



# DEFESA DE DISSERTAÇÃO

*Programa de Pós-Graduação em Ciência da Computação*

## UMA ANÁLISE COMPARATIVA ENTRE AS ABORDAGENS LINGUÍSTICA E ESTATÍSTICA PARA EXTRAÇÃO AUTOMÁTICA DE TERMOS RELEVANTES DE CORPORA

ALUNO: CARLOS ALBERTO DOS SANTOS

ORIENTADORA: PROFA. DRA. RENATA VIEIRA

BANCA EXAMINADORA: PROF. DR. RAFAEL HEITOR BORDINI (PPGCC/PUCRS) E  
PROF. DR. LEANDRO KRUG WIVES (UFRGS).

DATA: 27 DE ABRIL DE 2018

LOCAL: PRÉDIO 32, AUDITÓRIO 516.

HORÁRIO: 13:30

### RESUMO:

SABE-SE QUE O PROCESSAMENTO LINGUÍSTICO DE CORPORA DEMANDA GRANDE ESFORÇO COMPUTACIONAL DEVIDO A COMPLEXIDADE DOS SEUS ALGORITMOS, MAS QUE APESAR DISSO, OS RESULTADOS ALCANÇADOS SÃO MELHORES QUE AQUELES GERADOS PELO PROCESSAMENTO ESTATÍSTICO, ONDE A DEMANDA COMPUTACIONAL É MENOR. ESSA DISSERTAÇÃO DESCREVE UMA ANÁLISE COMPARATIVA ENTRE OS PROCESSOS LINGUÍSTICO E ESTATÍSTICO DE EXTRAÇÃO DE TERMOS. FORAM REALIZADOS EXPERIMENTOS ATRAVÉS DE QUATRO CORPORA EM LÍNGUA INGLESA, CONSTRUÍDOS A PARTIR DE ARTIGOS CIENTÍFICOS, SOBRE OS QUAIS FORAM EXECUTADAS EXTRAÇÕES DE TERMOS UTILIZANDO ESSAS ABORDAGENS. AS LISTAS DE TERMOS RESULTANTES FORAM REFINADAS COM O USO DE MÉTRICAS DE RELEVÂNCIA E STOP LIST, E EM SEGUIDA COMPARADAS COM AS LISTAS DE REFERÊNCIA DOS CORPORA ATRAVÉS DA TÉCNICA DO RECALL. ESSAS LISTAS, POR SUA VEZ, FORAM CONSTRUÍDAS A PARTIR DO CONTEXTO DESSES CORPORA E COM AJUDA DE PESQUISAS NA INTERNET. OS RESULTADOS MOSTRARAM QUE A EXTRAÇÃO ESTATÍSTICA COMBINADA AS TÉCNICAS DA STOP LIST E AS MÉTRICAS DE RELEVÂNCIA PODE PRODUZIR RESULTADOS SUPERIORES AO PROCESSO DE EXTRAÇÃO LINGUÍSTICO REFINADO PELAS MESMAS MÉTRICAS. CONCLUIU-SE QUE A ABORDAGEM ESTATÍSTICA COMPOSTA POR ESSAS TÉCNICAS PODE SER A OPÇÃO IDEAL PARA EXTRAÇÃO DE TERMOS RELEVANTES, POR EXIGIR POUCOS RECURSOS COMPUTACIONAIS E POR APRESENTAR RESULTADOS SUPERIORES ÀQUELES ENCONTRADOS NO PROCESSAMENTO LINGUÍSTICO.