

ATHOS

Ferramenta computacional para avaliação automatizada da escrita de projetos de pesquisa

Ricardo R. Meira¹, Eliseo Reategui²

¹Instituto Federal Sul-riograndense (IFSul) - Novo Hamburgo, RS - Brazil

²Faculdade de Educação, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS - Brazil

ricardomeira@ifsul.edu.br, eliseo.reategui@ufrgs.br

Abstract. This work presents the development and evaluation of the ATHOS tool, which integrates Natural Language Processing and Learning Analytics to assist university professors in assessing research proposals. The tool combines textual complexity analysis, structure analysis, and generative AI to extract key elements from proposals, generate comparative visualizations, and suggest customizable feedback. A multiple case study with professors of research methodology revealed that ATHOS partially optimizes the assessment process, allowing quick identification of critical points and freeing up time for substantive comments. Results indicate that the professor-centered approach, keeping them in control of the final feedback, is promising for integrating automated assessment technologies into pedagogical practice.

Resumo. Este trabalho apresenta o desenvolvimento e avaliação da ferramenta ATHOS, que integra Processamento de Linguagem Natural e Analítica da Aprendizagem para auxiliar professores universitários na avaliação de projetos de pesquisa. A ferramenta combina análises de complexidade textual, estrutural e IA generativa para extrair elementos chave dos projetos, gerar visualizações comparativas e sugerir feedbacks personalizáveis. Um estudo de casos múltiplos com professores de metodologia da pesquisa revelou que a ATHOS otimiza parcialmente o processo avaliativo, permitindo identificação rápida de pontos críticos e liberando tempo para comentários substantivos. Os resultados indicam que a abordagem centrada no professor, mantendo-o no controle do feedback final, é promissora para integrar tecnologias de avaliação automatizada na prática pedagógica.

1. Introdução

A habilidade de comunicar-se eficazmente por meio da escrita é fundamental na sociedade contemporânea, adquirindo importância ainda maior no ambiente acadêmico. A escrita acadêmica, particularmente no gênero projeto de pesquisa, exige dos estudantes não apenas o domínio do conteúdo, mas também a aderência a normas estruturais, estilísticas e argumentativas específicas. A avaliação desses projetos, por sua vez, impõe aos docentes um desafio considerável, demandando tempo para análises detalhadas e para a elaboração

de feedbacks que sejam simultaneamente precisos, construtivos e individualizados, essenciais para o desenvolvimento das competências discentes [Hattie e Timperley 2007], [Carless e Boud, 2018].

Contudo, a crescente demanda por avaliação em larga escala, associada à sobrecarga de trabalho docente, frequentemente limita a capacidade de fornecer esse retorno detalhado e frequente, essencial ao processo formativo [Sousa et al. 2018]. Nesse contexto, emergem as tecnologias educacionais, especificamente no campo da Analítica da Aprendizagem e da Analítica da Escrita, como potenciais aliadas. A aplicação de técnicas de Processamento de Linguagem Natural (PLN) e Inteligência Artificial (IA) oferece novas possibilidades para automatizar parte do processo avaliativo, visando otimizar o tempo do professor e fornecer aos alunos retornos mais ágeis sobre aspectos específicos de seus textos [Crossley, McNamara e McCarthy 2010].

Diversas ferramentas de avaliação automatizada da escrita (AWE) têm sido desenvolvidas, focando em aspectos como gramática, ortografia, coesão e complexidade textual. No entanto, a avaliação de projetos de pesquisa envolve nuances que vão além desses aspectos formais, abrangendo a coerência entre título, objetivos e metodologia, a adequação da revisão de literatura e a clareza na delimitação do problema e da abordagem metodológica. A simples automação pode não capturar a complexidade dessa avaliação, sendo necessária uma abordagem que combine a eficiência da máquina com a expertise pedagógica do professor.

Diante desse cenário, este trabalho investiga como uma ferramenta computacional, denominada Athos, que integra PLN, IA e métricas textuais específicas para o gênero projeto de pesquisa, pode auxiliar professores na avaliação e na elaboração de feedbacks formativos. A ferramenta foi desenvolvida com base em métricas de complexidade textual (NILC-Metrix) e de estrutura textual, adaptadas do Plugin MTA [Bastiani 2022], buscando identificar automaticamente elementos chave do projeto e avaliar sua adequação, gerando um feedback inicial que pode ser validado e complementado pelo professor.

O objetivo deste artigo é apresentar o desenvolvimento da ferramenta Athos e relatar os resultados de um estudo de caso conduzido com professores de disciplinas de metodologia científica em cursos de graduação. Buscou-se avaliar a percepção dos docentes sobre a utilidade da ferramenta, a qualidade dos feedbacks gerados e o impacto no processo de avaliação dos projetos de pesquisa de seus alunos. A investigação visa contribuir para a compreensão de como abordagens híbridas, que combinam avaliação automatizada e intervenção docente, podem otimizar o processo de feedback na escrita acadêmica, promovendo um suporte mais eficaz ao desenvolvimento dos estudantes.

2. Referencial Teórico

A avaliação da escrita acadêmica, especialmente em projetos de pesquisa de graduação, representa um desafio significativo no contexto educacional. A complexidade inerente a este gênero textual, aliada à necessidade de fornecer retornos construtivos e individualizados aos estudantes, demanda tempo e expertise por parte dos docentes. Nesse cenário, a Analítica da Aprendizagem, e mais especificamente a Analítica da Escrita, surge como um campo promissor, oferecendo abordagens e ferramentas baseadas em Processamento de Linguagem Natural (PLN) e Inteligência Artificial (IA) para apoiar tanto o processo avaliativo quanto o desenvolvimento das habilidades de escrita dos alunos.

2.1 A Importância do Feedback no Ensino da Escrita Acadêmica

O feedback desempenha um papel central no processo de ensino-aprendizagem da escrita acadêmica. Como destacado por [Carless e Boud 2018], ele oferece aos estudantes uma oportunidade crucial para refletir sobre suas produções, identificar pontos fortes e áreas que necessitam de aprimoramento, e receber direcionamentos claros para refinar suas habilidades. Ao fornecer insights sobre a adequação do conteúdo, a clareza da argumentação, a organização das ideias, a precisão vocabular e a conformidade com as normas acadêmicas, o feedback auxilia os alunos a compreenderem as expectativas e a ajustarem seu desempenho para alcançar melhores resultados.

A eficácia do feedback, contudo, está intrinsecamente ligada às suas características. [Hattie e Timperley 2007] argumentam que um feedback efetivo deve ser construtivo, específico, orientado para o desenvolvimento e fornecido em tempo hábil. Ele não deve apenas apontar erros, mas também sugerir caminhos para a melhoria, adaptando-se às necessidades individuais de cada estudante, considerando seu nível de proficiência e objetivos de aprendizagem. Estratégias como o feedback descritivo, que analisa detalhadamente o trabalho, e o feedback formativo, que acompanha o processo de escrita com foco no progresso contínuo, são fundamentais para promover a autorreflexão e a autocorreção [Nicol e Macfarlane-Dick 2006], [Higgins, Hartley e Skelton 2002].

O feedback mediado pelo professor, com sua experiência e conhecimento contextualizado, oferece uma análise aprofundada e personalizada, abordando nuances da organização textual, lógica argumentativa e adequação estilística que ferramentas automatizadas podem ter dificuldade em captar [Hattie e Timperley 2007]. Essa interação direta permite esclarecimentos e discussões que enriquecem o aprendizado. No entanto, a sobrecarga de trabalho docente, especialmente em turmas numerosas, frequentemente limita a capacidade de fornecer feedbacks detalhados e individualizados a todos os alunos [Sousa et al. 2018]. Essa limitação temporal pode comprometer a qualidade e a frequência do retorno, impactando o desenvolvimento da escrita discente.

2.2 Avaliação da Escrita, Ferramentas Automatizadas e Analítica da Escrita

A avaliação de projetos de pesquisa é um processo complexo que visa garantir a qualidade, originalidade e relevância do trabalho acadêmico [Lillis e Curr, 2010]. Ela envolve a análise crítica de múltiplos aspectos, desde a formulação do problema e objetivos até a metodologia e a fundamentação teórica. Dada a complexidade e o volume de trabalhos, surgem as ferramentas de avaliação automatizada da escrita (Automated Writing Evaluation - AWE) como um recurso complementar.

Estas ferramentas, impulsionadas por avanços em PLN e IA, utilizam algoritmos para analisar textos e fornecer feedback sobre diversos aspectos, como gramática, ortografia, estilo, coesão e coerência [Leal 2023]. A Analítica da Escrita, como subcampo da Analítica da Aprendizagem, foca especificamente na coleta e análise de dados do processo de produção textual para monitorar o progresso dos alunos e informar intervenções pedagógicas [Crossley, McNamara e McCarthy 2010].

Ferramentas baseadas em PLN, como tokenizadores, responsáveis por segmentar o texto em unidades menores, como palavras ou subpalavras e parsers, que analisam a estrutura gramatical das sentenças, são a base para análises mais complexas [Jurafsky e Martin 2018], [Bird, Loper e Klein 2009]. Técnicas como análise de sentimentos,

reconhecimento de entidades nomeadas (NER) e o uso de métricas textuais, como as calculadas por ferramentas como Nilc-Metrix, permitem avaliar a complexidade, leitura e outros atributos do texto [Liu 2012], [Nadeau e Sekine 2007], [Evers e Finatto 2016].

Estudos comparativos entre feedback automatizado e feedback do professor indicam que ambos podem ser benéficos, atuando de formas complementares. Enquanto o feedback do professor pode ter maior impacto em aspectos motivacionais e na compreensão de nuances contextuais [Taskiran e Goksel 2022], [Wang e Han 2022], [Gallien e Early 2005], o feedback automatizado se destaca pela rapidez, disponibilidade e objetividade em aspectos mais formais da escrita [Wilson et al. 2014], [Warschauer e Ware 2006].

A abordagem ideal, portanto, parece residir na integração inteligente entre a avaliação humana e a automatizada, configurando um modelo “human-in-the-loop” [Seo et al. 2021]. Nesse modelo, a ferramenta automatizada lida com aspectos mais padronizáveis da avaliação, liberando o professor para focar em elementos mais complexos e subjetivos do texto, além de personalizar o feedback gerado pela máquina e promover uma interação mais significativa com o aluno. Contudo, a implementação eficaz da Analítica da Escrita exige atenção a requisitos críticos como a precisão e confiabilidade das ferramentas, sua adaptabilidade a diferentes contextos e níveis, usabilidade e, fundamentalmente, a garantia da privacidade e segurança dos dados dos estudantes [Sherri e Burstein 2013], [Cope e Kalantzis 2016].

3. Metodologia e Desenvolvimento da Ferramenta Athos

Para abordar o desafio da avaliação de projetos de pesquisa e a necessidade de feedback eficaz, foi desenvolvida e avaliada uma ferramenta computacional denominada Athos. Esta seção detalha a metodologia empregada no desenvolvimento da ferramenta, a definição das métricas de análise textual incorporadas e o delineamento do estudo de caso conduzido para avaliar sua aplicação e aceitação por professores.

3.1 Desenvolvimento da Ferramenta Athos

A ferramenta Athos foi concebida como um sistema web para auxiliar professores na avaliação de projetos de pesquisa de graduação, focando na análise da estrutura textual e da complexidade linguística, e na geração de feedbacks automatizados. Sua arquitetura integra diversas tecnologias e recursos, incluindo técnicas de PLN, ferramentas de IA generativa e métricas validadas em pesquisas anteriores.

O fluxo de funcionamento da Athos inicia-se com a coleta dos trabalhos dos alunos, realizada através de um módulo de conexão com a plataforma Moodle, utilizando um token de identificação do professor. Os arquivos submetidos em formato .docx ou .pdf passam por uma etapa de pré-processamento automatizado. Nesta fase, elementos não textuais como tabelas e figuras são removidos, assim como seções padronizadas como capa, sumário, referências, etc. O texto remanescente é então segmentado em suas seções principais, introdução, referencial teórico, metodologia etc., através de algoritmos de PLN. O título do trabalho é extraído das páginas iniciais utilizando uma ferramenta de IA.

A figura 1 abaixo ilustra a sequência de funcionamento da ferramenta Athos, bem como apresenta os pontos de utilização das ferramentas Nilc-Metrix e plugin MTA.

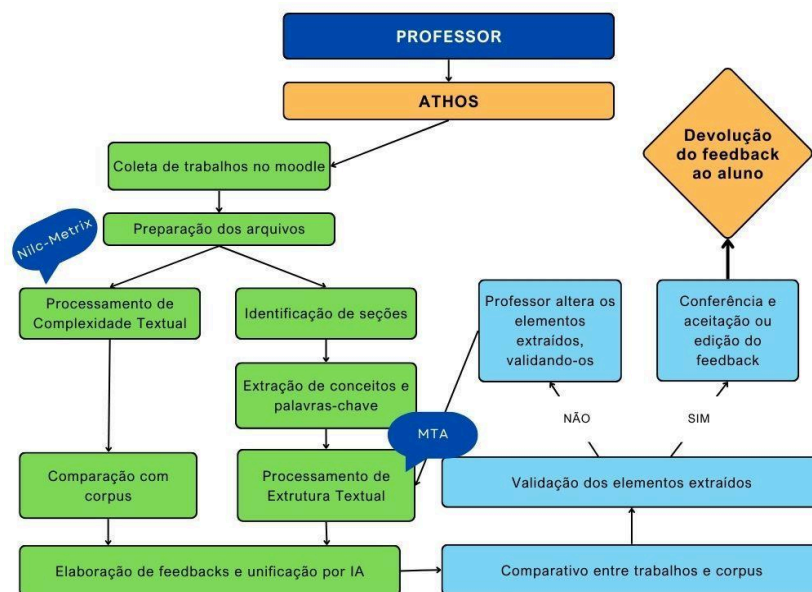


Figura 1: Diagrama com funcionamento da ferramenta Athos

Após o pré-processamento, o texto é submetido a dois módulos principais de análise. O primeiro, Processamento de Complexidade Textual, utiliza a ferramenta NILC-Metrix, baseada em trabalhos como os de [Scarton e Aluísio 2010] para calcular um conjunto de métricas linguísticas relacionadas à sofisticação léxica, complexidade sintática, coesão e tamanho do texto. Devido a limitações de tamanho de processamento da ferramenta NILC-Metrix, os textos mais longos são divididos em blocos para análise. Os resultados dessas métricas são comparados com corpora de referência previamente estabelecidos para contextualizar o desempenho do texto analisado.

O segundo módulo, Processamento de Estrutura Textual, baseia-se em uma versão adaptada e automatizada do Plugin MTA [Bastiani, 2022]. Este módulo avalia a coerência e adequação de elementos estruturais chave do projeto de pesquisa, como a relação entre título, objetivos e palavras-chave, a presença de elementos essenciais na metodologia e a clareza na definição dos objetivos. A extração inicial de elementos como objetivo e palavras-chave das seções identificadas foi realizada com o auxílio do ChatGPT 3.5, e a análise subsequente segue as regras e heurísticas validadas no trabalho de [Bastiani 2022], reimplementadas em Java para permitir o processamento automatizado fora do ambiente Google Docs.

Os resultados de ambos os módulos de processamento, Complexidade e Estrutura, disparam a geração de fragmentos de feedback pré-definidos, associados a cada métrica ou verificação estrutural. Estes fragmentos são então enviados à API da ferramenta de IA para serem consolidados em um texto de feedback unificado e coeso, destinado ao professor.

A interface da Athos permite ao professor visualizar os trabalhos dos alunos, acessar os resultados das análises, incluindo gráficos comparativos de desempenho em relação à turma e aos corpora de referência e, crucialmente, interagir com o processo de avaliação, essa interação é demonstrada na figura 1 através dos itens representados em fundo azul claro, dispostos no lado direito da figura. Antes de visualizar o feedback final, é

solicitado ao professor que este valide os elementos extraídos automaticamente pela IA, sendo título, objetivo e palavras-chave, podendo corrigi-los se necessário. Qualquer alteração nesses elementos aciona um reprocessamento do módulo de Estrutura Textual e a consequente regeneração do feedback. Finalmente, o professor visualiza o texto de feedback consolidado, podendo editá-lo livremente antes de enviá-lo ao aluno através da plataforma.

3.2 Desenvolvimento e Validação das Métricas Textuais

A ferramenta Athos foi desenvolvida para empregar um conjunto de métricas textuais divididas em duas categorias principais: Complexidade Textual e Estrutura Textual. As métricas de Complexidade Textual, calculadas pelo NILC-Metrix, abrangem diferentes dimensões linguísticas:

- **Sofisticação Léxica:** Avalia a riqueza e a especificidade do vocabulário utilizado, por meio de métricas como diversidade léxica, calculada através da relação type-token, frequência média das palavras de conteúdo, nível de concretude e polissemia das palavras.
- **Complexidade Sintática:** Analisa a estrutura das frases, utilizando métricas como o tamanho médio das sentenças e a frequência de diferentes tipos de estruturas sintáticas
- **Dispositivos de Coesão:** Mede a utilização de mecanismos que conectam as ideias no texto, como a incidência de conectivos lógicos e a sobreposição de argumentos entre sentenças adjacentes.
- **Tamanho e Extensão:** Quantifica aspectos básicos como o número de palavras por sentença e o tamanho relativo das seções principais.

As métricas de Estrutura Textual, focam na adequação do projeto de pesquisa aos requisitos do gênero:

- **Coerência:** Verifica a articulação entre diferentes partes do projeto, como a relação entre título e objetivo, a conexão entre palavras-chave e o referencial teórico, e a consistência entre título, objetivo e palavras-chave.
- **Objetivo:** Analisa a clareza e precisão na formulação do objetivo geral, avaliando o uso de verbos adequados, a quantidade de ações propostas, a ausência de aspectos metodológicos e de termos vagos.
- **Metodologia:** Avalia a completude e clareza da descrição metodológica, verificando a menção à modalidade da pesquisa, procedimentos de coleta de dados, classificação quanto aos objetivos, abordagem, definição de sujeitos/participantes e período de realização.

Para definir os limiares para a geração de feedback com base nas Métricas de Complexidade Textual e Estrutura Textual, foram realizados 3 estudos prévios, um analisando 2152 trabalhos de conclusão de curso de 6 grandes áreas do conhecimento de 8 instituições federais de ensino [Meira et. al. 2024], outro contendo 206 projetos de pesquisa retirados de disciplinas semestrais de metodologia da pesquisa de 2019 a 2023 e o último contendo avaliações de 738 redações de vestibular, todas corrigidas e divididas entre 50 temas [Meira et. al. 2023], esses estudos preliminares resultaram no estabelecimento de

médias e desvios-padrão para cada métrica. Feedbacks são gerados quando o valor de uma métrica em um trabalho específico se distancia significativamente da média do corpus de referência. Para as métricas de Estrutura Textual, os feedbacks são gerados com base na presença ou ausência de elementos esperados ou na identificação de inadequações, conforme as regras definidas por [Bastiani 2022].

Os três trabalhos serviram para balizar a construção da ferramenta, além da geração dos feedbacks, os dados obtidos foram utilizados para fornecer ao professor uma comparação das métricas obtidas em cada trabalho com a média de trabalhos de diferentes tipos, gêneros e áreas do conhecimento.

Uma análise de correlação de Pearson foi realizada entre as métricas de Estrutura Textual e as notas atribuídas por professores em um estudo preliminar [Meira et al. 2023]. Os resultados indicaram que métricas individuais possuem baixa correlação com a nota final, mas agrupamentos de métricas, especialmente as relacionadas à Metodologia, apresentaram correlação mais forte e estatisticamente significativa ($r = 0.79$, $p < 0.05$), corroborando achados da literatura [McNamara e Graesser 2012] de que categorias de métricas são mais preditivas que métricas isoladas.

3.3 Metodologia de Aplicação e Avaliação da Ferramenta

A avaliação da ferramenta Athos envolveu um total de 57 projetos de pesquisa, desenvolvidos por alunos distribuídos em cinco turmas, ao longo de um semestre letivo. Esses projetos foram acompanhados por três professores experientes, responsáveis por disciplinas de metodologia da pesquisa em cursos de licenciatura de diferentes áreas.

Os trabalhos dos alunos passaram por quatro fases distintas de entrega – Introdução, Revisão de Literatura, Metodologia e Versão Final – refletindo a evolução dos projetos ao longo do semestre. No processo, foram geradas 216 instâncias de feedback automático, vinculadas às métricas de Complexidade e Estrutura Textual, que os professores utilizaram para orientar e avaliar os estudantes.

Para cada trabalho avaliado em cada fase, foram coletados dados quantitativos através dos registros do sistema Athos, registrando como o professor interagiu com o feedback gerado automaticamente:

- Aceitação integral do feedback.
- Alteração parcial (até 50% do texto).
- Alteração substancial (mais de 50%).
- Rejeição completa do feedback.

Adicionalmente, os professores foram incentivados a registrar comentários qualitativos sobre os feedbacks gerados diretamente na interface da ferramenta durante o processo de avaliação.

Ao final do semestre, os dados de log e os comentários foram analisados. Os dados quantitativos sobre a aceitação/alteração dos feedbacks foram sumarizados para identificar padrões de uso. Os comentários qualitativos foram examinados por meio de análise de conteúdo [Saldaña 2013], utilizando um processo de codificação em dois ciclos realizado por dois codificadores independentes, seguido de conciliação, para identificar temas

recorrentes nas percepções dos professores sobre a qualidade e utilidade dos feedbacks automáticos.

Com base nessa análise inicial, foram elaborados roteiros para entrevistas semiestruturadas individuais com cada professor. As entrevistas buscaram aprofundar a compreensão sobre a experiência de uso da ferramenta, focando em aspectos como: a percepção sobre a agilidade e qualificação do processo avaliativo; a utilidade de funcionalidades específicas, como gráficos, validação de extrações e edição de feedback; as dificuldades encontradas; o impacto na qualidade e consistência do feedback fornecido; a percepção sobre a receptividade dos alunos; e sugestões de melhoria. Episódios específicos de uso identificados nos registros (aceitação total, alteração significativa, rejeição) foram discutidos para obter explicações contextuais. As entrevistas foram transcritas e submetidas à análise de conteúdo, seguindo o mesmo processo de codificação e categorização utilizado para os comentários, visando responder à questão central da pesquisa sobre como a ferramenta auxiliou no processo de avaliação e feedback.

4. Resultados e Discussão

A avaliação da ferramenta Athos no contexto real de disciplinas de metodologia científica forneceu dados quantitativos e qualitativos sobre sua aplicação e aceitação pelos professores participantes. Esta seção apresenta os principais resultados obtidos a partir dos registros de uso do sistema e das entrevistas semiestruturadas, discutindo suas implicações à luz do referencial teórico e dos objetivos da pesquisa.

4.1 Análise Quantitativa da Interação Professor-Ferramenta

A efetividade da ferramenta Athos no apoio à avaliação foi verificada na forma como os professores interagiram e utilizaram os feedbacks gerados automaticamente, a partir de 216 instâncias de feedback gerados para 57 projetos, associadas às métricas de Complexidade e Estrutura Textual. A análise dos registros revelou que, do total de feedbacks gerados, 150 foram aceitos integralmente pelos professores ou passaram apenas por ajustes mínimos, como pequenas alterações de redação ou contextualizações pontuais, antes de serem enviados aos alunos. Esse número corresponde a uma taxa geral de aceitação de aproximadamente 69,44%.

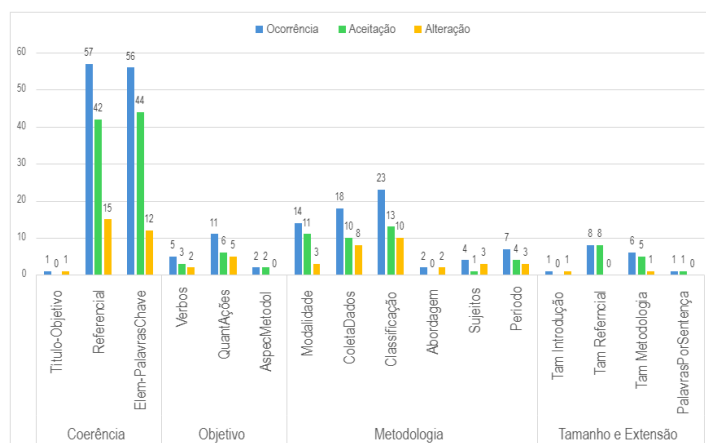


Figura 2 - Disposição dos feedbacks por métricas

Os 66 feedbacks restantes (30,56%) foram alterados mais significativamente ou completamente descartados e reescritos pelos professores. A Figura 2 detalha a frequência de geração e aceitação por métrica específica. Observou-se que métricas relacionadas à coerência estrutural, como a relação entre palavras-chave e referencial e a consistência entre título, objetivo e palavras-chave, foram frequentemente acionadas e apresentaram taxas de aceitação relativamente altas. Por outro lado, feedbacks relacionados a aspectos mais sutis da metodologia ou da formulação do objetivo tiveram taxas de alteração/rejeição um pouco maiores, sugerindo que a avaliação automatizada desses pontos ainda requer um olhar mais crítico e contextualizado do professor.

A figura 2 ainda mostra que a maioria das métricas foi acionada predominantemente para indicar problemas, ou seja, feedbacks corretivos, em vez de pontos positivos. Isso é esperado, dado o foco da ferramenta em auxiliar na identificação de áreas para melhoria. A taxa de aceitação de 69,44% sugere que a ferramenta foi capaz de gerar sugestões relevantes e úteis na maioria das vezes, aliviando parte do trabalho do professor na identificação de problemas recorrentes ou padronizáveis na estrutura e complexidade dos projetos.

4.2 Análise Qualitativa da Percepção Docente

As entrevistas semiestruturadas realizadas com os três professores participantes desta etapa do estudo, juntamente com os comentários registrados na plataforma, forneceram subsídios relevantes para a compreensão das percepções docentes em relação à ferramenta Athos. A análise de conteúdo revelou temas recorrentes, destacando-se a percepção de agilidade e otimização do tempo como um dos principais benefícios.

Os professores relataram que a ferramenta ajudou a acelerar o processo de correção, especialmente na identificação de problemas estruturais e formais comuns. Ao receber um feedback inicial gerado pela Athos, eles puderam concentrar sua atenção em aspectos mais complexos e substantivos do conteúdo, na argumentação e na originalidade do trabalho, em vez de gastar tempo excessivo apontando erros recorrentes de estrutura ou clareza que a ferramenta já sinalizava. Um professor comentou: *“Consegui focar mais na discussão das ideias do aluno, porque a parte mais ‘mecânica’ da estrutura já vinha apontada pela ferramenta. Isso otimizou bastante meu tempo”*.

Outro tema relevante foi a qualidade e especificidade do feedback. Embora a taxa de aceitação quantitativa tenha sido de cerca de 70%, as entrevistas revelaram que mesmo os feedbacks alterados ou reescritos muitas vezes partiram da sugestão inicial da ferramenta. Os professores valorizaram a capacidade da Athos de apontar problemas específicos relacionados às métricas (e.g., falta de clareza no objetivo, desconexão entre seções), servindo como um ponto de partida para um feedback mais elaborado e contextualizado. A função de validação das extrações (título, objetivo, palavras-chave) foi considerada crucial, garantindo que a análise automatizada partisse de premissas corretas e permitindo o ajuste fino antes da geração do feedback final, reforçando a importância da abordagem “human-in-the-loop”.

Os gráficos comparativos disponíveis na Athos foram mencionados como úteis para obter uma visão geral rápida do desempenho da turma e identificar alunos que poderiam necessitar de maior atenção, embora seu uso principal tenha sido mais diagnóstico do que diretamente na formulação do feedback individual.

As limitações também foram apontadas. Alguns feedbacks gerados foram considerados genéricos ou, em certos casos, não totalmente adequados ao contexto específico do trabalho do aluno, exigindo intervenção do professor. Isso reforça a ideia de que a avaliação da escrita acadêmica, especialmente de projetos de pesquisa, envolve uma subjetividade e uma compreensão contextual que a automação completa ainda não alcança plenamente. A necessidade de o professor revisar, validar e, frequentemente, complementar o feedback automático foi vista não como uma falha, mas como uma característica essencial do processo colaborativo entre humano e máquina.

Em relação ao impacto nos alunos, os professores não perceberam uma mudança drástica na receptividade geral ao feedback, mas alguns notaram que a maior agilidade no retorno e a clareza em apontar problemas estruturais podem ter contribuído para um processo de revisão mais focado por parte dos estudantes. A consistência proporcionada pela ferramenta ao aplicar os mesmos critérios a todos os trabalhos também foi vista como um aspecto positivo para a equidade da avaliação.

4.3 Discussão

Os resultados deste estudo corroboram a literatura sobre o potencial e os desafios das ferramentas de AWE e da Analítica da Escrita no apoio à avaliação e ao feedback [Wang e Han 2022], [Wilson et al. 2014]. A taxa de aceitação de 69,44% dos feedbacks gerados pela Athos indica uma capacidade significativa da ferramenta em identificar problemas relevantes nos projetos de pesquisa, alinhando-se com a premissa de que a automação pode auxiliar na avaliação de aspectos mais padronizáveis da escrita.

A percepção dos professores sobre a otimização do tempo e a possibilidade de focar em aspectos mais complexos do trabalho do aluno reforça o argumento de que essas ferramentas podem aliviar a sobrecarga docente. A Athos atuou como um “primeiro leitor”, sinalizando pontos de atenção e permitindo que o professor exercesse seu papel de mediador de forma mais estratégica [Barbosa 2017].

A importância da abordagem “human-in-the-loop” [Seo et al. 2021] foi claramente evidenciada. A necessidade de validação das extrações e a possibilidade de edição do feedback final foram cruciais para a aceitação da ferramenta pelos professores. Isso demonstra que a combinação da análise automatizada com a validação e o refinamento humano é fundamental para garantir a qualidade e a adequação do feedback em tarefas complexas como a avaliação de projetos de pesquisa. A ferramenta não substitui o professor, mas o potencializa, fornecendo informações e um ponto de partida para sua análise especializada.

A correlação mais forte encontrada entre o agrupamento de métricas de Metodologia e as notas, conforme estudo preliminar na tese, sugere que a análise estrutural realizada pela Athos, especialmente nesse quesito, captura aspectos relevantes para a qualidade percebida do projeto, indo além de simples verificações formais. Isso indica o potencial de métricas bem definidas e contextualizadas para o gênero textual em questão.

As limitações apontadas pelos professores, como a eventual generalidade de alguns feedbacks, destacam a necessidade contínua de aprimoramento dos algoritmos e das bases de conhecimento das ferramentas de AWE. A capacidade de adaptação a diferentes estilos e

níveis de proficiência, bem como a sensibilidade ao contexto específico de cada trabalho, permanecem como desafios importantes [Sherri e Burstein, 2013].

Em suma, os resultados sugerem que a ferramenta Athos, ao combinar análise automatizada de complexidade e estrutura textual com a intervenção ativa do professor, representa um avanço promissor no suporte à avaliação de projetos de pesquisa. Ela oferece um meio de otimizar o processo de feedback, mantendo o papel central do docente na orientação e no desenvolvimento das habilidades de escrita acadêmica dos estudantes.

5. Conclusão

Este trabalho abordou o desafio da avaliação de projetos de pesquisa na graduação, explorando o potencial da Analítica da Aprendizagem e do Processamento de Linguagem Natural para apoiar professores nesse processo complexo e demandante. O desenvolvimento e a avaliação da ferramenta Athos demonstraram a viabilidade de uma abordagem híbrida, que combina a análise automatizada de aspectos textuais com a validação e o refinamento pelo docente.

Os resultados indicaram que a Athos foi capaz de avaliar projetos de pesquisa e gerar feedbacks relevantes sobre estrutura e complexidade textual, com uma taxa de aceitação pelos professores participantes de aproximadamente 70%. A ferramenta mostrou-se útil ao otimizar o tempo dos docentes, permitindo que concentrassem sua atenção em aspectos mais substantivos e individualizados do trabalho dos alunos, em vez de se deterem em erros estruturais ou formais recorrentes. A abordagem “human-in-the-loop”, com a validação das extrações e a edição dos feedbacks, foi fundamental para a aceitação e eficácia da ferramenta, reforçando que a tecnologia deve atuar como um suporte à expertise pedagógica, e não como um substituto.

Os achados alinham-se à literatura que aponta os benefícios das métricas textuais e das ferramentas de AWE como complemento à avaliação tradicional [Crossley e McNamara, 2011], [McNamara e Graesser 2012], especialmente em contextos de larga escala. A Athos, ao integrar métricas de complexidade (NILC-Metrix) e estrutura e utilizar IA generativa para consolidar feedbacks, representa uma contribuição tecnológica para a área de Informática na Educação, oferecendo um recurso específico para o gênero projeto de pesquisa.

Contudo, a pesquisa também evidenciou limitações. A avaliação automatizada ainda enfrenta desafios na compreensão de nuances contextuais e na adaptação a diferentes estilos de escrita. A necessidade de aprimoramento contínuo dos algoritmos e das bases de conhecimento é evidente. Além disso, a dependência de ferramentas externas introduz questões de disponibilidade e custo que precisam ser consideradas em implementações futuras.

Como trabalhos futuros, vislumbra-se a expansão da Athos para outros gêneros textuais acadêmicos, a incorporação de sistemas de recomendação de conteúdo para auxiliar os alunos na melhoria de seus textos e o aprofundamento dos estudos sobre a correlação entre métricas textuais específicas e a qualidade percebida dos trabalhos em diferentes áreas do conhecimento. A investigação sobre o impacto direto no aprendizado e na autoeficácia dos alunos ao receberem feedbacks gerados ou mediados pela ferramenta também representa uma via importante para pesquisas futuras.

Em suma, este estudo reforça a importância do desenvolvimento de soluções tecnológicas que apoiem os processos educacionais, particularmente a produção textual acadêmica. Ferramentas como a Athos, que buscam auxiliar o professor em tarefas avaliativas, têm o potencial de qualificar o processo de feedback e, conseqüentemente, contribuir para o desenvolvimento das competências de escrita dos estudantes, com repercussões positivas para sua formação acadêmica e profissional.

6. Referências

- BASTIANI, Ederson; ZARTH, Antonio Miguel Faustini; REATEGUI, Eliseo. MTA: uma ferramenta apoiada em IA para aprimoramento da escrita de projetos de pesquisa acadêmica. *Renote*, v. 19, n. 1, p. 1-10, 2025. DOI: <https://doi.org/10.22456/1679-1916.141555>.
- BIRD, S., LOPER, E. e KLEIN, E. *Natural Language Processing with Python*. O'Reilly Media, Inc. 2009
- CARLESS, D.; BOUD, D. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, v. 43, n. 8, p. 1315-1325, 2018.
- CROSSLEY, S. A., MCNAMARA, D. S. e MCCARTHY, P. M. Learning analytics and writing instruction. *Computers and Composition*, 27(1), 34-47. 2010.
- CROSSLEY, S. A., ALLEN, D. B. e MCNAMARA, D. S. Text readability and intuitive simplification: A comparison of readability formulas and Coh-Metrix. *Reading in a Foreign Language*, 23(1), 84-101. 2011.
- COPE, B. e KALANTZIS, M. Big Data Comes to School: Implications for Learning, Assessment, and Research. *AERA Open*, 2(2). 2016. <https://doi.org/10.1177/2332858416641907>
- EVERS, Aline; FINATTO, Maria José Bocorny. Linguística de Corpus, Léxico-Estatística Textual e Processamento de Linguagem Natural: perspectiva para estudos de vocabulário em produções textuais. *Revista GTLex*, v. 1, n. 2, p. 271, 2016. DOI: <https://doi.org/10.14393/lex2-v1n2a2016-3>.
- GALLIEN, Tara; EARLY, Jody Oomen. Personalized versus collective instructor feedback in the online courseroom: does type of feedback affect student satisfaction, academic performance and perceived connectedness with the instructor? *International Journal on E-Learning*, v. 7, n. 3, p. 463-476, 2005.
- HATTIE, John; TIMPERLEY, Helen. The Power of Feedback. *Review of Educational Research*, v. 77, n. 1, p. 81-112, 2007. DOI: <https://doi.org/10.3102/003465430298487>.
- HIGGINS, R.; HARTLEY, P.; SKELTON, A. The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in Higher Education*, v. 27, n. 1, p. 53-64, 2002. DOI: <https://doi.org/10.1080/03075070120099368>.
- JURAFSKY, D. e MARTIN, J. H. "Speech and Language Processing" (3rd ed.). 2018. Disponível em: <<https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>>

- LEAL, S.E., DURAN, M.S., SCARTON, C., HARTMANN, N.S., e ALUÍSIO, S.M. NILC-Metrix: Assessing the complexity of written and spoken language in Brazilian Portuguese. *Language Resources and Evaluation*, p. 1-38, 2023. DOI: <https://doi.org/10.1007/s10579-023-09693-w>
- LILLIS, T. e CURRY, M. J. Academic writing in a global context: The politics and practices of publishing in English. Routledge. *J Bus Tech Commun*, v. 22, p. 179-198, 2010.
- LIU, Bing. Sentiment analysis and opinion mining. Cham: Springer, 2012. 167 p. (Synthesis Lectures on Human Language Technologies). DOI: <https://doi.org/10.1007/978-3-031-02145-9>.
- MCNAMARA, Danielle S.; GRAESSER, Arthur C. Coh-Metrix: an automated tool for theoretical and applied natural language processing. In: MCCARTHY, Philip; BOONTHUM-DENECKE, Chutima (Org.). *Applied natural language processing: identification, investigation and resolution*. Hershey: IGI Global, 2012. p. 18. DOI: <https://doi.org/10.4018/978-1-60960-741-8.ch011>.
- MEIRA, R. R.; WEIAND, A.; REATEGUI, E.; BIGOLIN, M.; MOTZ, R. A Análítica da Escrita para Identificação de Indicadores de Qualidade Textual. *RENOTE*, Porto Alegre, v. 21, n. 2, p. 342–351, 2023. DOI: 10.22456/1679-1916.137756
- MEIRA, R. R.; REATEGUI, E.; MOTZ, R. Análise de trabalhos de conclusão de curso utilizando técnicas de processamento de linguagem natural. *RENOTE*, Porto Alegre, v. 22, n. 1, p. 456–465, 2024. DOI: 10.22456/1679-1916.141571.
- NADEAU, David; SEKINE, Satoshi. A survey of named entity recognition and classification. *Linguisticæ Investigationes*, v. 30, n. 1, p. 3-26, jan. 2007. <https://doi.org/10.1075/li.30.1.03nad>
- NICOL, D. J. e MACFARLANE-DICK, D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, v. 31, n. 2, p. 199-218. 2006. DOI: <https://doi.org/10.1080/03075070600572090>
- SALDAÑA, J. The coding manual for qualitative researchers. London: Sage. 2013.
- SEO, K., TANG, J., ROLL, I. The impact of artificial intelligence on learner–instructor interaction in online learning. *Int J Educ Technol High Educ*, v. 18, n. 54. 2021. <https://doi.org/10.1186/s41239-021-00292-9>
- SOUSA, Luciano Dias de, ALMEIDA, Flávio Aparecido de, BARD, Lucimere Aleixo, CANCELA, Lucas Borcard. Os desafios enfrentados pelos professores no processo de avaliação no ensino superior. *Revista de Gestão e Avaliação Educacional*, v. 7, n. 16, p. 59-66, 2018.
- TASKIRAN, Ayşe e GÖKSEL, Nil. Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance Turkish Online Journal of Distance Education, v. 23, n. 2, p. 120-139, 2022. DOI: <https://doi.org/10.17718/tojde.1096260>

- WANG, Z.; HAN, F. The Effects of Teacher Feedback and Automated Feedback on Cognitive and Psychological Aspects of Foreign Language Writing: A Mixed-Methods Research. *Frontiers in Psychology*, v. 13, 2022. DOI: <https://doi.org/10.3389/fpsyg.2022.909802>
- WARSCHAUER, M. e WARE, P. Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research*, v. 10, n. 2, p. 157-180. 2006. DOI: <https://doi.org/10.1191/1362168806lr190oa>
- WILSON, Joshua; OLINGHOUSE, Natalie G.; ANDRADA, Gilbert N. Does Automated Feedback Improve Writing Quality? *Learning Disabilities: A Contemporary Journal*, v.12, n.1, p. 93-118. Mar 2014.