

**A. Ciências Exatas e da Terra - 2. Ciência da Computação - 6. Inteligência Artificial e Redes Neurais**

**Descoberta de sequência de acessos através da mineração de dados de arquivos de dados de log da web**

Rolf Pagotto Veiga, bolsista PIBIC/CNPq - DCC<sup>1</sup>

Marluce Rodrigues Pereira, orientadora - DCC<sup>1</sup>

Ahmed Ali Abdalla Esmin, orientador - DCC<sup>1</sup>

1. Universidade Federal de Lavras - UFLA

**RESUMO:**

Os problemas de mineração de dados envolvem grandes bases de dados e estão presentes em diferentes setores, tais como: bibliotecas digitais, comunidades virtuais, comércio eletrônico, bioinformática, detecção de mensagens indesejadas, padrões de navegação na web, entre outros.

Com a popularização do uso da Internet, a quantidade de dados disponível tornou-se extremamente grande, surgindo assim o interesse em projetar algoritmos para uma mineração de dados eficiente. Surge assim o conceito de mineração de dados na Internet ou Web Mining.

Web Mining pode ser definida como sendo técnicas para obter, compreender e extrair dados de documentos ou serviços voltados a Internet. A Web Mining muitas vezes pode ser confundida como sendo um backup, uma recuperação de informação, mas na verdade é um processo menos trivial que envolve meios de se recuperar ou dar um sentido para uma informação.

Uma importante tarefa da mineração de dados voltada para web é achar um padrão de navegação. Porém, os interesses dos usuários mudam periodicamente, não se mantendo eternamente. Isso dificulta a extração exata do padrão de navegação. Desta forma, um padrão pode ser encontrado através de um log de acessos a um site e sua identificação ocorre através da utilização de algoritmos de mineração de dados voltados para a web. Este trabalho apresenta a implementação de parte do algoritmo de descoberta de distância entre páginas, denominado Usage-Aware Average-Clicks, com um foco diferente. O objetivo é analisar logs de acesso a um site web para encontrar quais as páginas foram mais frequentemente acessadas, qual a sequência de acessos das páginas e quais os links mais acessados dentro de uma página. Este tipo de informação permite que seja identificada a melhor forma de organizar as páginas e os links dentro de uma página. Além disso, é possível encontrar a melhor forma de realizar recomendações de acesso para o usuário. A forma como esta informação é visualizada também facilita sua compreensão. Os experimentos foram realizados utilizando dados reais de um curso de pós-graduação Lato Sensu.

Instituição de Fomento: FAPEMIG

Palavras-chave: Web Mining, Mineração de Dados, Usage-Aware Average-Clicks.