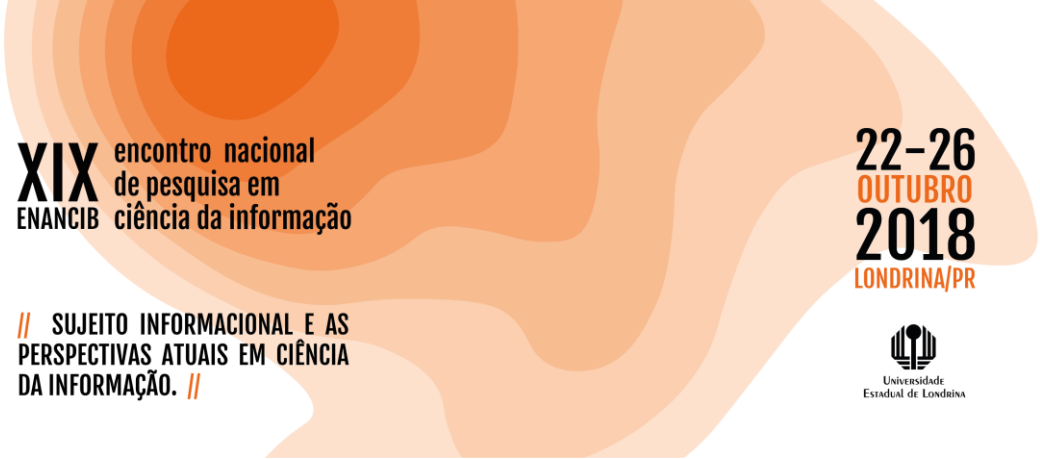




XIX encontro nacional
de pesquisa em
ciência da informação
ENANCIB

// 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

MINERAÇÃO DE TEXTO EM PEÇAS PRODUZIDAS NO INQUÉRITO POLICIAL RELATADO

Marcio Ponciano da Silva (Universidade Federal de Santa Catarina – UFSC)

Angel Freddy Godoy Viera (Universidade Federal de Santa Catarina – UFSC)

DISCOVERY OF KNOWLEDGE IN PIECES OF POLICE INVESTIGATION REPORTED

Modalidade da Apresentação: Pôster

Resumo: O inquérito policial é um procedimento suportado por solução tecnológica, no entanto há nele um acervo de dados textuais na qual pode-se realizar a exploração de padrões e conhecimentos. O uso das técnicas de mineração de textos serve para a descoberta de conhecimento em documentos não estruturados. O uso dessas técnicas sobre as peças de inquérito policial, de forma a tornar o processo de investigação mais eficaz, pode ser evidenciado em estudo do caso. O objetivo desse estudo é propor a utilização das técnicas de mineração de textos (MT) para extrair conhecimento das informações textuais produzidas no inquérito policial e armazenadas em base de dados. Os objetivos específicos são: identificar e delimitar a massa documental a ser minerada. O processo inclui realizar o pré-processamento dos documentos, processar utilizando algoritmos da mineração dos textos para extrair os padrões e avaliar esses padrões para aferir quais os que se aplicam a descoberta de conhecimento no âmbito de aplicação proposta. A amostra da pesquisa será formada por inquéritos que já foram relatados à justiça e não estão sob sigilo. O inquérito policial apresenta potencial de informação favorável à descoberta de conhecimento. Espera-se com essa proposta descobrir conhecimento, cuja finalidade é melhorar o processo de investigação policial.

Palavras-Chave: inquérito policial; mineração de texto; recuperação de informação; descoberta de conhecimento.

Abstract: The police investigation is a procedure supported by a technological solution, however there is a collection of textual data in which one can carry out the exploration of patterns and knowledge. The use of text mining techniques is for the discovery of knowledge in unstructured documents. The use of these techniques on pieces of police investigation, in order to make the investigation process more effective, can be evidenced in a case study. The purpose of this study is to propose the use of text mining techniques to extract knowledge of the textual information produced in the police investigation and stored in a database. The specific objectives are: to identify and to delimit the corpus

textual to be mined. The process includes performing documents preprocessing, using text mining algorithms to extract the patterns, and evaluating those patterns to gauge which ones apply to knowledge discovery in the proposed scope of application. The sample of the survey will be formed of inquiries that have already been reported to the court and they are not under secrecy. The police investigation presents potential information favorable to the discovery of knowledge. It is hoped with this proposal to discover knowledge, whose purpose is to improve the police investigation process.

Keywords: police investigation; text mining; information retrieval; knowledge discovery.

1 INTRODUÇÃO

Hodiernamente, muito se discute entre os operadores da segurança pública nacional acerca de propostas que ajudem na evolução do setor quanto ao emprego de procedimentos e técnicas. Essa preocupação visa encontrar melhores respostas aos problemas existentes, no uso de novas tecnologias voltadas a melhorar o tratamento, a guarda e a recuperação de informações. Os temas que surgem no âmbito dos inquéritos policiais no Brasil estão, imersos no contexto social, visto que respondem a questões impostas pela sociedade e pelo Estado.

Dada a peculiaridade do modelo de investigação criminal brasileira, pode-se afirmar que não há pesquisa científica semelhante. Esse estudo abordará pesquisas tangenciais que mais se aproximam do tema. A principal distinção está no fato de que o inquérito policial brasileiro não representa o processo criminal em si. No Brasil, o processo criminal inicia quando o Ministério Público oferece a denúncia ao Poder Judiciário. Por esta razão, o inquérito policial é um procedimento inserido na fase pré-processual.

Mesmo diante dessa realidade temática do inquérito policial brasileiro, que impõe uma discussão de cunho social, cabe como ponto de partida atentar para a unidade singular do inquérito policial. Essa unidade é o documento. O inquérito é composto de documentos nele produzidos e de documentos que a ele são anexados. Dentre esses últimos, estão documentos cuja origem se dá fora do órgão policial.

Cada documento constitui uma unidade singular no contexto do inquérito policial. De forma geral, os documentos produzidos no inquérito são chamados de “peças”. Esses documentos produzidos formam um aglomerado de documentos não estruturados e semiestruturados. O volume de documentos de um inquérito está relacionado ao resultado das apurações dos fatos. Esse volume abriga um acervo de informações.

O emprego de tecnologias tem se mostrado cada vez mais presente na resolução de questões sociais, conferindo papéis relevantes à Ciência da Informação, com a oferta de modelos de transformação, e à Ciência da Computação, por ofertar tecnologias aderentes a

esses modelos. Essa proposta se situa na fronteira entre essas Ciências. Um exemplo de que as tecnologias vêm ocupando fortemente espaço social é a iniciativa governamental que recentemente aprovou na Comissão de Ciência e Tecnologia, Comunicação e Informação¹, proposta de redução a zero da cobrança de taxas atualmente usuais sobre produtos caracterizados pelo padrão da Internet das Coisas.

Sendo o inquérito uma coleção de documentos, a descoberta de conhecimento com o uso da mineração de textos constitui uma análise promissora e caracteriza o emprego de procedimentos e técnicas da Ciência da Informação e da Ciência da Computação a serviço da missão social. Propõe-se ao caso do inquérito policial a aplicação de técnicas de recuperação da informação e mineração de texto, para a descoberta de conhecimento em peças de inquérito policial.

O objetivo dessa proposta é utilizar as técnicas de mineração de texto para extrair conhecimento das informações de textos de inquérito policiais armazenadas em base de dados. São objetivos específicos coletar a massa de dados a ser minerada para delimitar o seu conjunto, realizar o pré-processamento, processar algoritmo de mineração de textos para extrair padrões e avaliar os padrões extraídos para aferir os apropriados ao caso.

A proposta considera como etapas metodológicas a coleta, o pré-processamento, a indexação, a mineração de textos e análise que sintetizará a interpretação dos dados. O resultado esperado é descobrir informações relevantes e padrões desconhecidos, quais sejam as entidades representativas, as ações dessas entidades e os vínculos existentes entre elas, cuja finalidade é melhorar o processo de investigação documentado no inquérito policial.

2 O INQUÉRITO POLICIAL COMO UMA COLEÇÃO DE DOCUMENTOS

O inquérito policial é um instrumento utilizado pelas polícias brasileiras que registra, numa certa ordem pré-estabelecida e estruturado com um conjunto de elementos, os fatos apurados em diligências de ato tipificado como infração penal nas leis brasileiras. A função de registrar essas infrações penais e a sua autoria está no Código de Processo Penal, que subscreve: “Art. 9º Todas as peças do inquérito policial serão, num só processado, reduzidas a escrito ou datilografadas e, neste caso, rubricadas pela autoridade” (BRASIL, 1941).

¹ BRASIL. Câmara dos Deputados. **Comissão aprova proposta que zera taxas para incentivar o uso da “internet das coisas”**. Disponível em: < <http://www2.camara.leg.br/camaranoticias/noticias/ciencia-e-tecnologia/560458-comissao-aprova-proposta-que-zera-taxas-para-incentivar-o-uso-da-%e2%80%9cinternet-das-coisas%e2%80%9d.html>>. Acesso em: 01 jul. 2018.

As peças do inquérito policial, como traz o texto do Código de Processo Penal, são os documentos que o compõem. A expressão é um jargão jurídico para documentos. Para Luiz Flávio Gomes, o inquérito policial é um “conjunto de diligências que visa a apuração do fato punível e de sua autoria” (GOMES, 2005, p. 62).

Com isso, pode-se dizer que inquérito policial é uma coleção de documentos, posto que ao longo do tempo ele se tornou um volumoso conjunto de peças. Esse acervo documental está caracterizado por diversos tipos de documentos. Essas peças são documentos como portarias, autos (como auto de apreensão, auto de prisão em flagrantes, auto de qualificação e interrogatório), termos (de declaração, de depoimento, etc.), dentre outros. O próprio termo “autos” significa documentos. Um outro documento que se produz para o inquérito policial é o laudo pericial.

As informações de um inquérito policial estão ínsitas no volume de peças que o compõe. Não há um número certo de peças que formam o inquérito policial. Assim, os inquéritos policiais podem ter volumes diversos. De forma geral, os volumes processuais da Administração Pública são desmembrados em torno de duzentas e cinquenta folhas.² Na Polícia Federal, esse desmembramento está regulamentado com a Instrução Normativa nº 108/2016, que determina o desmembramento em volumes quando o procedimento policial atingir aproximadamente duzentas páginas. Nessa proporção, os volumes são formados, podendo o inquérito policial ter um ou vários volumes³.

Conforme se verifica na anotação de Misse (2010), é possível indicar minimamente como peças que compõem um inquérito policial as textuais: portaria, depoimentos, autos, relatórios e informação de investigação; as periciais: exame cadavérico, necropapiloscópico, exame de local de crime; os informativos: boletim de vida pregressa, fotos da vítima e do autor. Outro aspecto que deve ser levado em consideração para caracterizar documentos produzidos em inquérito policial diz respeito ao tamanho que cada documento pode ter. Esse tamanho é flexível, ele será determinado pela quantidade de informação que cada caso requer.

² BRASIL. Supremo Tribunal Federal. **Instrução Normativa nº 191**. Disponível em:

<<http://www.stf.jus.br/Arquivo/norma/instrucaonormativa191-2015.pdf>>. Acesso em: 01 jul. 2018.

³ FEDERAÇÃO NACIONAL DOS POLICIAIS FEDERAIS. **Boletim de Serviço nº 210/2016**. Disponível em:

<<http://www.fenapef.org.br/wp-content/uploads/2016/11/IN-nova-PJ-1.pdf>>. Acesso em: 23 jul. 2018.

3 DESCOBERTA DE CONHECIMENTO E MINERAÇÃO DE TEXTOS

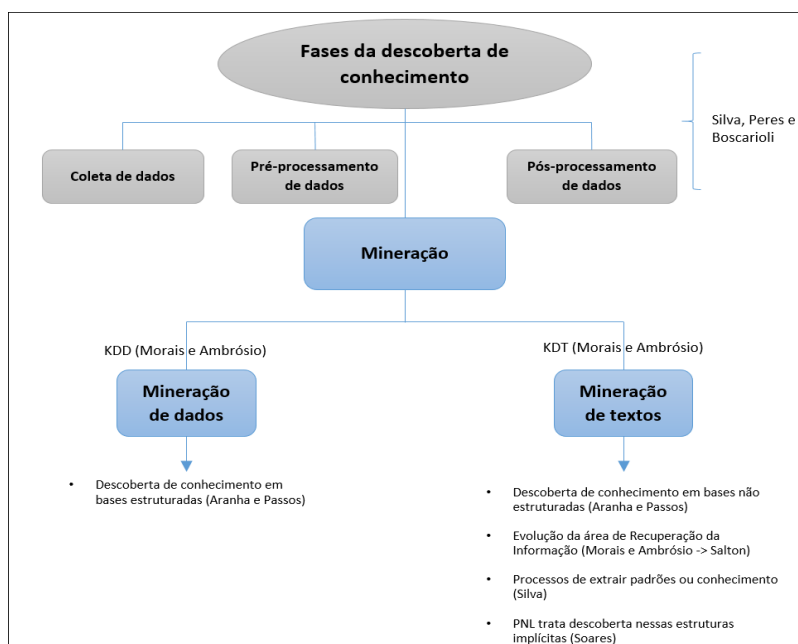
O processo de descoberta de conhecimento visa encontrar padrões em bases de dados. Para Silva, Peres e Boscardoli (2016), esse processo passa por quatro fases: a coleta de dados, pré-processamento de dados, a mineração de texto e o pós-processamento dos resultados provenientes da mineração. Os autores enfatizam que, diferente dos dados estruturados que já constam de certa organização, os dados não estruturados necessitam de uma preparação antes de submetê-las a análises exploratórias.

Segundo Aranha e Passos (2006), a mineração de textos pode também ser chamada de mineração de dados textuais ou descoberta de conhecimento de bases de dados textuais. Para eles, a mineração de textos pode ser vista como uma extensão da mineração de dados, sendo esta o processo de descoberta de conhecimento em bases de dados estruturadas, enquanto aquela o processo de descoberta de conhecimento em bases de dados não estruturadas.

Concordam Moraes e Ambrósio (2007, p.1 apud SALTON, 1983), pois registram a mineração de texto como uma evolução da área de Recuperação de Informações, acrescentando que é um Processo de Descoberta de Conhecimento, e que utiliza técnicas de análise e extração de dados a partir de textos. Para Silva (2005), a descoberta de conhecimento em textos pode ser definida como o processo de extrair padrões ou conhecimentos, interessantes e não-triviais, a partir de documentos textuais.

Moraes e Ambrósio (2007) explicam que existem basicamente duas abordagens na área de Descoberta de Conhecimento (KD), a Descoberta de Conhecimento a partir de Dados Estruturados – KDD (Knowledge Discovery in Databases) e a Descoberta de Conhecimento a partir de Dados não Estruturados – KDT (Knowledge Discovery from Text). A mineração de textos está representada nessa última abordagem. A figura 1 é uma demonstração das contribuições de diversos autores a respeito do processo de descoberta de conhecimento.

Figura 1: diagrama de conceitos da descoberta de conhecimento



Fonte: Autores – 2018.

Soares (2010) ratifica que é comum na bibliografia técnica afirmar que as bases de dados textuais se apresentam de forma não-estruturada, mas pondera que elas possuem uma estrutura implícita que necessita de técnicas especializadas para ser reconhecida por sistemas automatizados. Para ele, o Processamento de Linguagem Natural (PLN) trata exatamente da descoberta dessas estruturas implícitas. Também registraram Woszezenki e Gonçalves (2012, p. 92 apud ZHOU, 2008) semelhantemente, ao dizer que grande parte dessas atividades é desenvolvida com o uso das técnicas de Processamento de Linguagem Natural (PLN) e análise de estatística.

4 DESCOBERTA DE CONHECIMENTO EM PEÇAS DE INQUÉRITO POLICIAL

As peças textuais que são produzidas no inquérito policial brasileiro formam uma massa de dados não estruturada, armazenada em banco de dados relacional, que são utilizadas ao longo do curso do inquérito. Após ser relatado, essa massa constitui um dossiê digital. O objetivo desse dossiê é resgatar cópia das peças caso venham a ser solicitadas para atender a instrução do próprio inquérito. Essas peças são vistas na individualidade do inquérito que as gerou. Assim, essas peças, quase que absolutamente, não são comparadas com outras peças de outros inquéritos.

Essa realidade que individualiza a visibilidade das peças de um inquérito apenas em torno dele mesmo, demanda uma necessidade de exploração dessas peças levando em consideração a universalidade do conjunto dessa massa. Esse olhar como um todo permite uma exploração holística desse fenômeno, pondo as informações a serviço da descoberta de conhecimento. A discussão a respeito da exploração desse acervo documental coloca foco no cerne de uma importante área do conhecimento, qual seja a Recuperação da Informação.

Para Capurro e Hjørland (2007), o termo recuperação da informação é um dos mais importantes no campo do conhecimento da Ciência da Informação. A recuperação da informação não foi a única, mas foi a principal responsável pelo desenvolvimento da Ciência da Informação (SARACEVIC, 1996). Com esse viés, vê-se que a recuperação da informação em peças textuais de inquérito é uma promissora contribuição tanto da Ciência da Informação quanto da Ciência da Computação para a segurança pública nacional.

A análise de uma abordagem da Ciência da Informação, pelo empréstimo de seus métodos e técnicas, faz com que se reflita que o processo de descoberta e conhecimento em peças de inquérito policial permitirá a identificação de informações relevantes, capaz de agregar conhecimento ao processo de conhecimento e inteligência das polícias judiciárias. Em outro sentido, é possível que procedimentos e técnicas aplicados na prática no campo da segurança pública possam ser objeto de estudo da Ciência da Informação. Essa simbiose, já era defendida por Borko (1968), ao destacar que técnicas e procedimento empregados por bibliotecários e documentaristas deveriam ser baseados nos resultados teóricos da Ciência da Informação, bem como os teóricos dessa área deveriam estudar técnicas e procedimentos aplicados na profissão.

A essa altura, poder-se-ia ainda perguntar por que usar técnicas e procedimentos para exploração de peças textuais de inquéritos policiais. A resposta é direta, trata-se de um grande volume de informação, ao qual se presta o processo de descoberta de conhecimento pela mineração de dados textuais. Como já visto em Aranha e Passos (2006), esta é a aplicação do processo de descoberta de conhecimento em bases de dados não estruturados.

3 CONSIDERAÇÕES FINAIS

A proposta apresentada neste estudo faz parte da pesquisa de mestrado em Ciência da Informação aborda uma reflexão sobre o potencial de descoberta de conhecimento em peças de inquérito policial. O acervo documental a ser explorado é o conjunto de documentos textuais produzidos no inquérito policial, no âmbito da Polícia Federal, nos últimos dez anos.

Esse acervo reúne inquérito das vinte e sete unidades federativas do País. As peças textuais estão armazenadas em banco de dados. O volume de inquéritos na condição da pesquisa reúne mais de quinhentos mil registros, sendo que pode-se atribuir em média 7 documentos por inquérito, o que resulta numa base de documentos não-estruturados em torno de 3,5 milhões de documentos.

Como solução, propõe o emprego da recuperação de informação e a mineração de texto para descoberta de padrões que revelem novos conhecimentos. A análise do resultado dessa abordagem revelará quais técnicas de mineração de textos e modelos de recuperação de informação se aplicam em peças de inquérito policial para a descoberta de conhecimento.

Este estudo é um desafio de cunho inovador, vez que promove a possibilidade de contribuição da Ciência da Informação e da Ciência da Computação para o campo da segurança pública. Essa abordagem leva em conta os desafios da realidade nacional, que deposita no inquérito policial, lastreado por uma coleção de documentos, a missão de promover o processo de justiça social.

REFERÊNCIAS

ARANHA, Christian; PASSOS, Emmanuel. **A tecnologia de mineração de textos**. RESI-Revista Eletrônica de Sistemas de Informação, nº 2, 2006.

BORKO, H. **Information Science: What is it? American Documentation**, v.19, n.1, p.3-5, Jan. 1968. (Tradução Livre)

BRASIL. Câmara Dos Deputados. **Comissão aprova proposta que zera taxas para incentivar o uso da “internet das coisas”**. Disponível em:

<<http://www2.camara.leg.br/camaranoticias/noticias/ciencia-e-tecnologia/560458-comissao-aprova-proposta-que-zera-taxas-para-incentivar-o-uso-da-%e2%80%9cinternet-das-coisas%E2%80%9d.html>>. Acesso em: 01 jul. 2018.

BRASIL. **Decreto-Lei nº 3.689**, de 3.10.1941. Código de Processo Penal Brasileiro. DOU de 13.10.1941. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto-lei/Del3689Compilado.htm>. Acesso em: 10 jul. 2018.

BRASIL. Supremo Tribunal Federal. **Instrução Normativa nº 191**. Disponível em: <<http://www.stf.jus.br/arquivo/norma/instrucaonormativa191-2015.pdf>>. Acesso em: 01 jul. 2018.

CAPURRO, R.; HJORLAND, B. **O conceito de informação**. Perspectivas em Ciência da Informação, v. 12, n. 1, p. 148-207, 2007.

FEDERAÇÃO NACIONAL DOS POLICIAIS FEDERAIS. **Boletim de Serviço nº 120**. Disponível em: <<http://www.fenapef.org.br/wp-content/uploads/2016/11/IN-nova-PJ-1.pdf>>. Acesso em: 23 jul. 2018.

GOMES, Luiz Flávio. **Direito processual penal**. São Paulo: Revista dos Tribunais, 2005.

MISSE, Michel. **O inquérito policial no Brasil: uma pesquisa empírica**. Rio de Janeiro: NECVU/IFCS/UFRJ; BOOKLINK, 2010.

MORAIS, E. A. M; AMBRÓSIO, A. P. L. **Mineração de textos**. Technical report, Instituto de Informática – Universidade Federal de Goiás, UFG, Brasil, 2007.

SILVA, Jacqueline Uber. **Text mining com uma aplicação na validação dos registros de ocorrências policiais na região da grande Florianópolis**. p. 123. Dissertação (Mestrado) – Ciência da Computação, Universidade Federal de Santa Catarina, Florianópolis, 2005. Disponível em: <<https://repositorio.ufsc.br/bitstream/handle/123456789/102510/222367.pdf?sequence=1&isAllowed=y>>. Acesso em: 25 jun. 2018.

SILVA, Leonardo Augusto; PERES, Sarajane Marques; BOSCARIOLI, Clodis. **Introdução à mineração de dados – com aplicações em R**. Rio de Janeiro: Elsevier, 2016.

SOARES, Afrânio Martins. **A mineração de texto na análise de contas públicas municipais**. p. 85. Dissertação (Mestrado) - Informática Aplicada, Universidade Estadual do Ceará, Fortaleza, 2010. Disponível em: <www.uece.br/mpcomp/index.php/arquivos/doc_download/226-dissertacao-67>. Acesso em: 26 jun. 2018.

SARACEVIC, Tefko. **Ciência da Informação: origem, evolução e relações. Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

WOSZEZENKI, Cristiane; GONÇALVES, Alexandre Leopoldo. **Descoberta baseada em literatura: Estado da arte**. International Journal of Knowledge Engineering, Universidade Federal de Santa Catarina, Florianópolis, 2012, v.1, n.1, p.91-103. Disponível em: <<http://incubadora.periodicos.ufsc.br/index.php/IJKEM/article/view/1924/2231>>. Acesso em: 28 jun. 2018.