

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
MINAS GERAIS
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

DANIEL ELIAS DOS SANTOS

**PROCESSAMENTO DE LINGUAGEM NATURAL APLICADA A
TEXTOS JURÍDICOS**

PRÉ-PROJETO DE TRABALHO DE CONCLUSÃO DE CURSO

SABARÁ
2019

DANIEL ELIAS DOS SANTOS

PROCESSAMENTO DE LINGUAGEM NATURAL APLICADA A TEXTOS JURÍDICOS

Pré-Projeto de Trabalho de Conclusão de Curso de graduação, apresentado à disciplina de Trabalho de Conclusão de Curso 1, do Curso de Bacharelado em Sistemas de Informação - BSI - do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais - IFMG - Campus Sabará, como requisito parcial para a obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Daniel Bruno Fernandes Conrado

Coorientadora: Cristiane Norbiato Targa

SABARÁ
2019

1 PRÉ-PROJETO DE TRABALHO DE CONCLUSÃO DE CURSO

1.1 TÍTULO

Processamento de Linguagem Natural Aplicada a Textos Jurídicos.

1.2 MODALIDADE DO TRABALHO

Desenvolvimento de aplicação.

1.3 ÁREA DO TRABALHO

Inteligência Artificial; Processamento de Linguagem Natural; Análise de Sentimentos.

1.4 RESUMO

O Projeto “Amora” foi concebido com o objetivo de fortalecer a representatividade das mulheres no mercado de trabalho tecnológico, por meio da criação de uma plataforma de classificação de empresas de tecnologia quanto à amigabilidade ao gênero feminino, a partir de documentos jurídicos de reclamações trabalhistas sobre questões relacionadas a gênero. A extração automatizada de informações de documentos legais é um problema relevante, e que ainda não foi completamente resolvido. Neste sentido, um dos componentes da plataforma terá a função de extrair informações de interesse, tais como conceitos jurídicos e entidades nomeadas (locais, organizações, datas e referências), de documentos jurídicos de reclamações trabalhistas obtidos por meio de *crawlers*. Estas informações serão identificadas usando informações semânticas da saída de um analisador de linguagem natural, e a partir delas, os documentos serão classificados, utilizando técnicas de Análise de Sentimentos. Essa abordagem linguística é baseada nas propostas de [Quaresma e Gonçalves \(2010\)](#) e [Lopes \(2012\)](#), que obtiveram resultados promissores.

2 DESCRIÇÃO DA PROPOSTA

2.1 INTRODUÇÃO

2.1.1 Objetivo Geral

2.1.2 Objetivos Específicos

2.2 ESTADO DA ARTE

2.3 DIFERENCIAL TECNOLÓGICO

2.4 PROCEDIMENTOS METODOLÓGICOS/METODOLOGIA

2.5 CONCLUSÃO/CONSIDERAÇÕES FINAIS

2.6 PLANEJAMENTO DO TRABALHO

Quadro 1 – Cronograma de Atividades.

Atividades	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1. Revisão dos apontamentos da banca										
2. Revisão bibliográfica										
3. Redação do projeto de TCC			X	X						
4. Defesa do projeto de TCC					X					
5. Escrita da Monografia de TCC						X	X	X		
6. Elaboração da apresentação final								X	X	
7. Defesa final do TCC									X	

2.7 HORÁRIO DE TRABALHO

Quadro 2 – Horário de Trabalho.

Horário	Seg	Ter	Qua	Qui	Sex	Sab
07h30 - 08h20						
08h20 - 09h10						
09h10 - 10h00						
10h10 - 11h00						
11h00 - 11h50						
13h00 - 13h50		TCC	Orientação	TCC		
13h50 - 14h40		TCC	TCC			
14h40 - 15h30		TCC	TCC	TCC		
15h40 - 16h30		TCC	TCC	TCC		
16h30 - 17h20						
17h20 - 18h10						
18h50 - 19h40						
19h40 - 20h30						
20h30 - 21h20						
21h30 - 22h15						

Referências

LOPES, L. Extração automática de conceitos a partir de textos em língua portuguesa. **Pontifícia Universidade Católica do Rio Grande do Sul**, 2012. Citado na página [1](#).

QUARESMA, P.; GONÇALVES, T. Using linguistic information and machine learning techniques to identify entities from juridical documents. **Lecture Notes in Computer Science**, vol **6036**, 2010. Citado na página [1](#).