

CEX - CÂMARA DE CIÊNCIAS EXATAS E DOS MATERIAIS (PÔSTER)

NOME: ELTON ANGELO RABELO

TÍTULO: EXTRAÇÃO E ANÁLISE DE DADOS DA PLATAFORMA VIRTUAL DA REVISTA NATURE: UM ESTUDO SOBRE O RELACIONAMENTO DE PALAVRAS-CHAVE E O VOLUME DE PUBLICAÇÕES AO LONGO DOS ANOS

AUTORES: PATRÍCIA MASCARENHAS DIAS, ELTON ANGELO RABELO, ELTON ANGELO RABELO, JHONATAN FERNANDO DE OLIVEIRA, PATRICIA MASCARENHAS DIAS, THIAGO MAGELA RODRIGUES DIAS

AGÊNCIA FINANCIADORA (se houver): PAPq

PALAVRA CHAVE: PALAVRAS-CHAVE, TÍTULO, EXTRAÇÃO, RELACIONAMENTO

RESUMO

Com o crescimento dos repositórios de dados científicos impulsionados principalmente pela facilidade de acesso e publicação proporcionados pela Internet, a adoção de técnicas de mineração de dados e técnicas de análise de redes com o uso de grafos têm sido amplamente empregadas com o intuito de explorar tais dados e extrair conhecimento antes desconhecidos. A revista virtual da Nature, é hoje um dos mais importantes repositórios de dados científicos, criada para que os artigos publicados na revista Nature sejam disponibilizados em um ambiente online. Informações como autores de uma publicação, ano de publicação, títulos e palavras-chave de cada trabalho publicado na revista podem ser acessados, caracterizando-se como uma importante fonte de dados para compreender como a ciência tem evoluído ao longo dos anos. Neste trabalho, o principal objetivo é realizar a extração de todos os títulos das publicações dispostas na revista virtual da Nature, para que técnicas de mineração de dados possam ser empregadas juntamente com práticas de normatização, que possibilitem retirar as palavras não relevantes. De posse do conteúdo relevante, vai se buscar compreender como as palavras-chave que estão vinculadas se relacionaram ao longo do tempo, também é objetivo do trabalho verificar as palavras mais utilizadas ano a ano criando um ranqueamento de utilização das mesmas, verificar quais palavras continuam sendo utilizadas e quais caíram em desuso, além de uma análise estatística que possibilite verificar a quantidade de publicações científicas ano a ano, criando gráficos que mostrem o desenvolver das publicações ano a ano, possibilitando ver quedas e aumentos de publicações, além de buscar entender o porquê das quedas mais marcantes.