



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
Centro Tecnológico
Colegiado do Curso de Engenharia de Computação

COMUNICADO DE DEFESA DO PROJETO DE GRADUAÇÃO **(Anexo I – Resolução 02/2013-CCEC)**

Nome do Estudante: João Lucas Luz Lima Sarcinelli

Título: Estudo de Fatores que Influenciam a Propagação de Artigos da Wikipédia em Diferentes Idiomas

Data da apresentação: 17/08/2022

Horário: 13h

Local: meet.google.com/kqv-nfdd-jrd

Banca Examinadora:

Giovanni Ventorim Comarela: (Departamento de Informática) (Orientador)

Rodolfo Silva Villaça: (Departamento de Informática)

Alessandro Jose Queiroz Sarnaglia. (Departamento de Estatística)

Resumo do trabalho:

Nos últimos anos, tem se experienciado um forte aumento de usos de tecnologias digitais, dando um grande espaço para a conectividade global no papel da propagação de conhecimento.

De diversas plataformas que esse fenômeno é perceptível, a Wikipédia é uma das mais interessantes, principalmente pelo seu caráter colaborativo na sua estrutura, o que alcança diversas línguas e tópicos de diferentes níveis de interesse.

Neste trabalho, foi realizado um estudo investigativo dos artigos da Wikipédia, utilizando um conjunto de dados que traz informações sobre as edições em diferentes línguas de cada um de mais de 33 milhões de artigos da plataforma, bem como os tópicos relacionados a cada, a fim de encontrar fatores que influenciam suas propagações para diferentes idiomas.

Foram utilizadas duas abordagens por aprendizado não supervisionado, por meio de

algoritmos de agrupamento, separando o conjunto de dados original em subconjuntos de artigos similares quanto a diferentes características. Foram formados, também, subconjuntos de artigos que pertencem aos diferentes tópicos descritos pelo conjunto.

Por fim, os subconjuntos obtidos foram utilizados como dados de entrada para um modelo de aprendizado de máquina, uma rede neural baseada na arquitetura Transformer, que realizou a previsão da sequência de edições dos artigos. Os resultados do modelo foram avaliados utilizando duas métricas distintas.

Os desempenhos finais foram, então, comparados, mostrando a influência de cada característica analisada na qualidade da previsão obtida pela rede e, conseqüentemente, na previsibilidade da propagação de idiomas dos artigos que possuem tais características.

Foi possível constatar que existem, de fato, fatores que influenciam a propagação dos artigos para diferentes línguas. Dentre esses fatores, características como tópicos relacionados e popularidade de sequências de edições se mostraram relevantes para essas propagações.