5CONTECSI June 04 - 06, 2008 São Paulo, Brazil

5th International Conference on Information Systems and Technology Management
5º Congresso Internacional de Gestão da Tecnologia e Sistema de Informação
De 04 a 06 de Junho de 2008 - São Paulo - Brasil

PS-913

SELF-ORGANIZATION MAP (SOM) OF KOHONEN APPLIED TO THE ACADEMIC INFORMATION CLASSIFICATION FOR SAFETY LEVELS

José Antonio Corrales Torres (Universidade Presbiteriana Mackenzie, São Paulo – Brasil) - jcorrales@uol.com.br
Nizam Omar (Universidade Presbiteriana Mackenzie, São Paulo – Brasil) - omar@mackenzie.br

The information and knowledge grew in importance and have become the most valuable assets, space and time are less relevant and more vulnerable. The information classification is the primary requirement to adjust rules and procedures, the protection level and cost. The current process is manual and subject to imperfections. This study suggests a method to classify the information, regarding its confidentiality, using groups generated by an Artificial intelligence Neural Network. The implementation made use of a Neural Network, based on the Self-Organization Maps (SOM) of Kohonen. The study case objective was the implementation and it considered the information from universities, due to their various properties. The results indicated the similarity among the elements that composed the groups generated by the training of the Neural Network, complemented by calculations using the original weights. The viability of the application of the considered method to an organization was confirmed.

Keywords: information, security, information classification, artificial intelligence neural networks, self-organization maps.

MAPAS AUTO ORGANIZÁVEIS (SOM) DE KOHONEN APLICADOS À CLASSIFICAÇÃO DAS INFORMAÇÕES ACADÊMICAS POR NÍVEIS DE SEGURANÇA

A informação e o conhecimento representam os ativos, num cenário em que o espaço e o tempo, perderam a relevância e tornaram-se vulneráveis. A classificação das informações é fundamental para direcionar as medidas de proteção. Atualmente o processo é manual e sujeito a imperfeição. Propõe-se um método para classificar as informações, quanto à sua confidencialidade, em grupos gerados por uma Rede Neural de Inteligência Artificial. A implementação usou a Rede Neural, baseada nos Mapas Auto-Organizáveis (*SOM*) de Kohonen. O estudo contemplou as informações das universidades. Os resultados obtidos atestam à semelhança dos elementos que compõe os grupos gerados pelo treinamento da Rede Neural, complementado por cálculos que utilizam os pesos iniciais. Mostrou-se assim a viabilidade da aplicação do método proposto para uma organização.

Palavras-Chave: informação, segurança, classificação da informação, rede neural de inteligência, mapas auto-organizáveis.

1 - Problema, Motivação e Objetivo

Em face do avanço das inovações tecnológicas, as concepções de espaço e de tempo perderam a relevância, de forma que o virtual usa novos espaços e novas velocidades, problematizando e reinventando o mundo (Lévy, 1997). Como conseqüência, grande parte dos ativos perdeu a tangibilidade e passou a ser representado por informação e conhecimento.

O novo ambiente tem fronteiras sutis, fragilidade do controle e facilidade de dissimular a identidade dos usuários. Permite o desenvolvimento e a disseminação de ameaças ao seu conteúdo, assim como pode atender a interesses ilícitos (Lévy, 1997).

Os recursos atuais não afiançam a uniforme e adequada categorização da informação, cuja imprecisão provocará danos consideráveis. Se os níveis de confidencialidade e integridade forem determinados em proporções superiores à sua real demanda, onerará os custos computacionais e de segurança desnecessariamente. Caso contrário permitirá a exposição de dados imprudentemente, com conseqüências prejudiciais às informações classificadas. Portanto, a assertividade na classificação, atrelada ao eficiente gerenciamento dos dispositivos de segurança, estabelecem uma relação satisfatória e econômica.

A escassez de recursos financeiros disponíveis para inversão em segurança associada à diversidade de ofertas de serviços e dispositivos para esse fim, compele ao refinamento de separar o conhecimento, informações ou dados que efetivamente representem valor para as organizações, de maneira a preservar o patrimônio e a acurácia dos investimentos (Schneier, 2001).

Este trabalho se propõe a desenvolver um método destinado à classificação das informações, norteado pelo objetivo de segurança da confidencialidade.

2 - Segurança das Informações

Para Lévy (1997) a virtualização da economia é fortemente dependente de dois ativos primordiais e particulares: informação e o conhecimento. São considerados dessa maneira em razão de sua importância para a produção de riquezas; particulares porque se diferem de outros bens pela sua característica de compartilhamento, uma vez que sua cessão não implica em sua perda e o seu consumo não decorre na sua destruição.

O ponto de partida das metodologias, estruturas de trabalho, como o *Control Objectives for Information and Related Technology* - Cobit 4.1 (*ITGI*, 2007) ou padrões de segurança da informação, como a ABNT NBR ISO/IEC NBR 17799 (*ISO*, 2005) destacam a importância da classificação para revestir de maior controle e proteção às informações de maior relevância.

Será necessário analisar todos os impactos decorrentes desse novo cenário, uma vez que a ampliação de transações eletrônicas permite visualizar o fluxo de informação com abrangência jamais imaginada, como os dados de valores, datas, produtos adquiridos, estabelecimentos, freqüência de compras etc., cujo rastreamento consente descrever os hábitos e perfil das pessoas e empresas. Assim, poderá caracterizar-se a perda gradual de privacidade e desenvolver novos incômodos aos cidadãos.

De acordo com Ferraiolo *et al.* (2003) o controle de acesso visa determinar os direitos do usuário sobre determinado recurso, o período permitido para exercer seus direito e de que forma poderá fazê-lo. Assim, o controle de acesso se tornou uma das principais soluções para segurança computacional e para mitigar os riscos relacionados à informação. Os objetivos de segurança da informação podem ser categorizados em Confidencialidade, Integridade e Disponibilidade, cuja descrição, a definição de perda e impactos estão descritos no Tabela 1.

Tabela 1: Descrição Perda e Impacto dos Objetivos de Segurança

Objetivo de Segurança	Descrição	Definição de Perda	Impacto
Confidencialidade	Guarda segura e privada da informação, inclui qualquer situação (de segredo de estado a um memorando) e tipo de informação (financeira, de segurança etc.).	Sua perda implica na divulgação não autorizada da informação.	Pode resultar em desrespeito de privacidade, constrangimento, ações legais etc.
Integridade	Proteção da informação, quanto à alteração imprópria da informação, por pessoas ou grupos não autorizados.	Sua perda implica na destruição ou modificação, não autorizada, da informação.	Pode gerar a imprecisão dos dados, fraude, decisões incorretas e afeta a disponibilidade e a confidencialidade.
Disponibilidade	Garante que a informação é disponível para uso quando é requisitada.	Sua perda ocasiona a interrupção de acesso ou uso da informação e dos sistemas de informação.	Pode resultar na perda de funcionalidade e eficiência operacional da organização.

Fonte: Descrição (Ferraiolo et al., 2003); Definição de Perda. (NIST, 2004) e; Impacto (NIST, 2002).

Stallings (2005) adiciona aos objetivos de segurança citados, um quarto objetivo dedicado a Autenticidade, que define a exigência de um *host* ou um serviço que seja capaz de verificar e validar a identidade de um usuário.

A preservação dos objetivos de segurança é assegurada pela autorização e a autenticação dos usuários, desde que a validação mantenha a dependência entre ambos.

Autenticação é o processo que determina a legitimidade da reivindicação de uma identificação do usuário, portanto comprova que o usuário é realmente quem diz ser. Uma das formas mais comuns de autenticação é o uso de senhas, contudo o uso de equipamentos biométricos (utilizam características do ser humano para identificação, tais como: impressão digital, íris, voz, veias da palma da mão etc.), *smart cards* etc. vêm apresentando maior eficácia e eficiência. O principal requisito de autenticação é reconhecer algo que seja do conhecimento exclusivo do usuário, ou algo que o usuário tenha ou algo que represente uma característica física do usuário. Atualmente, é comum a sobreposição de recursos de autenticação (senha, cartão e outros etc).

Autorização é o ato de determinar se um usuário, seja uma pessoa, um grupo de pessoas ou um sistema computacional, têm o direito de executar determinada tarefa, tal como leitura ou alteração do conteúdo de um arquivo ou execução de um programa. Geralmente os usuários encontram-se divididos em diferentes grupos com características de direitos distintos (Kuong, 1974).

Autenticidade e autorização são sempre empregadas em conjunto, pois um usuário deve ser autenticado antes de poder executar tarefas que ele esteja autorizado a executar (Kuong, 1974).

3 - Administração de Risco no Ambiente de TI

As alterações originadas pelo processo de mudança sistemática e profunda provocam, na mesma intensidade, a alteração das variáveis que afetam os controles de proteção contra ameaças e ataques. O estudo da administração de risco destina-se a direcionar as ações que visam restabelecer o controle e a proteção esperados.

Em termos de estratégia organizacional, risco é a possibilidade de ocorrência de um determinado evento hostil que possa reduzir o valor dos seus ativos (Blakley, 2002). Este evento hostil, ou risco, representa um custo, mesmo quando envolve a segurança das informações, pode ser quantificado (Farahmand, 2003). Estes custos refletem tanto o impacto financeiro causado por ataques às informações, quanto pelos gastos na aquisição e implementação de recursos de segurança.

O gerenciamento de risco permite balancear os custos operacionais e econômicos das medidas de proteção e alcançar ganhos na missão de capacitar a proteção dos sistemas de TI e das informações. Este processo é aplicável a outras atividades operacionais e de negócio das organizações (NIST, 2002).

Devido à indisponibilidade de métricas consistentes de segurança e, por decorrência, na dificuldade de justificar os investimentos em segurança, o gerenciamento de risco busca quantificar a probabilidade de concretização das ameaças, as possíveis conseqüências e os custos de proteção (Geer *et al.*, 2003).

O ambiente computacional (hardware e software) continua se expandindo e atualizando, os sistemas de informação são substituídos ou atualizados. As mudanças de pessoas, das políticas e dos procedimentos também são freqüentes. Este conjunto de mudanças cria novos riscos que demandam avaliações constantes e evolutivas dos controles.

Risco é a função que apresenta a *Probabilidade* das *Ameaças* exercerem sua potencialidade sobre as *Vulnerabilidades*, que resultam em *Impacto* adverso. A probabilidade considera a conjugação das ameaças, as vulnerabilidades e os controles. Os impactos referem-se à magnitude dos eventuais danos.

4 - Redes Neurais

Os estudos preliminares do neurofisiologista McCulloch e do matemático Walter Pitts forneceram bases para a neuro-computação (Russel, 2004). Tal implicação da lugar a construção de um dispositivo baseado no cérebro humano. Esse modelo utiliza um conjunto de entradas, pesos, conexões sinápticas, limiares (*threshold*) e funções de ativação que norteiam o resultado de uma rede neural artificial.

Redes Neurais possuem algoritmos que podem pertencer a grupos distintos de aprendizado, nos modos supervisionados, não-supervisionados e construtivos (Neto; Nicoletti, 2005). No aprendizado supervisionado, para cada vetor de treinamento, são associadas classes (ou vetores de saída). Nesta fase, o algoritmo de treinamento da rede tenta ajustar os pesos das conexões de maneira que a saída da rede coincida com a classe (ou vetor de saída) associada ao exemplo do conjunto de treinamento (Neto; Nicoletti, 2005).

O aprendizado não-supervisionado é mais usado em sistemas de classificação, pois não existe uma saída desejada para cada entrada. Nesse aprendizado, a rede é treinada por excitações ou padrões de entrada e, de forma arbitrária, organiza a saída em padrões e categorias. Para cada entrada é fornecida uma resposta que apresente a classe a qual determinada entrada pertença, gerando uma nova classe, caso o padrão não coincida. Portanto, os padrões de entrada são agrupados à medida que os pesos vão sendo alterados conforme o tempo passa (Kohonen, 2001).

A escolha da técnica adaptativa de rede neural destinada à implementação do modelo envolveu o estudo das Redes Bayesianas, Sistemas Especialistas e os Mapas Auto-Organizáveis (*SOM*) de Kohonen.

Uma especialização das redes neurais são as redes de Kohonen (2001) ou *SOM* ou *Konets*, desenvolvida por Teuvo Kohonen no início da década de 80. (Kohonen, 2001; Suuronen, 2001). De acordo com Wangenheim (2006), Kohonen voltou-se à descoberta de um modelo de auto-organização de informações e um processo de aprendizado indutivo, capaz de ser usado como modelo de aprendizado e organização de informações no neocórtex cerebral de um animal superior. *SOM* são redes competitivas, de aprendizado não-supervisionado, não busca a solução ótima, apenas uma solução viável.

Treinamento da Rede Neural

Na superfície de entrada da rede, para cada parâmetro de saída representado, é atribuído um peso. Cada neurônio armazena um vetor de pesos, cada vetor de pesos correspondente a uma das entradas de um vetor de entradas. Quando surge uma nova entrada, cada neurônio da rede calcula seu nível de ativação através da definição (1):

$$\sqrt{\sum_{i=0}^{n}(weight_i - input_i)^2} \quad (1)$$

onde $weight_i$ é o *i*-ésimo elemento do vetor de pesos e, $input_i$ é a *i*-ésima entrada.

São três os processos envolvidos na formação dos mapas Competição, Cooperação e Adaptação Sinátiptica (Francisco, 2004; Haykin, 2001). Considere-se que o treinamento de uma rede de Kohonen é feito de modo competitivo e não supervisionado. O algoritmo atende as seguintes etapas (Kohonen, 2001; Roussinov, 2001):

- **Inicializar nós de entrada, saída e pesos**: A primeira etapa consiste em criar um *grid* bidimensional de *m* possíveis nós de saída. Organiza-se o *grid* como um grafo bipartido, inicializando os pesos $weight_{ij}$ das conexões entre cada um dos *i*-ésimos nós de entrada e cada j-ésimo nó do *grid* com valores aleatórios. Cada um dos j-ésimos nós de saída está associado a um vetor de pesos $weight_{ij}$.
- **Fornecer dados de entrada**: À medida que são fornecidos os dados de entrada e tem início a interação com o sistema, as informações sobre suas preferências e aspectos importantes são gradativamente apresentados à rede. Cada *i*-ésima entrada do usuário em um dado tempo *t* é representada por um vetor vi(t).
- **Calcular distância no espaço euclidiano**: A terceira consiste em computar a distância euclidiana *dj* entre cada um dos vetores de entrada *vi(t)* e o vetor de pesos $weight_{ij}$:

$$d_j = \sum_{i=0}^{n-1}(v_i(t) - weight_{ij}(t))^2 \quad (2)$$

- **Selecionar o nó vencedor *j** e atualizar os pesos de *j** e de seus vizinhos**: Nesta etapa é selecionado o nó vencedor *j** , que produz a menor distância *dj*. Também ocorre a atualização dos pesos que visa diminuir a distância de *j** e de seus vizinhos em relação a *vi(t)*:

$$weight_{ij}(t+1) = weight_{ij}(t) + n(t)v_i(t) - weight_{ij}(t) \quad (3)$$

onde *η* é um coeficiente de ajuste de erro entre 0 e 1 que diminui ao longo do tempo. Após essas atualizações, os nós na vizinhança de *j** estarão mais similares ao vetor de entradas $v_i(t)$.

- **Rotular regiões no mapa:** Após a etapa de treinamento, a cada saída atribui-se o maior peso como um termo de valoração, conhecido como “termo de vitória” (*winning term*). Todos os nós da vizinhança com o mesmo termo são agrupados em clusters, representando regiões conceitualmente próximas.

O *SOM* foi selecionado em razão de não demandar conhecimento prévio para sua execução e tampouco requerer a intervenção de profissionais, uma vez que se trata de aprendizado não supervisionado. Porém, o principal motivo de sua escolha recaiu pela sua especialidade na formação de clusters, ou agrupamento, principal objetivo deste estudo.

5 - Análise de Clusters

Análise de Clusters é uma função útil na mineração de dados para descobrir grupos e identificar distribuições e padrões de dados. Pode ser entendido como um determinado conjunto de dados em grupos (Cluster), de tal forma que os elementos contidos em um cluster são semelhantes entre si e diferentes dos elementos de outros clusters (Guha *et al.*, 1998).

Para Hair *et al* (1999), é a denominação ao grupo de técnicas que propõe agrupamento de objetos baseados em características próprias. Portanto classifica objetos, por similaridades aos objetos daquele cluster, respeitando critérios previamente definidos. Assim, a situação interna do cluster é homogênea e a situação externa ao cluster é heterogênea.

Em termos gerais, o agrupamento pode servir como um pré-requisito para o processamento de outros algoritmos, tais como a classificação, que posteriormente podem identificar novos clusters.

6 - Proposta de um Método para Classificar as Informações (Confidencialidade)

As regras e critérios para classificar as informações foram baseados na “Ordem Executiva” No. 12.958 – Classificação das Informações de Segurança Nacional, do Departamento de Defesa dos Estados Unidos da América, cuja emissão foi de responsabilidade do presidente Clinton em abril de 1995 (*Department of justice USA;* 2005). Também consideraram a abordagem do *National Institute of Standards and Techonology* (2000 e 2002) e os padrões recomendados pela norma ISO/IEC NBR 17799, que trata especificamente aspectos de segurança da informação.

As informações a serem classificadas devem: Pertencer à própria entidade ou tenha sido produzida sob sua responsabilidade ou, seja destinada para sua utilização; Fazer parte ou apresentar relevância equivalente a uma ou mais Categorias de Risco.

A classificação da informação deverá atender ao nível de confidencialidade requerido por suas características e pelo contexto ao qual está inserida. Para este estudo, foram criados três níveis de confidencialidade:

- “Altamente Secreta” será aplicado à informação, cuja divulgação desautorizada pode causar danos excepcionalmente graves à segurança;
- “Secreta” será aplicado à informação, cuja divulgação desautorizada pode causar danos à segurança;

- “Interna” será aplicado à informação, cuja divulgação desautorizada pode causar danos administráveis pelo gerenciamento de risco.

Há circunstâncias que justificam alterações sobre a classificação original da informação ou sua atribuição dos níveis de confidencialidade. Poderá existir uma data ou um evento específico que determine a re-classificação da informação. Por exemplo, as Demonstrações Financeiras de uma empresa de capital aberto, possui o nível de confidencialidade mais alto até à véspera de sua publicação, entretanto, a partir dessa data torna-se uma informação pública.

A perspectiva de uma re-classificação, quanto a Confidencialidade da informação, será indispensável para a atribuição dos pesos destinados ao treinamento da rede neural, pois elevará o valor do referido peso, devido à necessidade de optar pelo maior nível de impacto decorrente da exposição indevida daquela informação.

Categorização do Risco

As Categorias de Risco representam as circunstâncias em que dados ou informações estão sujeitos às ameaças e denota uma perspectiva de risco, o que compreende a possibilidade de materialização e os impactos motivados pela divulgação não autorizada das informações. Abrange os impactos legais, operacionais, financeiros ou de propriedade intelectual.

As Categorias de Risco foram baseadas na Classificação dos Riscos Operacionais sugeridos pelo *Bank International of Settlements*, para elaborar os riscos. decorrentes de falhas de controle no ambiente computacional, que causam impacto nas atividades de negócio. O *National Institute Standards and Technology* complementou com a descrição do Risco Técnico, que integra o risco técnico (infra-estrutura de TI) e do risco de negócio.

As definições das Categorias de Risco, que direcionam a formulação deste método, estão vinculadas aos estudos do gerenciamento de riscos, dos impactos, ameaças e vulnerabilidades. Assim, a formulação de uma tabela destinada a auxiliar na definição das Categorias de Risco, foi originada pela integração do: Fatores de Risco e Eventos de Perda (Adaptado pelo autor *BIS*, 2001); Descrição dos Tipos de Eventos de Risco Operacional (Adaptado pelo autor *BIS*, 2001); Ameaças humanas: Origem, Motivação e Ações (Adaptado do *NIST*, 2002) e Vulnerabilidades e Ameaças (Adaptado do *NIST*, 2002). Esta consolidação originou a Tabela 2.

TABELA 2: DEFINIÇÃO DAS CATEGORIAS DE RISCO

Objetivo da Ameaça	Motivação	Fatores de Vulnerabilidade	Ações	Categoria de evento	Definição da Categoria
Hacker e Cracker	Desafio Ego Rebelião	• Pessoas • Sistemas • Processos • Eventos Externos	Hacking Engenharia social Quebra dos sistemas de proteção de Intrusões Acesso desautorizado aos sistemas	• Fraudes internas	• Perdas devidas a atos com intenção de defraudar a instituição, violar regulamentos, a lei ou política interna (exclui discriminação), que envolvam ao menos uma parte interna.
Crime computacional	Destruição da informação Divulgação ilegal da informação Ganho financeiro Alteração ilegal de dados		Crime computacional Ato fraudulento Suborno Trapaça Sistema de intrusão	• Fraude externa	• Perdas devidas a atos com a intenção de defraudar a instituição, violar regulamentos, lei ou política interna (exclui discriminação), que sejam cometidos por uma terceira parte.
Terrorismo	Blackmail Destruição Explosão Vingança		Bombas terroristas Informações de guerra Ataque aos sistemas de informação Sistemas de falsificação	• Práticas empregatícias e segurança no ambiente de trabalho	• Perdas devidas a atos inconsistentes com as condições empregatícias. Violações de acordos sanitários ou de segurança trabalhista ou perdas com danos de acidentes de trabalho ou de ações de discriminação de qualquer tipo (inclui assédio sexual).
Espionagem industrial (empresas, governos etc.)	Vantagem competitiva Espionagem econômica		Exploração econômica Roubo de informação Intrusão à privacidade pessoal Engenharia social Penetração nos sistemas Acesso não autorizado aos sistemas	• Clientes, produtos e práticas de negócio	• Perdas oriundas de falhas em cumprir obrigações com clientes ou perdas por causa de desenhos/estruturas de produtos.
Ações internas (treinamento insuficiente, descontentamento, maldade, negligência, desonestidade ou funcionários demissionários)	Curiosidade Ego Inteligência Ganho monetário Vingança Erros não intencionais e omissões (erros de sistemas e falhas de infra-estrutura)		Assalto ou ataque por funcionários *Blackmail* Pesquisa sobre a propriedade das informações *Abuse computer* Fraude e roubo Suborno Input de informações falsas ou corrompidas Interceptação Códigos maliciosos Venda de informações pessoais Erros de sistemas Intrusão nos sistemas Sabotagem nos sistemas Acesso não autorizado aos sistemas	• Danos a ativos físicos • Interrupção de negócios e falhas nos sistemas • Execução, entrega e gestão de processos	• Perdas oriundas de danos a ativos físicos. • Perdas devidas a qualquer interrupção do negócio ou falhas em sistemas • Perdas oriundas de falha no processamento de transações, ou gestão de processos, de relações com parceiros comerciais e vendors

Fonte: Adaptado do BIS - sombreado (2001) e NIST (2002)

Treinamento da Rede Neural Aplicado ao Método

Inicialmente, é necessário contextualizar o objeto da classificação, Neste método, corresponde à seleção das informações. Estas devem ser representativas, quanto aos produtos, processos, atividades e entidades envolvidas. Posteriormente, são definidos os parâmetros de avaliação, representados pelas Categorias de Risco, que podem ser baseadas na Tabela 2.

As informações e as Categorias de Risco serão correlacionadas, originando os dados de entrada para o treinamento da rede neural. A Correlação compreende o conhecimento e experiência de profissionais que desempenham atividades vinculadas à segurança, através do preenchimento de uma planilha que atenderá aos seguintes critérios:

- Cada informação será associada a todas as Categorias de Riscos, considerando a hipotética divulgação indevida da informação. Para mensurar os possíveis impactos;
- Algumas informações podem ter alterado o requerimento de segurança, quando houver um evento ou prazo pré-estabelecido que assim o determine. Neste caso será considerada perspectiva de maior severidade;
- A representação do nível de impacto, a ser preenchido na planilha de correlação, será dada pela atribuição de valores inteiros, entre zero e dez, onde o maior valor representa o maior impacto e zero o menor impacto;
- O preenchimento pode relacionar cada Informação com todas as Categorias de Risco.

Após o preenchimento pelos profissionais, os dados serão consolidados em uma única versão através da média aritmética simples. Este resultado representa os dados de entrada para o treinamento da rede neural.

Para o desenvolvimento da rede *SOM*, Kohonen (2001) recomenda a utilização de softwares que tenham sido submetidos à ampla quantidade de testes para obter maior segurança dos resultados. Por isso, o software *Matlab* versão 6.5, complementado pelo recurso *SOM Toolbox for Matlab 6.5*, para o pré-processamento das informações, a geração inicial dos dados, uso de diversas topologias, na visualização e na análise das propriedades e das informações produzidas (Vesanto *et al*., 2000).

O produto alcançado pelo processamento obtido visualizou os agrupamentos das informações em grupos, por intermédio do recurso *SOM Toolbox* denominado U-Matriz, através da coloração das células ou hexágonos que identifica as distâncias Euclidianas Quadráticas, por conseqüência os grupos. Também é gerada uma tabela que mostra os hexágonos que possuem mais de um elemento e sua posição no plano reticulado.

Posterior a geração dos grupos, inicia-se a análise quantitativa, para proceder à distribuição dos grupos gerados em três níveis de Confidencialidade dessas informações, ou seja, Altamente Secreta, Secreta e Interna.

A análise numérica considerará, para cada grupo, a média dos pesos atribuídos aos seus componentes (informações), onde os grupos com as maiores médias demandam maior proteção quanto ao sigilo.

Contudo a atribuição do nível “Altamente Secreta” requer ao menos a pontuação média equivalente a cinco. Embora o valor cinco tenha considerado a média do intervalo de pesos que foram atribuídos, não apresenta fundamentação técnica, portanto, está sujeito a alterações evolutivas à medida que o método seja utilizado.

7 - Estudo de Caso para Informações Acadêmicas

O Estudo de Caso visa aplicar o método descrito em universidades, pois contemplam todas as atribuições de cunho administrativo e também as atividades pedagógicas e de pesquisa científica. Além disso, possui particularidades quanto ao perfil de seu público, as características de risco e dos impactos são diferenciados. Atendem à legislação específica do Ministério da Educação e outras normas voltadas à pesquisa cientifica.

Foi preciso levantar as principais rotinas, fluxos da informação e atuação dos principais agentes (funcionários, clientes, acionistas, governo etc.) para identificar as Categorias de Risco e as Informações destinadas ao treinamento da rede neural.

Categorização dos Riscos das Universidades

As Categorias de risco abrangem os impactos legais, operacionais, financeiros ou de propriedade intelectual, e podem referenciar-se na Tabela 2.

As funções pedagógicas e de pesquisa foram obtidas através da documentação dos procedimentos registrados no padrão ISO 9000, da entidade de ensino Instituto Paulista de Ensino e Pesquisa. A este trabalho, foi acrescida a indagação realizada junto a alguns professores e coordenadores de curso. Assim, foram criadas 28 Categorias de Risco, apresentadas na Tabela 3.

As questões de cunho estratégico, de imagem institucional e infra-estrutura de TI também compuseram a categorização utilizada, assim como eventuais ameaças que possam afetar a sociedade.

Seleção das Informações das Universidades

Considerou as atividades cotidianas de uma universidade, divididas em grupos de afinidade e, em cada grupo, relacionou as mais representativas, que alcancem a maioria dos produtos, processos, atividades e agentes.

Também foi usada a documentação dos processos no padrão ISO 9000 da entidade de ensino Instituto de Paulista de Ensino e Pesquisa, complementada com dados da Plataforma Lattes (Plataforma Lattes, 2007). O resultado deste levantamento deu origem às 35 informações, distribuídas em:

base cadastral (sombreadas de 1 a 11), pesquisa cientifica, atividades pedagógicas (sombreadas 17 a 22), informações administrativas e sobre a infra-estrutura de TI (sombreadas 33 a 35), retratadas no Tabela 4.

Tabela 3: Categorias de Risco

Categorias de Risco (Coluna)

1	Causaria impacto à segurança ou controle da Universidade
2	Causaria impacto na especificação de dispositivos de segurança
3	Causaria impacto à propriedade intelectual
4	Serve para transgressão legal
5	Serve para transgressão da ordem social
6	Serve para transgressão da integridade de pessoas
7	Serve para transgressão normativa
8	Serve para lesar a instituição ou o acionista
9	Serve para lesar o aluno
10	Serve para lesar os fornecedores
11	Serve para lesar Órgão regulador/fiscalizador
12	Causaria impacto operacional ao aluno
13	Causaria impacto operacional ao professor
14	Causaria impacto operacional ao funcionário
15	Causaria impacto operacional à universidade
16	Causaria impacto financeiro para o aluno
17	Causaria impacto financeiro para o professor
18	Causaria impacto financeiro para o funcionário
19	Causaria impacto financeiro para a universidade
20	Causaria impacto legal para o aluno
21	Causaria impacto legal para o professor
22	Causaria impacto legal para o funcionário
23	Causaria impacto legal para a administração da universidade
24	Causaria impacto na infra-estrutura de TI da universidade
25	Causaria impacto no plano estratégico da universidade
26	Causaria impacto no plano acadêmico de pesquisa
27	Causaria impacto no plano pedagógico
28	Causaria impacto à imagem institucional da universidade

Tabela 4: Informações Selecionadas

Informações Selecionadas (Linha - P)

1	Histórico e currículo professores
2	Histórico e currículo funcionários
3	Histórico e currículo alunos
4	Perfil ingresso aluno
5	Perfil egresso aluno
6	Histórico de avaliações Discentes
7	Histórico de avaliações Docente
8	Histórico de avaliações Curso e disciplinas
9	Histórico de avaliações Infra-estrutura
10	Histórico de avaliações Órgãos Fiscalizadores
11	Histórico de avaliações atividades diversas
12	Pesquisa de softwares de segurança
13	Pesquisa de recursos de criptografia
14	Pesquisa de produção de energia
15	Pesquisa de desenvolvimento de armas
16	Pesquisa que tabula informações de empresas
17	Objetivos Cognitivos da Universidade
18	Objetivos do curso
19	Ementa dos cursos
20	Quadro de competências
21	Quadro de bases tecnológicas
22	Quadro de pré-requisitos
23	Marketing – Campanhas publicitárias
24	Marketing – Verba orçamentária de publicidade
25	Cont – Demonstrações Contábeis Financeiras
26	Cont – Fluxo contábil
27	Fin – Disponibilidade e movimentação recurso
28	Fin – Lançamentos financeiros
29	Fiscal – Plano estratégico tributário
30	Fiscal - Informações para Órgãos Reguladores
31	RH – Proventos do corpo diretivo
32	RH – Dados sobre a folha de pagamento
33	Sist. Inf. – Parâmetros criptográficos
34	Sist. Inf. – Topologia da rede local
35	Sist. Inf. – Plano de contingência

O critério de seleção contemplou informações que podem alterar sua condição de confidencialidade, por um evento ou uma data específica.

Correlação entre as Categorias de Risco e as Informações para o Treinamento da Rede Neural

Os pesos obtidos na correlação traduzem o conhecimento de profissionais envolvidos nas atividades concernentes à segurança ou afins, no ambiente acadêmico. Para alcançar esse objetivo foram acionados quatro representantes de setores de controle, segurança, qualidade e coordenação de universidades diferentes.

Estes profissionais preencheram a planilha que representa a correlação das Categorias de Risco (colunas) e as Informações (linhas), com valores entre 0 e 10, crescente na razão do aumento da severidade pela

hipotética divulgação da informação. O resultado foi consolidado pela média aritmética simples, que originou a Tabela 5.

Tabela 5: Correlação das Categorias de Risco e as Informações

Correlação para Treinamento da Rede

L\C	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	5,5	3,5	2	3	2	5	3,5	3	1,5	1,5	1,5	3,5	5,5	1,5	4,5	1	4,5	1	4,5	1	4,5	1,5	4,5	2,5	5	6	5	4,5
2	3,5	1,5	1,5	5	3	3	3,5	3	1,5	2	3,5	4	2,5	4	5	1,5	3	5,5	4,5	1,5	2,5	5	4,5	1,5	4	1,5	1,5	2,5
3	5	2	4	3	2	3	3,5	3	5,5	3,5	5,5	6	1,5	1,5	4	4,5	3,5	3	6	6	3,5	1,5	7	4	6,5	5,5	6	7,5
4	3,5	3,5	5	4	1	2,5	3,5	3	7,5	1,5	1,5	3,5	5	1,5	5	2	1,5	1,5	1,5	3,5	2	1,5	3	2	3,5	5	5	8
5	4	2	4	4,5	1	3,5	3,5	3	5,5	2	2	6	4	2	4,5	2,5	1,5	1,5	5,5	5	1,5	4	5	1,5	4	4	5	9
6	3,5	1,5	5,5	7,5	2	2,5	5	5,5	6,5	1,5	2	7	6,5	2	6,5	4,5	5	1,5	6,5	5	4,5	1	7,5	1,5	5	5,5	5	7,5
7	3,5	2	5,5	7,5	2	2,5	5	4,5	3,5	2	3	5,5	7	1,5	8,5	3,5	8	1,5	6,5	3,5	7	1,5	5	1,5	7,5	7,5	7	5
8	5	5,5	5	9	4	2,5	5	5,5	5,5	2	3,5	6,5	6	1	9	3	3	3	4,5	5	1,5	1,5	9	3,5	6,5	7	7,5	9
9	4	7	4	4	4	6	6,5	6,5	4,5	4,5	4,5	6	4,5	4,5	5,5	3,5	3,5	4	4,5	4	3,5	3,5	4,5	7	4,5	3,5	4	6,5
10	5	4	3,5	4	4,5	4	6,5	6	3,5	3,5	5	6	4	3,5	5	4	4	4	5	3,5	3,5	3,5	4,5	4	6	5	5	6,5
11	1,5	2	1,5	1,5	2	1,5	4,5	2,5	1,5	1,5	1,5	2	2,5	2	2	1	2	1,5	2,5	1	1,5	1,5	2,5	1	2,5	1,5	2,5	4
12	5	4,5	3	3,5	3	3	3,5	5	2,5	3	1	1,5	1,5	1,5	2	1,5	1,5	1,5	3	2	2	1	3,5	5	2,5	2,5	1,5	3,5
13	4,5	5	3	2	2	2,5	4	4,5	2	1	1	1,5	1,5	1,5	2	1,5	1,5	1,5	2,5	1,5	2	1	2,5	4	2	2,5	1,5	2,5
14	4,5	4,5	3	5	5	5,5	5	4,5	2,5	1,5	1	2,5	1,5	1,5	2	1,5	1,5	1,5	2,5	2	2	1	3,5	4	3	2,5	1,5	3
15	5,5	4,5	3	5,5	5,5	5,5	4,5	4,5	2,5	1	1	5	5	5	5,5	1,5	1,5	1,5	5,5	5	3	1	5,5	4	4	2,5	1,5	5,5
16	5	4,5	3	5,5	4	3,5	5,5	5,5	2,5	2	1	2	2,5	2,5	3	1,5	1,5	1,5	4,5	5	3,5	1	5,5	4	2,5	2,5	1,5	5,5
17	3	1,5	3,5	1,5	1,5	1,5	4	4,5	4	2	1,5	4,5	4	2	5	2	1,5	1,5	3	1	1,5	1	2,5	2	5	5	8,5	8
18	3	3,5	5	1,5	1,5	1,5	6	3	3	1	1	4,5	4,5	1	4,5	3	3	1	4,5	3	3	1	1	1	5	5	8,5	8
19	4,5	5	5	1,5	1,5	1,5	4	1,5	3	1	1	5	5,5	3	5	4,5	3,5	1	3,5	3	3	1	3	3	5	5	8,5	8
20	3	3,5	3,5	3,5	1,5	1,5	4	1,5	3	1	1	4,5	5	1	3,5	3	5	1	5	1	3	1	3	1	5	5	8,5	8
21	4,5	5,5	5	2	1,5	1,5	3,5	2,5	1,5	1,5	1	5	5,5	5	5,5	3,5	1,5	1	5,5	1	1,5	1,5	2	5,5	5	5	8,5	8
22	1	1	3,5	1,5	1	1,5	3,5	1,5	1	1	1	7	4	4	4	1,5	1,5	1	1,5	1	1	1	1,5	1,5	4	3,5	7	6,5
23	7	7	5,5	6	4,5	1,5	3	4	5,5	2,5	2	4,5	4	5,5	4,5	2	1,5	1,5	7,5	3,5	1,5	1,5	6,5	4	9	3,5	3,5	4,5
24	3,5	4,5	3	2,5	2,5	2	4,5	7	2	2	3	1,5	2	3,5	6	2	2	2	7	1	1	1	2,5	3,5	8	4	3,5	4,5
25	6,5	7	1,5	9	2	2	7	8	1,5	4	5	1	1	1	5,5	1,5	1,5	1,5	6,5	1	1	1	9	2	6	4	4	6
26	5,5	6	1	6	1,5	1,5	6,5	7,5	1,5	3,5	5	1,5	1,5	1,5	5,5	1,5	1,5	1,5	6	1	1	1	6	1,5	6	3,5	3,5	6
27	5,5	6	1	6	2,5	3	6	8,5	2	3,5	5	2	2,5	2	6	1,5	1,5	1,5	7	1	1	1	6,5	1,5	5,5	4	3,5	6
28	5,5	6	1	8,5	2,5	3	6	8,5	1,5	5	5	2	2	2	6	1,5	1,5	1,5	5	1	1	1	8,5	1,5	5,5	4	3,5	6
29	5	6	1	7,5	1,5	2	6	8	1,5	4,5	5	1,5	1,5	1,5	6	1,5	1,5	1,5	6,5	1,5	1,5	1	8,5	1,5	8	4	3,5	6
30	5	6	1,5	8	2	2	7,5	8,5	4	5	7	2,5	2,5	2,5	7	2,5	2,5	2,5	8	2,5	2,5	2,5	9	1,5	6,5	4,5	4	6
31	9	6	1	9	5	5,5	7,5	9	1	3	4,5	1	1	1	6	1,5	1,5	1,5	6,5	1	1	1	9	1,5	5	3,5	3,5	6
32	8,5	6	1	9	5	5	7,5	8	2	4	4,5	1,5	2	2	5,5	1	2	2	6	1	2,5	2,5	9	1,5	5	3,5	3,5	6
33	8	6	6,5	5,5	3,5	2	5	5	3,5	3,5	1,5	4,5	4,5	5	7,5	4,5	2	2	7,5	2,5	2,5	2,5	5	8	7,5	3,5	3	7,5
34	8	9,5	7	3,5	3	3	4,5	3	3,5	3,5	2	4,5	6,5	6,5	7	4,5	2,5	2,5	7,5	2,5	2,5	2,5	5	8	6,5	3,5	3	7,5
35	8	9,5	5,5	2,5	2	2	3,5	4,5	3,5	4	3	7	7	7	5,5	5	2,5	2,5	8,5	2	2	2	4,5	9	6,5	4	6	7

Fonte: O próprio autor

Desenvolvimento do *Script* em *Matlab* e Treinamento da Rede Neural

Os valores da Tabela 5 são a entrada para o programa de treinamento da Rede Neural.

Na etapa preliminar ao processamento, a normalização dos dados seria necessária para equalizar os valores que retratam os espaços das variáveis, com a finalidade de evitar distorções no computo da distância Euclidiana. Em face da formatação da planilha de correlação (inteiros e limitados ao intervalo de zero a dez), não foi necessário submetê-los à função de normalização do *SOM Toolbox*.

O treinamento da rede inicia-se pela determinação do neurônio vencedor, através do processo competitivo.

No processo de aprendizado cooperativo os neurônios de saída estão topologicamente próximos, ativam-se reciprocamente para aprender com as informações de entrada (Kohonen, 2001). O neurônio vencedor no processo

competitivo determina o centro desta vizinhança topológica de neurônios cooperativos (Haykin, 2001).

O algoritmo para a criação do *SOM* utiliza o modelo de entrelaçamento do vetor hexagonal, pois tem um melhor resultado visual (Kohonen, 2001) e os seis vizinhos do neurônio têm a mesma distância. O processo de "inicialização linear" usa os dados da Tabela 5, o que permite convergir mais rapidamente (Kohonen, 2001; Haykin, 2001). Os mesmos dados realizaram os processos de Competição, Cooperação e Adaptação Sináptica, e formação do *SOM* (Haykin, 2001).

Visualização dos Resultados

Por intermédio do recurso U-Matriz, foram gerados os *Labels* ou hexágonos apresentados na Figura 1, onde estão destacadas, em letras vermelhas, as Informações Selecionadas identificadas pelos respectivos números, mencionados na Tabela 4, acrescido da letra *P*. Também apresenta as respectivas matrizes de intensidades de cores, onde os hexágonos com a menor distância Euclidiana estão representados pela cor azul (concentração). Enquanto que os vermelhos mostram as maiores distâncias entre os demais hexágonos em sua volta (dispersão). A proporção das Distâncias Euclidianas pode ser analisada pelas cores da barra de escala vertical apresentada na Figura 2. Onde, as informações com as características de risco de maior semelhança, foram delimitadas em seis grupos.

Estas se destacaram devido a sua maior influência para a formação dos clusters, obtidas por intermédio do processamento do *Best Matching Unit* (BMU), pois criaram o ponto de atração dos neurônios vencedores, que norteou o cálculo da distância Euclidiana.

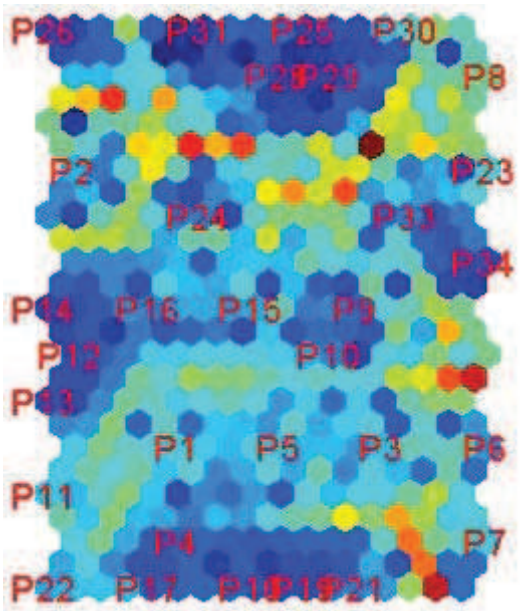

Figura 1: Produto Inicial do Treinamento

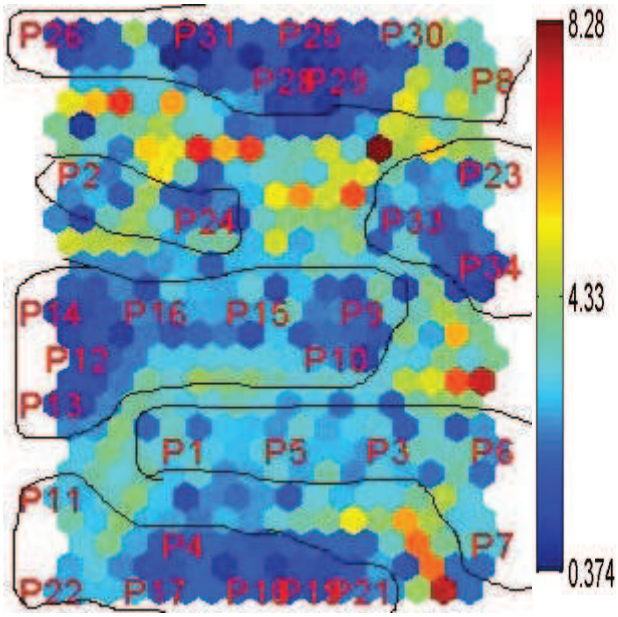

Figura 2: Produto Agrupado do Treinamento e Escala

Note-se que algumas informações não foram retratadas nas Figuras 1 e 2 devido ao fato de ocuparem o mesmo hexágono de outras informações. A Tabela 6 mostra as faltantes, as compartilhadas e a posição do hexágono.

Tabela 6: Informações Agrupadas no Treinamento

Número do Hexágono	Informação Apresentada	Informação Agrupada
1	P26	P27
7	P12	P14
27	P31	P32
52	P18	P20
53	P25	P28
78	P19	P21
110	P34	P35

Fonte: O próprio autor

A primeira coluna da Tabela 6, “Número do Hexágono”, mostra a posição do hexágono no plano reticulado e auxilia a visualização da distância física entre as informações. A coluna “Informação Apresentada” relaciona aquelas mencionadas nas Figuras 1, 2 e 3 e, a última coluna mostra as informações que não foram apresentadas anteriormente, por ocuparem o mesmo hexágono das informações da segunda coluna.

A Figura 3 apresenta os *Labels*, sem as cores das distâncias Euclidianas, mostra as sobreposições (quantidade de informações no Hexágono), entre os parênteses. A Figura 4 expõe todas as posições do plano reticulado de maneira a facilitar a visualização e o posicionamento dos hexágonos.

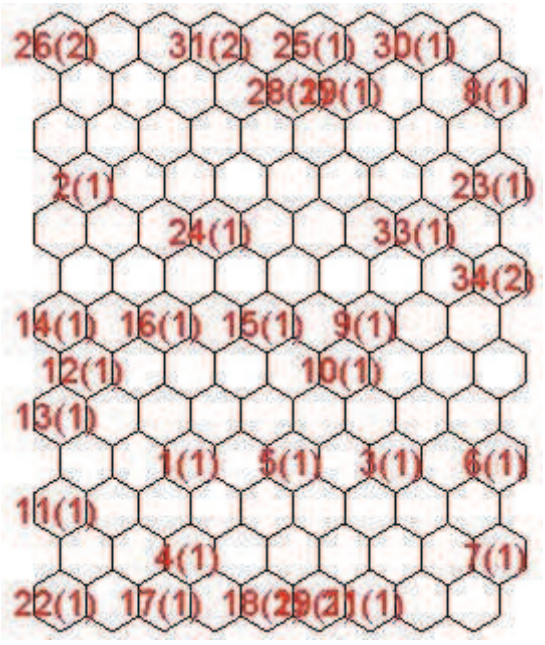

Figura 3: *Labels* e Sobreposição

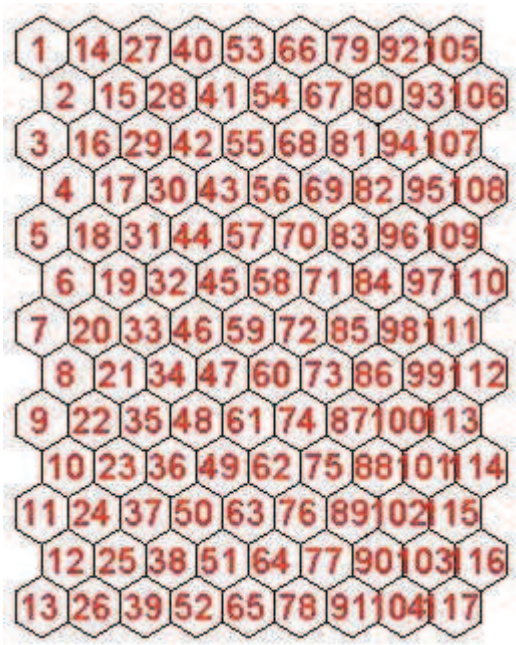

Figura 4: Ocupação do Plano Reticulado

Estes recursos permitiram agrupar as informações, dando origem aos clusters representados por letras, conforme a Tabela 7.

Tabela 7: Formação dos Clusters

Clusters	Ref.	Informação
A	1	Histórico e currículo professores
A	3	Histórico e currículo alunos
A	5	Perfil egresso aluno
A	6	Histórico de avaliações Discentes
A	7	Histórico de avaliações Docente
B	2	Histórico e currículo funcionários
B	24	Marketing – Verba orçamentária de publicidade
C	4	Perfil ingresso aluno
C	11	Histórico de avaliações atividades diversas
C	17	Objetivos Cognitivos da Universidade
C	18	Objetivos do curso
C	19	Ementa dos cursos
C	20	Quadro de competências
C	21	Quadro de bases tecnológicas
C	22	Quadro de pré-requisitos
D	8	Histórico de avaliações Curso e disciplinas
D	25	Cont – Demonstrações contábeis Financeiras
D	26	Cont – Fluxo contábil
D	27	Fin – Disponibilidade e movimentação recurso
D	28	Fin – Lançamentos financeiros
D	29	Fiscal – Plano estratégico tributário
D	30	Fiscal – Informações para Órgãos Reguladores
D	31	RH – Proventos do corpo diretivo
D	32	RH – Dados sobre a folha de pagamento
E	9	Histórico de avaliações Infra-estrutura
E	10	Histórico de avaliações Órgãos Fiscalizadores
E	12	Pesquisa de softwares de segurança
E	13	Pesquisa de recursos de criptografia
E	14	Pesquisa de produção de energia
E	15	Pesquisa de desenvolvimento de armas
E	16	Pesquisa que tabula informações de empresas
F	23	Marketing – Campanhas publicitárias
F	33	Sist. Inf. – Parâmetros criptográficos
F	34	Sist. Inf. – Topologia da rede local
F	35	Sist. Inf. – Plano de contingência

A apuração do nível de confidencialidade, entre os seis grupos, consistiu no cálculo das médias aritméticas simples de cada cluster, que considerou os valores atribuídos às informações do cluster (Tabela 5), que indica a pontuação relativa aos impactos decorrentes da possível efetivação das ameaças. Os clusters estão ordenados decrescentemente, enunciando a comparação entre as médias obtidas por cada cluster, quanto ao nível de Confidencialidade, no Tabela 8.

Tabela 8: Níveis de Confidencialidade

Clusters	Informação	Média	Confidenc.
F	Marketing – Campanhas publicitárias		
F	Sist. Inf. – Parâmetros criptográficos	5,72	Altamente Secreta
F	Sist. Inf. – Plano de contingência		
F	Sist. Inf. – Topologia da rede local		
D	Cont – Demonstrações contábeis Financeiras		
D	Cont – Fluxo contábil		
D	Fin – Disponibilidade e movimentação recurso		
D	Fin – Lançamentos financeiros		
D	Fiscal – Informações para Órgãos Reguladores	4,94	Altamente Secreta
D	Fiscal – Plano estratégico tributário		
D	Histórico de avaliações Curso e disciplinas		
D	RH – Dados sobre a folha de pagamento		
D	RH – Proventos do corpo diretivo		
A	Histórico de avaliações Discentes		
A	Histórico de avaliações Docente		
A	Histórico e currículo alunos	4,19	Secreta
A	Histórico e currículo professores		
A	Perfil egresso aluno		
E	Histórico de avaliações Infra-estrutura		
E	Histórico de avaliações Órgãos Fiscalizadores		
E	Pesquisa de desenvolvimento de armas		
E	Pesquisa de produção de energia	3,90	Secreta
E	Pesquisa de recursos de criptografia		
E	Pesquisa de softwares de segurança		
E	Pesquisa que tabula informações de empresas		
C	Ementa dos cursos		
C	Histórico de avaliações atividades diversas		
C	Objetivos Cognitivos da Universidade		
C	Objetivos do curso	3,65	Interna
C	Perfil ingresso aluno		
C	Quadro de bases tecnológicas		
C	Quadro de competências		
C	Quadro de pré-requisitos		
B	Histórico e currículo funcionários	3,63	Interna
B	Marketing – Verba orçamentária de publicidade		

Este resultado retrata o produto obtido pelo processamento da Rede Neural *SOM* de Kohonen, adicionado aos cálculos da média apresentados no método proposto neste estudo. A definição pelos níveis de “Altamente Secreta” para os clusters F e D se deve à condição de ambos apresentarem pontuação próxima ao valor cinco, conforme proposto anteriormente.

8 - Conclusão sobre os Resultados alcançados no Estudo de Caso

O propósito deste trabalho foi descobrir os grupos que mantenham similaridade dos seus elementos por intermédio de um padrão voltado às suas características de risco. O processamento da Rede Neural de Inteligência Artificial realizou a tarefa de identificar os grupos que possuem a mesma natureza. Esta similaridade foi norteada pelas Categorias de Risco aplicadas para as informações tratadas no âmbito das universidades. Logo, esta meta inicial foi cumprida.

Contudo, a análise do desenvolvimento, da implementação e dos resultados obtidos no Estudo de Caso nos permite a formulação de algumas considerações que visam à evolução deste método.

As informações destinadas ao treinamento da Rede Neural, quando formuladas estavam distribuídas inicialmente em: base cadastral, pesquisa cientifica, atividades pedagógicas, administrativas e de infra-estrutura de TI. Esta

distribuição envolveu naturalmente um nível de similaridade entre os elementos, uma vez que há poucos hexágonos de coloração vermelha escura, que expõem as maiores distâncias Euclidianas. A seguir, estão apresentadas as similaridades observadas em cada grupo gerado após o treinamento da rede:

- A: composto somente por informações da base cadastral;
- B: duas informações e grupos diferentes;
- C: maioria dos elementos faz parte das atividades pedagógicas;
- D: maior parte das informações pertence às funções administrativas;
- E: maioria dos dados está relacionada à pesquisa cientifica; e
- F: predominância das informações é de infra-estrutura de TI.

Os grupos formados ratificam a pré-existente similaridade parcial entre os seus elementos, embora esta não tenha sido prevista na seleção das informações para o treinamento da rede. Embora se apresente em menor escala, se observa a mescla de informações originadas em atividades diferentes, nos novos grupos formados.

O resultado também abrangeu os riscos decorrentes de todas as atividades desempenhadas pelas universidades, o que envolve a administração, pedagogia e pesquisa cientifica. Estes riscos compreendem as propriedades, características e requerimentos legais inerentes aos processos das universidades, as quais devem ser refletidas na correlação entre as Informações e as Categorias de Risco. O produto dessa correlação é preponderante para formação dos grupos ou clusters e também para a determinação dos níveis de Confidencialidade.

Por essa razão, a formulação da correlação necessita de amplo conhecimento, pelos responsáveis de sua confecção, das diretrizes estratégicas, processos e atividades de uma universidade, acrescidos de vivência e percepção das ameaças existentes e aquelas que possam se materializar e, principalmente dos possíveis impactos às entidades que se relacionam com o ambiente acadêmico. Embora os impactos possam originar-se na infra-estrutura computacional ou nos sistemas de informação, serão contabilizados ou sensibilizarão as atividades operacionais, financeiras, administrativas etc.

No desenvolvimento do Estudo de Caso, se verificou a necessidade de ampliar as explicações e as discussões junto às pessoas entrevistadas, de forma a auxiliar a visão de risco e impacto a cada correlação realizada. Portanto, o sucesso da aplicação do método descrito neste trabalho mantém dependência direta da capacitação e conhecimento dos profissionais que constroem a correlação.

Outro fator relevante na formulação dos pesos na correlação é a possibilidade de alterar a condição de sigilo em face da ocorrência de um determinado evento ou datas pré-estabelecidas. Esta circunstância foi sugerida aos profissionais pesquisados, pois prevaleceu, quando identificada, a demanda de sigilo de maior nível.

O Estudo de Caso demonstrou que a aplicação do método apresentado é factível através dos Mapas Auto-Organizáveis de Kohonen, com a adequada e assertiva categorização dos riscos. Os resultados alcançados pelo treinamento da Rede Neural se mostraram compatíveis nos grupos que foram gerados.

O produto da classificação das informações concernentes ao nível de Confidencialidade observados neste caso prático leva a conclusão que sua aplicação pode estender-se a outros segmentos, na indústria de base, no segmento financeiro, no comércio, organizações não governamentais, governos etc. Dessa maneira atingindo aos objetivos de apresentar um método para classificação das informações que seja baseado no conhecimento e experiência acumulada de profissionais, seja uniforme, sistemático e assertivo.

O emprego deste método traz como benefício à condição de aprimoramento do dimensionamento dos investimentos em segurança e, principalmente, prover os mecanismos de proteção que sejam mais apropriados às características de cada informação, obtendo maior efetividade na mitigação dos riscos dos sistemas de informação e da infra-estrutura computacional.

9 - Trabalhos Futuros

No cenário em que os processos de negócio não conseguem acompanhar a velocidade de inovação tecnológica, quanto a sua aplicabilidade, a própria tecnologia propicia condições que além dos benefícios esperados, também serve de insumo para ameaças e ataques inusitados.

As organizações se preparam com investimentos maciços e crescentes para prover maior proteção para seus ativos, por intermédio de recursos técnicos e metodológicos de segurança, contudo, de eficácia e eficiência questionáveis em razão de não haver instrumentos para medição.

O método de classificação implementado pelo SOM de Kohonen atende a primeira etapa de um processo de segurança porque separa as informações relevantes, que requerem maior proteção, daquelas que possuem características que não demandam a preservação de sigilo. Portanto, doravante surge à perspectiva de explorar as novas etapas de implementação de segurança, o que representa um conjunto de trabalhos que abrangem:

- Desenvolver um método, derivado deste apresentado, que trate simultaneamente todos os objetivos de segurança (confidencialidade, disponibilidade e integridade) de aplicação integrada e maior alcance na mitigação dos riscos;
- Em complemento a implementação deste modelo, sugere-se desenvolver e implantar novos modelos que possam ser Integrados a este primeiro, destinados a direcionar os equipamentos e recursos de segurança à medida do nível de classificação das informações. De forma que as informações definidas como “Altamente Secreta” recebam indicações de uma gama de

dispositivos de segurança mais adequados à sua condição, assim como nos outros dois níveis de confidencialidade;
- Em face da importância da fase de categorização dos riscos, seria necessário e útil o desenvolvimento de método ou mecanismos que auxiliem identificação e quantificação na composição dos riscos.

REFERÊNCIAS BIBLIOGRÁFICAS

BIS-*Bank for International Settlements* (2001). "*Overview of The New Basel Capital Accord*". *Technical report, Bank for International Settlements*, Disponível em: http://www.bis.org, Acesso em: 18/04/2007.

BLAKLEY B., E. MCDERMOTT, and D. GEER. "*Information Security is Information Risk Management*". Communications of the ACM, 2002.

DoJ-Departament of Justice USA (1995), Executive, "*Order 12,958 - Classified National Security Information*", Disponível na Internet <http:www.usdoj.gov> , Acesso em 22/12/2007;

FARAHMAND F., S. B. Navathe, G. P. Sharp, and P. H. Enslow. Managing "*Vulnerabilities of Information Systems to Security Incidents*". Communications of the ACM, 2003.

FERRAIOLO D. F., Kuhn D. R., Chandramouli R., *Computer Security Series Role-Based Access Control,* Artech House Inc, 2003;

FRANCISCO, C. A. C., Rede de Kohonen: Uma ferramenta no estudo das relações tróficas entre as espécies de peixes. Curitiba: Universidade Federal do Paraná, 2004;

GEER D. Jr, Hoo K. S., and Jaquith A. "*Information Security: Why the Future Belongs to The Quants*", IEEE Security & Privacy, July 2003.

GUHA, S., Rastogi, R., and Shim K. (1998). CURE: An Efficient Clustering Algorithm for Large Databases. In Proceedings of the ACM SIGMOD Conference.

HAIR J.F.; Anderson R. E., Tatham R. L.; Black W.C.; "*Multivariate Data Analisys*", EdHaykin, S. Neural Networks: "*Comprehensive Foundation*", New Jersey: Prentice-Hall, 1999.

HAYKIN, S., "Redes Neurais: Princípios e prática", Ed. Bookman, 2001;

ISO-International Organization for Standardization; 2005, Disponível em: <http://www.iso.org>. Acesso em: 22/7/2007.

ITGI-IT Governance Institute, *Framework Control Objectives. Management Guidelines Maturity Models* (2007), Disponível em: <http://www.itgi.org>, Acesso em: 17/04/2007.

KOHONEN, T., "*Self-Organizing Maps*", Berlin: Springer-Verlag, 2001.

KUONG J. F., "*Computer Security, Auditing and Controls*", Management Advisory Publications Series on, 1974;

LÉVY, P. "O Que é Virtual". São Paulo: Ed. 34, 1997.

NETO, Luis Garcia Palma; Nicoletti, Maria do Carmo, Introdução às Redes Neurais Construtivas. São Carlos: Edufscar, 2005;

NIST- *National Institute of Standards and Technology* (2002), "*Risk Management Guide for Information Technology System*", Disponível em: <http://www.nist.gov>, Acesso em 20/10/2007

NIST- *National Institute of Standards and Technology* (2004), "*Standards for Security Categorization of Federal Information and Information System*", Disponível em: <http://www.nist.gov>, Acesso em 24/10/2007

PLATAFORMA LATTES (2007); Disponível em: <http://lattes.cnpq.br>; Acesso em 3/2/2007.

ROUSSINOV, D. G.; CHEN, H., "Information navigation on the web by clustering and summarizing query results", *Information Processing and Management*, 2001

RUSSELL, S. J. and Norvig, P., Inteligência Artificial, 2. ed., Campus, 2004.

SCHNEIER, B., Segurança.com: Segredos e Mentiras sobre a Proteção Digital; Rio de Janeiro; Campus, 2001

STALLINGS, W., *Redes e Sistemas de Comunicação de Dados*, 5ª.edição, São Paulo, Elsevier Editora, 2005

SUURONEN, Tomi. *Java2 Implementation of Self-Organizing Maps based on Neural Networks utilizing XML based Application Languages for Information Exchange and Visualization*. Espoo: Vantaa Institute of Technology Departament, 2001.

VESANTO, J.; Himberg, J.; Alhoniemi, E.; Parhankangas. J. *SOM Toolbox for Matlab 5*, Espoo, Helsinki University of Technology, 2000

WANGENHEIM, Aldo Von. Reconhecimento de Padrões. Florianópolis: Universidade Federal de Santa Catarina. Disponível na Internet em: http://www.inf.ufsc.br/~awangenh/RP/subsimbolicas1.pdf. Acesso em: 23/11/2007