

XIX encontro nacional
de pesquisa em
ENANCIB ciência da informação

// SUJEITO INFORMACIONAL E AS
PERSPECTIVAS ATUAIS EM CIÊNCIA
DA INFORMAÇÃO. //

22-26
OUTUBRO
2018
LONDRINA/PR



XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT-8 – Informação e Tecnologia

Jairo Brandão Santana, Universidade Federal de Santa Catarina

Vinícius Farias Ramos, Universidade Federal de Santa Catarina

Diefferson Moro, Universidade Federal de Santa Catarina

Gustavo Medeiros de Araujo, Universidade Federal de Santa Catarina

DETECÇÃO DE VÍNCULOS EM RELATÓRIOS DE INTELIGÊNCIA FINANCEIRA

LINK DETECTION IN FINANCIAL INTELLIGENCE REPORTS

Modalidade da Apresentação: Pôster

Resumo: A Polícia Federal tem, entre suas competências, analisar os Relatórios de Inteligência Financeira (RIF), gerados pelo Conselho de Controle de Atividades Financeiras (COAF). Essa análise verifica se existe algum indício de crime de lavagem de dinheiro e, se for o caso, inicia uma investigação. Essa análise é realizada de forma manual e, dependendo da experiência que o analista possui, pode obter resultados discrepantes. Para minimizar esse problema e tornar mais ágil essa análise, esse projeto vem propor que esse trabalho seja realizado de forma automatizada, utilizando as técnicas de Mineração de Texto, Processamento em Linguagem Natural e Aprendizado de Máquina. A ideia é que o analista, após o COAF enviar o RIF, que é um arquivo em formato PDF, submeta esse arquivo a um sistema computacional que faça sua leitura detalhada e retorne as informações contidas no relatório, como entidades, operações, valores, datas e vínculos entre as entidades. As informações encontradas serão armazenadas em uma base de dados e disponibilizadas de forma estruturada para que o analista possa submeter os vínculos detectados a uma ferramenta de diagramação e análise de redes. Além disso, a organização e armazenamento dessas informações, também irá permitir o cruzamento dos vínculos de diversos RIFs e manter o seu histórico.

Palavras-Chave: Lavagem de dinheiro; Relatório de Inteligência Financeira; Mineração de texto; Processamento em linguagem natural; Aprendizado de máquina.

Abstract: The Federal Police has, among its powers, to analyze the Financial Information Reports (RIF), generated by the Financial Activities Control Council (COAF). This analysis checks if there is any

indication of a money laundering crime and, if it is the case, initiates an investigation. This analysis is performed manually and, depending on the experience the analyst has, may have divergent results. To minimize this problem and to make this analysis more agile, this project proposes that this work be performed in an automated way, using the techniques of Text Mining, Natural Language Processing and Machine Learning. The idea is for the analyst, after COAF to send the RIF, which is a PDF file, to submit this file to a computer system that performs a detailed reading and returns the information contained in the report, such as entities, operations, values, dates and links between entities. The information found will be stored in a database and made available in a structured way so that the analyst can submit the detected links to a network diagramming and analysis tool. In addition, the organization and storage of this information will also allow you to cross-link the various RIFs and maintain their history.

Keywords: Money laundry; Financial information report; Text Mining; Natural language processing; Machine learning.

1 INTRODUÇÃO

A Polícia Federal é uma das principais autoridades competentes responsáveis pela investigação de crimes de lavagem de dinheiro. Recebe do Conselho de Controle de Atividade Financeiras (COAF) o Relatório de Inteligência Financeira (RIF), que relata as transações suspeitas (COAF, 2015). Atualmente as áreas de inteligência da PF analisam o RIF de modo manual e identificam entidades¹, valores e operações realizadas (POLÍCIA FEDERAL, 2013). Essas informações são tabuladas, permitindo a sua leitura por ferramentas de Análise de Vínculos e a geração de diagramas de relacionamentos entre as entidades, facilitando a detecção de Organizações Criminosas.

Figura 1: Detecção de Vínculos.



Fonte: Elaborado pelo autor.

Esse trabalho vem propor que a leitura do RIF e detecção dos vínculos² a serem tabulados sejam realizadas de modo automatizado, através da utilização de técnicas de Mineração de Texto, facilitando e padronizando o trabalho de análise do RIF.

¹ Pessoas Físicas ou Jurídicas.

² No contexto do Relatório de Inteligência Financeira, vínculos são as ligações entre as entidades informadas concretizadas através de algum tipo de operação financeira.

2 LAVAGEM DE DINHEIRO

Podemos entender por lavagem de dinheiro como “a atividade de investir, ocultar, substituir ou transformar e restituir o dinheiro de origem ilícita aos circuitos econômico-financeiros legais, incorporando-o a qualquer tipo de negócio como se fosse obtido de forma lícita”, (CALLEGARI, 2001 apud LUSTOSA, 2009). Outra definição sobre o termo diz que se trata de uma “forma genérica de referir-se ao processo ou conjunto de operações de ocultar a origem de dinheiro ou dos bens resultantes das atividades delitivas e integrá-los no sistema econômico ou financeiro, em operações capazes de converter o dinheiro sujo em dinheiro limpo”, (LUSTOSA, 2009).

A lavagem de dinheiro é um processo complexo no qual pode ser identificado três fases, ocultação, dissimulação e integração. Somente no fim da década de 1980 foram tomadas medidas concretas para o combate da lavagem de dinheiro em âmbito internacional. Capitaneada pela ONU, foi realizada a Convenção de Viena em 1988, com o objetivo de combater o tráfico internacional de entorpecentes (ONU, 1988). E em 1989, a cúpula do G-7³, que foi realizada em Paris, criou o Grupo de Ação Financeira contra a Lavagem de Dinheiro e o Financiamento do Terrorismo - GAFI/FATF, com o objetivo de elaborar políticas de proteção ao sistema bancário e às instituições financeiras e identificar suas vulnerabilidades.

Em 1998, no Brasil foi promulgada a lei 9.613 chamada de “Lei de Lavagem de Capitais” ou também de “Lei de Lavagem de Dinheiro” com o intuito de disponibilizar uma lei específica para esse tipo de crime e também adequar o país à Convenção de Viena e às recomendações do GAFI (OBREGON, 2001). Essa lei também instituiu a criação do Conselho de Controle de Atividade Financeiras (COAF) que tem como função a prevenção e combate à lavagem de dinheiro e ao financiamento do terrorismo. O COAF atua em conjunto com diversos órgãos e entidades financeiras com o objetivo de evitar a utilização desses diversos setores sejam usados nas práticas de lavagem de dinheiro e, quando detectam operações atípicas⁴ realizadas pelos seus usuários, têm a obrigação de informar ao COAF. O controle deve ser realizado não apenas sobre as pessoas físicas, mas também em qualquer empresa, grupo ou conglomerado (MARQUES, 2014).

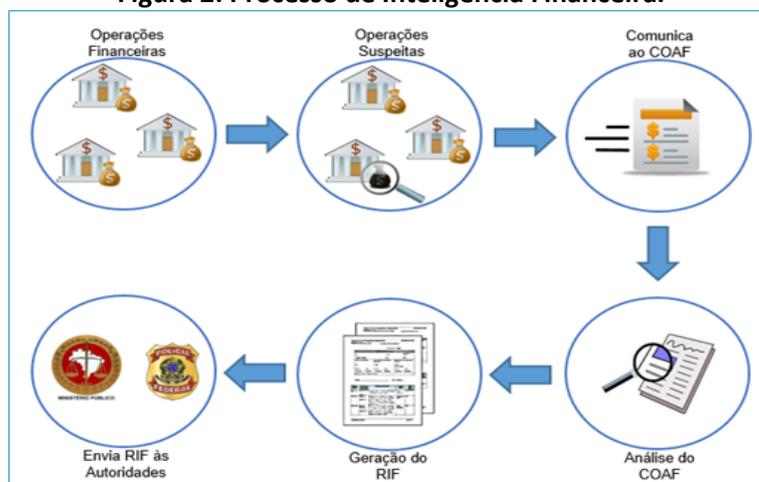
³ Grupo dos sete países mais ricos do mundo – Alemanha, Canadá, Estados Unidos, França, Itália, Japão e Reino Unido.

⁴ O Banco Central do Brasil define um rol de hipóteses de operações suspeitas através das cartas-circulares 2.826/1998 e 3.098/2003.

Recebida a comunicação de operações suspeitas, o COAF realizará a análise para detectar se existem evidências de crime de lavagem de dinheiro e, se realmente forem detectadas, será feito um intercâmbio de informações com as autoridades competentes (MARQUES, 2014), conforme é disciplinado pelo artigo 15 da Lei nº 9.613/1998: “O COAF comunicará às autoridades competentes para a instauração dos procedimentos cabíveis, quando concluir pela existência de crimes previstos nesta Lei, de fundados indícios de sua prática, ou de qualquer outro ilícito” (BRASIL, 1998).

São consideradas as autoridades competentes para receberem os informes do COAF o Ministério Público Federal, o Ministério Público do Estado respectivo e a Polícia Federal, os quais poderão proceder com o bloqueio da operação financeira suspeita, iniciar uma investigação criminal e, se for o caso, propor a ação penal (ARAS, 2007). O informe disponibilizado pelo COAF às autoridades competentes é chamado de Relatório de Inteligência Financeira (RIF). Podemos observar abaixo, as etapas percorridas por esse processo, que também é conhecido como “Inteligência Financeira”.

Figura 2: Processo de Inteligência Financeira.



Fonte: Elaborado pelo autor.

3 DETECÇÃO DE VÍNCULOS NO RIF PELA POLÍCIA FEDERAL

No âmbito da Polícia Federal, o RIF será recebido pela área de repressão a crimes financeiro, a qual realizará uma avaliação preliminar para indicar qual será a delegacia responsável pela análise do documento. Estando na delegacia específica, o RIF deverá ser analisado minuciosamente para detectar se existem os indícios necessários para retratar uma prática criminosa e, conseqüentemente, a instauração de um procedimento investigativo formal (POLÍCIA FEDERAL, 2013). O RIF é um documento em formato PDF – *Portable Document*

Format, e é escrito em linguagem natural, não estruturado, com relatos descritos pelos analistas do COAF de onde deverão ser extraídos as referências aos possíveis vínculos.

Dependendo do nível de complexidade do RIF, a análise poderá ser realizada apenas com a sua leitura e, nos casos mais complexos, os vínculos detectados deverão ser armazenados em planilhas eletrônicas ou em banco de dados e, posteriormente, a utilização de softwares analíticos para a análise desses vínculos.

Na fase de detecção e tabulação de vínculos, é essencial que seja realizada com o maior grau de precisão possível pois, caso contrário, irá comprometer uma correta análise na fase posterior. Assim, esse estudo propõe uma solução automatizada para leitura do RIF e detecção dos vínculos, que são as entidades e eventos identificados utilizando principalmente técnicas de *Text Mining* (Mineração de Textos), *Natural Language Processing* (Processamento de Linguagem Natural) e *Machine Learning* (Aprendizado de Máquina) e posterior armazenamento dos resultados em base de dados estruturada.

A Mineração de Textos consiste na extração de informações úteis em textos não estruturadas escritos em linguagem natural e envolve várias áreas da informática como mineração de dados, aprendizado de máquina, recuperação de informação e linguagem computacional, para poder transformar os textos analisados em algo compreensível para o computador (MACHADO et al., 2010). A metodologia proposta por Aranha, Vellasco e Passos (2007), divide a mineração de textos em cinco etapas distintas: extração, pré-processamento, transformação, mineração e análise.

O Processamento de Linguagem Natural (PLN), em sentido amplo, tem a função de realizar a comunicação entre o computador e a linguagem humana tratando, no nível computacional, os seus diversos aspectos, como sons, palavras, sentenças e discursos, sendo considerados os seus formatos e referências, estruturas e significados, contextos e usos. Essa comunicação poderá ocorrer nos diversos níveis de entendimento ou geração, fonético, morfológico, sintático, semântico e pragmático (GONZALEZ; LIMA, 2003).

E Aprendizado de Máquina é uma área da Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais sobre o aprendizado bem como a construção de sistemas capazes de adquirir conhecimento de forma automática. Um sistema de aprendizado é um programa de computador que toma decisões baseado em experiências acumuladas por meio de solução bem-sucedida de problemas anteriores (Michalski, et. all, 2013).

XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018
22 a 26 de outubro de 2018 – Londrina – PR

4 METODOLOGIA

A mineração de texto realizada no RIF deverá identificar, extrair as informações e estruturá-las conforme o padrão exemplificado na tabela abaixo:

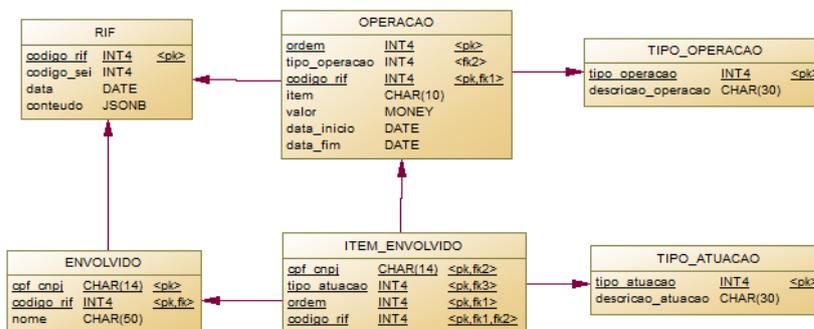
Tabela 1: Informações do RIF tabuladas.

ORDEM	RIF	ITEM RIF	ORIGEM ou DEPOSITANTE CPF/CNPJ	ORIGEM ou DEPOSITANTE NOME	DESTINO ou SACADOR CPF/CNPJ	DESTINO ou SACADOR NOME	RESPONSÁVEL pelo DEPÓSITO ou SAQUE CPF/CNPJ	RESPONSÁVEL pelo DEPÓSITO ou SAQUE NOME	TIPO OPERAÇÃO Seleção	VALOR (EM R\$) somente números	DATA/PERÍODO
01	12345	R-1.1.1	01.222.888/0001-70	Auto Posto 1 Ltda	19.555.999/0001-01	X1 Fomento Mercantil Ltda			Transferência	RS 4.411.827,00	01/09/2014 a 15/10/2014
02	12345	R-1.1.1	01.555.999/0001-30	Auto Posto 2 Ltda	19.555.999/0001-01	X1 Fomento Mercantil Ltda			Transferência	RS 3.855.000,00	01/09/2014 a 15/10/2014
03	12345	R-1.1.1	01.777.222/0001-52	Auto Posto 3 Ltda	19.555.999/0001-01	X1 Fomento Mercantil Ltda			Transferência	RS 1.963.150,00	01/09/2014 a 15/10/2014
04	12345	R-1.1.1	01.888.444/0001-88	Auto Posto 4 Ltda	19.555.999/0001-01	X1 Fomento Mercantil Ltda			Transferência	RS 3.124.624,00	01/09/2014 a 15/10/2014
05	12345	R-1.1.1	01.222.777/0001-01	Auto Posto 5 Ltda	19.555.999/0001-01	X1 Fomento Mercantil Ltda			Transferência	RS 4.300.049,00	01/09/2014 a 15/10/2014
06	12345	A-1.1.	19.555.999/0001-01	X1 Fomento Mercantil Ltda	858.756.543-53	Fabiana Siqueira	087.654.321-00	Ricardo Gil	Saque	RS 100.000,00	06/10/2014
07	12345	R-1.1.2	19.555.999/0001-01	X1 Fomento Mercantil Ltda	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda			Transferência	RS 8.572.000,00	01/09/2014 a 15/10/2014
08	12345	R-1.1.2	19.555.999/0001-01	X1 Fomento Mercantil Ltda	06.999.555/0001-06	Logs Logística e Serviços Ltda			Transferência	RS 8.331.317,00	01/09/2014 a 15/10/2014
09	12345	R-2.1	858.756.543-53	Fabiana Siqueira	02.333.777/0001-13	Zel Petróleo Ltda	02.333.777/0001-13	Zel Petróleo Ltda	Depósito	RS 130.000,00	05/11/2012
10	12345	R-3	06.999.555/0001-06	Logs Logística e Serviços Ltda	02.333.777/0001-13	Zel Petróleo Ltda			Transferência	RS 4.734.930,00	01/09/2014 a 15/10/2014
11	12345	R-3	06.999.555/0001-06	Logs Logística e Serviços Ltda	15.151.151/0001-09	Contan Logística Ltda			Transferência	RS 496.614,00	01/09/2014 a 15/10/2014
12	12345	R-4	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	02.333.777/0001-13	Zel Petróleo Ltda			Transferência	RS 4.637.200,00	01/09/2014 a 15/10/2014
13	12345	R-5	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	011.444.333-40	Cicrano Ferreira	011.444.333-40	Cicrano Ferreira	Saque	RS 200.000,00	01/10/2014
14	12345	R-5	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	011.444.333-40	Cicrano Ferreira	011.444.333-40	Cicrano Ferreira	Saque	RS 100.000,00	03/10/2014
15	12345	R-5	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	011.444.333-40	Cicrano Ferreira	011.444.333-40	Cicrano Ferreira	Saque	RS 250.000,00	07/10/2014
16	12345	R-5	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	011.444.333-40	Cicrano Ferreira	011.444.333-40	Cicrano Ferreira	Saque	RS 150.000,00	10/10/2014
17	12345	R-6	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	099.222.555-87	Beltrano de Sousa	099.222.555-87	Beltrano de Sousa	Saque	RS 200.000,00	08/10/2014
18	12345	R-6	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	099.222.555-87	Beltrano de Sousa	099.222.555-87	Beltrano de Sousa	Saque	RS 100.000,00	09/10/2014
19	12345	R-6	07.555.222/0001-40	Logs Distribuidora de Petróleo Ltda	099.222.555-87	Beltrano de Sousa	099.222.555-87	Beltrano de Sousa	Saque	RS 100.000,00	10/10/2014

Fonte: Elaborado pelo autor.

As informações então serão armazenadas em banco de dados, conforme modelo de entidade e relacionamento proposto abaixo, permitindo que possam ser acessadas posteriormente, realizar cruzamentos entre diversos RIFs e pesquisas sobre as entidades identificadas.

Figura 3: Modelo de Entidade e Relacionamento.



Fonte: Elaborado pelo autor.

Os testes iniciais para a mineração de texto no RIF foram realizados com a biblioteca SpaCy⁵, que possui um modelo em português, pt_core_news_sm. A biblioteca SpaCy tem um analisador sintático (*parser*) que realiza a análise de um texto e determina sua estrutura gramatical. Primeiro utilizamos a biblioteca Pdfminer⁶ para transformar o documento PDF em

⁵ <https://spacy.io/>

⁶ <https://pypi.org/project/pdfminer/>

arquivo texto. Posteriormente, utilizamos o analisador sintático da biblioteca SpaCy para identificar as entidades CPF, CNPJ, Localidade e Data (em dois formatos, por extenso ou com barras). Com os primeiros resultados obtidos, alimentamos à biblioteca SpaCy com os seus próprios resultados de treinamento, com o intuito de melhorar a acurácia.

5 RESULTADOS

Os resultados preliminares da fase de mineração de texto do documento RIF, utilizando a biblioteca SpaCy e as realimentações, podem ser observados na tabela 1. As métricas avaliadas foram: i) a quantidade de palavras marcadas corretamente como entidade, ii) a quantidade de palavras marcadas indevidamente como entidade e iii) a quantidade de entidade omitidas. Foram utilizados 4 modelos diferentes para realizar a mineração do texto:

- Modelo 1 - modelo original disponibilizado no SpaCy, versão utilizada 2.0.0;
- Modelo 2 - modelo gerado com base no original, acrescido do treino de 500 iterações dos 'dados de treino' e 20 iterações nos dados dos arquivos de Localidades;
- Modelo 3 - modelo gerado com base no original, acrescido do treino de 20 iterações dos 'dados de treino';
- Modelo 4 - modelo em branco, acrescido do treino de 20 iterações dos 'dados de treino'.

Os resultados podem ser vistos na Tabela 1. O Modelo 1 teve acurácia de 14%, enquanto o Modelo 2 não conseguiu nenhuma marcação. Os modelos Modelo 3 e Modelo 4 tiveram uma performance superior às demais, obtendo acurácia de 43% e 50% respectivamente. Notadamente, houve um avanço na acurácia, de acordo com a mudanças no modelo original da biblioteca SpaCy. Entretanto, mesmo utilizando a biblioteca SpaCy com as alterações realizadas foi insatisfatória para o cumprimento do objetivo de retirar as entidades do texto.

Tabela 1: Resultados da Identificação.

Modelo	Acertos(%)	Erros de Marcação (%)	Erros por Omissão (%)
Modelo 1	14	53	33
Modelo 2	0	3	97
Modelo 3	43	33	24
Modelo 4	50	45	5

Fonte: Elaborado pelo autor.

6 CONSIDERAÇÕES FINAIS

Podemos concluir que, teoricamente a proposta de automatização na detecção de vínculos é viável, mas na prática o modelo em português da biblioteca SpaCy ainda deixa muito a desejar. Os modelos treinados, apesar de uma melhor performance, também não são satisfatórios. Serão necessários mais aprimoramentos com os treinamentos, pesquisa de outras alternativas de modelos em português ou mesmo mudar a estratégia utilizada, realizando pesquisas heurísticas para alguns tipos padrões de entidades.

REFERÊNCIAS

ARANHA, C. N.; VELLASCO, M. M. B. R.; PASSOS, E. P. L. **Uma Abordagem de Pré-Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional**. [s.l.] Universidade Católica do Rio de Janeiro, 2007.

ARAS, V. Sistema nacional de combate à lavagem de dinheiro e de recuperação de ativos. **Revista Jus Navigandi**, nº 1411, Teresina, 2007.

BRASIL. **Lei da Lavagem de Dinheiro**, 1998. Disponível em:
<http://www.planalto.gov.br/ccivil_03/Leis/L9613.htm. Acesso em jun 2018.>

CALLEGARI, A. L. **Imputação Objetiva. Lavagem de dinheiro e outros temas do Direito Penal**. Porto Alegre: Livraria do Advogado, 2001.

COAF. **Relatório de Inteligência Financeira - RIF**. Disponível em:
<<http://coaf.fazenda.gov.br/menu/a-inteligencia-financeira/relatorio-de-inteligencia-financeira-rif>>. Acesso em: 1 jul. 2018.

GONZALEZ, M.; LIMA, V. L. S. DE. Recuperação de Informação e Processamento da Linguagem Natural. **XXIII Congresso da Sociedade Brasileira de Computação**, p. 3, 2003.

LUSTOSA, D. S. DE M. Aspectos gerais do crime de lavagem de dinheiro (Lei 9.613/98). **Âmbito Jurídico, Rio Grande, XII, n. 70**, 2009.

MACHADO, A. P. et al. Mineração de Texto em Redes Sociais Aplicada à Educação a Distância. **Colabor@ - A Revista Digital da CVA-RICESU**, nº 23, v. 6, 2010.

MARQUES, N. J. F. O papel do COAF no combate ao crime de lavagem de dinheiro. **Conteúdo Jurídico, Brasília-DF**, 2014.

OBREGON, S. R. D. G. P. **Lavagem de Dinheiro. Argumentum (UNIMAR)**, v. 1, p. 75–82, 2001.

Michalski, Ryszard S., Jaime G. Carbonell, and Tom M. Mitchell, eds. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

ONU. CONVENÇÃO DE 20 DE DEZEMBRO DE 1988 Convenção das Nações Unidas contra o Tráfico Ilícito de Estupefacientes e Substâncias Psicotrópicas, Viena - Áustria, 1988.

POLÍCIA FEDERAL. Manual Prático de Combate à Lavagem de Dinheiro e aos Crimes Financeiros (reservado) Brasília - DF, 2013.