. Logo, o uso de mecanismos (semi) automatizados pode reduzir o esforço realizado nesse processo.

O Processamento de Linguagem Natural (PLN) é uma área da Inteligência Artificial que tem como um dos objetivos proporcionar aos computadores o entendimento dos textos escritos ou falados. Através de algoritmos que promovem o entendimento e geração de linguagem natural, o PLN pode prover recursos importantes para automatizar processos realizados manualmente por humanos. Para isso, um conjunto de tarefas, com auxílio de técnicas, são realizadas para alcançar o objetivo da tarefa em questão [Indurkha and Damerau 2010]. Logo, observa-se que o PLN pode contribuir nas atividades realizadas por professores durante a avaliação dos artefatos de texto ou fala criado por alunos no ambiente educacional.

Diante disso, por meio de um Mapeamento Sistemático da Literatura (MSL), este trabalho visa investigar o uso de técnicas de Processamento de Linguagem Natural em artefatos textuais no contexto da Informática na Educação brasileira. Diferentemente do estudo realizado por [Ferreira et al. 2017], este não se limita ao uso de PLN em fóruns, dando cobertura a outros artefatos textuais. Além disso, não são considerados relatos de uso de uma ferramenta baseada em PLN, sendo isso um critério de exclusão, diferenciando-se do trabalho citado. Assim, este MSL foi planejado e conduzido utilizando as bibliotecas digitais do CTRL+E, ENIAC, RBIE, RENOTE, SBIE, WCBIE e WIE, que resultaram em 47 publicações candidatas, sendo 12 foram selecionadas para responder às questões de pesquisa.

O restante deste artigo está organizado da seguinte forma. Na Seção 2 é apresentado o uso de Estudos Secundários em Informática na Educação; na Seção 3, é apresentado o planejamento do MSL; a Seção 4 apresenta a análise dos dados em resposta às questões de pesquisa e, por fim, a Seção 5 apresenta as considerações finais do trabalho.

2. Estudos Secundários em Informática na Educação

A realização de Estudos Secundários é uma prática comum entre os pesquisadores. Estes estudos, dependendo da sua finalidade, têm como objetivo caracterizar uma área de pesquisa (Mapeamento Sistemáticos da Literatura - MSL) ou identificar, interpretar, avaliar e comparar experiências realizadas em um determinado tópico de pesquisa (Revisão Sistemática da Literatura - RSL) [Cabrejos et al. 2018].

Na área da Informática na Educação, esta prática é comum, principalmente, quando observa-se a publicação desses estudos com destaque para o planejamento, condução e análise dos resultados em eventos da área. No Brasil, observa-se uma diversidade desses estudos, o que auxilia pesquisadores no entendimento das áreas de pesquisa e na definição de possíveis contribuições. Ressalta-se, ainda, que esses estudos podem ser conduzidos de forma mais restrita ou abrangente, a depender do seu planejamento.

Em [Aguiar et al. 2014] é apresentada a condução de um estudo secundário de forma mais abrangente. Com o objetivo de averiguar a evolução histórica de publicações que contemplem Sistemas de Recomendação Educacional (SRE) e como são realizadas as avaliações desses sistemas, os autores, conduziram a pesquisa utilizando apenas o termo “*recom**”, ou seja, o radical do termo “Recomendação”. Tal planejamento proporcionou a consulta por trabalhos indexados em 4 línguas contempladas pelos veículos utilizados

para busca. Como resultado, 333 publicações foram retornadas, porém, 34 foram selecionadas para análise após aplicação dos critérios de seleção.

No trabalho de [dos Santos et al. 2019], por exemplo, é apresentado um exemplo da realização de um estudo com escopo mais reduzido, através da condução de um Mapeamento Sistemático da Literatura envolvendo as áreas de Processamento de Linguagem Natural (PLN) e Mapas Conceituais (MC) com a finalidade de identificar como o PLN pode contribuir para a criação dos MC. Inicialmente, 438 publicações foram retornadas, porém, após a aplicação de critérios de seleção, 23 foram selecionadas para análises.

3. Mapeamento Sistemático da Literatura

O planejamento utilizado nesse Mapeamento Sistemático da Literatura (MSL) baseou-se no processo apresentado por [Cabrejos et al. 2018]. Nessa seção, são apresentadas as questões de pesquisas que norteiam o estudo, a definição das estratégias de busca e seleção das publicações e, por fim, é descrito o processo de extração dos dados para análises.

3.1. Questões de Pesquisa

A Questão Principal (QP) que norteia este estudo centra-se na necessidade de investigar **“Como as técnicas de Processamento de Linguagem Natural têm sido empregadas nos artefatos textuais produzidos na educação presencial e/ou mediada por computador”**. A partir dessa questão, foram derivadas algumas Questões de Pesquisa Secundárias (QS), conforme mostra a Tabela 1.

3.2. Estratégia de Busca e Critérios de Seleção

Na definição da *string* de busca foram considerados, principalmente, os termos “Artefatos Textuais” e “Processamento de Linguagem Natural”, definidos na QP. Embora a população, “educação presencial e/ou mediada por computador”, não tenha sido especificadas nos termos, a delimitação das bibliotecas digitais exclusivas da IE resultará na cobertura da mesma. Após testes, foram definidos os termos e construída a *string* de busca com o uso de expressões booleana, resultando em: *((((Processamento OR Mineração) AND (“Linguagem Natural” OR “Língua Natural” OR Text*)) OR PLN OR NLP) AND ((Produção AND Text*) OR Escrita OR Redaç* OR Diálogo OR Monólogo OR Debate OR Discussão OR Mensag* OR Chat* OR Fórum OR Foruns))*. Também foram considerados termos equivalentes em inglês.

Para a execução das buscas automáticas, foram utilizadas as bibliotecas digitais brasileiras que indexam as publicações do **Simpósio Brasileiro de Informática na Educação**¹, **Workshop de Informática na Escola**², **Revista Brasileira de Informática na Educação**³, **Revista Novas Tecnologias na Educação**⁴ e **Workshops do Congresso Brasileiro de Informática na Educação**⁵. Ressalta-se que não foram utilizados artigos de controle durante a execução dos testes, porém, os títulos das publicações retornadas na primeira página foram analisados a cada execução.

¹<http://www.br-ie.org/pub/index.php/sbie/issue/archive>

²<https://www.br-ie.org/pub/index.php/wie/issue/archive>

³<https://www.br-ie.org/pub/index.php/rbie/issue/archive>

⁴<https://seer.ufrgs.br/renote/issue/archive>

⁵<https://www.br-ie.org/pub/index.php/wcbie/issue/archive>

Tabela. 1. Questões de Pesquisa Secundárias

Nº	Pergunta	Justificativa
(1)	Qual a quantidade de estudos realizados por ano e biblioteca?	Espera-se apresentar uma análise das buscas realizadas, com destaque para quantidade de publicações por biblioteca e ano.
(2)	Qual o objetivo principal dos estudos realizados?	Tal identificação proporcionará uma visão da contribuição do uso da PLN aplicado ao contexto educacional, nos estudos selecionados.
(3)	Em qual nível educacional o trabalho foi realizado?	O planejamento do MSL não delimitou escopo em nível de ensino (fundamental, médio ou superior). Logo, essa questão mostrará, se possível, qual o nível de ensino abordado na realização do trabalho.
(4)	Qual a abordagem de ensino na qual o artefato foi aplicado?	Essa questão visa identificar se o estudo é aplicado na EaD ou presencial/suporte presencial.
(5)	Qual o artefato textual utilizado?	Diferentes artefatos são produzidos no contexto educacional. Essa questão apresentará os artefatos analisados nos estudos selecionados.
(6)	Quais as técnicas de PLN empregadas?	A literatura apresenta uma diversidade de técnicas de PLN. Logo, como resposta, espera-se destacar as técnicas empregadas no processamento dos artefatos textuais.
(7)	O estudo apresenta alguma contribuição para a área de PLN? Se sim, qual?	Esta pergunta tem como finalidade verificar se o estudo apresenta alguma contribuição área de PLN. Caso positivo, a contribuição será apresentada.

Além das buscas automáticas, descrita anteriormente, foram consideradas buscas manuais nos anais do **Congresso sobre Tecnologias na Educação**⁶ (Ctrl+e) e **Encontro Nacional de Inteligência Artificial e Computacional**⁷ (ENIAC), totalizando 7 bibliotecas digitais. Para a condução das buscas manuais, foi realizada a leitura de título e resumo dos estudos que apontavam o uso de PLN ao contexto educacional.

A seleção das publicações identificadas nas buscas baseou-se nas análises de critérios de inclusão (CI) e de exclusão (CE). Quanto aos CI, considerou-se: (i) a publicação aborda o uso de técnicas de PLN em artefatos textuais diretamente relacionado ao escopo do estudo; e (ii) os artefatos textuais devem ser produzidos pelos discentes. Em relação aos CE, foram definidos: (i) publicação não disponível por motivos de indisponibilidade da biblioteca ou necessidade de pagamento; (ii) publicação duplicada, ou seja, já

⁶Ctrl+e 2016 (<http://ceur-ws.org/Vol-1877>), Ctrl+e 2017 (<http://ceur-ws.org/Vol-1667>), Ctrl+e 2018 (<http://ceur-ws.org/Vol-2185>) e Ctrl+e 2019 e 2020 (<https://sol.sbc.org.br/index.php/ctrl+e/issue/archive>)

⁷A indisponibilidade da Biblioteca Digital Brasileira de Computação, impossibilitou consultas de 2010 – 2017. Os anais de 2018 e 2019 estão disponíveis em: <https://sol.sbc.org.br/index.php/eniac/issue/archive>

selecionada por outra base ou abordada em outra publicação de maneira semelhante; (iii) a publicação não descreve o uso de técnicas de PLN em artefatos textuais ou apenas cita o uso de uma ferramenta de processamento de textos; (iv) trabalho não publicado no período entre 2010 – 2020; e (v) a publicação não é um artigo científico (estudo primário) ou é apresentado de forma resumida. Ressalta-se que a aplicação desses critérios ocorreram em duas fases: leitura de títulos e resumo, e leitura completa das publicações candidatas. Como resultado, tem-se o conjunto de publicações que são analisadas para responder às questões de pesquisa.

3.3. Estratégia para Extração dos Dados

As publicações selecionadas são catalogadas com base nos dados da publicação (ano, título, veículo de publicação e URL) e são analisadas para responder às questões de pesquisa. Um formulário⁸ é utilizado para auxiliar o processo de extração dos dados. A próxima seção apresenta a análise dos dados extraídos com base nas questões de pesquisa a serem respondidas.

4. Condução e Análise do MSL

Nessa seção é apresentada a execução dos processos definidos (Seção 3) e a análise dos dados extraídos. Para facilitar a apresentação, as Questões Secundárias de pesquisa serão referenciadas pela sigla **QS-#**, onde “#” indica o número da questão analisada.

A execução automática (uso da *string*) e manual das buscas resultou, respectivamente, em 45 e 2 publicações, totalizando 47 estudos que, posteriormente, foram submetidos aos CI e CE, reduzindo a quantidade total para 12 publicações⁹. Tal redução ocorreu, principalmente, em decorrência da aplicação dos CE (*iii*, *iv* e *v*) e da existência de publicações com mesmo conteúdo apresentado em trabalhos diferentes [Panceri and de Menezes 2014, Panceri and de Menezes 2015], caracterizando o CE (*ii*). A condução das etapas de aplicação de CI e CE, é apresentada na Figura 1(a).

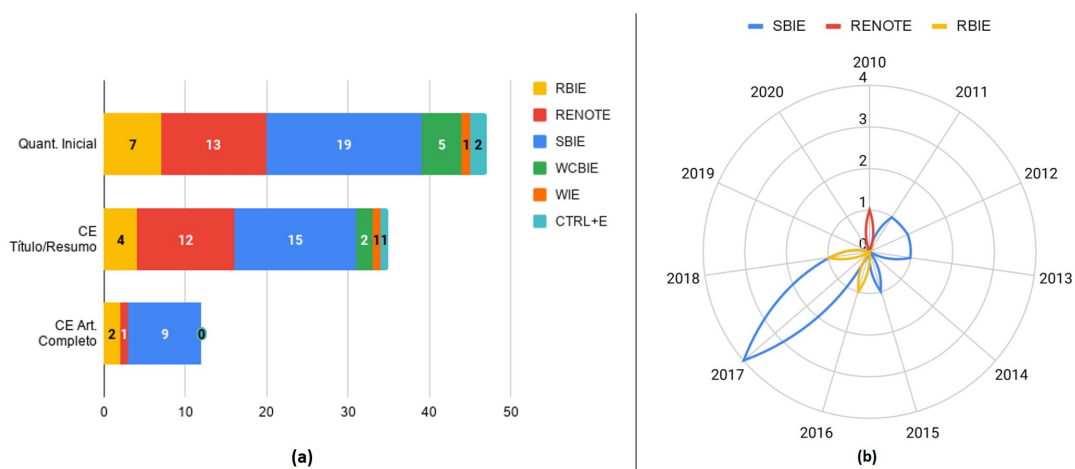


Figura 1. Quantidade de publicações: (a) etapas de seleção; (b) biblioteca/ano

⁸Disponível em: <https://forms.gle/cMctF1gnvuZSuw9y7>

⁹Lista das publicações selecionadas. Disponível em: <https://www.encurtador.com.br/eghx5>

A Figura 1(b) mostra a quantidade de publicações por biblioteca/ano (em resposta à **QS-1**). Apenas as bibliotecas RENOTE (1), RBIE (2) e SBIE (9) tiveram trabalhos selecionados para análise. Como esperado, juntos, a RBIE e SBIE possuem grande parte dos trabalhos analisados, sendo este último o que apresenta maior quantidade de publicações, mantendo-se constante entre 2011-2013, 2015 e 2018 (1 publicação/ano), apresentando a máxima quantidade em 2017 (4 publicações). Acredita-se que essa quantidade é influenciada pela natureza do evento e existência da trilhas que favorecem a submissão de publicações que englobam a temática do PLN.

O mapa (Figura 2) mostra as contribuições alcançadas com o uso do PLN aplicado ao processamento de artefatos textuais identificados no estudo. Em resposta à **QS 2**, verifica-se, com base nas legendas internas dos quadrantes, os objetivos dos trabalhos selecionados. Há diferentes objetivos, porém, de modo geral, verifica-se que os estudos que processam as mensagens dos Fóruns fundamentam-se na realização de Análise de Sentimentos ou Opinião e na análise do conteúdo da mensagem postada. Por outro lado, os estudos que analisam a produção textual, objetivam avaliar o que é escrito para identificação de erros gramaticais, preconceito e poder argumentativo.



Figura 2. Contribuição do Uso de PLN por Artefato Textual

Com relação aos artefatos do tipo “Fórum” (9 estudos), nota-se que os trabalhos buscam identificar o comportamento dos discentes por meio de análise das postagens em busca de termos que indiquem, por exemplo: a adequação entre o conteúdo da postagem e um recurso externo de vídeo que deveria ser utilizado como material de estudo [Machado et al. 2019]; estados de ânimo dos estudantes dentro do curso (satisfeito, insatisfeito, animado e desanimado) [Longhi et al. 2010]; a colaboração entre os participantes (presença de grupo, elogios, compartilhamento de informações e novas perspectivas) [Dionísio et al. 2017]; sinais de possível evasão estudantil [Azevedo et al. 2017]; plágio interno de mensagens postadas por estudantes [Cavalcanti et al. 2017] – com 1 estudo/cada; além de polarizações de mensagens como positiva/negativa para indicar dúvidas, esclarecimentos, respostas e conteúdos negativos [Rigo et al. 2013, de Oliveira Júnior et al. 2011, de Oliveira Júnior and Esmin 2012, Ferreira et al. 2018], totalizando mais 4 estudos.

Sobre artefatos do tipo “Produção textual” (3 estudos), nota-se que aluno e professor podem ser beneficiados através do auxílio na avaliação ortográfica-gramatical de redações possibilitando a coesão na escrita e correção de termos com grafias incorretas

[Santos et al. 2016] e a identificação de frases preconceituosas com uso de dicionários de termos em português [Neto et al. 2017], violando um parâmetro de avaliação do ENEM. Além disso, o trabalho de [Panceri and de Menezes 2014] apresenta uma contribuição a partir de uma abordagem que analisa, por exemplo, o poder argumentativo em um debate de teses, auxiliando a construção do conhecimento pelo aluno e a análise argumentativa por parte do professor.

Quanto aos objetivos dos trabalhos, sabe-se que eles podem ser aplicados em diferentes níveis de ensino, porém, para resposta à **QS-3**, analisou-se o contexto de aplicação dos experimentos relacionados. Com isso, constatou-se que 83% foram realizados no contexto do ensino superior e 17% do ensino médio (redações para o ENEM). Devido à natureza dos artefatos identificados, verifica-se, em resposta à **QS-4**, que os artefatos do tipo “Fórum” são utilizados para comunicação na EaD (9 estudos). Os artefatos de “produção textual”, por suas vezes, exigiram apenas a criação de modelo ou uso de ferramenta para coleta dos dados analisados, ou seja, provendo apoio a atividades fora do contexto de EaD (3 estudos).

A Figura 2 também responde à **QS-5**, quando agrupa os trabalhos selecionados pelo tipo do artefato: Fóruns (75%), agrupando mensagens síncronas ou assíncronas trocadas entre participantes; e Produção Textual (25%), referenciando os textos, redações, criados para análises. Em resposta à **QS-6** foram identificadas as técnicas de PLN, não excludentes, utilizadas no processamento dos artefatos (Figura 3).

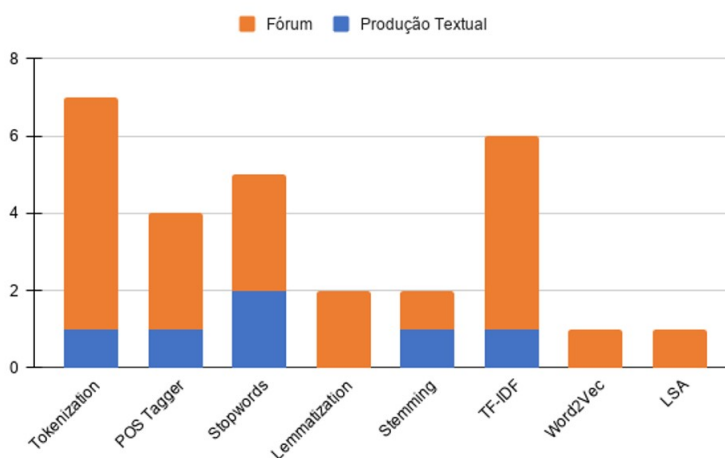


Figura 3. Técnicas de PLN aplicada aos artefatos

De modo geral, com base na Figura 3, a *Tokenization* – criação de um conjunto de *tokens* a partir do caractere espaço – é a mais utilizada em ambos os artefatos. Ao analisar os grupos de artefatos, verifica-se que nos Fóruns o uso da estatística TF-IDF (que pondera a importância de um termo em um documento ou corpus) é consideravelmente alto, o que justifica seu uso quando observa-se os objetivos dos trabalhos que analisam a presença de termos em um contexto com a finalidade de caracterizar o comportamento dos discentes. Ainda sobre os Fóruns, o uso de *word2vec* (técnica que associa um grande conjunto de textos com a finalidade de identificar sinônimos) foi utilizada por [Cavalcanti et al. 2017] para identificar a similaridade entre textos na detecção de plágio.

Com relação à “Produção textual”, o uso do *POS Tagger* – técnica que etiqueta

os termos com relação à sua classe gramatical – foi essencial para [Santos et al. 2016] no contexto das correções ortográficas-gramaticais das redações do ENEM, principalmente quando o termo não era reconhecido, dando indícios de um erro de escrita. Por fim, uso de LSA (*Latent Semantic Analysis*) – técnica que analisa a relação entre texto e documento – exclusivo por esse grupo, foi realizado por [Dionísio et al. 2017] para analisar a argumentação nos debates, também foi utilizada a TF-IDF.

Para responder à **QS.7**, durante a análise das publicações, foi verificada a existência de uma proposta ou criação de mecanismos que contribuíssem para área de PLN que pudesse ser utilizada em outros estudos. Em resposta, das 12 publicações analisadas, 2 apresentaram contribuições, são elas:

1. Em [Cavalcanti et al. 2017] foi apresentada a criação de uma *feature* baseada em uma Matriz de Similaridade Binária, na qual a similaridade entre sentenças é determinada com base na média dos pesos das similaridade entre as palavras das sentenças. Os pesos indicam presença (1) ou ausência (0) de similaridade entre palavras. Os autores justificam a criação alegando que *feature* de similaridade baseada na técnica *word2vec*, também utilizada no classificador, apresenta valores altos mesmo que as sentenças apresentem similaridade baixa. Logo, o novo cálculo favorece essas sentenças. Para validação, foram realizados experimentos que mostraram resultados promissores.
2. A contribuição de [Neto et al. 2017] surge com a criação de Léxico de termos preconceituosos para o idioma português brasileiro. Os autores descrevem que o processo de criação baseou-se em 500 postagens do Twitter, 200 contendo referências preconceituosas, que foram pré-processados (tokenização, remoção de *stopwords* e *stemming*) e identificados os termos que possuem maior representatividade no conjunto analisado. Foram identificados, por exemplo, os termos: racista, negro, preto, cabelo, gay, peso, branco, racismo, preconceito, Brasil, luta, homofobia, cor e mulher, dentre outros. Somado aos termos identificados, foram acrescentados termos preconceituosos traduzidos do idioma Alemão.

5. Considerações Finais

O Mapeamento Sistemático com base nos 12 estudos primários teve as sete perguntas definidas respondidas. Inicialmente, verifica-se que nas bibliotecas digitais selecionadas há diversas publicações que abordam os termos utilizados na *string* de busca, porém, quando aplicado os critérios de inclusão e exclusão determinados no escopo da pesquisa, vimos uma redução considerável da quantidade de pesquisas, indicando que, no contexto analisado, o uso de PLN ainda pode ser explorado por pesquisadores para as mais diversas aplicações.

Ao realizar buscas manuais, constatou-se uma limitação devido a indisponibilidade na divulgação dos trabalhos em anos anteriores. Além disso, verificou-se que há uma quantidade considerável de trabalhos que apenas relatam o uso de uma ferramenta que usa a PLN, não sendo objeto de análise para esse trabalho.

Ressalta-se que, como era esperado, as bibliotecas da RBIE e SBIE apresentaram um maior volume de publicações, principalmente devido à natureza dos trabalhos. A RENOTE, por sua vez, apresentou um grande número de publicações, porém, com um escopo de análise de uso de ferramentas no contexto educacional.

Quanto à natureza dos “artefatos textuais” analisados, chama atenção quando verifica-se a concentração em dois grandes grupos: Redações/Produções Textuais e Fóruns. Quanto aos Fóruns, é quase unânime o uso das mensagens compartilhadas para análise do comportamento de seus participantes, caracterizando uma Análise de Sentimento/Opinião. Em relação às redações, a análise da escrita (ortografia, gramática e identificação de termos racistas) ganha destaque. Além disso, destaca-se a existência de uma publicação que analisa o poder argumentativo em um debate de teses (ponto de vista), contribuindo para análise do comportamento argumentativo dos envolvidos.

Em relação às técnicas de PLN, de forma não excludentes, nota-se a diversidade de uso de técnicas de pré-processamento textual (*Tokenization, Stopwords, Lemmatization*, dentre outras), além das que identificam a função sintática dos termos (*POS Tagger*) e análise da sua importância em um corpus (TF-IDF). Contudo, questões semânticas, no contexto analisado, foram exploradas por apenas uma publicação (uso do *word2vec*). Além disso, destaca-se o uso e criação de léxicos para análise dos conteúdos abordados nos textos.

Além disso, as análises mostram que a grande maioria dos trabalhos são realizados no contexto da EaD e aplicados ao ensino superior, principalmente no contexto de troca de mensagens ou atividades de escrita. Logo, a expansão de aplicações da PLN ao ensino pode gerar contribuições válidas para a área de Informática na Educação.

Por fim, acredita-se que esse trabalho auxiliará pesquisadores das áreas Educação e Computação no tocante ao uso de PLN na automação do processamento de atividades que envolvam textos (escrita ou fala transcrita), produzidos pelos discentes no contexto presencial ou mediado por computador, através de um panorama das publicações identificadas no contexto em questão.

Referências

- Aguiar, J. J. B., Santos, S. I. N., Fachine, J. M., and Costa, E. B. (2014). Um mapeamento sistemático sobre iniciativas brasileiras em sistemas de recomendação educacionais. *Simpósio Brasileiro de Informática na Educação*.
- Azevedo, D., Ferreira, R., Mendonca, V., and Miranda, P. (2017). Aplicação de análise de sentimento em fóruns educacionais para prevenir evasão. *Simpósio Brasileiro de Informática na Educação*.
- Cabrejos, L. J. E. R., Viana, D., and dos Santos, R. P. (2018). Planejamento e execução de estudos secundários em informática na educação: Um guia prático baseado em experiências. *Jornada de Atualização em Informática na Educação*.
- Cavalcanti, A., Ferreira, R., Dionísio, M., Neto, S., Passero, G., and Miranda, P. (2017). Uma nova abordagem para detecção de plágio em ambientes educacionais. *Simpósio Brasileiro de Informática na Educação*.
- de Lima, L., Haguenaer, C., and de Lima, A. (2011). Ferramentas e-mail, chat e fórum: A percepção do professor. *Associação Brasileira de Educação a Distância - ABED*.
- de Oliveira Júnior, R. L. and Esmin, A. A. A. (2012). Monitoramento automático de mensagens de fóruns de discussão usando técnica de classificação de texto. *Simpósio Brasileiro de Informática na Educação*.

- de Oliveira Júnior, R. L., Esmín, A. A. A., Coelho, T. A., Araújo, D. L., Silva, L. A., and Giroto, R. (2011). Uma ferramenta de monitoramento automático de mensagens de fóruns em ambientes virtuais de aprendizagem. *Simpósio Brasileiro de Informática na Educação*.
- Dionísio, M., Ferreira, R., Cavalcanti, A., Carvalho, R., and Neto, S. (2017). Mineração de texto aplicada à identificação de colaboração em fóruns educacionais. *Simpósio Brasileiro de Informática na Educação*.
- dos Santos, V., Érica Ferreira de Souza, Felizardo, K. R., Watanabe, W. M., Vijaykumar, N. L., Aluizio, S. M., and Júnior, A. C. (2019). Criação conceitual de mapas a partir do processamento de linguagem natural: um mapeamento sistemático. *Revista Brasileira de Informática na Educação*.
- Ferreira, J. A. D., Mello, R. F. L., Garrozi, C., Rolim, V. B., and Cavalcanti, A. P. (2018). Um sistema baseado em pln e ag para apoiar a mediação pedagógica em fóruns de discussão. *Revista Brasileira de Informática na Educação*.
- Ferreira, M. A. D., Cavalcanti, A. P., Ferreira, R., Miranda, P., Neto, S., and Oliveira, A. (2017). Mineração de textos em fóruns educacionais: uma revisão da literatura. *Simpósio Brasileiro de Informática na Educação*.
- Indurkha, N. and Damerau, F. J. (2010). Handbook of natural language processing.
- Longhi, M., Simonato, G., Behar, P., and Bercht, M. (2010). Um framework para tratamento do léxico afetivo a partir de textos disponibilizados em um ambiente virtual de aprendizagem. *Revista Novas Tecnologias na Educação*.
- Machado, C., Maciel, A., Rodrigues, R., and Menezes, R. (2019). A análise de relevância temática de postagens em fóruns de discussão em relação ao uso de vídeos como recurso didático. *Simpósio Brasileiro de Informática na Educação*.
- Neto, S., Cavalcanti, A., Costa, E., Ferreira, R., and Dionísio, M. (2017). Uma abordagem computacional de análise de opinião para identificação de preconceito em redações. *Simpósio Brasileiro de Informática na Educação*.
- Panceri, S. S. and de Menezes, C. S. (2014). Alpes: Um sistema multiagente para análise de produções textuais no contexto de um debate de teses. *Simpósio Brasileiro de Informática na Educação*.
- Panceri, S. S. and de Menezes, C. S. (2015). Apoio a mediação pedagógica em um debate de teses utilizando técnicas de processamento de texto. *Simpósio Brasileiro de Informática na Educação*.
- Rigo, S. J., Flores, E. M., Abech, M., Barbosa, J. L. V., Costa, C., Martins, L. L., Oliveira, T. B., and Araujo, D. A. (2013). O papel do processamento de língua natural e da representação de conhecimento na extração de informações em mensagens textuais na educação a distância. *Simpósio Brasileiro de Informática na Educação*.
- Santos, J. J., Paiva, R., and Bittencourt, I. I. (2016). Avaliação léxico-sintática de atividades escritas em algoritmo genético e processamento de linguagem natural: Um experimento no enem. *Revista Brasileira de Informática na Educação*.