

Identificação e Avaliação Automática da Proposta de Intervenção em Textos Dissertativos-Argumentativos: Uma Revisão Sistemática da Literatura

Jonathan Nau^{1,2}, Aluizio Haendchen Filho^{1,2}, Rudimar Luís Scaranto Dazzi¹

¹ Laboratório de Inteligência Aplicada (LIA)
Universidade do Vale do Itajaí (UNIVALI) – Itajaí, SC – Brasil

² Núcleo de Inteligência Artificial e Sistemas Inteligentes (NIASI)
Centro Universitário de Brusque (UNIFEBE) – Brusque, SC – Brasil

jonathan_nau@live.com, aluizio.h.filho@gmail.com, rudimar@univali.br

Abstract. *An dissertative-argumentative essay writing has a structure composed of three fundamental parts: definition of a thesis, argumentation and intervention proposal. This structure must be articulated in a way that the thesis is formulated in the introduction, the argumentation in the development and the proposal of intervention in the conclusion. We examined 87 articles from three digital repositories, following a methodology composed of research questions, search expression and inclusion and exclusion criteria. After the selection of the articles, a critical analysis of the techniques used and their applicability to the context of automatic correction were carried out. The objective is to identify the best practices used for the identification and evaluation of intervention proposals in essays.*

Resumo. *Uma redação de natureza dissertativa-argumentativa apresenta uma estrutura composta por três partes fundamentais: definição de uma tese, argumentação e proposta de intervenção. Essa estrutura deve estar articulada de forma que a tese é formulada na introdução, a argumentação no desenvolvimento e a proposta de intervenção na conclusão. Examinou-se 87 artigos de três repositórios digitais, seguindo uma metodologia composta por perguntas de pesquisa, expressão de busca e critérios de inclusão e exclusão. Após a seleção dos artigos, realizou-se uma análise crítica das técnicas utilizadas e sua aplicabilidade para o contexto de correção automática. O objetivo é identificar as melhores práticas utilizadas para a identificação e a avaliação de propostas de intervenção em redações.*

1. Introdução

No ENEM (Exame Nacional do Ensino Médio), a avaliação dos textos tem por objetivo verificar competências linguísticas do aluno na dimensão textual. Dados recentes do ENEM de 2017 mostram uma grave deficiência no ensino de produção textual, especificamente de redação, no qual mais de 300 mil alunos obtiveram nota zero na redação. Além disso, de 2016 para 2017 houve um aumento de 5,9% na quantidade de redações com nota zero e em contrapartida uma queda de 45,3% na quantidade de redações com nota mil [INEP 2018]. Tal constatação demonstra que as práticas de ensino relacionadas com a produção e a interpretação textual não estão surtindo o efeito desejado, e esta deficiência pode inclusive afetar o desempenho em outras disciplinas.

As redações permitem avaliar as habilidades de leitura, interpretação e produção de texto do candidato. Portanto, além de ter o conhecimento do assunto questionado, o candidato precisa compreender bem o tema, para que a produção textual da redação esteja adequada ao que foi proposto. Redações permitem ao candidato desenvolver raciocínio e espírito crítico em relação ao texto desenvolvido. A utilização de redações no processo de ensino-aprendizagem permite avaliar processos cognitivos mais elevados, além de treinar e aprimorar a capacidade de interpretação textual, melhorando a qualidade da aprendizagem.

Segundo INEP (2017), para uma boa redação é preciso compor uma tese a partir de um tema – uma opinião a respeito do tema proposto –, que deve ser apoiada em argumentos consistentes, estruturados com coerência e coesão, formando uma unidade textual. Além disso, é necessário elaborar uma proposta de intervenção¹ social para o problema apresentado no desenvolvimento do texto que respeite os direitos humanos.

No entanto, além de uma boa redação é necessária uma correção consistente com critérios bem definidos, em geral, esse processo demanda tempo e custo. Com isso, a correção automática busca automatizar o processo de correção, sem precisar de um corretor humano para a tarefa. Conforme Dikli (2006), no processo de correção automática, problemas como tempo, custo, confiabilidade e subjetividade podem ser minimizados.

Este trabalho apresenta uma revisão sistemática da literatura para analisar o estado da arte relacionada a identificação e avaliação automática de propostas de intervenção. Para isto, realizou-se uma revisão contemplando artigos publicados entre 01/2012 e 12/2017 em três repositórios diferentes, sendo estes: (i) ACL; (ii) Scopus e (iii) ScienceDirect.

Dentro deste contexto, esse trabalho tem como objetivo contribuir na área da computação por meio da seleção das melhores técnicas de processamento de linguagem natural, que serão utilizadas para o desenvolvimento da identificação e avaliação automática de propostas de intervenção em textos dissertativos-argumentativos.

O trabalho está dividido nas seguintes seções: (i) metodologia, que apresenta os critérios utilizados para realizar a revisão sistemática; (ii) resultados, tomando por base os artigos selecionados e as informações obtidas; e (iii) considerações finais, com as contribuições.

2. Metodologia

A metodologia aplicada neste estudo consiste na delimitação dos repositórios utilizados para a revisão da literatura, os critérios de inclusão, exclusão e a definição das palavras-chave utilizadas na busca dos artigos. Os artigos obtidos devem responder as seguintes perguntas de pesquisa:

1. Quais as características do corpus de redações?
2. Identifica e/ou avalia a proposta de intervenção?
3. Quais técnicas são utilizadas para identificar e/ou avaliar a proposta de intervenção?

¹ Ações que podem contribuir para amenizar ou solucionar os problemas levantados pelo autor durante a redação.

A fonte de busca para responder às perguntas do estudo foram selecionados periódicos e anais de conferências internacionais. O Quadro 1 mostra os repositórios utilizados para execução deste estudo. As bases foram selecionadas pela proximidade com o tema, a base da Associação para Linguística Computacional (ACL, do inglês: Association for Computational Linguistics) contém diversos jornais e conferências relacionados aos problemas de linguística computacional e processamento de linguagem natural (PLN). As bases Scopus e Science Direct foram escolhidas por serem genéricas e contemplarem diversas áreas, incluindo o PLN.

Quadro 1. Lista de repositórios

Repositório	Endereço do Repositório
ACL	https://www.aclweb.org/portal/
Scopus	https://www.scopus.com/
Science Direct	https://www.sciencedirect.com/

Baseado nos repositórios do Quadro 1, foram selecionadas palavras-chave que expressassem o assunto e que não apresentassem redundância na busca, tendo em vista, contemplar o maior número de artigos encontrados. Além disso, buscou-se uma estrutura que formasse um filtro para que houvesse retorno dos trabalhos mais relevantes para o tema, para atender aos objetivos desta pesquisa. Utilizou-se a seguinte expressão de busca para cada repositório:

essay AND (scoring OR identifying) AND ("discourse element" OR "discourse analysis" OR "discourse structure" OR "conclusion statements") AND (nlp OR "natural language processing")

Após a criação da expressão de busca foram incluídos critérios de inclusão e exclusão, com a finalidade de selecionar apenas os mais relevantes para a área de pesquisa do trabalho e desta forma reduzir os artigos selecionados pela expressão para leitura. Foram definidos os seguintes critérios:

1. Inclusão: (i) trabalhos que corrijam automaticamente a redação ou identifiquem a proposta de intervenção; (ii) trabalhos publicados entre 01/2012 e 12/2017; e (iii) filtros devem estar contidos nos títulos, resumos e palavras-chaves;
2. Exclusão: (i) resumos estendidos; (ii) trabalho sem revisão por pares; (iii) conferência com aceitação de artigo com análise apenas do resumo curto; (iv) artigos que não estejam na língua portuguesa ou inglesa.

Com a filtragem dos trabalhos nos repositórios selecionados, realizou-se a seleção dos artigos por meio da leitura dos seguintes tópicos: (i) título; (ii) palavras-chave; (iii) resumo; (iv) introdução; e (v) conclusão. A seleção desses tópicos, foram úteis para realizar a exclusão de trabalhos que não são relevantes ao projeto, desta forma, permitiu minimizar o esforço na leitura e seleção de trabalhos que respondem às perguntas do tema de pesquisa.

Na sequência, realizou-se o processo de sumarização dos trabalhos selecionados, extraindo as características mais relevantes, como: título, autores, ano de publicação, e a resposta de cada uma das perguntas de pesquisa. Portanto, com base nos dados sumarizados, é possível analisar e interpretar as informações, no qual serão fundamentais para o desenvolvimento de novos trabalhos nesta área de pesquisa.

3. Análise dos Resultados

Após a aplicação da expressão de busca e execução dos critérios de inclusão e exclusão em cada repositório, foram descobertos 87 trabalhos. A Figura 1 apresenta os artigos localizados e agrupados por repositório.

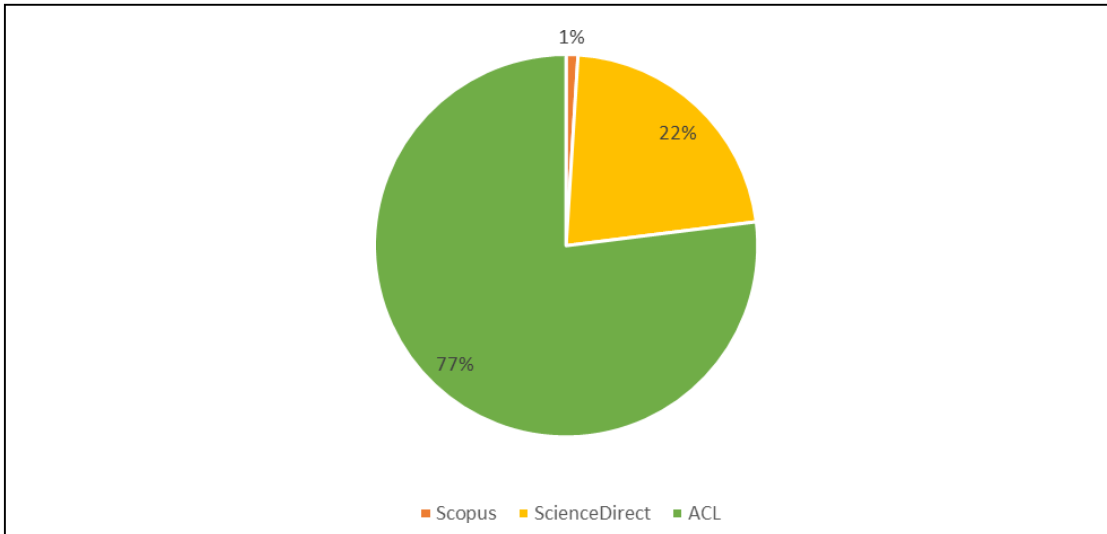


Figura 1. Trabalhos localizados na primeira etapa por repositório

Conforme o gráfico, é possível observar que o repositório ACL tem grande impacto nos trabalhos selecionados, correspondendo a 77% dos trabalhos encontrados. Esse elevado número se dá pelo fato desse repositório ser especificamente de linguística computacional, no qual estão relacionadas às tarefas de processamento de linguagem em diversas aplicações.

Na segunda etapa, realizou-se a seleção dos artigos descobertos na primeira etapa. Após esta seleção, o número de artigos reduziu para 5, ou seja, aproximadamente 6% dos artigos localizados na primeira etapa apresentaram alguma relação direta com o tema proposto. Realizou-se a seleção lendo o título, palavras-chave, resumo, introdução e conclusão de cada trabalho. A Figura 2 ilustra um gráfico de barras com os resultados obtidos após essa seleção. As barras verdes indicam os trabalhos localizados e podem ser comparados com as barras em azul que são os trabalhos selecionados.

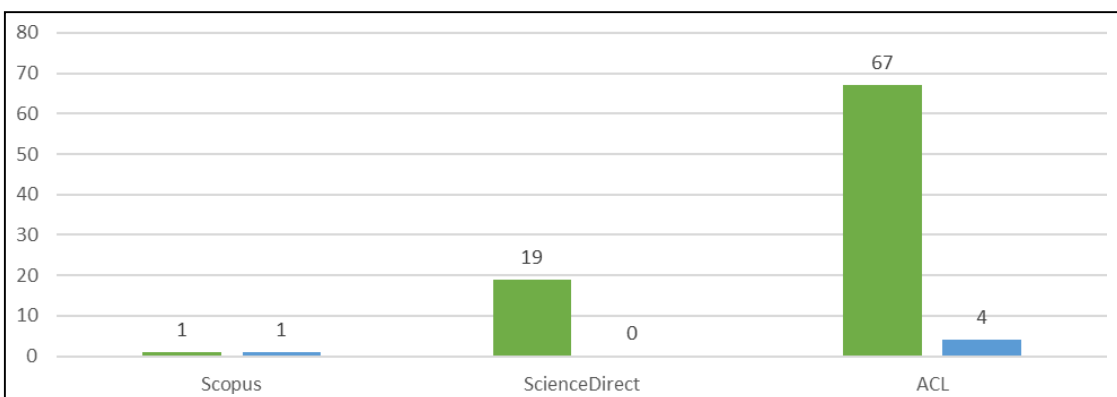


Figura 2. Trabalhos da primeira em comparação com trabalhos da segunda etapa

Com base no gráfico é possível observar que muitos trabalhos localizados não contribuem com o problema de pesquisa. Alguns trabalhos não foram selecionados pois apresentaram as redações apenas como exemplos de aplicação e a abordagem proposta era em outras estruturas textuais, como textos em fóruns de discussão e textos jornalísticos.

Na terceira etapa realizou-se a sintetização dos trabalhos selecionados na segunda etapa. O Quadro 2 exibe os artigos selecionados, a lista está ordenada pela quantidade de citações. O quadro exibe os detalhes de cada trabalho, tais como o ID (utilizado para enumerar e identificar os artigos), número de citações, repositório, ano, título e autores. Neste quadro são apresentados todos os trabalhos selecionados que formam a lista final da revisão sistemática da literatura.

Quadro 2. Lista de trabalhos obtidos na segunda etapa

ID	Citações	Repositório	Ano	Título	Autores
1	122	ACL	2014	Identifying argumentative discourse structures in persuasive essays	Christian Stab; Iryna Gurevych
2	23	ACL	2016	End-to-end argumentation mining in student essays	Isaac Persing; Vincent Ng
3	8	Scopus	2014	Identifying thesis and conclusion statements in student essays to scaffold peer review	Mohammad Hassan Falakmasir; Kevin D. Ashley; Christian D. Schunn; Diane J. Litman
4	4	ACL	2015	Discourse element identification in student essays based on global and local cohesion	Wei Song; Ruiji Fu; Lizhen Liu; Ting Liu
5	1	ACL	2017	A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese	Evelin Carvalho Freire de Amorim; Adriano Veloso

O Quadro 3 ilustra a sumarização das características de cada trabalho apresentado no Quadro 2 para responder às perguntas de pesquisa. Neste quadro é respondido, para cada artigo selecionado, qual o tamanho do corpus utilizado, se o trabalho identifica e/ou avalia a conclusão e qual o algoritmo utilizado para a tarefa.

Quadro 3. Reposta as perguntas de pesquisa do projeto

ID	1. Corpus de redação utilizado	2. Identifica e/ou Avalia?	3. Qual algoritmo utilizado?
1	90 Redações	Identifica	Máquina de vetores de suporte (SVM, do inglês: <i>support vector machine</i>)
2	90 Redações	Identifica	Maximum Entropy
3	432 redações	Identifica	Naive Bayes, Árvore de Decisão e SVM

4	910 Redações	Identifica	Campos aleatórios condicionais (CRF, do inglês: Conditional Random Fields)
5	1840 Redações	Avalia	Regressão linear simples

No trabalho *Identifying argumentative discourse structures in persuasive essays* [Stab and Gurevych 2014] os autores realizam a identificação da estrutura do discurso em dois passos, o primeiro identifica os componentes do argumento e o segundo realiza a classificação das relações do argumento em suporte ou não suporte. Os testes foram realizados em 90 redações anotadas e o treinamento feito com a máquina de vetores de suporte (SVM, do inglês: support vector machine) [Steinwart and Christmann 2008]. Na identificação dos componentes de argumentos os autores alcançaram 0,773 de acurácia e na identificação da relação conseguiram 0,863 de acurácia.

Assim como os autores anteriores, o trabalho *End-to-end argumentation mining in student essays* [Persing and Ng 2016] realiza as duas tarefas de identificação da estrutura do discurso e o mesmo corpus, porém, para o treinamento utilizam o algoritmo de Entropia Máxima [Berger, Pietra, and Pietra 1996]. A precisão da identificação dos componentes dos argumentos foi de 53,4 e a identificação das relações foi de 12,7.

No trabalho *Identifying thesis and conclusion statements in student essays to scaffold peer review* [Falakmasir et al. 2014], os autores propõem a identificação de teses e conclusões em redação, com a finalidade de preparar a reflexão do revisor. Os autores utilizaram 3 classificadores no RapidMiner [Mierswa et al. 2006]: (i) Naïve Bayes; (ii) Árvore de decisão; e (iii) SVM. Na identificação da tese a acurácia foi 0,70, 0,82 e 0,82 respectivamente, na identificação da conclusão os algoritmos alcançaram 0,65, 0,49 e 0,60 respectivamente e no nível da redação o resultado foi de 0,63, 0,75 e 0,62 respectivamente. Portanto o melhor algoritmo do trabalho para identificar tese e conclusão no nível da sentença é o SVM e no nível da redação a árvore de decisão.

O trabalho *Discourse element identification in student essays based on global and local cohesion* [Song et al. 2015] realiza a classificação dos elementos discursivos, o trabalho classificou as sentenças em: (i) introdução; (ii) tema; (iii) tese; (iv) ideia principal; (v) ideias de suporte; (vi) conclusão; e (vii) outros. Os autores testaram outros algoritmo e compararam com o SVM, o algoritmo campos aleatórios condicionais (CRF, do inglês: conditional random fields) [Hirohata et al. 2008] apresentou os melhores resultados e foi utilizado para apresentação dos resultados. Para a conclusão os autores apresentaram 87,2 de precisão em features básicas e com mais features de coesão a precisão subiu para 89,1.

O único trabalho que avalia a conclusão é o *A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese* [Amorim and Veloso 2017], os autores utilizaram um corpus de 1840 redações escritas em português, um conjunto de features e regressão linear simples para avaliar redações. O trabalho apresenta como resultado o coeficiente kappa, que na proposta de solução apresentou 0,1393, que significa uma correlação muito baixa entre o algoritmo e o avaliador humano.

6. Discussão

Os filtros e critérios de busca aplicados nesta pesquisa selecionaram trabalhos relevantes e recentes sobre o tema de pesquisa, o que caracteriza a atualidade do tema. Além disso, é possível observar que a técnica mais utilizada para a identificação de conclusões em redações foi o SVM, mas não foi exclusividade, o que permite dizer que a identificação de conclusão apresenta lacunas a serem preenchidas por novas pesquisas. Na avaliação da conclusão foi encontrado apenas um trabalho, o que permite a aplicação de novas abordagens e experimentos.

O corpus de redação dos trabalhos selecionados foi bastante diversificado, variando entre 90 a 1840 redações. Cabe ressaltar que o corpus de 1840 redações apresenta apenas a nota da conclusão, isto é, não há qualquer anotação na redação, além disso nesse corpus as redações foram escritas em português. Nos demais corpora são identificados as conclusões no texto, porém as redações não foram escritas em português.

Dentre os trabalhos analisados, *Identifying argumentative discourse structures in persuasive essays* [Stab and Gurevych 2014] e *End-to-end argumentation mining in student essays* [Persing and Ng 2016] contribuíram nas questões relacionadas com utilização dos componentes de argumentos de forma global em textos argumentativos. As técnicas utilizadas podem ser úteis se os marcadores discursivos utilizados forem adaptados para identificar características do discurso na língua portuguesa.

Em *Identifying thesis and conclusion statements in student essays to scaffold peer review* [Falakmasir et al. 2014] os autores apresentam uma análise específica para declarações de Conclusão e apresentam testes com mais de um algoritmo. Dentre os trabalhos analisados, é o que tem mais aderência com a proposta da pesquisa. As técnicas utilizadas podem contribuir na construção da solução para identificação e avaliação de proposta de intervenção em textos dissertativos-argumentativos.

No artigo *Discourse element identification in student essays based on global and local cohesion* [Song et al. 2015] foram apresentadas ideias para classificação dos elementos discursivos, que podem ser usados para identificar a introdução, elementos do desenvolvimento e conclusão de textos dissertativos-argumentativos. O principal desafio é adaptar os marcadores discursivos que foram utilizados no idioma chinês. Um aspecto interessante foi a constatação de que o algoritmo SVM foi superado pelo CRF.

Dentre os trabalhos analisados, o *A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese* [Amorim and Veloso 2017] trouxe uma contribuição importante na Questão de Pesquisa 2, relacionada com a identificação e avaliação da proposta de intervenção. Apesar do baixo índice de correlação entre o algoritmo e o avaliador humano, as features utilizadas com técnica de regressão linear podem ser ampliadas para melhorar a correlação. As mesmas features acrescidas de outras podem também ser testadas com outros algoritmos, tais como SVM (Support Vector Machine) e o Lasso [Tibshirani 1996].

7. Considerações Finais e Trabalhos Futuros

A correção automática de redações é um trabalho complexo, mas oferece vantagens em relação aos sistemas tradicionais, tais como: (i) tempo de espera de feedback do aluno; (ii) redução da carga de trabalho do professor; e (iii) diminuição dos custos para as

instituições. Portanto, conclui-se que este trabalho é útil na tarefa de correção de redações, permitindo a melhora dos sistemas tradicionais de correção.

A principal contribuição deste trabalho é disponibilizar informações obtidas em um trabalho de revisão sistemática da literatura, selecionando os melhores trabalhos para a tarefa de identificação e avaliação de conclusão em redações. Além disso, demonstrando as técnicas e os corpora utilizados nessa tarefa. Desta forma, contribui com um ponto de partida para os pesquisadores realizarem novas pesquisas nesta área de estudo.

Considerando essa pesquisa sistemática, e os resultados obtidos como ponto de partida, os passos seguintes são (i) compor um corpus de redações em português, que seja anotado as propostas de intervenção e com uma nota atribuída, (ii) implementação dos algoritmos encontrados na literatura e a realização de testes; e (iii) sumarização dos resultados e a comparação com a literatura apresentada neste trabalho.

Referências

- Amorim, E. C. F e Veloso, A. (2017). A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese. 15th Conference of the European Chapter of the Association for Computational Linguistics, pages 94–102, Valencia, Spain.
- Berger, A. L., Pietra, V. J. D., and Pietra, S. A. D. (1996). A maximum entropy approach to natural language processing. *Comput. Linguist.*, 22(1):39–71.
- Brasil Escola (2018). “Banco de Redações”.
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *Journal of Technology, Learning, and Assessment*, 5(1).
- Falakmasir, M. H., Ashley, K. D., Schunn, C. D., and Litman, D. J. (2014). Identifying thesis and conclusion statements in student essays to scaffold peer review. In Trausan-Matu, S., Boyer, K. E., Crosby, M., and Panourgia, K., editors, *Intelligent Tutoring Systems*, pages 254–259, Cham. Springer International Publishing.
- Hirohata, K., Okazaki, N., Ananiadou, S., and Ishizuka, M. (2008). Identifying sections in scientific abstracts using conditional random fields. In *In Proc. of the IJCNLP 2008*
- INEP (2017). “Redação no ENEM 2017: Cartilha do Participante”.
- INEP (2018). “Resultados Dos Participantes De 2017”.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06*, pages 935–940, New York, NY, USA. ACM.
- Persing, I. and Ng, V. (2016). End-to-end argumentation mining in student essays. In *Proceedings of Human Language Technologies: The 2016 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1384–1394.
- Song, W., Fu, R., Liu, L., and Liu, T. (2015). Discourse element identification in student essays based on global and local cohesion. In *Proceedings of the 2015*

Conference on Empirical Methods in Natural Language Processing, pages 2255–2261. Association for Computational Linguistics.

Stab, C. and Gurevych, I. (2014). Identifying argumentative discourse structures in persuasive essays. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 46–56. Association for Computational Linguistics.

Steinwart, I. and Christmann, A. (2008). Support vector machines. Springer, New York.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), pages 267–288.

UOL Educação (2018). “Banco de Redações”.