



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: Felipe de Aquino Fernandes

Título: Uma Aplicação Web para Classificação de Documentos PDF

Data da apresentação: 01/04/2022

Horário: 13:00

Local: meet.google.com/bqw-arqr-mmw

Banca Examinadora:

Profa.: Patrícia Dockhorn Costa (Depto. de Informática) (Orientadora)

Prof.: Celso A. Saibel Santos (Depto. de Informática) (Co-orientador)

Profa.: Roberta Lima Gomes (Depto. de Informática)

Prof.: José Gonçalves Pereira Filho (Depto. de Informática)

Resumo do trabalho:

Palavras-chave sumarizam o conteúdo de um texto de forma simples e objetiva e permitem que um leitor tenha conhecimento sobre o assunto retratado. Com o crescente volume de informações, as palavras-chave têm se provado ferramentas úteis para buscar, selecionar e classificar textos. Grande parte do conteúdo textual está disponível no formato PDF, incluindo artigos científicos, revistas, livros e manuais. Em todos estes exemplos, os conteúdos podem estar vinculados a palavras-chave, permitindo a seleção daqueles cujo leitor tem maior interesse. Nem sempre as palavras-chave são disponibilizadas pelos autores, ou, ainda, podem estar mal definidas, principalmente quando a quantidade de informações no documento é muito grande. Portanto, encontrar alternativas para identificar palavras-chave pode ampliar a visão geral sobre o conteúdo ou tema.

O objetivo principal deste trabalho é a criação de um sistema Web capaz de extrair palavras-chave de documentos PDF, especialmente de artigos científicos, com a intenção de classificá-los. Esse sistema é constituído de duas partes, o servidor e o cliente, ambas escritas usando tecnologias Javascript, tais como Nodejs, Expressjs e Angular, almejando que o sistema tenha um bom desempenho e que possua usabilidade simples. Os usuários do sistema podem criar contas e submeter seus documentos PDF, usando ou não seus currículos Lattes para obtenção de palavras-chave sugeridas a fim de compará-las com as extraídas pela aplicação. Foram usadas nuvens de palavras para exibir de maneira intuitiva as palavras-chave, onde as mais importantes possuem maior destaque. Conceitos de Web Semântica foram usados para categorizar os documentos e ter uma visão mais ampla das áreas do conhecimento abordadas por eles.