

SALÃO DE
INICIAÇÃO CIENTÍFICA
XXIX SIC

UFRGS
PROPESQ



múltipla 
UNIVERSIDADE
inovadora  inspiradora

Evento	Salão UFRGS 2017: SIC - XXIX SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2017
Local	Campus do Vale
Título	Rede de Relacionamentos Criminais na DeepWeb
Autor	ALEXANDRE ALBUQUERQUE DE SOUZA
Orientador	SEBASTIAN GONCALVES

Título do Projeto: Redes de Relacionamento Criminais na DeepWeb

Nome: Alexandre Albuquerque de Souza

Orientador: Sebastian Gonçalves

Instituto de Física - UFRGS

Mineração de textos é um processo em que se utiliza de algoritmos computacionais para analisar uma enorme coleção de documentos texto com a finalidade de extrair informações valiosas para os mineradores. Muitos algoritmos utilizam da abordagem estatística para que palavras importantes sejam capturadas destas coleções. Atualmente, a mineração de texto, ou dados (forma mais geral), é importante pelo imenso volume de dados para serem analisados. O objetivo final do trabalho é analisar um conjunto de textos resultantes de uma operação da Polícia Federal sobre pedofilia na DeepWeb. O conjunto dos textos ocupam aproximadamente 1 Terabyte de dados, de tal forma que, manualmente, esta tarefa não poderia ser realizada. O programa SOBEK, desenvolvido por pesquisadores do Instituto de Informática da UFRGS, é uma das ferramentas que analisa textos e relaciona as palavras identificadas como palavras importantes, construindo uma rede semântica entre elas como resultado final do programa. Esta ferramenta mostrou-se interessante, porém seu uso é limitado a pequenos textos inseridos em um formulário Java on-line. Por isso, na primeira fase do trabalho, usamos Python para replicar os resultados do programa SOBEK em diversos textos testes, obtendo a mesma rede semântica final. Apresentamos as regras e passos usados na replicação do programa e a comparação dos resultados entre nosso programa e a interface on-line SOBEK. Em uma segunda parte aplicamos o método a textos que excedem a capacidade máxima do SOBEK on-line como são os discursos do impeachment. Finalmente discutimos a vantagem de usar uma ferramenta mais poderosa e de uso crescente para criar redes semânticas que é o “word2vec”.