

## Uma Medida de Similaridade Textual para Identificação de Plágio em Fóruns Educacionais

Anderson Pinheiro Cavalcanti<sup>1,2</sup>, Rafael Ferreira Leite de Mello<sup>1,2</sup>

<sup>1</sup>Centro de Informática – Universidade Federal de Pernambuco (UFPE)

<sup>2</sup>Departamento de Estatística e Informática  
Universidade Federal Rural de Pernambuco (UFRPE)

apc@cin.ufpe.br, rafael.mello@ufrpe.br

### Resumo

O uso de Ambiente Virtual de Aprendizagem (AVA) como ferramenta de apoio educacional tem aumentado nos últimos anos. Estes ambientes disponibilizam várias ferramentas que possuem um grande potencial para gerar conteúdo, o que pode ser usado para auxiliar no processo de ensino-aprendizagem. Porém, devido a grande quantidade de interações entre os alunos e o professor, torna-se difícil para o professor avaliar e acompanhar todo o material que é disponibilizado pelos alunos. Uma ferramenta que se destaca em relação à geração de conteúdo colaborativo é o fórum. Contudo, devido a grande quantidade de dados postado na ferramenta, é difícil para o professor identificar problemas nas postagens, como por exemplo a detecção de plágio [Cavalcanti et al. 2017b]. A base fundamental para a criação de sistemas automáticos de detecção de plágio é a criação de uma medida de similaridade que possa medir a relação existente entre dois textos. Várias medidas de similaridade entre textos já foram criadas; entretanto, em geral, elas são dependentes de idioma. No caso do português, poucas medidas foram encontradas e a maioria utiliza apenas técnicas estatísticas, não levando em consideração aspectos semânticos dos textos. Diante disso, este trabalho propõe uma medida que calcula a similaridade existente entre sentenças escritas em português levando em consideração a semântica dos textos [Cavalcanti et al. 2017a]. A medida proposta alcançou resultados melhores que o primeiro colocado da competição atingindo 0,70 de correlação de Pearson e 0,47 de erro quadrático médio [Cavalcanti et al. 2017c]. Além desta avaliação, foi realizado um estudo de caso para avaliação de similaridade em postagens de fóruns educacionais em uma disciplina de Ciência da Computação. Os resultados foram avaliados pelos professores da disciplina que confirmaram a eficácia da ferramenta.

### Referências

- Cavalcanti, A., Ferreira, R., Dionísio, M., Neto, S., Passero, G., and Miranda, P. (2017a). Uma nova abordagem para detecção de plágio em ambientes educacionais. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 28, page 1177.
- Cavalcanti, A. P., de Mello, R. F. L., Ferreira, M. A. D., de Miranda, P. B. C., Rolim, V. B., and da Silva Neto, S. R. (2017b). O plágio em ambiente educacional virtual: Uma revisão da literatura. *RENOTE*, 15(2).
- Cavalcanti, A. P., de Mello, R. F. L., Ferreira, M. A. D., Rolim, V. B., and Tenório, J. V. S. (2017c). Statistical and semantic features to measure sentence similarity in portuguese. In *Intelligent Systems (BRACIS), 2017 Brazilian Conference on*, pages 342–347. IEEE.