

IDENTIFICAÇÃO DE AUTORIA EM MANUSCRITOS UTILIZANDO CARACTERÍSTICAS GRAFOMÉTRICAS: UM MAPEAMENTO SISTEMÁTICO

Nayane Batista Costa (PIC/UEM), Aline Maria Malachini Miotto Amaral (Orientadora).
E-mail: ra117189@uem.br.

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

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

Palavras-chave: computação forense; características grafométricas; critérios para extração.

RESUMO

A escrita como elemento biométrico tem sido alvo de muitas pesquisas. Neste contexto, diferentes soluções computacionais para identificação de autoria em documentos manuscritos vêm sendo apresentadas na literatura, cada uma delas focando em características específicas da escrita, bem como em aspectos referentes à imagem dos manuscritos. Características grafométricas são aquelas utilizadas pelos peritos forenses durante a análise de um manuscrito questionado. Assim sendo, o suporte automatizado a extração e análise de tais características pode ser uma importante ferramenta de apoio ao trabalho dos peritos. Um dos grandes desafios enfrentados na identificação de autoria em manuscritos se refere a definição das características e agrupamento de características que podem ser utilizadas com o objetivo de obter as melhores taxas de acerto, ao mesmo tempo em que tem sua aceitação no âmbito jurídico. Nesse sentido, este trabalho teve como foco levantar o estado da arte com relação às características grafométricas e sua efetividade na identificação de manuscritos. Espera-se que os dados coletados pela pesquisa sirvam para auxiliar no desenvolvimento de trabalhos relacionados a área, podendo trazer novas ideias e pontos de vista para pesquisadores e atuantes no mesmo, além de permitir uma visão geral do estado da arte, até o presente momento, do tema investigado.

INTRODUÇÃO

De acordo com Mendes (2003), a “documentoscopia é a parte da criminalística que estuda os documentos para verificar se são autênticos e, em caso contrário, determinar a sua autoria”. Observa-se nesta área um grau elevado de subjetividade, uma vez que diferentes peritos podem chegar a diferentes conclusões sobre os mesmos documentos. Dessa forma, o uso de ferramentas computacionais que automatizem e padronizem todo ou parte do processo de identificação adotado pelos peritos tem se tornado campo de interesse da computação.

No contexto da identificação de autoria em manuscritos, pode-se destacar o uso de um certo grupo de características, que são aquelas que utilizam em sua definição os mesmos aspectos usados pelos peritos, denominadas como características grafométricas. Elas possuem como principal vantagem a aceitação no âmbito jurídico, uma vez que seus métodos de extração implicam na automatização das ações já realizadas pelos peritos forenses. O uso combinado de características grafométricas podem melhorar os resultados, em termos de taxa de acerto (Amaral *et al.*, 2013).

Nesse sentido, torna-se extremamente relevante identificar o estado da arte com relação às diferentes pesquisas que utilizam características grafométricas, para que estudos mais avançados sobre sua combinação possam ser subseqüentemente realizados. O objetivo do projeto foi realizar um mapeamento sistemático sobre os desafios para a identificação autoria em manuscritos baseada em características grafométricas, a fim de identificar publicações científicas e/ou projetos de *software* promissores relacionados a área, permitindo com isso, a aquisição de conceitos necessários para a construção e aprimoramento de dados que possam responder às questões da pesquisa.

MATERIAIS E MÉTODOS

Este projeto, através do mapeamento sistemático, buscou responder quatro questões levantadas sobre o referido tema, sendo elas, quais características grafométricas são utilizadas em trabalhos de identificação de autoria em manuscritos; quais desafios existem com relação ao uso de características grafométricas; quais combinações de características grafométricas são utilizadas em trabalhos de identificação de autoria em manuscritos; quais são as bases de dados utilizadas em trabalhos de identificação de autoria em manuscritos.

Por meio de um mapeamento sistemático é possível ter uma noção geral do estado da arte referente a uma certa questão de pesquisa ou fenômeno de interesse. Segundo Kitchenham (2004), um mapeamento sistemático envolve várias atividades discretas e o processo pode ser dividido em três fases principais: planejamento do mapeamento, realização do mapeamento e *review*.

Para a realização do mapeamento sistemático, foi utilizada a ferramenta *StArt* e as quatro bibliotecas digitais pesquisadas para a coleta de artigos foram ACM, IEEE, *Springer* e *Science Direct*. Estas bibliotecas foram selecionadas pela relevância que possuem na área de pesquisa em Ciência da Computação, bem como, pelo número de estudos relacionados que apresentam.

Previamente a execução do mapeamento, foi definida uma *string* de busca: (“*graphometric characteristics*”) AND (“*authorship identification*” OR “*handwritten identification*” OR “*writing identification*”) AND (“*computer forensics*” OR “*forensic science*” OR “*forensic features*”). A pesquisa de trabalhos nessas bibliotecas foi executada por meio de buscas utilizando as ferramentas de pesquisa oferecidas por cada uma delas, além da seleção e classificação dos estudos retornados em cada base de dados. Estudos categorizados como relevantes foram lidos na íntegra, buscando-se informações relacionadas à pesquisa.

RESULTADOS E DISCUSSÃO

Na primeira etapa do mapeamento, foi encontrado um total de 145 trabalhos que de alguma forma estavam relacionados com a *string* de busca aplicada. Através da leitura de seus títulos e resumos, foram selecionados 23 artigos na etapa posterior. Na terceira etapa, foi realizada a leitura da introdução e conclusão dos 23 trabalhos, para realizar uma terceira filtragem. Por fim, na quarta e última etapa foi realizada a leitura na íntegra de cada um dos 12 trabalhos selecionados na etapa três.

Tabela 1

ACM	IEEE	Springer	Science Direct	Total
29	54	21	41	145

Quantidade de trabalhos por base.

Tabela 2

ACM	IEEE	Springer	Science Direct	Total
6	13	2	2	23

Quantidade de trabalhos por base após primeira seleção.

Tabela 3

ACM	IEEE	Springer	Science Direct	Total
5	5	2	0	12

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

Após a leitura dos doze artigos, foi feita uma relação entre o conteúdo abordado pelo estudo com as quatro questões de projeto. Para a primeira questão, os doze artigos trouxeram características grafométricas distintas entre si, contudo, constatou-se que as mais empregadas são aquelas referentes a estrutura geral das letras, como tamanho, formato, inclinação e curvatura.

Observou-se que as questões “quais combinações de características grafométricas são utilizadas em trabalhos de identificação de autoria em manuscritos?” e “quais desafios existem com relação ao uso de características grafométricas?” foram as menos versadas nos artigos analisados, demonstrando que o estado da arte da referida área de pesquisa permanece sem mudanças expressivas. Já que poucos estudos abordaram essas questões de forma explícita, isso indica uma oportunidade para pesquisas futuras que explorem a combinação e o efeito do uso de múltiplas características, além da possibilidade de destacarem as dificuldades práticas e teóricas enfrentadas pelos pesquisadores, como a variabilidade natural da escrita ou a necessidade de aprimorar os algoritmos de análise.

Para a quarta questão de projeto, sete dos doze artigos não especificaram as bases de dados que foram usadas. Percebe-se que o mapeamento sistemático revelou avanços significativos na utilização de características grafométricas para a identificação de autoria em manuscritos, porém existem lacunas importantes a serem preenchidas, como a necessidade de explorar mais combinações de características e a discussão dos desafios enfrentados, buscando aprimorar as técnicas de análise de características grafométricas, além de serem mais transparentes em relação às bases de dados utilizadas, contribuindo para a replicabilidade dos resultados.

CONCLUSÕES

O processo de extração de características grafométricas e quais podem ser utilizadas perante um processo judicial, mostrou-se uma área de exploração recente na Computação Forense. Este trabalho realizou um mapeamento sistemático para identificar como se encontra o corpo de conhecimento sobre a identificação de autoria em manuscritos a partir do uso de características grafométricas. Pode-se observar que há diversas abordagens que atendem ao mesmo problema. A técnica que se mostrou mais presente nas análises dos trabalhos são as abordagens baseadas em reconhecimento de padrões, onde se monta uma base de dados bem estruturada com um número relevante de textos manuscritos, que serão utilizados para estudo, observação e extração das características grafométricas. É conclusivo que muitos dos trabalhos ainda estão em fase de aperfeiçoamento de suas abordagens e investigações, o que reitera que a área da Computação Forense se mostrou bem promissora para realização de pesquisas futuras.

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.