

CEX - CÂMARA DE CIÊNCIAS EXATAS E DOS MATERIAIS (COMUNICAÇÃO COORDENADA)

NOME: THIAGO MAGELA RODRIGUES DIAS

TÍTULO: UMA ANÁLISE BASEADA EM MÉTRICAS DE REDES SOCIAIS PARA VISUALIZAÇÃO DA PRODUÇÃO CIENTÍFICA MUNDIAL

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

AGÊNCIA FINANCIADORA (se houver): PAPq - UEMG

PALAVRA CHAVE: ANÁLISE DE REDES, EXTRAÇÃO, RECUPERAÇÃO DE INFORMAÇÕES

RESUMO

Com o crescimento dos repositórios de dados científicos impulsionados principalmente pela facilidade de acesso e publicação, tendo como facilitador a Internet. A adoção de técnicas de mineração de dados e técnicas de análise de redes têm sido amplamente empregadas nesses conjuntos de dados com o intuito de 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, contendo informações sobre autores, títulos e palavras-chave de cada trabalho publicado na revista, bem como ano de publicação, 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 a proposta de uma plataforma que realize a extração de todos os títulos das publicações dispostas no ambiente virtual da Nature, para que técnicas de mineração de dados possam ser aplicadas no conjunto de dados extraído. Para isso, os títulos das publicações anteriormente extraídos são normalizados através de um processo de radicalização das palavras que o compõem e com isso, é possível obter um conjunto consistente de palavras-chave que compõem um título. Aliado ao processo de radicalização as palavras sem valor semântico também são eliminadas com o intuito de se reduzir o conjunto a ser analisado mantendo somente palavras que podem representar tópicos de pesquisa a serem investigados.

Posteriormente, todas as palavras-chaves de uma determinada publicação são utilizadas para a criação de grafos onde palavras-chave serão os vértices destes grafos e uma aresta será caracterizada quando uma determinada palavras for utilizada em conjunto com outras em uma mesma publicação, ou seja, os grafos de palavras-chave serão construídos através da inserção de cliques por justaposição. Com os grafos caracterizados, diversas métricas de análise de redes são aplicadas para que dessa forma possam ser analisados detalhes de como termos que podem representar tópicos de pesquisa estão sendo utilizados em conjunto, como estes tópicos estão distante ou vem sendo investigados ao longo dos anos, bem como diversas análises estatísticas que demonstram como a produção científica tem evoluído. Logo, é possível compreender como as palavras-chave do conjunto analisado estão vinculadas e como elas se relacionam ao longo do tempo. Além disso, também são verificadas as palavras mais frequentes ao longo dos anos, resultando em uma visão sobre como os tópicos de pesquisa têm evoluídos. É possível concluir com as análises realizadas que existem tópicos de pesquisa, representadas por palavras-chave que são aplicados em conjunto com diversos outros tópicos em publicações de áreas distintas, mostrando um bom índice de interdisciplinaridade, além disso, destaca-se que as crises históricas tem um impacto negativo sobre as pesquisas e a quantidade de publicações científicas, como nos períodos de guerra e crises financeiras.