Produção de Provas Digitais a partir de Rastreamento em Relacionamentos por *e-mails*

Jackson Mallmann , Cinthia O. A. Freitas, Altair Olivo Santin

Programa de Pós Graduação em Informática (PPGIa)
Pontifícia Universidade Católica do Paraná (PUC-PR)
Caixa Postal 80215-901 - Curitiba - PR - Brasil

{jackson.mallmann, cinthia, santin}@ppgia.pucpr.br

Abstract. *It is obvious nowadays that cybercrime is a phenomenon with a global dimension. The necessary evidence hopefully is found by the experts, but sometimes it is not possible to be found or it is a hard-working . Without evidences the suspect cannot be charged and prosecuted. In a business context, there are many types of potential evidence (log files, e-mails, and personal computers). Thus, this paper presents a mechanism for clustering and classification the conversations by e-mails. The experimental results, which reach correct rates close to 98%, demonstrated that it is possible to use the e-mail tracking to production of digital evidence.*

Resumo. *Atualmente é evidente que o cibercrime é um fenômeno de dimensão global. Espera-se que as provas necessárias sejam encontradas pelos peritos, mas algumas vezes é difícil encontrá-las ou isto se torna um trabalho árduo. Sem evidências o suspeito não pode ser acusado e processado. Em um contexto de negócios, existem muitos tipos de evidências em potencial. Citam-se os arquivos de log, e-mails e computadores pessoais. O presente artigo apresenta um mecanismo de agrupamento e classificação das conversações por e-mails. Os resultados experimentais, que atingem taxas de acerto próximas de 98%, demonstraram que é possível usar o rastreamento em relacionamentos por e-mails para a produção de provas digitais.*

1. Introdução

O *e-mail* é uma das aplicações de rede mais antigas [Peterson e Davie 2004], constituindo-se num método para enviar e receber mensagens mediante emprego de sistemas eletrônicos de comunicação [Limeira 2007]. O crescimento no uso desta ferramenta de comunicação tem sido contínuo ao longo dos anos, visto que pesquisa feita por [Group 2009] durante o ano de 2009 registrou aproximadamente 1.4 bilhões de usuários de *e-mails* e, ainda, um número de 247 bilhões de mensagens enviadas por dia. Tal pesquisa, estima a existência de 1.9 bilhões de usuários de *e-mail* até o ano de 2013, totalizando assim 507 bilhões de *e-mails* enviados por dia.

Além da grande utilização da ferramenta eletrônica *e-mail*, observa-se também o aumento de crimes realizados por meio de serviços virtuais, ou seja, as denominadas condutas criminosas no *cyber* espaço ou, ainda, os *cibercrimes* [Cia 2004] [Broadhurst 2006] [Pinheiro 2009] [Surendran et al. 2005]. Cita-se como exemplo, a propagação de crimes de pedofilia na *Internet* [Chopra 2006] e casos de assédio sexual [Sipior 1999]. Segundo a *SaferNet*, durante o mês de dezembro de 2009 [SaferNet 2010], foram

registradas 3.492 denúncias de crimes virtuais ligados ao *site* de relacionamentos *Orkut* [Orkut 2010]. Além destas, outras 1.567 denúncias estavam relacionadas a outros domínios.

Quando um crime virtual ocorre e é denunciado formalmente, os órgãos competentes o tipificam de acordo com a legislação onde o mesmo ocorreu. Após a tipificação, normalmente é necessário que um profissional especializado realize uma perícia nas evidências do crime, para que seja comprovada a ligação do *cibercrime* cometido com o seu autor, gerando o nexo causal [Delmanto 2000] [Jesus 2002] [Noronha 1998].

Considerando o exposto, este artigo apresenta um mecanismo que visa auxiliar os peritos na produção de provas digitais a partir do *e-mail*. Realiza-se o rastreamento em informações textuais no corpo (*body*) dos *e-mails* para identificar contextos (agrupamentos) de palavras (sentenças) nas conversações (trocas de mensagens) por *e-mails*. Mediante a identificação e análise dos agrupamentos, visa-se a produção de provas digitais e constituição do nexo causal. Entende-se que o nexo de causalidade está relacionado com o vínculo entre a conduta ilícita e o dano, ou seja, o dano deve decorrer diretamente da conduta ilícita praticada por um indivíduo, sendo então consequência única e exclusiva dessa conduta. Assim, o nexo causal é elemento necessário para se configurar a responsabilidade civil do agente causador do dano.

O presente trabalho propõe e descreve um mecanismo para auxiliar no rastreamento de relacionamentos por *e-mails*, para que as provas digitais possam ser apresentadas pelos peritos aos membros da ciência jurídica, os quais por sua vez irão formalizar o nexo causal no Processo Judicial. Assim, para validação do mecanismo proposto, o presente trabalho analisa a taxa de acerto proporcionado pelo mecanismo com base na aplicação de diferentes técnicas de agrupamento e classificação. E, ainda, apresenta indicações do tempo de execução do mecanismo, visando a comparação com o tempo de trabalho de um perito, o qual realiza manualmente a análise de bases de *e-mails*. Os resultados experimentais demonstram a aplicabilidade do mecanismo proposto no contexto de produção de provas digitais.

O artigo está organizado da seguinte forma. A Seção 2 descreve conceitos que serão aplicados para auxiliar na determinação do nexo causal pela produção de provas digitais a partir do rastreamento de relacionamentos em *e-mails*. A Seção 3 discute trabalhos relacionados. A Seção 4 apresenta a proposta de rastreamento de *e-mails*. A Seção 5 detalha aspectos técnicos da implementação, dos experimentos realizados e, dos resultados obtidos. Finalmente, a Seção 6 discute os resultados e apresenta as conclusões.

2. Técnicas de Processamento de Textos

A seguir serão abordados brevemente aspectos teóricos sobre o pré-processamento textual e a extração de características, a representação dos atributos textuais, e métodos de agrupamento e de classificação.

2.1. Pré-processamento Textual e Extração de Características de Documentos

No texto localizado no corpo e cabeçalho de *e-mails*, identificam-se, entre outros, diversas palavras. Por meio do procedimento de normalização, o qual integra o pré-processamento do conjunto textual de *e-mails*, visa-se facilitar a localização de palavras

que caracterizem nos *e-mails* algum tipo de ação criminosa [Tam e Lourenço 2008].

Algumas técnicas utilizadas neste processo são [Neto et al. 2000]:

- *Stop words*: retiram-se do texto palavras de pouco valor semântico, como por exemplo, artigos e preposições;
- *Case foldering:* todas as palavras que representam o texto de um *e-mail* são convertidas para sua forma minúscula;
- *N-grams*: cada palavra do texto é dividida em pequenas partes. Cita-se como exemplo a palavra "azul", que ao se utilizar a técnica de *tri-grams*, gerará termos com até 3 letras oriundas desta palavra, obtendo-se todas as possibilidades de subsequências a partir da palavra em questão: " az", "azu", "zul" e "ul_".

Entende-se para o presente trabalho que palavras que caracterizam um conjunto de *e-mails* são chamadas de características ou atributos. Os atributos dividem-se em duas categorias: quantitativos e qualitativos [Jain et al. 1999]. Os atributos quantitativos são aqueles que podem ser medidos em uma escala quantitativa, ou seja, apresentam valores numéricos que fazem sentido, podendo ser contínuos ou discretos. Por outro lado, os atributos qualitativos são aqueles que não possuem valores quantitativos, mas, representam uma classificação dos indivíduos ou no caso em questão permitem identificar características do conjunto de *e-mails* analisado.

A partir do conjunto de atributos, pode-se realizar uma representação do mesmo. Mediante análise do conjunto *D* (Equação 1) verifica-se que cada atributo é representado por x_i, definindo, assim, *n* atributos, sendo que "i = 1, 2, ..., n". Desta forma, *n* equivale a quantidade total de atributos escolhidos para representação dos *e-mails*.

$$D = \{x1, x2, x3, \dots, xn\} \quad (1)$$

2.1.1. Representação de Atributos Textuais

Feito o pré-processamento dos textos a serem investigados e definido o conjunto de atributos relevantes para sua caracterização, pode-se obter a representação dos atributos em vetores, matrizes, sequências (*codebooks*), contagem e verificação (*assertions*) e, finalmente, grafos [Trier et al. 1996]. Dentre as técnicas utilizadas para representação dos atributos dos *e-mails*, destacam-se: tf-idf (*term frequency–inverse document frequency*) e *bag-of-words*. Resumidamente, tais técnicas, referem-se à:

- Tf-idf: Técnica utilizada como medida estatística para avaliar o quanto é importante o atributo de um texto, atribuindo um peso a cada atributo. Esta medida na verdade é a representação da proporção de vezes que um atributo ocorre em um texto em relação ao número de textos em que o mesmo atributo foi localizado [Salton e Buckley 1988] [Rodrigues 2009], tal qual definido pela Equação 2:

$$tf - idf = \frac{Freq_{atributo.texto}}{DocFreq_{atributo}} \quad (2)$$

- *Bag of words*: É uma técnica utilizada na área de processamento da linguagem natural e recuperação da informação. Nesta técnica, um texto (como uma frase ou um documento) é representado como uma coleção não ordenada de palavras, ignorando a gramática e até mesmo a ordem das palavras, representando todos

os atributos do texto em um único vetor. Assim, pode-se verificar a quantidade de repetições dos atributos existentes no conjunto textual [Hotho et al. 2005].

2.2. Agrupamento

Define-se a técnica de agrupamento como a exploração de padrões entre objetos, no caso em questão os *e-mails*. Os objetos são então reunidos em grupos, sendo que os grupos não são previamente padronizados (treinados). Por não existir o treinamento, a técnica de agrupamento é considerada uma abordagem não supervisionada [Jain et al. 1999].

Esta técnica pode ser utilizada para verificar se dois *e-mails* pertencem a um mesmo contexto, no caso, se estabelecem uma sequência de mensagens trocadas durante uma conversação. Assim, contextos de mesma similaridade identificarão um mesmo grupo. A medida de similaridade pode ser definida durante a parametrização da técnica de agrupamento.

A técnica de agrupamento é implementada por um algoritmo e necessita seguir algumas etapas para seu funcionamento. Determinam-se os padrões que simbolizam um conjunto de *e-mails* [Jain et al. 1999]. Em seguida, os padrões são extraídos dos *e-mails* em forma de atributos. Feita a extração de atributos, estipula-se uma função para cálculo da similaridade que irá calcular a distância entre os diferentes padrões, sendo que esta distância é calculada por métodos específicos, por exemplo, a distância *Euclidiana*. Outros métodos existem para este cálculo, podendo-se citar: Manhattan, Chebychev, Camberra, Cossine, entre outros [Skeoch 2006].

Desta forma, o algoritmo de agrupamento poderá analisar um conjunto de *e-mails* teste. Baseado nos atributos informados, o algoritmo avaliará os resultados do cálculo de distância entre os atributos dos *e-mails* e quanto menor o valor da distância, maior a similaridade entre *e-mails*, ou que pertençam ao mesmo contexto, identificando seu grupo.

Dentre os algoritmos de agrupamento, o mais comumente utilizado é o *k-means* [Jain et al. 1999] [Hotho et al. 2005]. O algoritmo *k-means* (método particional) realiza a divisão do conjunto de *e-mails*, em *K* grupos. O valor de *K* deve ser informado na inicialização do algoritmo, por isso é um método supervisionado. O número *K* será representado por centróides (pontos geométricos). O algoritmo inicia os centróides em um hiperplano, e a partir da verificação do cálculo da distância entre *e-mails*, calcula a média da distância entre todos os e-*mails* do grupo formado, identificando um novo centróide. Assim, os *e-mails* mais próximos do centróide estarão contidos em um mesmo grupo/contexto. Esses centróides são atualizados do início ao fim do processo de agrupamento, e a técnica para cálculo da distância utilizada normalmente é a *Euclidiana* [Jain et al. 1999].

Após formação dos agrupamentos, procede-se com a identificação dos *e-mails* agrupados de acordo com seu contexto. Cada *e-mail* terá um rótulo fornecido pelo algoritmo de agrupamento, que informa a qual grupo este pertence. Embora realizada a verificação dos *e-mails* agrupados, torna-se interessante a validação dos grupos originados. Esta validação pode ser realizada por meio da verificação da taxa de acertos quando concluído o agrupamento de *e-mails* conhecidos. E, ainda, por meio de métodos estatísticos pode-se analisar a qualidade dos grupos formados, por exemplo, analisando-se a matriz de similaridade.

A seguir apresenta-se a classificação de *e-mails* através de aprendizagem supervisionada, de modo complementar às técnicas de agrupamento.

2.3. Classificação

A classificação consiste em rotular textos em classes [Tam et al. 2005] [Rodrigues 2009] [Sebastiani 2002]. Os textos podem corresponder a *e-mails*, que podem ser classificados na classe criminosa ou não, por exemplo, como é o interesse do presente artigo. A classificação é designada mediante um treinamento supervisionado, no qual treina-se o classificador para reconhecer os padrões e as classes a serem utilizadas, formalizando-se, assim, um modelo de classificação. As classes *c* pertencem ao conjunto *C*, composto por *y* classes, conforme Equação 3:

$$C = \{c1, c2, \ldots cy\} \quad (3)$$

Feito o treinamento, o modelo de classificação estará pronto para realizar a caracterização de *e-mails* teste de forma automática. Existem diferentes métodos de classificação, sendo sua aplicação ou não ligada a natureza do problema a ser resolvido, por exemplo: verificação de assinaturas, reconhecimento de palavras manuscritas, reconhecimento de dígitos manuscritos, entre outros [Britto Jr. et al. 2001]. O algoritmo escolhido deverá gerar modelos que consigam classificar com a maior exatidão possível o conjunto de *e-mails* de teste nas classes definidas [Tam et al. 2005].

Posterior a classificação, avalia-se o desempenho do classificador mediante a utilização de diferentes métricas, podendo-se citar a análise da precisão resultante da classificação. Nesta métrica, a avaliação da precisão do modelo de classificação baseia-se nas taxas de erro e acerto do classificador. Esses valores são anotados e apresentados em uma matriz de confusão [Tam et al. 2005]. Assim é possível a realização de comparações entre diferentes classificadores na solução de um mesmo problema.

Dentre os vários algoritmos para a tarefa de classificação, detalham-se os mais comumente utilizados, SVM (*Support Vector Machine*), NB (*Naïve Bayes*) e DT (*Decision Tree*) [Tam et al. 2005]:

- SVM binário é um classificador estatístico com o objetivo de encontrar um hiperplano ótimo que estabeleça geometricamente (linear, polinomial, exponencial, entre outros) dois conjuntos baseados no padrão dos textos de treinamento [Rodrigues 2009] [Lorena e Carvalho 2007] [Hotho et al. 2005]. Cada lado do hiperplano identifica uma classe. Os *e-mails* utilizados durante o teste e analisados pelo classificador são caracterizados em um dos lados do hiperplano, identificando a classificação do *e-mail*. Para o trabalho em questão a decisão de 1 hiperplano se apresenta devido ao fato de termos duas classes para os *e-mails*: criminoso ou não criminoso;
- NB é um classificador probabilístico fundamentado no Teorema de *Bayes*, baseado em padrões e classes designadas durante o treinamento, sendo que este classificador indica a classe de maior probabilidade para cada novo *e-mail* teste [Hotho et al. 2005] [Rodrigues 2009] [Sebastiani 2002];
- DT consiste em um conjunto de regras que são aplicadas de maneira seqüencial, finalizando com uma decisão, caracterizando-se por um classificador estatístico [Hotho et al. 2005]. As várias possibilidades são organizadas na forma de uma

árvore de decisão hierarquicamente estruturada, sendo que uma DT também pode ser representada por um conjunto de regras condicionais do tipo "se-então" [Rodrigues 2009].

3. Trabalhos Relacionados

Na literatura técnica existem estudos de processos que automatizam o trabalho de usuários da ferramenta *e-mail.* A seguir são descritos estudos que apresentam alguns métodos que podem ser utilizados na constituição do nexo causal relacionado com o rastreamento de *e-mails*.

No estudo de [Nagwani e Bhansali 2010] é proposto um modelo que utiliza o algoritmo de agrupamento *k-means* (similaridade do Cossine) para agrupamento de 310 *e-mails* existentes na base da *Enron Corpus*. O modelo calcula a similaridade entre *e-mails*, considerando como atributos os textos dos campos do cabeçalho, corpo e destinatário. Este trabalho obteve uma precisão de 78% na utilização da técnica *10-fold Cross Validation* [Rabiner e Juang 1993]. O modelo proposto pode ser utilizado para classificação dos *e-mails* de um usuário armazenados em pastas do sistema de arquivos, por exemplo.

No trabalho realizado por [Cselle 2006] os autores aplicaram algoritmos de classificação e agrupamento com o objetivo de organizar 3 bases de *e-mails* por assuntos. Os autores testaram 1.346 *e-mails* pré-organizados em 76 assuntos. Os resultados experimentais mostraram que: o classificador NB obteve uma taxa de acertos de 55,1%; o classificador baseado em regras, 52,0% e o classificador fundamentado no remetente da mensagem, 47,4%. Além destes classificadores, os autores testaram as bases com um algoritmo de agrupamento (*Single Link Cluster* - tf-idf), alcançando uma taxa média de acertos de 82%.

Em [Dredze et al. 2006] foram utilizados 149 *e-mails* pré-definidos em 27 classes para treinamento em dois classificadores distintos. Estas classes foram utilizadas na classificação automática de uma base composta por 1.146 *e-mails*. No primeiro classificador foi aplicada como métrica de similaridade os contatos (destinatário e remetente) identificados no cabeçalho de *e-mails*. No outro classificador, foi identificada a similaridade entre o conteúdo do corpo de um *e-mail* teste com os *e-mails* de treinamento, usando uma variação do algoritmo LSI (*Latent Semantic Indexing*) para medir a similaridade. Para ambos os classificadores projetados, quanto maior a similaridade, maior a probabilidade de um *e-mail* que está sendo testado ser classificado corretamente em uma classe. Foi realizada também uma comparação com o classificador NB. Os resultados experimentais demonstraram que os classificadores propostos alcançaram uma taxa de acerto de 94% dos *e-mails* testados. Entretanto, o algoritmo NB não obteve bons resultados, visto que os testes demonstraram que tal algoritmo obteve uma taxa de erros de 50% na classificação dos *e-mails*.

Nos trabalhos de [Cselle 2006] e [Dredze et al. 2006] foi utilizado o classificador NB, sendo que este classificador se mostrou inadequado para realizar agrupamento/classificação de *e-mails*. Porém, quando se trata de somente duas classes o desempenho do algoritmo NB é bom, o que pode ser constatado no trabalho de [Balamurugan e Ranjaram 2008].

Foi proposto por [Balamurugan e Ranjaram 2008] uma classificação de *e-mails* em duas classes: ameaçador e não ameaçador. Os autores apresentam comparações entre

os classificadores DT, SVM e NB com duas bases de *e-mails,* com 2.099 e 3.893 *e-mails*, respectivamente. Os resultados dos testes utilizando as palavras do corpo dos *e-mails* mostraram que o algoritmo DT apresentou 97,4%, o algoritmo SVM 95,6% e o NB 95,4% de acerto na classificação de *e-mails*. Para [Balamurugan e Ranjaram 2008] o classificador DT é uma abordagem promissora para detecção automática em *e-mail*; tendo melhor desempenho que o SVM e sendo menos complexo que o classificador NB.

Em paralelo, no estudo de [Nunes et al. 2009], palavras que caracterizam assédio moral no ambiente de trabalho foram coletadas de diversas maneiras, passando inclusive por pré-processamento. No pré-processamento aplicou-se a técnica de 3-*grams* sobre duas bases de dados contendo 25 *e-mails* e 512 palavras de um dicionário de assédio moral. Para efeito de testes, foram comparados todos os 3-*grams* de uma palavra do *e-mail* com os 3-*grams* de todas as palavras do dicionário, individualmente. Na classificação dos 25 *e-mails,* por similaridade, utilizando o dicionário de assédio moral, os autores conseguiram uma taxa de acerto de 90,91%, com uma taxa de falso positivo igual a 9%, sendo estes os *e-mails* identificados manualmente com indícios de assédio e não identificados corretamente pelo sistema.

Nesta Seção foram apresentados diversos trabalhos com bons resultados na aplicação de métodos para classificação de *e-mails*. Embora haja conflitos entre os resultados apresentados e o tamanho das bases de *e-mail* utilizadas para testes em cada trabalho, como por exemplo, resultados do classificador NB entre diferentes autores como [Dredze et al. 2006] e [Balamurugan e Ranjaram 2008], é visível a utilização dos conceitos apresentados na produção de provas digitais relacionada ao rastreamento de *e-mails*. Ademais, os trabalhos apresentados são limitados a utilização de apenas um método, os quais não foram utilizados exaustivamente como aplicações da área forense, seja em problemas de classificação ou de agrupamento de *e-mails*. Deste modo, o mecanismo aqui proposto utiliza muitos dos conceitos apresentados.

4. Proposta

Objetiva-se um mecanismo que produza provas digitais a partir da análise de evidências de *cibercrime* no conjunto de conversações deflagrados nas mensagens textuais dos *e-mails* de um suspeito. O mecanismo visa auxiliar a análise que peritos realizam em *e-mails*, haja vista o grande esforço e tempo que profissionais despendem na realização deste tipo de trabalho. A Figura 1 apresenta as etapas da proposta deste mecanismo, e no decorrer desta Seção suas devidas explicações são descritas.

Inicialmente, procede-se com a cópia "bit-a-bit" (imagem) da base de *e-mails* de um suspeito (Etapa 1 - Figura 1). Evidências digitais em formato de *e-mails* não podem sofrer modificações durante procedimentos de análise forense, mantendo-se assim a integridade dos *e-mails* do suspeito [Craiger 2007] [McKemmish 1999] [Kruse e Heiser 2002].

Posteriormente aos *e-mails* serem copiados, estes são submetidos as técnicas de pré-processamento (Etapa 1 - Figura 1). Neste conjunto textual aplicam-se as técnicas de c*ase foldering* e *stop words.* Removem-se também os códigos HTML (*HyperText Markup Language*), URLs (*Uniform Resource Locator*), símbolos e números. Isto devido ao fato de que objetiva-se que elementos (palavras, símbolos e códigos) encontrados nos *e-mails* sejam separados do conjunto textual investigado, tornando as

palavras resultantes do pré-processamento, os atributos que representem os *e-mails* do suspeito.

Feita a extração dos atributos, realiza-se uma primeira Representação dos Atributos ainda na Etapa 1 mediante o uso de vetores. Estes vetores identificam a similaridade entre *e-mails*, pois similaridade entre dois *e-mails* é indício de pertencerem à mesma conversação. Desta forma, quanto maior a quantidade de atributos repetidos entre os *e-mails*, maior a evidência de pertencerem à mesma conversação.

Como visto anteriormente, existem várias técnicas para representação dos atributos de um texto. No presente trabalho são aplicadas as técnicas *bag of words* e tf-idf [Rodrigues 2009].

Utilizando os atributos representados, o mecanismo aplica o Agrupamento (Etapa 2 - Figura 1) por meio do método *k-means*. Este é o algoritmo mais utilizado no agrupamento de dados textuais [Nagwani e Bhansali 2010]. Na utilização do algoritmo de *k-means* optou-se pelo uso da técnica da distância *Euclidiana*, visto que os trabalhos relacionados descritos anteriormente indicam a aplicação desta técnica. E, ainda, que tais algoritmos foram aplicados com sucesso por [Huang 2008] no estudo do agrupamento de 7 documentos textos.

Efetuado o agrupamento, os resultados são submetidos à Análise dos Resultados (Etapa 2 - Figura 1). Evidencia-se a quantidade e quais os *e-mails* pertencentes à mesma conversação entre usuários (remetentes e destinatário). Conversações entre usuários, mesmo na utilização de diferentes endereços eletrônico de *e-mail* também são identificadas. Os *e-mails* agrupados corretamente em suas conversações são assim contabilizados e os valores registrados.

Identificados os *e-mails* das conversações, realiza-se o procedimento de produção de provas digitais encontradas nesses *e-mails*, ou seja, os *e-mails* são classificados (Etapa 3 – Figura 1) como criminosos ou não. Efetua-se esta etapa do trabalho mediante a aplicação de diferentes algoritmos de classificação. Porém, primeiramente deve-se proceder com uma nova extração de atributos.

Na segunda Extração de Atributos (Etapa 3 - Figura 1) para classificação criminal dos *e-mails*, são utilizadas as palavras de um dicionário contextualizado como base para extração de tais atributos dos *e-mails*. Assim, aplicam-se as técnicas de 3-*grams* e 4-*grams* nas palavras encontradas nos textos suspeitos já pré-processados. Os termos resultantes são então comparados às subsequências geradas pela utilização dos algoritmos 3-*grams* e 4-*grams* nas palavras do dicionário. Identifica-se assim, as palavras dos *e-mails* que tenham um grau de similaridade ao do dicionário que serve de base para identificação do que é criminoso ou não (atributos).

O grau de similaridade é denotado por um limiar ou *threshold*. Quanto maior o valor deste limiar, maior a necessidade da existência de *x* termos *n-grams* para serem comparados com as palavras dos *e-mails* e com as palavras do dicionário. Identificando uma palavra do *e-mail* que tenha *threshold x*, identifica-se o atributo como positivo. Assim, a existência de assédio moral nos *e-mails* ocorre através da semelhança entre as palavras existentes no corpo dos *e-mails* avaliados e aquelas do dicionário utilizado pelo método. Por isso considera-se a utilização de diferentes *n-grams* para analisar similaridade entre palavras. Deste modo, durante os experimentos utilizou-se um dicionário contendo 539 palavras de crimes de assédio moral formalizado no trabalho de

[Nunes et al. 2009]. O detalhamento deste dicionário está além do escopo do presente artigo e informações podem ser encontradas em [Nunes et al. 2009].

Uma vez identificadas as palavras criminosas existentes nos *e-mails* (extração dos atributos), estas são representadas em vetores (Etapa 3 – Figura 1), tendo por base a utilização de uma modificação da técnica *bag of words.* Para o problema em questão, as palavras encontradas no texto dos *e-mails* são identificadas, originando assim uma representação binária.

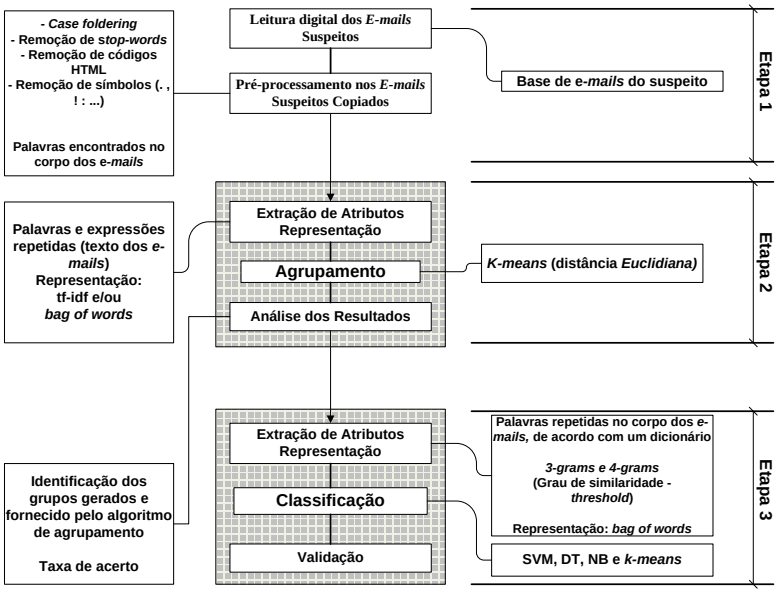

Figura 1. Fluxo de Trabalho do Mecanismo Proposto

Esta representação (vetores) é usada na classificação (criminoso e não criminoso) dos *e-mails* (Etapa 3 – Figura 1). Optou-se por avaliar os algoritmos de agrupamento *k-means* com distância *Euclidiana* e os algoritmos: SVM, DT e NB. A escolha de tais algoritmos tem por base os trabalhos relacionados já descritos anteriormente, o que os apontam como opções quando o problema envolve agrupamento [Nagwani e Bhansali 2010] [Huang 2008] e classificação de textos [Tam et al. 2005]. Ademais, apresentaram bons resultados, a exemplo do ocorrido no trabalho de [Balamurugan e Ranjaram 2008], sendo que os autores obtiveram taxas na ordem de 95% para tais classificadores. Deve-se lembrar que estes classificadores foram aplicados na detecção de textos de *e-mails* ameaçadores e não ameaçadores ou, seja, um problema com duas classes dentro de mesmo contexto de crimes virtuais.

Realizada a classificação dos *e-mails* pertencentes às conversações, realiza-se a Validação (Etapa 3 – Figura 1) do processo de classificação, mediante análise das taxas de acerto e erro dos métodos aplicados.

5. Estudo de Caso

Inicialmente, foi implementada uma ferramenta denominada *Reader*, a qual foi desenvolvida em linguagem *Java* [Java 2010]. Para esta ferramenta a base de *e-mails* do usuário suspeito deve estar no formato do *software* de *e-mail Windows Live Mail* [Windows 2010]. Assim, *Reader* realiza a leitura dos textos encontrados no corpo dos *e-mails* suspeitos e informados seguindo notações padronizadas [IETF 2008].

Reader procede com a etapa de Pré-processamento dos *e-mails* do suspeito, assim como a Representação de Atributos para a Agrupamento e Classificação dos *e-mails* (Figura 1). Terminadas estas etapas, faz-se uso do *software Weka* [Weka 2010] para as etapas de Agrupamento e Classificação. Este *software* possibilita o uso de vários algoritmos de Aprendizagem de Máquina. Dentre eles, o algoritmo de agrupamento *k-means* e os algoritmos classificadores: SVM, NB e DT.

Apresentam-se a seguir a descrição das bases de *e-mails* utilizadas, os resultados dos experimentos, assim como, as comparações entre os resultados obtidos.

5.1. Bases Testadas

Nos experimentos foram utilizadas duas bases de *e-mails* seguindo as características das bases apresentadas e estudadas na Seção Trabalhos Relacionados. A primeira base contém 179 *e-mails* coletados entre os autores, conforme Tabela 1. Estes *e-mails* foram mapeados em conversações. Todas as conversações foram identificadas e quantificadas. A quantidade total de palavras encontradas nas conversações é igual a 120.119 palavras, tendo uma média de 671 palavras por *e-mail*.

Tabela 1. Primeira Base – 179 *e-mails*

Conversação	Quantidade de *e-mails*	Quantidade de palavras	Conversação	Quantidade de *e-mails*	Quantidade de palavras
1	11 *e-mails*	12.149	7	8 *e-mails*	3.591
2	2 *e-mails*	820	8	5 *e-mails*	4.196
3	2 *e-mails*	739	9	7 *e-mails*	3.331
4	6 *e-mails*	2.649	10	15 *e-mails*	43.764
5	3 *e-mails*	2.549	11	108 *e-mails*	32.137
6	12 *e-mails*	14.194	**Total**	**179 *e-mails***	**120.119**

Para verificação de uma conversação, foi analisada a existência da troca de mensagens de *e-mail* entre usuários (destinatário e remetente). As mensagens de uma conversação envolvem o mesmo contexto (conjunto de palavras). Dois usuários podem ter várias conversações, mesmo na utilização de diferentes endereços de *e-mail*.

Os *e-mails* identificados como não pertencentes a uma conversação, foram caracterizados na Conversação “11”, como por exemplo, *e-mails* recebidos e não respondidos ou *spam*.

Ademais, todos os *e-mails* das conversações foram utilizados durante a aplicação do método de agrupamento proposto pelo mecanismo, visando atender a etapa de identificação de conversações. Os resultados destes agrupamentos serão mostrados na Seção 5.2.

A segunda base de dados contém 40 *e-mails* característicos de crime de assédio moral [Nunes et al. 2009]. Estes *e-mails* foram divididos na razão 80:20 visando a criação de outros dois conjuntos de *e-mails*: treinamento e teste. Estes conjuntos auxiliaram nos métodos de classificação da primeira base.

No conjunto de treinamento, composto por 32 *e-mails* criminosos, adicionou-se também 32 *e-mails* não criminosos. Procedeu-se de forma similar para o conjunto teste. Formado por 8 *e-mails* criminosos, somado a 8 *e-mails* não criminosos.

Finalmente, uniu-se o conjunto de teste (16 *e-mails*) à primeira base de *e-mails* (179 *e-mails*) para realização de testes de classificação. Totalizando desta forma uma

base de *e-mails* para classificação equivalente a 195 *e-mails*. Entre os 195 *e-mails*, esperava-se encontrar 12 *e-mails* (Conversação "6") e outros 8 *e-mails* criminosos que foram adicionados a partir da base de teste elaborada.

5.2. Resultados Experimentais

Inicialmente foi testado o agrupamento da primeira base de *e-mails* no uso do algoritmo *k-means* com distância *Euclidiana*. Os resultados estão disponíveis na Tabela 2. Nesta Tabela, apresenta-se a quantidade de *e-mails* da base agrupada, e a taxa de acertos provenientes de *k-means* na utilização da representação via técnica *bag of words* e tf-idf.

Tabela 2. Taxa de Acertos - Agrupamento – *k-means* – *Euclidiana*

179 *e-mails*	Técnica ***bag of words***	Técnica tf-idf
	75,00%	74,45%

Observando os resultados apresentados na Tabela 2, verifica-se que muitos *e-mails* foram agrupados corretamente. *E-mails* que pertencem a Conversação "11" (Tabela 1) foram agrupados, em sua grande maioria, neste mesmo grupo. E ainda, confirma-se que quanto maior a quantidade de atributos significativos no conjunto de *e-mails*, maior a facilidade e precisão no agrupamento correto dos *e-mails*.

Tendo-se os resultados do agrupamento dos *e-mails*, realizaram-se testes de classificação. Desta forma, facilita-se a determinação de conversações criminosas visando a produção de provas digitais.

Os *e-mails* da primeira base, somados aos *e-mails* de teste foram submetidos aos algoritmos de classificação SVM com *kernel Polynomial*, DT e NB, assim como, ao algoritmo de agrupamento *k-means*, na utilização da distância *Euclidiana*. Disponibiliza-se na Tabela 3 a quantidade de *e-mails* classificada (195 *e-mails*), o resultado dos classificadores utilizados, bem como, as técnicas *n-grams* empregadas.

Na classificação foram usadas as técnicas de 3-*grams* e 4-*grams* para comparação entre palavras existentes no corpo dos *e-mails* com as palavras do dicionário de assédio moral. A detecção de palavras criminosas foi considerada com base nesses atributos, tendo sido a representação dos atributos realizada mediante a técnica *bag of words*. Usaram-se também as técnicas de *n-grams* com *threshold* equivalente a 67% e 100%.

Observa-se através dos resultados expostos na Tabela 3 que o classificador SVM obteve os melhores resultados na classificação dos *e-mails* quando fez uso das técnicas de 3-*grams* com *threshold* equivalente a 100% e 4-*grams* com *threshold* de 67% e 100%. Entretanto com *threshold* de 67%, o classificador SVM apresentou taxas de acerto inferior ao classificador NB.

O classificador NB comparado ao SVM resultou em menor taxa de acertos, concordando com [Dredze et al. 2006] [Cselle 2006]. E também, ao contrário dos resultados apresentados em [Balamurugan e Ranjaram 2008], haja vista que o SVM teve melhor desempenho que NB e DT. Assim, de acordo com os experimentos realizados, o classificador SVM apresenta-se como um classificador interessante para estudos de classificação de *e-mails* no contexto de crimes virtuais.

Tabela 3. Taxa de Acertos – Classificação

		SVM *Polynomial*	DT	NB	*k-means Euclidiana*
195 *e-mails*	3-*grams* 67%	51,29%	90,26%	32,31%	87,18%
	3-*grams* 100%	**98,98%**	90,31%	96,43%	83,16%
	4-*grams* 67%	87,18%	94,07%	86,67%	82,06%
	4-*grams* 100%	**98,98%**	90,31%	96,43%	83,16%

Além dos testes de precisão realizados, no uso de um computador provido de processador Intel (R) Core (TM) Duo CPU 2.20 GHz, com 4 G de RAM, mediu-se o tempo para o processamento do mecanismo para análise da base de 195 *e-mails*, comparando este valor com a processo manual que um perito utilizaria nesta prática. Na execução do mecanismo (análise criminosa de conversações), necessitou de aproximadamente 14 minutos, com média de 4,3 segundos para processamento de cada *e-mail*. Na análise manual da mesma base, foi necessário em média, 40 segundos para realizar a leitura, agrupamento e classificação dos *e-mails* de cada conversação, totalizando 130 minutos. Resumidamente, o mecanismo reduziu o tempo de análise em aproximadamente 89%, o que é de suma importância para os peritos, pois em situações reais o volume de *e-mails* analisados pode estar na ordem de grandeza de GBytes.

6. Conclusões e Direções Futuras

Este artigo apresentou um mecanismo para produção de provas digitais a partir do rastreamento de relacionamentos em *e-mails*. Foram aplicados para tal, métodos de agrupamento e classificação de *e-mails* e os resultados experimentais demonstram ser possível obter evidências digitais quanto à autoria e/ou quanto à materialidade de possível delito. Assim, as análises realizadas podem ser formalizadas pelos membros da ciência jurídica de modo a caracterizar o nexo causal (vínculo entre a conduta ilícita e o dano).

Na investigação da base de *e-mails* de um usuário suspeito contextualizam-se conversações praticadas por *e-mail*. Assim, *e-mails* de um usuário que possuam o mesmo contexto, podem ser agrupados. As conversações foram agrupadas pela aplicação do algoritmo *k-means*. Ademais, os *e-mails* das conversações são posteriormente classificados objetivando-se a produção de provas digitais. Foram aplicados os algoritmos: DT, NB, SVM e *k-means*. Ao se obter a classificação dos *e-mails* das conversações em criminoso ou não, têm-se os resultados comprobatórios das evidências digitais. Assim, o sistema proposto serve como framework para análise de qualquer crime que tenha disponível um dicionário formalizado. Neste trabalho foi utilizado o dicionário com palavras de assédio moral.

Dentre os algoritmos utilizados neste trabalho, *k-means* apresentou taxa de acerto de 75% no agrupamento de conversações, e o algoritmo SVM alcançou 98% para classificação dos *e-mails* agrupados, sobre uma base de 195 *e-mails*. E, ainda, obteve-se uma redução de 89% do tempo necessário para um perito produzir as provas digitais usando a proposta em relação a realização do processo manual para a mesma análise.

Os métodos utilizados por esses algoritmos podem ser modificados em futuros trabalhos na busca por melhores resultados computacionais, como por exemplo, no uso de diferentes técnicas para cálculo da distância utilizado pelo algoritmo de agrupamento, assim como, um diferente método de treinamento dos classificadores.

Espera-se ainda estruturar uma apresentação gráfica (grafos) dos relacionamentos criminosos existentes na base de *e-mails* investigada.

Referências

Balamurugan, S.A.A.; Ranjaram, R. (2008) "Learning to classify threatening e-mail", Int. J. Artificial Intelligence and Soft Computing, V. 1, N. 1, p.39-51.

Britto Jr, A. S.; Freitas, C. O. A.; Justino, E.; Borges, D. L.; Facon, J.; Bortolozzi, F.; Sabourin, R. (2001) "Técnicas em Processamento e Análise de Documentos Manuscritos", Revista de Informática Teórica e Aplicada, Porto Alegre - UFRGS, V. 8, N. 2, p. 47-68.

Broadhurst, R. (2006) "Developments in the global law enforcement of cyber-crime", Policing: An Int. Journal of Police Strategies & Management. V. 29, N. 3, p.408-433.

Chopra, M.; Martin, M. V.; Rueda, L.; Hung, P. C. K. (2006) "Toward new Paradigms to Combating Internet Child Pornography", Canadian Conference on Electrical and Computer Engineering, CCECE, Ottawa, Maio, p.1012-1015.

Cia, S. Ó. (2004) "An Extended Model of Cybercrime Investigations", Int. Journal of Digital Evidence, V. 3, 1ª. Edição.

Craiger, J.P. (2007) "Computer Forensics Procedures and Methods", To appear in H. Bigdoli (Ed.), Handbook of Information Security. John Wiley & Sons.

Cselle, G. (2006) "Organizing email", Master's thesis, ETH Zurich.

Delmanto, C. (2000) "Código penal comentado", Editora Renovar. 5ª. Edição atual. e ampl., Rio de Janeiro.

Dredze, M.; Lau, T.; Kushmerick, N. (2006) "Automatically Classifying Emails into Activities", Proc. of the 11th International Conference on Intelligent User Interfaces, Sydney, Janeiro, p.70-77.

Group, The Radicati. (2009) "Email Statistics Report, 2009-2013", http://www.radicati.com/?p=3237.

Hotho, A.; Nürnberger, A.; Paaβ, G. (2005) "A Brief Survey of Text Mining", Journal for Computational Linguistics and Language Technology, 20, 1, p.19-62.

Huang, A. (2008) "Similarity Measures for Text Document Clustering", New Zealand Computer Science Research Student Conference (NZCSRSC), p.49-56.

IETF. (2008) "Internet Message Format", http://www.ietf.org/rfc/rfc5322.txt.

Jain, A. K.; Murty, M. N.; Flynn, P. J. (1999) "Data Clustering: A Review", ACM Computing Surveys, V. 31, N. 3.

Java (2010) "Linguagem de Programação Open Source", http://java.sun.com.

Jesus, D. E. (2002) "Direito Penal", Editora Saraiva. São Paulo.

Kruse, W.G.; Heiser, J.G. (2002) "Computer Forensics: incident response essentials", Indianapolis: Addison-Wesley.

Limeira, T. M. V. (2007) "E-Marketing", Editora Saraiva. 2ª. Edição, São Paulo.

Lorena, A. C.; Carvalho, A. C. P. L. F. (2007) "Uma introdução às Support Vector Machines (*In Portuguese*)", Revista de Informática Teórica e Aplicada, V. 14, p. 43.

McKemmish, R. (1999) "What is Forensic Computing?" Australian Institute of Criminology Trends and Issues, N. 118, http://www.aic.gov.au/publications/tandi/ti118.pdf.

Nagwani, N. K.; Bhansali, A. (2010) "An Object Oriented Email Clustering Model Using Weighted Similarities between Emails Attributes", Int. Journal of Research and Reviews in Computer Science (IJRRCS), V. 1. N. 2, p.1-6.

Neto, J. L.; Santos, A. D.; Kaestner, C. A. A.; Freitas, A. A. (2000) "Document Clustering and Text Summarization", http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.4634, Abril de 2010.

Noronha, E. M. (1998) "Curso de Direito Processual Penal", Editora Saraiva. 26ª. Edição, São Paulo.

Nunes, A. V.; Freitas, C. O. A.; Paraíso, E. C. (2009) "Detecção de Assédio Moral em e-mails", In: I Student Workshop on Information and Human Language Technology, São Carlos. I Student Workshop on Information and Human Language Technology - 7th Brazilian Symposium in Information and Human Language Technology. POA : SBC. V. 1. p.01-05.

Orkut. (2010) "Site de relacionamentos", http://www.orkut.com.

Peterson, L. L.; Davie, B. S. (2004) "Redes de Computadores: uma abordagem de sistemas", Editora Elsevier. 3ª. Edição, Rio de Janeiro.

Pinheiro, P. P. (2009) "Direito Digital", Editora Saraiva. 3ª. Edição, São Paulo, 2009.

Rabiner, L.; Juang, B.H. (1993) "Fundamentals of speech recognition", Prentice Hall Inc., London, UK. p.506.

Rodrigues, J. P. (2009) "Sistemas Inteligentes Híbridos para Classificação de Texto", Dissertação de Mestrado. Universidade Federal de Pernambuco, Recife.

SaferNet. (2010) "Central Nacional de Denúncias de Crimes Cibernéticos", http://www.safernet.org.br/site/indicadores.

Salton, G.; Buckley, C. (1988) "Term Weighting Approaches in Automatic Text Retrieval", Information Processing and Management, 24, 5, p.513-523.

Sebastiani, F. (2002) "Machine Learning in Automated Text Categorization", ACM Computing Surveys, V. 34, N. 1, Março, p.01-47.

Sipior, J. C.; Ward, B. T. (1999) "The Dark Side of Employee Email", Communications of the ACM, V. 42, N. 7, Julho, p.88-95.

Skeoch, A. (2006) "An Investigation into Automated Shredded Document Reconstructing using Heuristic Search Algorithms", Dissertação: Bachelor of Science in the Department of Computer Science, University of Bath, Reino Unido.

Surendran, A. C.; Platt, J. C.; Renshaw, E. (2005) "Automatic Discovery of Personal Topics to Organize Email", Proc. 2nd Conference on Email and Anti-Spam, CEAS.

Tam, P-N.; Steinbach, M.; Kumar, V. (2005) "Introduction to Data Mining", Addison-Wesley.

Tam, T.; Lourenço, A. (2008) "Estudo Exploratório para a Organização Automática de Mensagens de Correio Electrónico", JETC'08, ISEL.

Trier O.; Jain A.K.; Taxt T. (1996) "Feature extraction methods for character recognition", Pattern Recognition, V. 29, N. 4, p.641–662.

Weka. (2010) "Data Mining Software in Java", http://www.cs.waikato.ac.nz/ml/weka/.

Windows, Windows Live Mail. (2010) "Software de correio eletrônico", http://explore.live.com/windows-live-mail.