

UMA ANÁLISE COMPARATIVA DE FERRAMENTAS COMPUTACIONAIS PARA AUXÍLIO À VERIFICAÇÃO DE PLÁGIO EM DOCUMENTOS

Thiago Monteiro Tuxi – tmtuxi@ime.eb.br

Marcelo de Araújo Oliveira – oliveira.marcelo@ime.eb.br

Paulo César Pellanda – pellanda@ime.eb.br

Antonio Eduardo Carrilho da Cunha – carrilho@ime.eb.br

Instituto Militar de Engenharia

Praça General Tibúrcio, 80, Praia Vermelha

22.290-270 – Rio de Janeiro – RJ

Resumo: Uma análise comparativa de seis ferramentas computacionais para auxílio à verificação de plágio em documentos é apresentada. Os resultados dessa análise são discutidos com base em aspectos qualitativos e quantitativos, bem como em relação à sensibilidade ao parafraseamento e à detecção de textos traduzidos.

Palavras-chave: Plágio, Verificação de plágio, Ferramentas computacionais.

1. INTRODUÇÃO

O plágio em trabalhos acadêmicos tornou-se um assunto de discussão relevante recentemente no meio científico. Com a evolução tecnológica, o que em décadas atrás demandava procura e leitura em diversas fontes e locais diferentes, hoje em dia pode ser feito em tempo relativamente menor, por meio de consulta em base de dados especializadas disponíveis na internet. Além de se encontrar a informação de forma mais rápida, as mídias digitais permitem a reprodução de conteúdos de forma facilitada, cuja fonte deve ser referenciada para não se caracterizar o plágio. Há vinte anos, em (PINHEIRO; KOURY; MEDEIROS, 1997), os autores comentavam que a grande quantidade de informações disponíveis na Internet tornava-a um instrumento poderoso para o ensino de engenharia. Entretanto, já ressaltavam naquela época que "a facilidade de pesquisa na Internet está estimulando os alunos a plagiar textos na redação de trabalhos escolares".

O plágio em trabalhos acadêmicos se caracteriza pela reprodução parcial ou total de conteúdos de outros trabalhos, sem a devida citação da fonte pesquisada. Além de ser crime previsto em lei no Brasil, o plágio é antiético e compromete o desenvolvimento científico. Segundo (RABELLO, 2006), "a inadmissão do plágio no ambiente acadêmico deve-se, além da questão legal, à credibilidade dos trabalhos, fundamental para a evolução da ciência no país. [...] sem honestidade intelectual não existe evolução científica".









Como a questão do plágio está diretamente relacionada ao aspecto ético, é importante que as Instituições de Ensino, em particular as de engenharia, adotem medidas para evitálo e que as disciplinas que tratam da questão da ética profissional discutam sobre o plágio em trabalhos acadêmicos e na pesquisa de um modo geral.

Reavaliar as concepções contemporâneas e os projetos do amanhã são tarefas da engenharia. Para isso é importante que os cursos de engenharia se articulem com os mais diferentes segmentos, assumindo sua função de formadores de engenheiros cidadãos e conscientes de sua ação social. [...] É importante que os Engenheiros adotem uma atitude ética, de reflexão quanto às consequências de sua atividade, de modo a buscar elevar ao máximo a sociedade no desenvolvimento de sua atuação profissional. [...] Urge, no período presente, que os engenheiros adotem uma atitude ética onde haja, predominantemente, o respeito à vida e, sobretudo, do homem como sujeito cidadão e não, exclusivamente, como um cliente em potencial (MENESTRINA et al., 2014).

Por outro lado, a própria evolução tecnológica que facilita a cópia de conteúdos de trabalhos também oferece recursos para fiscalizar a ocorrência do plágio. De forma institucional ou individual, professores e universidades vem adotando ferramentas computacionais para avaliar a ocorrência de plágio em trabalhos acadêmicos. Por exemplo, em (OLIVEIRA; PELLANDA, 2017), apresenta-se uma abordagem didática para tratar a questão do plágio em trabalhos acadêmicos, que pode ser aplicada em cursos de engenharia. Este artigo apresenta uma análise comparativa de algumas das ferramentas computacionais disponíveis que, supostamente, avaliam a existência de plágio em documentos.

Este trabalho está organizado como segue. Na seção 2, são apresentados os conceitos preliminares pertinentes ao tema. Na seção 3, é apresentada uma análise comparativa das ferramentas computacionais disponíveis e os resultados são discutidos com base em aspectos qualitativos e quantitativos, bem como em relação à sensibilidade ao parafraseamento e à detecção de textos traduzidos. Por fim, são apresentadas considerações finais na seção 4.

2. CONCEITOS PRELIMINARES

2.1. Definição de plágio e suas implicações legais

O plágio em trabalhos acadêmicos se caracteriza pela reprodução parcial ou total de conteúdos de outros trabalhos, sem a devida citação da fonte pesquisada. De acordo com o Escritório Federal de Política de Ciência e Tecnologia Norte Americano, "plágio é a apropriação das ideias, processos ou palavras de outra pessoa sem o devido crédito, incluindo aqueles obtidos através da revisão confidencial de outras propostas de pesquisa e manuscritos" (ELSEVIER, 2017). No Brasil, o plágio é considerado crime e o Código Penal estabelece o seguinte no seu Art. 184: "Violar direitos de autor e os que lhe são conexos: Pena - detenção, de 3 (três) meses a 1 (um) ano, ou multa".

2.2. Métodos de detecção de plágio

De acordo com (MAURER; KAPPE; ZAKA, 2006), existem três métodos de detecção de plágio. O primeiro método realiza a comparação, palavra a palavra, de um docu-









mento em relação a um grupo de outros documentos, que podem estar localmente ou remotamente disponíveis. O segundo utiliza a análise de um parágrafo característico, procurando-o por intermédio de um motor de busca tal como o *Google*. O terceiro emprega a estilometria, isto é, busca analisar o estilo de escrita do autor, comparando certo documento com documentos anteriores desse mesmo autor. Dos três métodos, o primeiro é o mais usado de todos (MAURER; KAPPE; ZAKA, 2006). Por esta razão, este trabalho foca o primeiro método.

A técnica mais comum na comparação de documentos é a que compara palavras de tamanhos moderados, denominadas de *impressões digitais*, que são buscadas em uma base de dados e as similaridades são classificadas. O resultado é um índice de similaridade entre o documento submetido e os documentos analisados (MANBER, 1994 apud MAURER; KAPPE; ZAKA, 2006). A Figura 1 ilustra um sistema típico de verificação de plágio em documentos.

Submissão de Documentos

Leitura do Documento
Geração das impessões digitais

Algoritmos de consulta e classificação, detectando similaridades entre diversos fragmentos de texto e de fontes indexadas.

Organização das similaridades, Identificação das fontes

Resultados

Figura 1 – Metodologia de verificação de plágio em documentos.

Fonte: Adaptada de (MAURER; KAPPE; ZAKA, 2006).

3. ANÁLISE DAS FERRAMENTAS PARA VERIFICAÇÃO DE PLÁGIO EM DOCUMENTOS

Esta seção apresenta uma análise comparativa de algumas ferramentas para verificação de plágio em documentos.

3.1. Características gerais

Atualmente, existem diversas ferramentas computacionais para a verificação de plágio em documentos, podendo estas serem de custo elevado, de baixo custo e até mesmo gratuitas. Algumas possuem uma plataforma de trabalho online, em que os documentos são enviados ao servidor para o devido processamento. O resultado geralmente pode ser obtido diretamente na página da plataforma ou, em alguns casos, serem enviados para o e-mail cadastrado. Outras ferramentas necessitam ser instaladas no computador e realizam o processamento localmente. Vale lembrar que, mesmo sendo o processamento local, a conexão à internet se faz necessária para a realização da pesquisa.







Joinville/SC – 26 a 29 de Setembro de 2017 UDESC/UNISOCIESC

"Inovação no Ensino/Aprendizagem em Engenharia"



Devido a limitações de espaço e custo, este trabalho analisou apenas seis ferramentas para verificação de plágio gratuitas ou de baixo custo, quais sejam, *New jester*¹, *Plagiarism Checker*², *Plagius*³, *Farejador de Plágio*⁴, *Plagiarism Detector*⁵ e *COPYLEAKS*⁶. Os programas profissionais, tais como *Turnitin*⁷ ou *Ithenticate*⁸, não foram analisados devido ao seu alto custo e à não disponibilização de versões de demonstração.

Os programas escolhidos foram analisados levando-se em consideração aspectos qualitativos e quantitativos, detalhados a seguir:

Aspectos qualitativos:

Aspectos quantitativos:

- Tecnologia utilizada;
- Custo;
- Tipos de arquivos suportados.
- Quantidade máxima de palavras por busca.

3.2. Resultados

New Jester

A ferramenta *New jester*, anteriormente denominada *Doc Cop*, é gratuita e opera em uma plataforma *online*. Permite a verificação de plágio apenas entre documentos submetidos, não comparando-os com outros de uma determinada base de dados, sendo mais indicado, por exemplo, para verificação de plágio entre trabalhos acadêmicos de uma mesma classe de alunos. Permite analisar até oito documentos, contendo até 15.000 palavras cada um. Envia um relatório para o e-mail cadastrado em um período médio de uma hora. A Figura 2 apresenta um exemplo de um relatório de dois textos submetidos para a análise. Nota-se que os trechos em que há possibilidade de plágio, são enfatizados com cores diferentes.

Plagiarism Checker

Semelhante ao *New Jester*, o *Plagiarism Checker* também é uma ferramenta gratuita e opera em uma plataforma online. Verifica a existência de plágio no documento ou texto submetido por meio de comparação com uma base de dados da internet. Os textos podem ser inseridos diretamente na plataforma ou submetidos em arquivos do tipo .docx. Permite analisar no máximo 1.000 palavras por busca, e informa a possibilidade de existência de plágio no mesmo ambiente da plataforma, indicando as fontes das similaridades encontradas. É possível comparar os textos encontrados diretamente na plataforma, selecionando a fonte apresentada. Os trechos suspeitos são enfatizados com cor vermelha. A Figura 3 apresenta um exemplo de um resultado de uma da análise de um texto submetido.

⁸http://www.ithenticate.com/







¹https://newjester.com/

²https://www.duplichecker.com/

³http://www.plagius.com/

⁴http://www.plagiarismcombat.com/

⁵https://plagiarism-detector.com/index.php

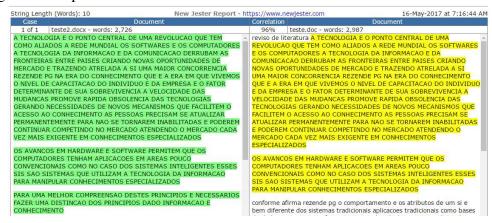
⁶https://copyleaks.com/

⁷http://turnitin.com/pt_br/

Engenharia"

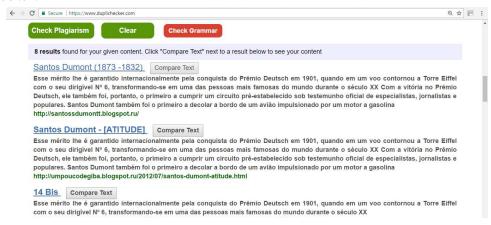


Figura 2 – Captura de tela do relatório da ferramenta New Jester.



Fonte: Autoria própria.

Figura 3 – Captura de tela do resultado de análise da ferramenta *Plagiarism Checker*.



Fonte: Autoria própria.

Plagius

Dentre as ferramentas analisadas e que não são gratuitas é a que possui o menor custo de aquisição. Opera no próprio computador do usuário. A licença do tipo pessoal custa R\$ 36,90 e, a profissional, R\$ 65,90 (valor de consulta realizada em 22/05/17), para um período de uso de um ano. Assim como a ferramenta anterior, realiza a análise de textos ou documentos por meio de comparação com uma base de dados da internet. Os textos podem ser inseridos diretamente na interface ou lidos diretamente de arquivos de extensão .doc, .docx, .odt, .pdf, .html ou .txt. Segundo o desenvolvedor do software, não existe limitação quanto à quantidade de palavras por busca. Após a análise, é gerado um relatório em formato .html, conforme exemplifica a Figura 4, em que os trechos suspeitos são indicados juntamente com as fontes encontradas.



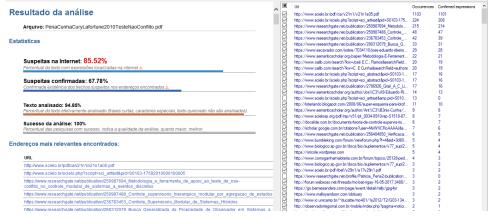




"Inovação no Ensino/Aprendizagem em Engenharia"



Figura 4 – Captura de tela do relatório da ferramenta *Plagius*.

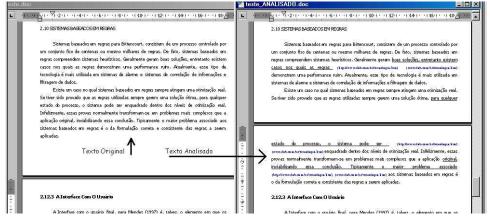


Fonte: Autoria própria.

Farejador de Plágio

Trata-se de uma ferramenta de custo relativamente baixo e que opera no próprio computador do usuário. A licença custa \$ 19,90 USD, para um período de uso de dois anos. A renovação custa \$ 9,90 USD (valores de consulta realizada em 22/05/17). Considerando o valor anual e a cotação atual do dólar, esta ferramenta possui custo semelhante à anterior. Realiza a análise da existência de plágio em documentos por meio de comparação com uma base de dados da internet. Os textos são lidos diretamente de arquivos de extensão .doc, .docx, .html ou .txt. O desenvolvedor do software não cita nenhuma restrição quanto à quantidade máxima de palavras por busca. Após a análise, o software indica na própria plataforma a quantidade de ocorrências e possíveis fontes de plágio. Por fim, é gerado um novo arquivo em formato .doc com os links para as fontes encontradas. A Figura 5 ilustra um exemplo de análise efetuada.

Figura 5 – Captura de tela do resultado de análise da ferramenta *Farejador de Plágio*.



Fonte: Autoria própria.









Plagiarism Detector

Engenharia"

Plagiarism Detector é uma ferramenta paga, que opera no computador do usuário. Custa de \$ 39 a 99 USD, dependendo do tipo de licenças e do número de computadores a serem registrados (consulta realizada em 22/05/17). A licença é por tempo indeterminado. Realiza a análise da existência de plágio em documentos por meio de comparação com uma base de dados da internet. Os textos são lidos diretamente de arquivos de extensão .doc, .docx, .pdf, .rtf, .html, .htm, .php, .ppt, .pptx ou .txt. Na página do desenvolvedor não há informação quanto à quantidade máxima de palavras por busca. Após a análise, é gerado um relatório que pode ser salvo em formato .pdf, .html, .csv. Os trechos com suspeita de plágio são indicados juntamente com as fontes encontradas. A ferramenta verifica se há plágio de fato, ou se o trecho com suspeita de plágio foi devidamente referenciado. A figura 6 ilustra uma parte do relatório gerado pela ferramenta.

Plagiarism Detection Chart:

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism

Plagiarism Detection Chart:

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Plagiarism (2.50%)

Referenced 69% / Linked 0%
Original - 10% / 22% - Plagiarism

Figura 6 – Captura de tela de parte de relatório da ferramenta *Plagiarism Detector*.

Fonte: Autoria própria.

A ferramenta possui, como opção de análise, a possibilidade de avaliação de parafraseamento. Nessa opção, avalia-se a similaridade de conteúdos em textos escritos com palavras distintas. Por exemplo, o texto a seguir foi analisado nas duas opções e a Figura 7 ilustra ambos os resultados.

Texto Original:

Santos Dumont também foi o primeiro a decolar a bordo de um avião impulsionado por um motor a gasolina. Em 23 de outubro de 1906 voou cerca de 60 metros a uma altura de dois a três metros com o Oiseau de Proie' (francês para «ave de rapina»), no Campo de Bagatelle, em Paris. Menos de um mês depois, em 12 de novembro, diante de uma multidão de testemunhas, percorreu 220 metros a uma altura de 6 metros com o Oiseau de Proie III. Esses voos foram os primeiros homologados pelo Aeroclube da França de um aparelho mais pesado que o ar, e possivelmente a primeira demonstração pública de um veículo levantando voo por seus próprios meios, sem a necessidade de uma rampa para lançamento.

Texto Parafraseado:

Santos Dumont também foi o primeiro a decolar a bordo de um avião com motor a combustão. Em 23 de outubro de 1906 voou cerca de 60 metros a uma altura de dois metros com o Oiseau de Proie, o Campo Bagatelle em Paris. Menos de um mês depois, no dia 12 de novembro, diante de uma multidão de testemunhas, sobrevoou 220 metros a uma altura de 6 metros com Oiseau de Proie III. Estes voos foram os primeiros aprovados pelo Clube Aero da França de uma máquina um mais pesada do que o ar, e possivelmente a primeira demonstração pública de um voo por seus próprios meios, sem a necessidade de rampa para o lançamento.

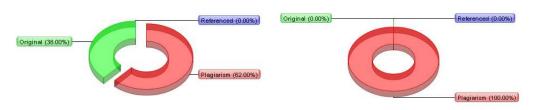








Figura 7 – Capturas de tela de partes dos relatórios.



- (a) Opção de análise palavra a palavra.
- (b) Opção de análise com parafraseamento.

Fonte: Autoria própria.

COPYLEAKS

É uma ferramenta paga, que opera em uma plataforma online. A tarifação é feita por um sistema de créditos, em que cada crédito possibilita a análise de até 250 palavras. Existem três opções: a gratuita, com possibilidade de análise de até 10 créditos por mês; a pré-paga, variando de \$ 12,99 USD (100 créditos) até \$ 1839,99 USD (20 mil créditos) e assinatura com pagamento mensal, variando de \$ 10,99 USD (100 créditos) até \$ 1299,99 USD (20 mil créditos) (consulta realizada em 22/05/2017). Os textos a serem submetidos podem estar nas seguintes extensões: .doc, .docx, .pdf, .rtf, .html, .htm, .php, .ppt, .pptx ou .txt. Na página do desenvolvedor não há informação quanto a quantidade máxima de palavras por busca. Como resultado, informa a possibilidade de existência de plágio no mesmo ambiente da plataforma, indicando as fontes das similaridades encontradas. Os trechos com suspeita de plágio são indicados juntamente com as fontes encontradas. A Figura 8 ilustra o resultado de uma análise efetuada.

Figura 8 – Captura de tela do relatório do software *COPYLEAKS*.



Fonte: Autoria própria.

3.3. Análise comparativa dos softwares

A Tabela 1 apresenta as principais características das ferramentas analisadas.







Tabela 1 – Análise comparativa dos softwares para auxílio à verificação de plágio em documentos.

Nome Do Software	New Jester	Plagiarism Checker	Plagius	Farejador de Plágio	Plagiarism Detector	COPYLEAKS
Custo	Gratuito	Gratuito	R\$ 36,90 - 65,90 (1 ano)	\$ 19,90 USD (2 anos)	\$ 39 - 99 USD	12,99 - 1899,99 USD (por crédito)
Tecnologia	Plataforma Online	Plataforma Online	Plataforma em desktop, compatível com Windows 7,8,10	Plataforma em desktop, compatível com Windows 7,8,10	Plataforma em desktop, compatível com Windows 7,8,10	Plataforma Online
Arquivos suportados	.doc	Textos diretamente na plataforma ou .doc, .docx	Textos diretamente na plataforma ou .doc, .docx, .odt, .pdf, .html, .txt	.doc, .docx, .html, .txt	Textos diretamente na plataforma ou .doc, .docx, .odt, .pdf, .html, .txt	.doc, .docx, .odt, .pdf, .html, .txt
Limite de palavras	15000	1000	Não informado	Não informado	Não informado	Dependente do numero de créditos
Relatório	Enviado por e-mail	Apresentado diretamente na Plataforma	Apresentado diretamente na Plataforma	Apresentado diretamente na Plataforma, podendo ser salvo no formato .pdf, .html, .csv	Apresentado diretamente na Plataforma, podendo ser salvo no formato .pdf, .html, .csv	Apresentado diretamente na Plataforma









4. CONSIDERAÇÕES FINAIS

Engenharia"

Este trabalho analisou seis ferramentas computacionais para auxílio à verificação de plágio em documentos, gratuitas ou de baixo custo. Serve como referência para a adoção de uma dessas ferramentas por professores ou até mesmo instituições de ensino de engenharia. As ferramentas foram comparadas considerando aspectos qualitativos e quantitativos, com o intuito de auxiliar o leitor na escolha da mais adequada a seus objetivos e orçamento. Considera-se que tais ferramentas, de fato, auxiliam bastante no trabalho de verificação de plágio em documentos. Entretanto, para uma decisão final da existência de plágio, ainda se faz necessária a intervenção humana para inspeção dos textos indicados como suspeitos nas ferramentas analisadas.

4.1. Agradecimentos

O primeiro autor é bolsista CAPES-DS.

REFERÊNCIAS BIBLIOGRÁFICAS

ELSEVIER. *Get Published Quick Guide*. 2017. Author Services. Acessado em 27 de maio de 2017. Disponível em: (https://www.elsevier.com/authors/author-services).

MANBER, U. Finding similar files in a large file system. Winter USENIX Technical Conference, San Francisco, CA, USA, 1994.

MAURER, H.; KAPPE, F.; ZAKA, B. Plagiarism - a survey. *Journal of Universal Computer Science*, v. 12, n. 8, p. 1050–1083, ago. 2006.

MENESTRINA, T. C. et al. Ética e cidadania, uma visão sobre ciência, tecnologia e sociedade nos cursos de engenharia. *XLII Congresso Brasileiro de Ensino de Engenharia (COBENGE-14). Juiz de Fora: UFJF*, 2014.

OLIVEIRA, M. de A.; PELLANDA, P. C. Plágio em trabalhos acadêmicos: abordagem baseada em estudos de casos reais aplicada em cursos de engenharia. *COBENGE*, 2017.

PINHEIRO, P. C. da C.; KOURY, R. N. N.; MEDEIROS, E. B. Internet: Uma ferramenta para o ensino de engenharia. Salvador: UFBA, 1997.

RABELLO, C. *Idéias roubadas*. 2006. UNB Notícias. Acessado em 27 de maio de 2017. Disponível em: (http://www.secom.unb.br/unbagencia/ag0706-27.htm).

A COMPARATIVE ANALYSIS OF COMPUTATIONAL TOOLS FOR AIDING THE VERIFICATION OF PLAGIARISM IN DOCUMENTS

Resumo: A comparative analysis of six computational tools for aiding the verification of plagiarism in documents is presented. The results of this analysis are discussed considering qualitative and quantitative aspects, as well as with relation to the sensibility to paraphrasing and the detection of translated texts.

Palavras-chave: Plagiarism, plagiarism verification, computational tools.





