

CARACTERÍSTICAS GRAFOMÉTRICAS E NÃO GRAFOMÉTRICAS E SUA RELEVÂNCIA PARA IDENTIFICAÇÃO DE MANUSCRITOS: UM MAPEAMENTO SISTEMÁTICO.

Juliana Naomi Kawakami (PIC/UEM), Aline Maria Malachini Miotto Amaral (Orientadora). E-mail: ammmamaral@uem.br

Universidade Estadual de Maringá, Centro de Tecnologia, Maringá, PR.

Área e subárea do conhecimento: Ciências Exatas e da Terra / Metodologia e Técnicas de Computação.

Palavras-chave: Computação forense; identificação de autoria; algoritmos de classificação

RESUMO:

Na criminalística, o ramo da documentoscopia é focado para a análise de documentos a fim de observar se um manuscrito é ou não daquele autor. Para isso, são utilizadas características para a identificação de autoria, sendo elas características grafométricas e não grafométricas. A primeira está relacionada àquelas que remetem à estrutura da escrita, usualmente utilizada pelos peritos forenses. A segunda é aquela que utiliza as informações da imagem no documento, como por exemplo sua textura. Dessa forma, esta pesquisa teve por objetivo principal estudar o estado da arte em relação ao uso de ambas as características citadas e observar a precisão de cada uma na identificação de manuscritos. Para isso, foi realizada um mapeamento sistemático da literatura, selecionando artigos de acordo com os critérios de seleção e extração para, por fim, realizar a leitura na íntegra. Espera-se que os dados coletados na pesquisa auxiliem como base para trabalhos relacionados à área.

INTRODUÇÃO:

Segundo Mendes (2003), a documentoscopia é um campo da criminalística que analisa documentos para verificar sua autenticidade e identificar seu autor, reconhecendo a subjetividade nas conclusões dos peritos. Dada essa subjetividade, há um crescente interesse em desenvolver ferramentas computacionais que automatizem e padronizem o processo de identificação.

Na identificação de manuscritos, utilizam-se características grafométricas, que capturam aspectos únicos de cada escritor, e não grafométricas, que analisam elementos da imagem do documento, como textura. O mapeamento sistemático investigou o estado da arte dessas características na identificação de autoria de manuscritos em artigos entre os anos de 2020 a 2024 (KITCHENHAM, 2004; AMARAL, 2013).

As questões norteadoras dessa pesquisa foram: Quais características foram usadas para a identificação de manuscritos?; Quais trabalhos utilizavam características grafométricas e quais eram utilizadas?; Qual era a taxa de identificação de autoria? e Qual era o algoritmo de classificação utilizado?

Dessa maneira, o objetivo deste projeto é realizar um mapeamento sistemático acerca do estado da arte da identificação de manuscritos fazendo uso das características grafométricas e não grafométricas. Com esse mapeamento sistemático será possível identificar publicações científicas que possam responder a essas questões, proporcionando uma fundamentação teórica para futuros projetos de pesquisa sobre o tema.

MATERIAIS E MÉTODOS:

Para conduzir o mapeamento sistemático, utilizou-se a ferramenta StArt (*State of the Art through Systematic Review*), desenvolvida pela Universidade Federal de São Carlos (UFSCAR). Foram selecionadas quatro bibliotecas digitais para a coleta de artigos: ACM, IEEE, Springer e Science Direct. Essas bibliotecas foram escolhidas por sua relevância na área de Ciência da Computação e pela quantidade significativa de estudos relacionados. Antes de iniciar a segunda fase, que envolve o mapeamento sistemático, definiu-se a *string* de busca e os critérios de seleção dos trabalhos. A *string* de busca formulada foi:

“computer forensic” AND (“handwriting identification” OR “writer identification” OR “forensic handwriting identification”).

O processo de busca e seleção de trabalhos em bibliotecas digitais foi realizado em várias etapas. Primeiro, utilizou-se ferramentas de filtragem para identificar estudos relevantes através da busca de palavras-chave nas seções de título, resumo e palavras-chave. Os títulos e resumos dos estudos encontrados foram lidos para avaliar sua relevância em relação à pergunta do mapeamento sistemático, e apenas os que atendiam aos critérios de seleção foram considerados para as etapas seguintes. Em seguida, a introdução e conclusão dos trabalhos foram analisadas para verificar a contribuição dos estudos.

Após essa triagem inicial, os estudos selecionados foram lidos na íntegra para coletar informações pertinentes à pesquisa. Os critérios de inclusão foram rigorosos, considerando apenas artigos sobre Computação Forense, revisados por pares, completos e com resultados concretos sobre a identificação de autoria. Foram excluídos artigos publicados antes de 2020, não em inglês, puramente teóricos, fora do domínio da computação forense, ou que não estivessem disponíveis na íntegra.

RESULTADOS E DISCUSSÕES:

Na Etapa 1, foram identificados 551 trabalhos que estavam relacionados à string de busca utilizada. Após a leitura dos títulos e resumos, 71 artigos foram selecionados para a próxima etapa. Na Etapa 3, a leitura das introduções e conclusões desses 71

trabalhos resultou na seleção de 23 artigos para uma análise mais aprofundada.

Tabela 1. Quantidade de trabalhos por base.

IEEE	ACM	Sci. Direct	Springer	Total
12	310	105	124	551

Tabela 2. Quantidade de trabalhos por base após seleção respeitando os critérios impostos.

IEEE	ACM	Sci. Direct	Springer	Total
9	7	21	34	71

Tabela 3. Quantidade de trabalhos por base após extração a partir da introdução e conclusão.

IEEE	ACM	Sci. Direct	Springer	Total
5	6	5	7	23

Na quarta e última etapa, foi feita a leitura completa dos 23 artigos escolhidos na terceira etapa. No entanto, alguns desses trabalhos, particularmente aqueles da biblioteca virtual ACM, não estavam disponíveis na íntegra, e outros 2 artigos da IEEE não abordavam explicitamente a questão da taxa de identificação de autoria. Por fim, criou-se uma tabela para responder às questões norteadoras da pesquisa (explicitadas na seção Introdução) em relação a cada um dos quinze artigos restantes. Observou-se que existe uma ampla diversidade de características sendo utilizadas, totalizando 35 características. Somado a isso, das quinze publicações selecionadas pela metodologia do artigo, foi possível observar que 87% desses dados eram compostos por publicações de características não grafométricas. Como já explicado anteriormente, as características não grafométricas possuem níveis altos de identificação correta de manuscritos e tal fato se confirmou, já que a menor taxa de identificação encontrada foi de 87%. Além disso, foi possível observar que muitas das publicações utilizam como algoritmos de classificação CNNs, SVMs ou algoritmos baseados nos mesmos. Por fim, pode-se concluir que houveram muitas pesquisas relacionadas às características não grafométricas nos anos de 2020 a 2024. Todavia, não se pode concluir que houve um aumento dos anos passados até 2020, por esta pesquisa não abranger tais anos. Para futuras pesquisas, sugere-se a pesquisa do uso de características grafométricas e não grafométricas anterior ao ano de 2020.

CONCLUSÃO:

Este estudo sistematizou o uso de características grafométricas e não grafométricas na identificação de manuscritos em artigos publicados entre 2020 e 2024. Descobriu-se que a maioria das pesquisas concentrou-se em características não grafométricas,

que demonstraram altas taxas de acerto na identificação de autoria, com destaque para algoritmos como CNNs e SVMs, amplamente utilizados nessas análises. Embora os resultados reforcem a eficácia das técnicas não grafométricas, o estudo identifica a necessidade de explorar pesquisas anteriores a 2020 para uma compreensão mais completa da evolução dessas abordagens. Os dados coletados servem como base para futuras investigações e melhorias nas metodologias de identificação forense de manuscritos.

REFERÊNCIAS:

AMARAL, A. M. M. M.; FREITAS, C. O. A.; BORTOLOZZI, F. **Combining multiple features based on graphometry for writer identification as part of Forensic Handwriting Analysis**. In: INTERNATIONAL DOCUMENT IMAGE PROCESSING, 2013. Proceedings of International Document Image Processing. Patras-Greece: International Association for Pattern Recognition (IAPR), v.1. p. 23-30, 2013c.

KITCHENHAM, B., (2004). **Procedures for Performing Systematic Reviews**, Joint Technical Report Software Engineering Group, Department of Computer Science Keele University, United King and Empirical Software Engineering, National ICT Australia Ltd, Australia.

MENDES, L.B. **Documentoscopia**. Campinas: Millennium, 2003. 344 p.