

MINERAÇÃO DE DADOS APLICADA A ANÁLISE DE CONTEÚDOS DE ARMAZENAMENTO DE MASSA

JOÃO VITOR FREITAS, SIBELIUS LELLIS VIEIRA

joaoovfreitas@gmail.com

Objetivo: Analisar e identificar padrões de interesse e informações implícitas baseados em arquivos de texto de notícias de temas diferenciados utilizando a técnica de mineração de dados e análise de clusters K-Means. **Método:** Aplicado à metodologia de conhecimento em base de dados para garantir a qualidade dos resultados, foram selecionadas cem notícias de variados temas e palavras chave que as caracterizam e/ou fazem parte do tema. Os dados obtidos foram registrados em arquivos texto para formar a base de dados, sendo posteriormente transformado para o formato ARFF que é utilizado no software Weka e após a transformação executar o algoritmo K-Means. Os resultados foram interpretados manualmente, utilizando apenas a ferramenta Qt creator para gerar um software que indica a frequência das palavras chave selecionadas e obter melhores interpretações. **Resultados:** Como resultados o arquivo texto organizado em uma tabela e transformado no formato ARFF para posterior consulta e edição, a frequência de cada atributo foi calculado pelo software criado com o Qt Creator, o tempo de 0,02 segundos de execução com o software Weka e os padrões de interesse encontrados com a técnica, sendo que apresenta melhor resultado com treze clusters (grupos) dissimilares. **Conclusão:** O método K-Means mostrou bom desempenho e resultado, sendo aplicado a uma base dados diversificada, agrupando as notícias de acordo com os atributos de maior frequência, alguns atributos que possuem menor frequência e possivelmente relacionados a outros atributos não se tornaram característica visível de um cluster.

Palavras-chave: Mineração de Dados. Agrupamento. K-Means.