FELIPE NICOLIELLO DERRICO

# Aplicação de técnicas de redes neurais artificiais e mineração de dados para busca de informação na Internet

SÃO PAULO
2010

FELIPE NICOLIELLO DERRICO

# Aplicação de técnicas de redes neurais artificiais e mineração de dados para busca de informação na Internet

Trabalho apresentado ao PECE USP para conclusão do curso de MBA de Tecnologia da Informação.

SÃO PAULO
2010

FELIPE NICOLIELLO DERRICO

# Aplicação de técnicas de redes neurais artificiais e mineração de dados para busca de informação na Internet

Trabalho apresentado ao PECE USP para conclusão do curso de MBA de Tecnologia da Informação.

Área de concentração:
Tecnologia da informação

Orientador: Prof. Stephan Kovach

SÃO PAULO
2010

FICHA CATALOGRÁFICA

**Derrico, Felipe Nicoliello**
**Aplicação de técnicas de redes neurais artificiais e mineração de dados para busca de informação na internet / F.N. Derrico. -- São Paulo, 2010.**
**p.**

**Monografia (MBA em Tecnologia da Informação) - Escola Politécnica da Universidade de São Paulo. Programa de Educação Continuada em Engenharia.**

**1. Redes neurais 2. Mineração de dados 3. Tecnologia da informação 4. Internet I. Universidade de São Paulo. Escola Politécnica. Programa de Educação Continuada em Engenharia II. t.**

# DEDICATÓRIA

Dedico este trabalho à Deus pela vida. Aos meus pais, irmãos, namorada pelo apoio, compreensão e carinho que me deram durante a realização deste trabalho.

# AGRADECIMENTOS

Agradeço primeiramente a Deus, pois sem ele nada seria possível.

Aos meus queridos pais, por sempre terem colocado a minha educação em primeiro lugar, e por sempre terem me apoiado e se sacrificado para que eu pudesse realizar os meus objetivos.

Aos meus queridos irmãos por serem fonte de minha inspiração e apoio.

À minha querida namorada pelo amor e por compreender e apoiar-me nos momentos de dedicação a esta monografia.

Aos meus familiares e amigos pelo carinho e presença sempre constante.

Aos meus mestres do MBA, pelas horas de estudo a nós dedicadas e esforço para uma boa orientação acadêmica e profissional, em especial ao meu orientador, professor Stephan Kovach, pelos ensinamentos, direcionamentos e irrestrito apoio no desenvolvimento deste trabalho.

A todos que colaboraram direta ou indiretamente na execução deste trabalho.

# RESUMO

Hoje em dia o grande volume de dados e informações disponíveis em diversos meios de armazenamento torna a busca de uma informação específica em um desafio a ser superado. Por isso é cada vez maior a necessidade de aperfeiçoar continuamente os mecanismos de busca de informações relevantes nesses bancos de dados, onde estão envolvidos grandes volumes de dados como a Internet.

As Redes Neurais Artificiais com sua capacidade de paralelismo e aprendizado surgem como alternativa para resolução de diversos problemas computacionais. A Mineração de Dados surge também como uma etapa importante para identificar de forma automática os padrões e correlações entre os dados e informações.

Neste trabalho, o conceito de Redes Neurais Artificiais em conjunto com técnicas de mineração de dados é apresentado tendo como objetivo o aperfeiçoamento de busca por informações relevantes. São abordados os principais conceitos e técnicas de Redes Neurais Artificiais e Mineração de dados e como eles podem ser utilizados para aperfeiçoar buscas em bases de dados como as existentes na Internet.

Palavras chave: Redes Neurais Artificiais. Mineração de dados. Internet.

# ABSTRACT

Nowadays the large volume of available data and information in a variety ways of storage turns a search of specific information into a challenge to be overcome. That is why the need for continuous improvement of the relevant information search engines in the databases is growing up, where a large volume of data are involved, as in the Internet.

The Artificial Neural Networks with its capacity of parallelism and learning comes up with an alternative to solve the different computational problems. The Data Mining also arise as an important step to identify in an automatic way the patterns and correlations between data and information.

In this work, the Artificial Neural Networks concept combined with Data Mining techniques are proposed in order to improve the search of relevant information. The main concepts and techniques of Artificial Neural Networks and Data Mining are proposed and how they can be used to improve searches in databases, as exists in the Internet.



Keywords: Artificial Neural Networks, Data Mining, Internet.

# LISTA DE ILUSTRAÇÕES

# LISTA DE TABELAS

# LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| RNA | Redes Neurais Artificiais |
| PDP | Parallel Distributed Processing |
| KDD | Knowledge Discovery in Database |
| URL | Uniform Resource Locator |

# SUMÁRIO

# 1 INTRODUÇÃO

Atualmente o acesso e o compartilhamento do grande volume de informações disponíveis em diversas bases de dados possibilitou para a sociedade um crescimento ainda maior de dados, informações e conhecimento. Essas bases de dados representam desde pequenas bases em computadores pessoais até as de maior amplitude e importância global como a Internet. Mas para se obter resultados satisfatórios na busca dessas informações é necessário aumentar a eficiência na busca e recuperação, assim como, melhorar a qualidade dos resultados encontrados.

As informações podem estar armazenadas de forma organizada e estruturada como em pequenos bancos de dados domésticos ou podem estar disponíveis em grandes conglomerados virtuais como a própria Internet. Existem diversas técnicas que podem ser utilizadas para a busca e recuperação das informações nestes bancos de dados. Dentre as técnicas utilizadas está a mineração de dados que permite a extração de informações das bases de dados de maneira mais automatizada possível e sua posterior análise.

No entanto, a extração de informações relevantes e em tempo satisfatório torna-se um desafio uma vez que os resultados podem ser amplos e irrelevantes para o contexto de busca ou em tempo inviável segundo parâmetros definidos. E é por esse motivo que técnicas de inteligência artificial são consideradas como uma das alternativas para melhorar a eficiência na busca e aumentar a qualidade das respostas, como modelos de redes neurais artificiais para adaptá-los às tarefas a serem resolvidas através de processamento paralelo e capacidade de aprendizagem.

## 1.1 Objetivo

O Objetivo do trabalho é apresentar uma alternativa para otimizar a busca de informações na Internet utilizando técnicas de redes neurais artificiais e mineração de dados.

## 1.2 Estrutura do Trabalho

Este trabalho está estruturado conforme segue:

Capítulo 1 – Introdução

No primeiro capítulo (presente) são apresentadas as considerações iniciais do trabalho, os objetivos, justificativa e a estrutura geral.

Capítulo 2 – Revisão Teórica

No segundo capítulo é apresentada a revisão teórica do trabalho, abrangendo os conceitos de redes neurais artificiais e as técnicas que serão aplicadas. Além disso, traz um estudo sobre mineração de dados, com conceitos e atividades relacionadas.

Capítulo 3 – Aperfeiçoando buscas na Internet com redes neurais e mineração de dados

No terceiro capítulo é realizado o estudo e aplicação das técnicas e métodos de redes neurais artificiais e mineração de dados para aperfeiçoamento de buscas de informações na Internet.

Capítulo 4 – Conclusão

No quarto capítulo são apresentadas as contribuições do trabalho, suas limitações e proposições e as conclusões do trabalho, assim como os trabalhos futuros que podem ser desenvolvidos.

# 2 REVISÃO TEÓRICA

## 2.1 Redes Neurais Artificiais

### *2.1.1 Conceitos Básicos*

Segundo Braga, Carvalho e Ludermir [1] o cérebro humano é responsável pelo que se chama de emoção, pensamento, percepção e cognição, assim como pela execução de funções sensoriomotoras e autônomas. Além disso, a sua rede de conexões entre os neurônios, células básicas que o compõem, tem a capacidade de reconhecer padrões e relacioná-los, usar e armazenar conhecimento por experiência e interpretar observações. Redes Neurais Artificiais também chamadas de redes neuronais artificiais, redes neurais, modelos conexionistas de computação ou sistemas de Processamento Paralelo Distribuído (PDP) [2] são técnicas computacionais, dentro da área de Inteligência Artificial, que se fundamentam nos estudos sobre a estrutura do cérebro humano com o objetivo de simular sua forma inteligente de processar informação, incluindo aprendizado e generalização.

No entanto, segundo Kovacs [3], na literatura científica pode ser identificada como uma classe de modelos matemáticos para solução de problemas de classificação de informação e reconhecimento de padrões, como uma parte da teoria conexionista dos processos mentais (estudo da mente por uma perspectiva computacional) ou uma categoria de modelos em ciência da cognição.

Ainda de acordo com Braga, Carvalho e Ludermir [1], as Redes Neurais Artificiais são modelos matemáticos que se assemelham às estruturas neurais biológicas e que tem capacidade computacional adquirida por meio de aprendizado e generalização.

Alguns estudos da neurofisiologia atribuem ao grande número de neurônios interconectados por uma rede complexa de sinapses, o poder computacional do cérebro. A quantidade estimada de neurônios existentes no cérebro humano é de cerca de $10^{11}$ a $10^{14}$ e cada um destes está conectado através de $10^{3}$ a $10^{4}$

sinapses, em média. Porém, a velocidade de processamento destes neurônios individualmente é baixa em comparação aos computadores atuais.

Algumas características que são importantes para simulação em Redes Neurais Artificiais incluem paralelismo, aprendizagem, robustez e tolerância à falhas e processamento de informação incerta (informação incompleta, afetada por ruído ou parcialmente contraditória).

### *2.1.2 Neurônio Biológico*

As redes neurais são formadas por unidades de processamento, comumente chamadas de nós, neurônios ou células, interconectadas por arcos unidirecionais, também chamados de ligações, conexões ou sinapses. Sinapse é a região onde dois neurônios entram em contato e através da qual os impulsos nervosos são transmitidos entre eles. De acordo com Kovacs [3] o neurônio é delimitado por uma fina membrana, como qualquer célula biológica, que além de suas funções biológicas normais, possui propriedades que são essenciais para suas características elétricas como célula nervosa.

Os neurônios podem ser divididos basicamente por três partes:

- Dendritos: que é um conjunto de terminais que recebem os impulsos de entrada.
- Soma: que é o corpo da célula, onde está armazenada a memória local e onde são realizadas operações de processamento de informação localizada.
- Axônio: é uma única saída da célula que envia impulsos ou sinais de saída. Ela pode se ramificar em muitas ligações colaterais.

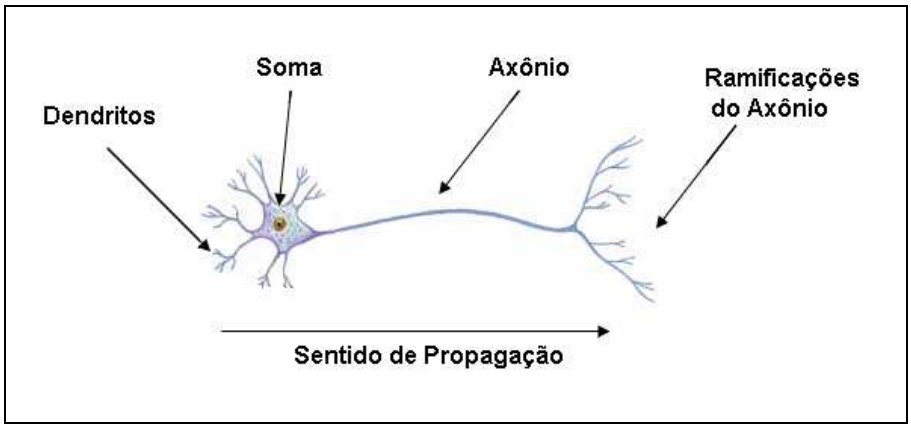


Figura 1 - Esquema geral de um neurônio.

### *2.1.3 O Neurônio Artificial*

As redes neurais artificiais são compostas por várias unidades de processamento de funcionamento simples. Essas unidades de processamento são representações dos neurônios biológicos, isto é, neurônios artificiais. No entanto, devido ao fato do conhecimento atual sobre o neurônio ser relativamente incompleto e nosso poder computacional ser limitado, os modelos existentes são apenas idealizações de redes reais de neurônios.

A figura 2 traz uma representação de um neurônio artificial, com estruturas correspondentes a um neurônio biológico. Os elementos adicionais constantes nesse modelo são a soma das entradas recebidas de outras unidades de processamento e a existência de um valor limite para enviar o sinal através do axônio, respeitando o sentido de propagação que é a direção que o sinal é enviado.

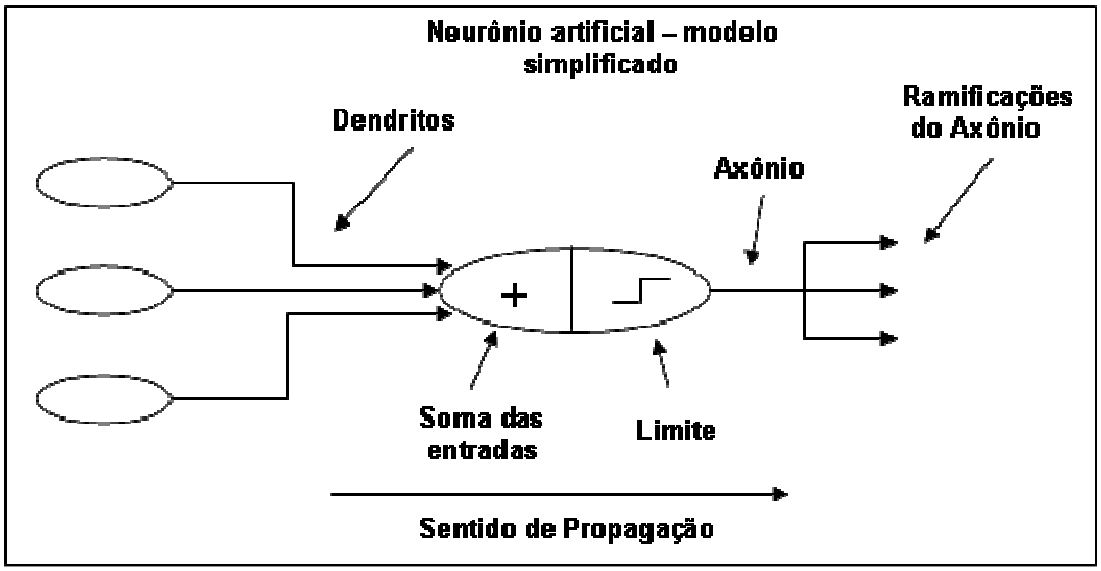


Figura 2 - Representação de um neurônio artificial.

### *2.1.3 Componentes básicos das RNAs*

Uma rede neural artificial consiste de um grupo de unidades de processamento simples (neurônios) que se comunicam através da troca mútua de sinais entre os elementos com um grande número de conexões (sinapses).

De acordo com Rumelhart [2], uma rede neural artificial pode ser descrita por oito elementos principais:

*i) Um conjunto de Unidades de Processamento:* Uma rede neural artificial é formada por um conjunto de unidades de processamento.

*ii) Estado de Ativação*: o estado de ativação das unidades de processamento especifica o que está sendo representado na rede em um determinado instante. O conjunto de valores de ativação de cada unidade de processamento em um determinado instante define o estado de ativação do sistema como um todo. Diferentes modelos possuem premissas distintas sobre os valores de ativação permitidos para uma unidade de processamento. Por exemplo, pode ser que em determinado modelo a unidade de processamento possua valor de ativação restrito aos valores “0” ou “1” onde o valor “1” usualmente significa que a unidade está ativa e “0” que está inativa.

*iii) Função de Saída*: as unidades de processamento se interagem através da transmissão de sinais. Associada a cada unidade de processamento há uma função de saída que mapeia (transforma) o estado de ativação corrente para um sinal de saída. A função de saída é assumida como um tipo de função limite que possibilita uma unidade de processamento afetar outra unidade de processamento (transmitindo um sinal) desde que o seu estado de ativação exceda um valor mínimo estabelecido.

*iv) Padrão de Interconexão*: representa as conexões entre as unidades de processamento. No caso de um neurônio natural seriam as sinapses. Geralmente cada conexão é definida por um peso, o qual determina o efeito que o sinal de saída de uma unidade tem sobre as entradas das unidades vizinhas. É esse padrão de interconexão que constitui o que um sistema sabe e determina como ele irá responder para uma determinada entrada, ou seja, representa o “conhecimento” de uma rede neural. Em sistemas biológicos o aprendizado envolve ajustes nas conexões sinápticas que existem entre os neurônios. Em redes neurais artificiais é o ajuste nestes padrões de interconexão que determinam o aprendizado. Essas conexões podem ser positivas ou excitatórias, que indicam o reforço na ativação de um neurônio, e negativas ou inibitórias, que indicam inibição na ativação do neurônio. Conexões excitatórias e inibitórias podem ser ainda de diferentes tipos. Porém, em muitos casos assume-se que cada unidade provê uma contribuição aditiva para as entradas das unidades aos

quais estão conectadas. Desta forma, o valor total da entrada para uma unidade é simplesmente a soma ponderada dos sinais recebidos das unidades de processamento vizinhas.

*v) Regra de Propagação*: é uma regra que determina a efetiva entrada em uma Unidade de Processamento a partir de suas entradas externas recebidas, ou seja, a combinação das entradas recebidas (sinais) das unidades de processamento conectadas a ela na rede através dos padrões de interconexão. Por exemplo, caso uma Unidade de Processamento receba sinais diferentes de três unidades de processamento vizinhas através dos padrões de interconexão, a combinação desses três sinais irá gerar um valor que será internalizado na unidade de processamento, após passar por uma regra de propagação.

*vi) Regra de Ativação*: é a regra que combina as entradas em uma unidade de processamento passando pela regra de propagação, com o estado de ativação atual da unidade, para produzir um novo valor de ativação para a unidade. Além de substituir o valor atual do estado de ativação, esse novo valor irá gerar um sinal de saída caso exceda o limite de ativação da unidade de processamento. Nos casos mais simples, a regra de ativação e a regra de propagação podem ser igualadas, isto é, existe apenas uma regra. Em geral, no entanto, a regra de ativação é uma função determinística. Por exemplo, se o limite de ativação de uma unidade de processamento é igual a “1” para emitir o sinal e “0” para não emitir o sinal, o valor recebido pela regra de propagação somada ao estado de ativação atual é que irá definir esse valor através da regra de ativação.

*vii) Regra de Aprendizado*: é a modificação do processamento ou do conhecimento de uma rede neural que envolve a alteração do seu padrão de interconexão (pesos). Indica como os padrões de interconexão são alterados pela experiência. Em princípio isto pode ser desenvolvido através de três tipos de modificações:

*a.* O desenvolvimento de novas conexões: um padrão de interconexão entre unidades de processamento inicialmente com peso zero se for alterada para peso “1”, por exemplo, significa desenvolver essa conexão.

*b.* A perda de conexões existentes: um padrão de interconexão entre unidades de processamento com peso diferente do valor zero, se for alterada para um peso zero, significa a perda da conexão entre as unidades.

*c.* A modificação dos pesos das conexões já existentes: os dois itens anteriores podem ser simulados através desse item, que também abrange a alteração dos pesos entre as unidades de processamento para valores intermediários.

*viii) Ambiente*: é o ambiente onde a rede deve funcionar. O ambiente irá fornecer os sinais de entrada e, se necessário, sinais de erros quando uma resposta for diferente da esperada (em uma fase de treinamento, por exemplo).

A figura 3 apresenta os elementos básicos de uma rede neural artificial em um determinado ambiente. Há um grupo de unidades de processamento (*u*), sendo que cada uma contém um valor de ativação (*a*), que é o estado de ativação da unidade. Esse valor de ativação é passado por uma função de saída (*f*) para produzir um valor de sinal de saída. Este valor de sinal de saída pode ser transmitido para o grupo de unidades de processamento que fazem conexão com a primeira. Para cada padrão de interconexão (*p1, p2.. pn*) entre as unidades de processamento é associado um peso ou força da conexão que determina o efeito que a primeira unidade tem sobre a segunda. Todas as entradas devem ser então combinadas por um operador (geralmente de adição) passando para a Regra de Propagação (*r*) que combinado com o valor atual de ativação determina, via Regra de Ativação (*F*), um novo valor de ativação. Caso este valor exceda a função de saída (f) irá repassar o sinal para as próximas conexões, se existirem.

Estes sistemas são vistos como flexíveis de modo que o padrão de interconexão não é fixo o tempo todo. Em cada sistema, a alteração dos valores desses padrões depende da regra de aprendizado utilizada, e é desta maneira que o sistema pode aprender. Todo esse processo acontece dentro de um determinado ambiente, tendo uma regra de aprendizado associado à Rede Neural Artificial.

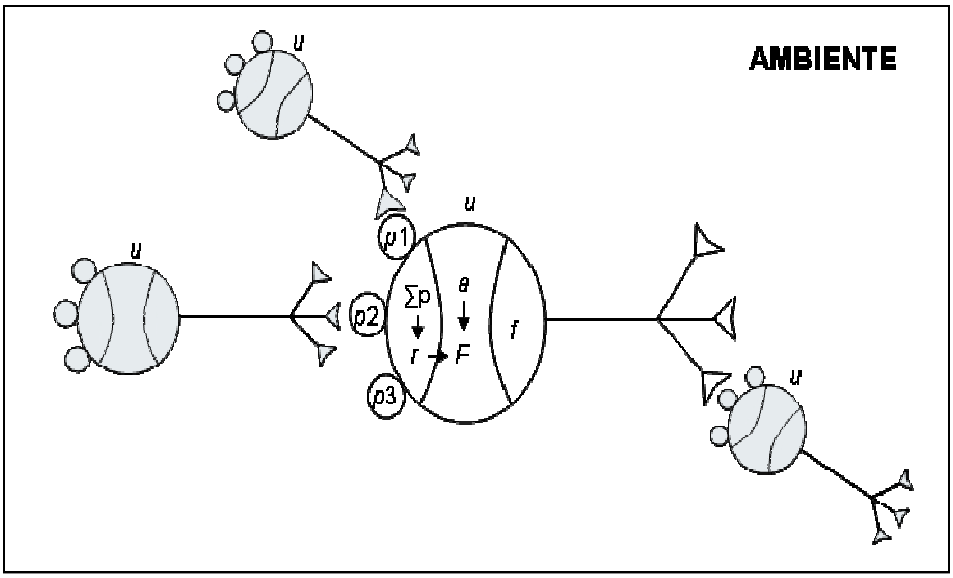


Figura 3 – Os componentes básicos de uma Rede Neural Artificial. Adaptado de [2].

### *2.1.4 Exemplo de treinamento supervisionado de uma Rede Neural Artificial*

A seguir é apresentado um exemplo muito simples para ilustrar como uma RNA aprende. Existem inúmeras formas e cálculos que podem ser utilizados. Neste caso, quando uma entrada é apresentada à rede, uma saída será produzida e posteriormente comparada com a saída desejada. Se a saída produzida for diferente da saída desejada, um ajuste de pesos deverá ser realizado na unidade de processamento de forma que a rede aprenda conforme desejado. É dessa forma que uma RNA é capaz de aprender e generalizar.

A figura 4 representa a rede neural artificial com três unidades de processamento (neurônios artificiais) que tem como objetivo somar ao peso atual o erro gerado pela rede e, dessa forma, corrigir o valor do peso. Esta rede tem como objetivo aprender a enviar uma resposta baseada nas entradas recebidas. Caso as entradas sejam iguais ao valor “0” a saída também deverá ser, se os valores de entrada forem iguais a “1” a saída deverá ser “1”.

Nesta figura pode-se encontrar as informações abaixo:

- Entrada do sinal a partir das unidades de processamento (*u1 e u2*);
- p: peso atual relativo às entradas (padrão de interconexão com as unidades de processamento*, p1 e p2*);
- r: regra de propagação que irá somar todas as entradas multiplicadas pelos seus respectivos pesos;

- a: o estado de ativação das unidades de processamento será assumido como igual a 1 (ativado);
- F: Regra de ativação. Neste caso simples, a regra de ativação será igual à regra de propagação.
- f: Função de saída:
    - Se soma das entradas = 0, valor de saída = 0;
    - Se soma das entradas > 0; valor de saída = 1;

Para a realização do cálculo deve-se considerar as seguintes informações:

- Erro (da saída da rede) = saída desejada – saída obtida;
- Correção associada à entrada = Erro * Entrada do sinal (1 e 2);
- Novo peso (aplicado nas conexões entre as unidades) = Peso atual relativo às entradas (1 e 2) + correção associada.

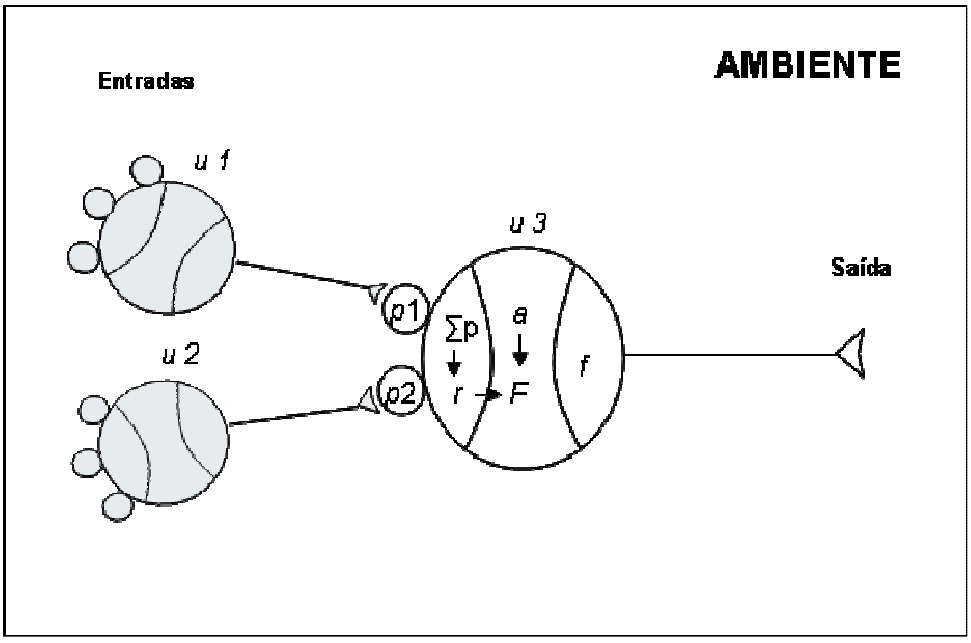


Figura 4 - Exemplo de Rede Neural Artificial.

Nesse exemplo serão usadas apenas três unidades de processamentos com duas entradas binárias ( 0 ou 1) e o aprendizado a ser atingido é o seguinte:

- Caso a entrada seja (1,1) a saída a ser produzida é (1).
- Caso a entrada seja (0,0) a saída a ser produzida é (0).

Em uma rede neural artificial não treinada os valores assumidos inicialmente

para os pesos são iguais a (0).

A tabela abaixo mostra os três passos para o treinamento dessa unidade de processamento observando os elementos acima citados. No passo “1” não houve correção, pois a saída desejada é igual à saída obtida. No passo “2” a saída desejada é diferente da saída obtida, portanto há uma correção. E no passo “3”, a rede neural artificial “aprende” e produz a saída desejada igual à saída obtida.

| Passo | Entrada do sinal 1 e 2 | Peso atual da Entrada 1 | Peso atual da Entrada 2 | Soma das entradas | Saída desejada | Saída obtida | Erro | Correção associada | Novo peso Entrada 1 | Novo peso Entrada 2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | (0,0) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | (1,1) | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 |
| 3 | (1,1) | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 1 |

Tabela 1 – Passos para treinamento de uma Unidade de Processamento.

### *2.1.5 Topologia de Redes*

Segundo Kröse e Smagt [4], os padrões de conexão entre as unidades de processamento e a direção da propagação dos dados definem as topologias básicas paras as RNAs. Vide figura 5.

Para esses padrões de conexão, as principais topologias básicas podem ser classificadas em:

- Redes em camadas (*Feed-Forward*): o fluxo de dados das entradas para as saídas das unidades de processamento seguem estritamente uma única direção. O processamento dos dados pode ser realizado em múltiplas camadas de unidades de processamento, mas não há realimentação nas conexões presentes. Ou seja, não há ligações entre unidades de processamento no mesmo nível ou nível inferior, apenas para níveis superiores. São exemplos de redes em camadas o Perceptron proposto por Rosenblatt [5] e Adaline (Adaptive Linear Element) proposto por Widrow e Hoff [6].

- Redes recorrentes (*Feed-Back*): redes que contém conexões retroalimentadas, ou seja, apesar de terem a mesma estrutura das redes em camadas, são permitidas conexões entre as saídas das unidades de processamento de um nível superior e as entradas de nós (realimentação) de um nível inferior. São exemplos de redes recorrentes as que são apresentadas por Kohonen [7] e Hopfield [8].

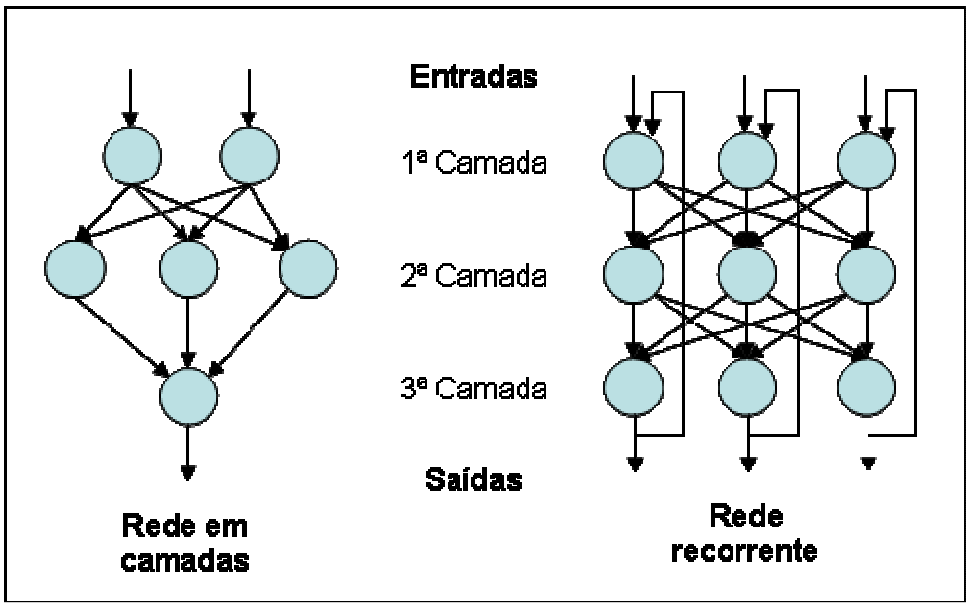


Figura 5 - Exemplos de Topologia de Redes Neurais Artificiais.

### *2.1.6 Treinamento das Redes Neurais Artificiais*

Por aprendizado entende-se fazer as redes neurais artificiais aprenderem a simular o comportamento de um sistema. Isto pode ser feito através de treinamento. Uma rede neural artificial deve ser treinada de modo que a aplicação de um conjunto de dados de entrada produza um conjunto de dados de saída desejado. Por exemplo, no caso de uma rede neural artificial para reconhecimento de fraudes em cartões de crédito, milhões de entradas são fornecidas para treinamento inicial. Assim, uma vez treinada, essa RNA deverá apresentar a partir de uma transação real a indicação se determinada transação é fraudulenta ou não de acordo com o seu “conhecimento” adquirido. Segundo Kröse e Smagt [4] existem vários métodos para modificar pesos das interconexões das unidades de processamento de forma

que uma rede “aprenda”. Uma forma é modificar os pesos explicitamente, usando um conhecimento pré-definido. Outra forma é “treinar” uma rede neural através do fornecimento de dados de entrada e permitindo que os pesos sejam alterados de acordo com alguma regra de aprendizado.

A idéia básica é que se duas unidades de processamento são ativadas simultaneamente, a interconexão (padrão de interconexão) entre elas deverá ser reforçada, com o aumento do valor do peso referente a conexão entre as unidades de processamento.

De uma forma geral, as situações de aprendizado podem ser categorizadas em dois tipos distintos:

- Aprendizado supervisionado ou associativo: no qual uma RNA é treinada fornecendo-a entradas e comparando as saídas de acordo com padrões pré-definidos ou desejados. Estes pares entrada-saída podem ser providos por um “professor” externo, ou por um sistema que contém a rede (auto-supervisionado). Neste caso, dispõe-se de um comportamento de referência preciso que deve ser ensinado para a rede. Para cada padrão de entrada submetido à rede, compara-se a resposta calculada (obtida) com a resposta desejada, ajustando-se os pesos das conexões para minimizar o erro existente, caso ocorra. Exemplos de aprendizados supervisionados são a *regra delta* também chamada *de regra Widrow-Hoff* [6] e a sua generalização para redes de múltiplas camadas, o algoritmo *backpropagation* [2].
- Não supervisionado ou auto-organizado: na qual uma unidade de processamento é treinada para fornecer grupos de padrões como resposta apenas fornecendo as entradas. Diferente do tipo supervisionado, não há um grupo pré-estabelecido (*a priori*) no qual os padrões são classificados. O próprio sistema deverá desenvolver sua própria representação do estímulo de entrada, classificando-as com algum critério de semelhança. Neste caso as unidades de processamento são utilizadas como classificadores dos dados de entrada, que são os elementos a serem classificados. Exemplos de aprendizados não supervisionados são Kohonen [9] e *Counterpropagation* [10].

## 2.2 Mineração de Dados

Mineração de dados é a exploração e análise de forma automática ou semi-automática de bases de dados com o objetivo de descobrir padrões e regras [13]. Existem diversas aplicações que podem se beneficiar com as técnicas de mineração de dados como: detecção de fraudes, segmentação de mercado, melhoramento de processos e serviços, análise de mercado, entre outros.

### *2.2.1 Descoberta de Conhecimento em Banco de Dados*

Segundo Braga [11], a Mineração de Dados está inserida em um processo maior denominado “Descoberta de Conhecimento em Banco de Dados” (*Knowledge Discovery in Database* (KDD)). O KDD é definido como um processo de descoberta de conhecimentos úteis previamente desconhecidos a partir de grandes bancos de dados.

O processo de KDD é interativo e iterativo dependendo constantemente da interferência de um especialista [12]. As principais etapas do KDD executadas são apresentadas abaixo:

*i) Conhecimento do domínio da aplicação:* inclui o conhecimento relevante anterior e a identificação do problema. Este passo utiliza o domínio do especialista em KDD para identificar problemas importantes e os itens necessários para resolvê-los.

*ii) Criação de um Banco de Dados alvo:* definir o local de armazenamento e selecionar um conjunto de dados para realização da busca.

*iii) Pré-processamento:* inclui operações básicas para tratamento dos dados como remover ruídos ou subcamadas, identificar e retirar valores inválidos inconsistentes ou redundantes.

*iv) Transformação de dados e projeção:* inclui encontrar formas práticas para representar dados e métodos de transformação para reduzir o número de variações que deve ser levado em consideração no contexto.

*v) Mineração de dados:* consiste na busca por padrões nos dados através da aplicação de algoritmos e técnicas computacionais adequadas.

*vi) Interpretação:* consiste na análise dos padrões descobertos e de uma possível visualização dos padrões extraídos, removendo aqueles redundantes ou irrelevantes e traduzindo os úteis em termos compreendidos pelos usuários.

*vii) Utilização do conhecimento obtido:* inclui a necessidade de incorporar este conhecimento obtido através do processo de KDD para apoio às decisões, ou simplesmente documentando e reportando este conhecimento para grupos interessados.

A figura abaixo representa as principais etapas do processo de KDD descrito.

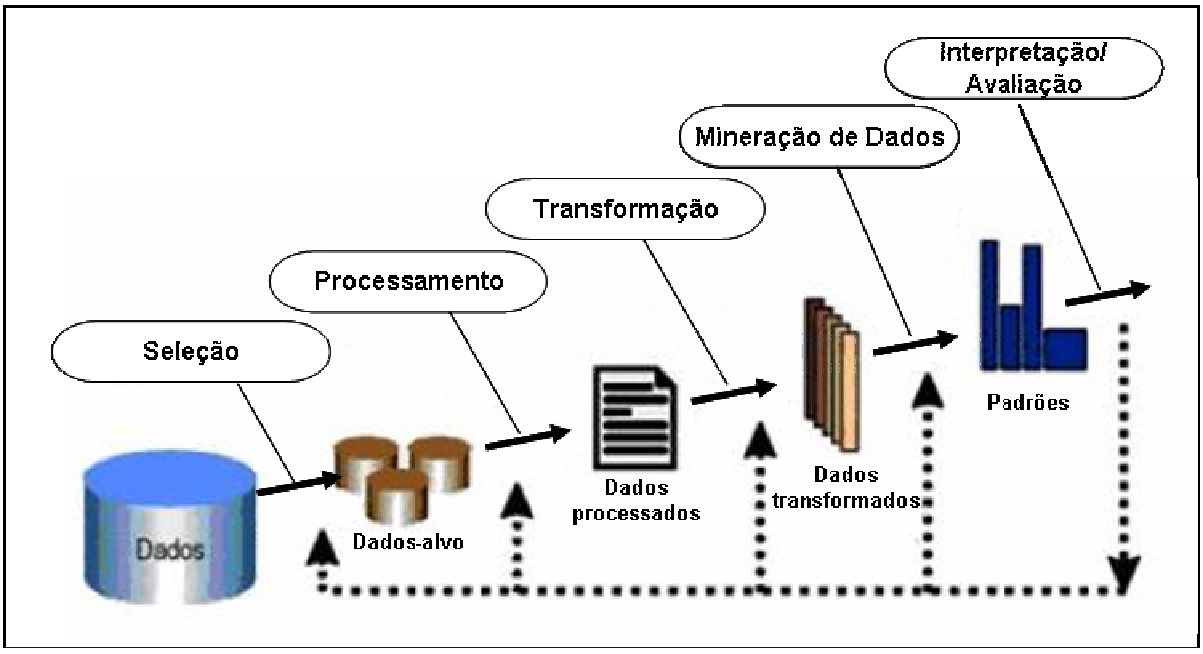


Figura 6 - Processo de KDD, adaptado de [13].

### *2.2.2 Tarefas de Mineração de Dados*

Existem diversas tarefas de mineração de dados encontradas em várias aplicações e pesquisas reconhecidas. Essas tarefas podem extrair diferentes dados, informações e conhecimento de acordo com o interesse que se deseja obter. Abaixo são relacionadas e descritas as principais tarefas de mineração de dados:

*i) Análise de Regras de Associação*

Uma regra de associação é uma regra que caracteriza a implicação ou influência que um conjunto de itens em uma base de dados tem em um ou mais itens diferentes nesta mesma base, ou seja, uma regra para encontrar todas as associações relevantes entre esse conjunto de itens aplicados aos demais itens. Um exemplo de uma regra de associação pode ser um constatação que 98% dos consumidores que compram pneus em uma loja de acessórios automotivos também utilizam serviços automotivos dessa mesma loja. Uma das grandes características da regra de associação é que elas permitem encontrar tendências para a exploração e entendimento de padrões nos bancos de dados [14].

*ii) Classificação e Predição*

Classificação é um processo de descoberta de propriedades ou características semelhantes em diferentes dados (entidades) numa base de dados e classificá-los em classes distintas. Os resultados são geralmente expressos em forma de regras, chamadas regras de classificação. Essas regras de classificação são utilizadas para a predição de dados ainda não classificados. Por exemplo, a tarefa de classificação e predição pode indicar dentro de um contexto de análise de crédito que, se clientes que tem idade abaixo de 30 anos tem renda acima de R$ 50.000,00 ao ano e não são estudantes, então são bons clientes tomadores de crédito [15,16,17]. São exemplos de técnicas de classificação e predição, árvores de decisão e redes neurais artificiais.

*iii) Análise de Clusters (Agrupamentos)*

Analisa os agrupamentos existentes na base de dados ou identifica as classes existentes onde os objetos são similares entre si. Um exemplo de aplicação são agrupamentos de páginas da *web* que possuem documentos tematicamente semelhantes ou com termos semelhantes [18,19].

*iv) Análise de Outliers (Análise de exceções)*

A análise de *outliers* foca em um percentual muito pequeno de dados que não se enquadraram nas regras de associação, classificação e predição, e nem nos agrupamentos. Geralmente esses dados podem ser descartados ou simplesmente ignorados. Por exemplo, em comércio pela Internet (*e-commerce*) são esperados várias transações de baixo valor monetário. No entanto, são os casos excepcionais (que podem ser monetários, tipo de compra, localização do comprador, etc.) que pode ser alvo de interesse na análise para detecção de fraudes [20].

# 3 APERFEIÇOANDO BUSCAS COM REDES NEURAIS E MINERAÇÃO DE DADOS

Segundo Schons [21], a Internet pode ser definida como uma vasta e onipresente rede global, que interconecta vários computadores em todo o mundo. Devido as características de sua própria estrutura funcional como desregulamentação, descentralização, aberta e não-hierárquica, o acesso e a disponibilização de informações na Internet foram favorecidas. Por conta disso, o processo atualmente observado na Internet é um grande acervo informacional em expansão, que a princípio é positivo, porém origina um grande problema na medida em que o excesso de informação dificulta a pesquisa e o seu resultado. Dessa forma, a grande quantidade de informações disponíveis na Internet e seu crescimento acelerado, a inexistência de ordem e a infinidade de temas, sob os mais diferentes enfoques e idiomas, bem como os mecanismos de busca gerais que indexam o conteúdo sem compreender o tema da página representam uma solução parcial para o problema [22].

Uma vez fundamentados os conceitos de Redes Neurais Articiais e Mineração de Dados este capítulo apresenta a aplicação das técnicas e métodos de redes neurais artificiais e mineração de dados para aperfeiçoamento de buscas de informações na Internet.

## 3.1 Utilização de Redes Neurais Artificiais em conjunto com Mineração de Dados

Segundo Amo [24], para a tarefa de classificação na mineração de dados existem alguns conceitos que podem ser aplicados. O conceito que será tratado neste trabalho é o de RNA e para isso será considerado para a entrada de dados um banco de dados de treinamento e a saída de dados será uma RNA treinada. O objetivo desta RNA é aprender como classificar novos dados, como por exemplo, quando recebe um termo de busca não presente na RNA. Assim, o primeiro passo é estabelecer a topologia que deverá ser utilizada para a rede neural artificial com o número de camadas intermediárias e das unidades de processamento em cada camada. O próximo passo consiste em definir qual deve ser a regra de aprendizado

e com isso inicializar os parâmetros da rede que incluem os pesos das interconexões entre as unidades de processamento e os parâmetros envolvidos nas regras associadas nas unidades de processamento, como a regra de propagação e a função de saída (ativação). Todos esses parâmetros de inicialização geralmente assumem valores pequenos (entre -1 e 1).

Neste contexto dispõe-se de um banco de dados para treinamento com as amostras já classificadas, composto de uma única tabela. Uma das colunas desta tabela corresponde ao Atributo-Classe (seriam os “títulos” das colunas), que na camada de saída de rede neural corresponde às classes (valores) possíveis para classificação. Na tabela abaixo é apresentado um exemplo de banco de dados de treinamento, onde o Atributo-Classe da camada de entrada é “Termos de busca” e os valores que no caso de busca seriam digitados pelo usuário são “Redes” e “Neurais”. Assim como na camada de saída é o título “Documentos” e os valores para esse Atributo-Classe que estão presentes são “Monografia” e “Tutorial Mineração de Dados”.

| Termos de busca | Termos de Indexação | Documentos |
|---|---|---|
| Redes | Redes Neurais Artificiais | Monografia |
| Redes | Redes Neurais Artificiais | Tutorial Mineração de Dados |
| Neurais | Redes Neurais Artificiais | Monografia |
| Neurais | Redes Neurais Artificiais | Tutorial Mineração de Dados |

Tabela 2 – Banco de dados de treinamento.

O treinamento da rede é realizado através de um banco de dados de treinamento, cujos elementos são chamados de amostras ou exemplos. Essas amostras já são classificadas e durante o treinamento são fornecidas para a rede na camada de entrada, uma de cada vez (cada uma das linhas da tabela).

Uma vez recebida a amostra, a RNA irá tratar essa amostra através de suas unidades de processamento. Cada unidade de processamento irá tratar um elemento componente da amostra (ou linha da tabela) de acordo com os padrões de interconexão existente entre as unidades, suas regras de propagação e função de ativação.

Ao final de uma iteração, a RNA irá fornecer uma resposta para a amostra. Caso a resposta da RNA seja diferente da resposta desejada, um processo inverso (chamado *Backpropagation*) é iniciado quando os pesos das conexões são ajustados.

Esse processo se repete a cada amostra até a totalidade existente no banco de dados de treinamento.

Cada iteração completa do banco de dados recebe o nome de época e na prática são necessárias diversas épocas para que um treinamento seja considerado satisfatório.

Para definição da topologia de uma RNA para a tarefa de classificação devem ser considerados os itens a seguir:

*i) Número de unidades de processamento na camada de entrada:* corresponde ao número de atributos que as classes contém. Por exemplo, se tivermos três classes com dois atributos cada, devemos ter seis unidades de processamento de entrada.

*ii) Número de unidades de processamento na camada de saída:* corresponde ao número de classes existentes para classificação.

*iii) O número de camadas intermediárias e de unidades de processamento nas camadas intermediárias:* de uma forma geral é utilizada uma única camada intermediária e não há regra clara para determinar o número de unidades de processamento na camada.

A determinação da topologia da RNA não é imutável, ou seja, uma vez que uma rede neural foi treinada e o grau de acertos de sua atividade de classificação não é considerado bom na fase de testes, é comum repetir todo o processo de aprendizado com uma RNA com topologia diferente.

## 3.2 Execução passo a passo de redes neurais artificiais na busca de uma informação: um exemplo simples

Neste capítulo será apresentado um exemplo com a execução passo a passo de uma rede neural artificial na busca de uma informação com o objetivo de mostrar como as informações passam de um neurônio para outro. Este exemplo será baseado no modelo Mozer que de acordo com Ferneda [23] tem algumas limitações como não utilizar a habilidade de aprender das redes neurais por meio da

alteração dos padrões de interconexão (pesos) entre os nós, pois não foi definida uma Regra de Aprendizado.

De acordo com Ferneda [23], a busca de informação lida basicamente com documentos, termos de indexação e as expressões de buscas dos usuários. As expressões de busca dos usuários são os termos de busca que os usuários inserem ao realizar uma busca. Os termos de indexação são palavras eleitas por um especialista que representam um documento [25] ou informação.

A estrutura do sistema de recuperação de informação poderia ser vista como uma rede neural artificial em três camadas: os termos de busca seriam a camada de entrada, a indexação (com os termos de indexação) seria a camada intermediária e a terceira camada que contém os documentos seria a saída da rede neural [23] e [26].

No entanto, uma vez assumida essa estrutura da RNA, definido os atributos que representam cada unidade de processamento da RNA e realizando o treinamento adequadamente é possível estabelecer uma estrutura cujos termos de busca poderiam ser por exemplo “Redes” e “Neurais”, os termos de indexação poderiam ser por exemplo “Redes Neurais Artificiais” e “Mineração de dados” e como documentos podemos assumir como exemplo esse trabalho e “Tutorial Mineração de dados”. Os nós dessa RNA e suas conexões seriam estabelecidos através do próprio treinamento realizado, com pesos distribuídos conforme a necessidade.

Os termos de busca poderiam ser ligados a zero ou mais termos de indexação, que por sua vez poderiam ser ligados a zero ou mais documentos, isso de acordo com a necessidade de aplicação da RNA e após a realização de treinamento adequado.

As conexões entre os termos de indexação e os documentos teriam um peso, que multiplicaria os sinais enviados pelos termos de indexação aos documentos. Os pesos dos nós da RNA e os sinais enviados seriam representados por valores numéricos.

Por sua vez, os documentos que receberam os sinais seriam “ativados” e enviariam os sinais recebidos aos termos de indexação.

Ao receberem estes sinais, os termos de indexação enviam novos sinais aos documentos, repetindo o processo e a cada iteração o sinal fica mais fraco por conta

da multiplicação dos pesos em cada nó (que geralmente variam de 0 a 1) e consequente diminuição do valor do sinal até que a propagação eventualmente pára.

De uma forma simplificada, o resultado final de uma busca utilizando essa estrutura seria um conjunto de documentos que foram ativados durante o processo e o nível de ativação poderia ser comparado a relevância do documento em relação à busca do usuário.

Na figura 7 abaixo, é apresentado um exemplo de uma RNA estabelecida e já treinada com um banco de dados de amostras adequado onde estão representados dois termos de busca na primeira coluna representados por círculos numerados “1” e “2” (os nós da RNA), inseridos por um usuário com valores “Redes” e “Neurais”. Na segunda coluna, existem “n” “Termos de indexação”, sendo o termo de indexação “1” com valor “Rede Neural Artificial” e “2” com o valor “Mineração de Dados”. Na terceira coluna existem “n” documentos em uma rede neural artificial já treinada, sendo o documento “1” representando este trabalho e o documento “2” um “Tutorial Mineração de dados”. As conexões entre os nós cujo valor do peso seja diferente de zero estão representadas pelas setas direcionais, as conexões entre os nós que possuem peso igual a zero não foram representadas para simplificação do modelo.

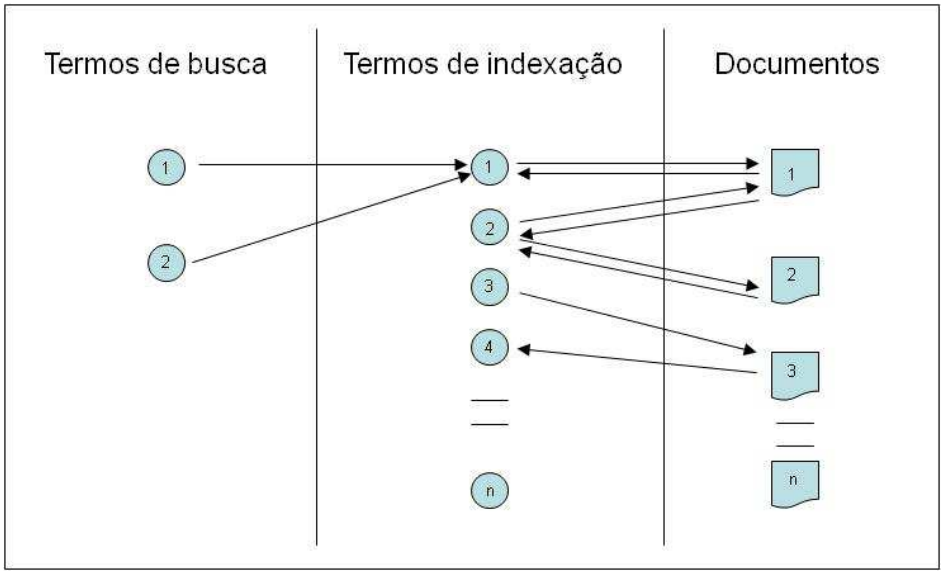


Figura 7 – Representação de uma RNA, adaptado de [23].

A figura 8 apresenta uma sequência com oito passos que representa um exemplo de funcionamento de uma rede neural artificial aplicada à recuperação de informação de acordo com o modelo citado acima. Neste exemplo assume-se que a

função de saída irá transmitir um sinal caso ele tenha valor maior que “0,5”. Caso seja menor ou igual, o sinal não será transmitido. O sinal de saída de cada nó da RNA será resultado da multiplicação entre os sinais de entrada recebidos e os seus respectivos pesos da conexão entre os nós. Cada nó pode estar ligada a zero ou mais nós nas diferentes camadas, neste exemplo apenas a camada de saída (documentos) terá uma conexão com característica recorrente, ou seja, os nós dessa camada estarão ligados nos dois sentidos de propagação da rede. As setas representadas neste exemplo indicam a transmissão e a direção dos sinais em determinado instante na RNA. Os sinais serão propagados de um nó para outro desde que tenham suas conexões estabelecidas.

No passo “1”, após inserir os termos de busca “1” e “2” (“Redes” e “Neurais”) dessa expressão iniciam o processo “ativando” o termo de indexação estabelecido “1” (“Redes Neurais Artificias”), representado pela seta direcional.

No passo “2”, o termo de indexação ativado envia por sua vez, um sinal para a próxima camada, a de “Documentos” para os documentos que estão conectados através de um valor (valor “1” por exemplo) que é multiplicado pelo peso (p1, que também possue valor “1” neste caso). Na função de saída, o valor resultante da multiplicação do peso p1 e o sinal recebido é igual a “1”.

No passo “3”, o documento ativado “1” retorna o sinal de ativação para os termos de indexação que estão conectados aos nós “1” e “2” e são multiplicados pelos pesos “p2” que possuem valor “0,75” (também estabelecido em tempo de treinamento da RNA). Na função de saída o valor resultante da multiplicação do peso p2 e o sinal recebido é igual a “0,75”.

Nos passos “4” e “5”, o sinal se propaga e se enfraquece a cada iteração, por conta da multiplicação dos pesos, até que eventualmente pára no passo “7” (o sinal assume valor menor que “0,5” que para o exemplo não ativa o sinal para a função de saída estabelecida).

Os resultados da busca são os documentos ativados pelos sinais, sendo que os mais ativados, no caso “Rede Neural Artificial” ( houve cinco ativação desse nó durante o processo) e “Tutorial Mineração de Dados” ( duas ativações ) significam os documentos com maior relevância para a busca realizadas.

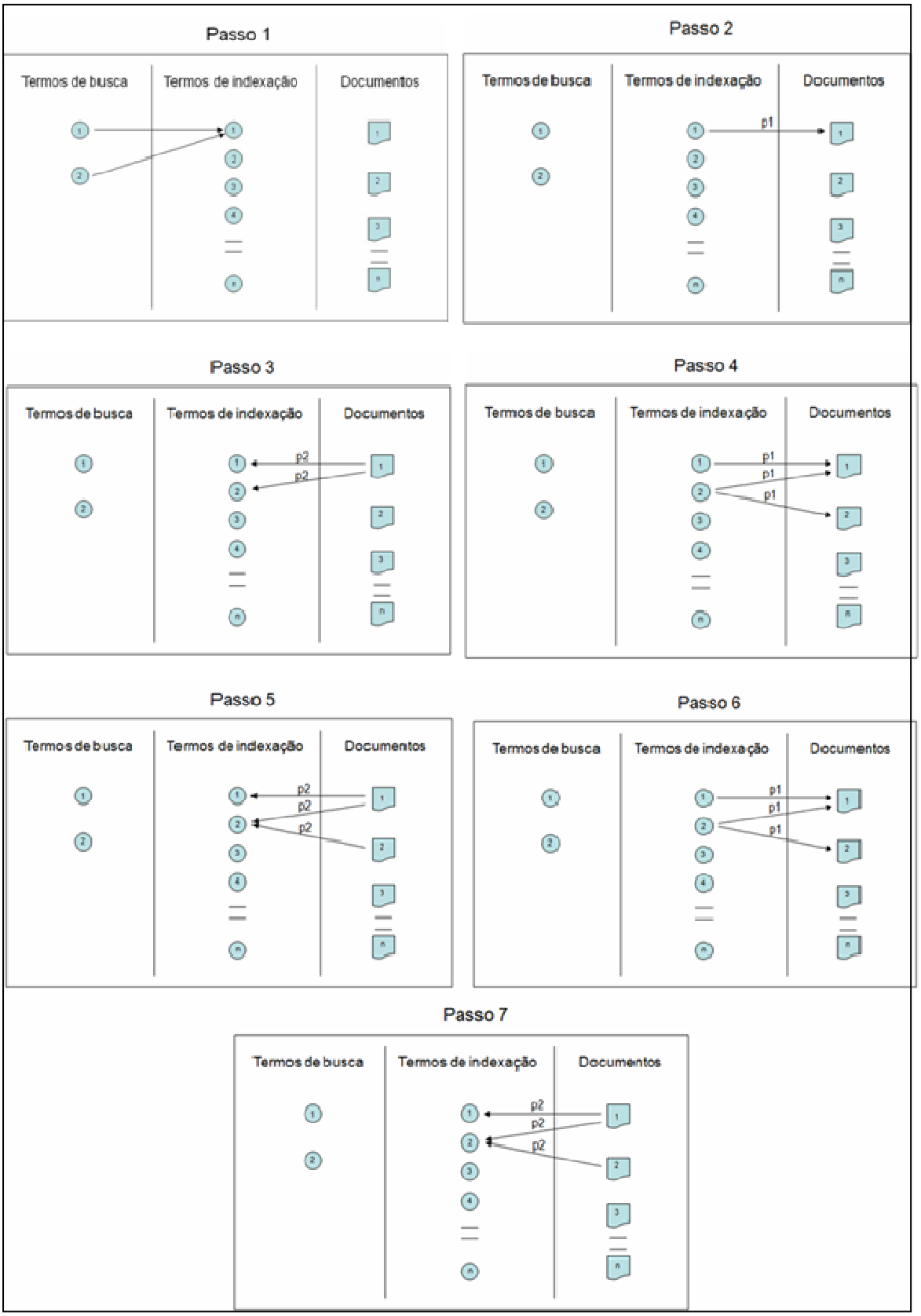


Figura 8 – Passos para representação de uma RNA.

Outros modelos foram propostos como o desenvolvido por Belew [28] que foram adaptados para explorar a habilidade de aprender das RNAs, denominado Adaptative Information Retrieval (AIR). Esse modelo é composto também de três camadas na RNA e fornece uma interface para que o usuário possa atribuir um grau de relevância para os itens recuperados. Com base nisso, os padrões de interconexão entre os neurônios artificiais em cada uma das camadas nessa interação são alterados de acordo com os parâmetros definidos na Regra de Aprendizado. Uma vez que há alteração e evolução da RNA como um todo, esse modelo pode ser visto como um processo contínuo de aprendizagem otimizando presumivelmente o desempenho para ambientes homogêneos.

### 3.3. Redes Neurais Artificiais e Mineração de dados na Internet

A recuperação de informação na Internet é facilitada pelos mecanismos de busca (*search engines*), que coletam e indexam uma parte da imensa quantidade de páginas disponíveis na Web [23]. Assim, utilizando esses mecanismos junto com uma possível adaptação do modelo apresentado por Ferneda [23] de aplicação de redes neurais artificiais poderia ser uma alternativa viável para a otimização de buscas na Internet. Essa adaptação seria a substituição da camada de "Documentos" por uma camada de URL. Portanto, a estrutura do sistema de recuperação de informação do ambiente Internet poderia ser vista como uma rede neural artificial em três camadas: os termos de busca seriam a camada de entrada, a indexação seria a camada intermediária e a terceira camada que contém as URLs, seria a saída da RNA. Aplicando esta idéia no exemplo da figura 8, teríamos "Redes" e "Neurais" como termos de busca, "Redes Neurais Artificiais" e "Mineração de dados" como termos de indexação, e como URLs poderíamos citar, por exemplo, "http://www.din.uem.br/ia/neurais" e "http://www.lsi.usp.br/icone/", considerando que essas classificações tenham sido realizadas a partir de tarefas de Mineração de Dados, como "Classificação e Predição".

Na busca de informações, os termos de busca poderiam ser ligados a zero ou mais termos de indexação, que por sua vez poderiam ser ligados a zero ou mais URLs, com pesos que multiplicam os sinais enviados pelos termos de indexação às

URLs. Por sua vez as URLs que receberem os sinais seriam “ativados” e enviariam os sinais recebidos aos termos de indexação.

Ao receberem estes sinais, os termos de indexação enviam novos sinais às URLs, repetindo o processo e a cada iteração o sinal fica mais fraco até que a propagação eventualmente pára.

De uma forma simplificada, um resultado final de uma busca utilizando essa estrutura seria um conjunto de URLs ativados durante o processo e o nível de ativação poderia ser comparada à relevância da URL em relação à busca do usuário.

Na figura 9 abaixo, é apresentado um exemplo onde estão representados dois termos de busca na primeira coluna representados por círculos numerados “1” e “2” (os nós da RNA). Na segunda coluna, existem de “1” a “n” “Termos de indexação”. Na terceira coluna existem de “1” a “n” URLs em uma rede neural artificial.

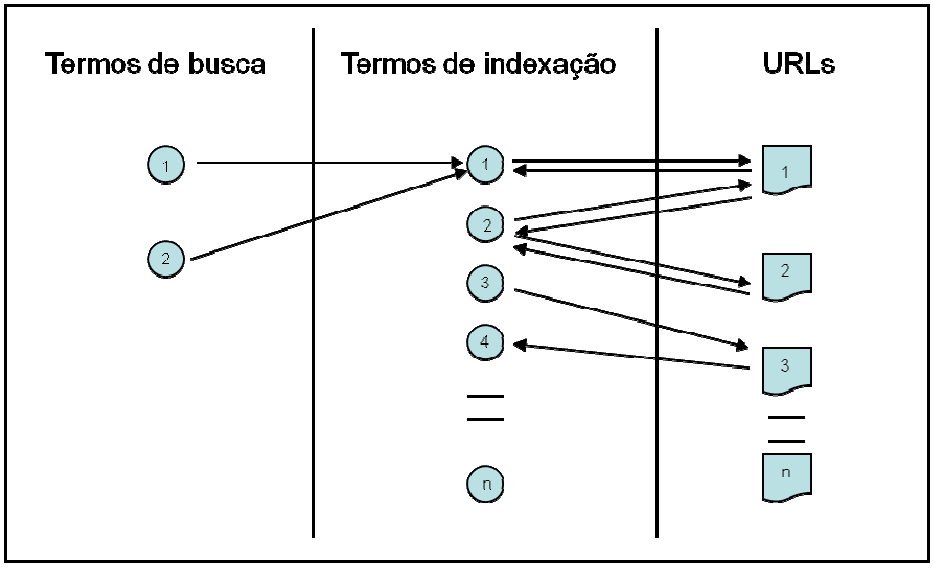


Figura 9 – Representação de uma RNA aplicada à Internet.

O exemplo acima apresenta uma rede neural artificial adaptada para o ambiente da Internet com os nós (neurônios artificiais) representados pelos círculos numerados e os ícones de documento e as conexões e o sentido de propagação dos sinais representadas pelas setas, fornecendo respostas através de URLs relevantes. Este exemplo possui as mesmas limitações apresentadas pelo modelo de Mozer citado, ou seja, caso não seja definida uma Regra de Aprendizado, a RNA não irá

“aprender” e a limitação de aplicação neste ambiente heterogêneo que é a Internet não seria superada.

Uma das alternativas que poderiam ser adotadas para contornar essas limitações é através da interação entre o usuário que está realizando a busca e a RNA. Dessa forma, ao realizar uma busca o usuário receberia os resultados ordenados conforme relevância estabelecida pela RNA. Ao acessar um resultado, o usuário estaria atribuindo uma relevância maior no resultado de sua pesquisa através de um retorno automático à RNA (com os parâmetros definidos pela Regra de Aprendizado), que alteraria seu padrão de interconexão (pesos) entre os termos de indexação e as URLs, como na fase de treinanemento da rede. Dessa forma, da próxima vez que um usuário realizar uma pesquisa, as URLs que tiveram seus padrões de interconexão (pesos) “aumentados” (a RNA estaria aprendendo), ou seja, as URLs que foram acessadas na pesquisa anterior, teriam chances maiores de serem listados com maior relevância nos resultados de busca, otimizando assim o sistema como um todo através da experiência de utilização de outros usuários.

A figura 10 apresenta um exemplo de alteração da estrutura da RNA apresentada na figura 9 após a interação de um usuário que escolhe a URL “2” acessando este *link*. Essa escolha irá alterar a estrutura da RNA através da alteração dos pesos das conexões existentes entre os nós da rede. Considerando uma regra de aprendizado onde as alterações ocorrem no peso da conexão do nó “2” da camada URL para o nó “1” da camada de termos de indexação, passando de valor zero para o valor “0,75” (com destaque em tracejado) e no peso da conexão entre o nó “2” da camada URL com o nó “2” da camada de termos de indexação para o valor “0,9” (com destaque em traço e ponto ). Dessa forma, a próxima vez que o usuário inserir os termos de busca “Redes” e “Neurais” o resultado apresentado será uma lista de URLs onde a URL “http://www.lsi.usp.br/icone/” irá aparecer como primeira opção ( com a maior relevância ) e “http://www.din.uem.br/ia/neurais” como a segunda opção.

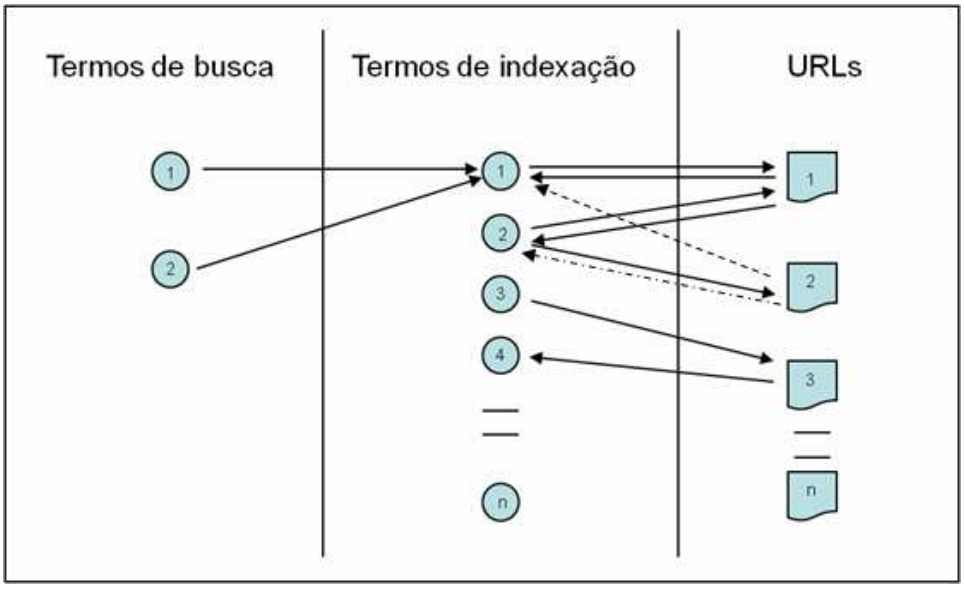


Figura 10 – Representação de uma RNA aplicada à Internet alterada com Regra de Aprendizado.

# 4 CONCLUSÃO

Este trabalho teve como objetivo apresentar uma alternativa para aperfeiçoar a busca de informações na Internet utilizando técnicas de redes neurais artificiais e mineração de dados.

Para o conceito de rede neural artificial foi apresentado um exemplo onde o entendimento do funcionamento de uma rede neural artificial simples pode ser facilitado.

Para a aplicação de redes neurais artificiais em busca de informações foi apresentado um exemplo específico que contribui para o entendimento de uma aplicação de redes neurais artificiais para busca de informações.

Foram feitas análises sobre a aplicação de redes neurais artificiais com mineração de dados na Internet além de apresentada uma proposta de aplicação com o objetivo concreto de otimização nas buscas por informações em ambientes heterogêneos.

Dentre as principais limitações para aplicação dos conceitos apresentados neste trabalho pode-se citar a necessidade de altos investimentos em infra-estrutura, apoio de especialistas das áreas envolvidas neste processo multidisciplinar. Além disso, as redes neurais devem ser previamente treinadas com grande quantidade de dados reais de busca e o sistema deverá prever um mecanismo de revalidação automático de endereços URL para eliminar resultados com endereços já inexistentes e redundantes.

Portanto, ao mesmo tempo em que a expansão da Internet se mostra como uma tendência irreversível e a busca de informações relevantes um problema crescente a alternativa apresentada e suas conseqüentes derivações viabilizam esse crescimento de forma sustentável, através de novas formas de recuperar eficientemente e eficazmente a informação.

## 4.1 Trabalhos Futuros

Conceitos como Redes Neurais Artificiais, Mineração de Dados e Internet são muito ricos. Sua combinação pode gerar diversos trabalhos futuros como a

implementação e testes do modelo proposto de otimização para busca de informação na Internet com a aplicação de redes neurais artificiais e mineração de dados. Além disso, poderiam ser utilizadas outras estratégias para receber e utilizar as respostas dos usuários em caso de buscas de informações como, por exemplo, níveis diferentes de relevância de busca para usuários conhecidos do sistema.

Além dessas modificações, poderiam ser levadas em consideração outras técnicas de mineração de dados como “Análise de *Clusters*” e “Análise de *Outiliers*” em conjunto com redes neurais artificiais para buscas de informações. [24]

Outras técnicas e conceitos poderiam ser agregados em novos estudos como o conceito de “Web Semântica”, que segundo [29] interliga significados de palavras para atribuir sentido aos conteúdos publicados na Internet.

# REFERÊNCIAS

[1] Braga, A. P.; Carvalho, A. C.; Ludermir, T. B. (2000). Redes Neurais Artificiais: Teoria e aplicações. 1a. ed. Rio de Janeiro, RJ - LTC.

[2] Rumelhart, D. E., McClelland, J. (1986), Parallel Distributed Processing, vol. 1, MIT Press.

[3] Kovács, Zsolt L. (1996). Redes Neurais Artificiais: Fundamentos e aplicações. São Paulo , SP – Livraria da Física Editora.

[4] Kröse, B. J. A.; Smagt, V.; Patrick, P (1993). An Introduction to Neural Networks, University of Amsterdam.

[5] Rosenblatt, F. (1959). Principles of Neurodynamics. New York: Spartan Books.

[6] Widrow, B.; Hoff, M. E. (1960). Adaptive switching circuits. DUNNO.

[7] Kohonen, T. (1977). Associative Memory: A System-Theoretical Approach. Springer-Verlag.

[8] Hopfield, J. J. (1982). Neural Networks and Physical Systems With Emergent Collective Computational Abilities, Proceedings of the National Academy of Sciences USA.

[9] Kohonen, T. (1989). Self-Organization and Associative Memory, Springer-Verlag, 3rd edition.

[10] Wasserman, P. D. (1989). Neural Computing, Theory and Practice, Van Nostrand Reinhold, New York.

[11] Braga, L. P. V. (2005). Introdução à Mineração de Dados. 2ª. Ed. E-papers.

[12] Berry, M. and Linoff, G., (2000). Mastering Data Mining - Art and Science of Customer Relationship Management, Ed. Wiley.

[13] Fayyad, Usama M. (1996). Advances in knowledge discovery and data mining. Menlo Park: Mit Press.

[14] Agrawal, R.; Srikant, R. (1994). Fast Algorithms for Mining Association Rules. Proc. 20th Int Conf. Very Large Data Bases, VLDB.

[15] Mehta, M.; Agrawal, R.; Rissamen, J. (1996) SLIQ: A Fast Scalable Classifier for Data Mining. Proc. of the Fifth Int'l Conference on Extending Database Technology, Avignon, France.

[16] Lu, H.; Setiono, R.; Liu, H. (1995). Neurorule: A connectionist approach to data mining. In Proc 1995 Int. Conf. Very Large Data Bases (VLDB'95), 478-489, Zurich, Switzerland.

[17] Han, J.; Kamber, M. (2000). Data Mining: Concepts and Techniques. Series Editor Morgan Kaufmann Publishers.

[18] Ester, M.; Kriegel, H-P; Sander, J.; Xu, X. (1996). A density-based Algorithm for Discovering clusters in Large Spatial Databases with Noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, KDD.

[19] Kriegel, H.-P., Kröger, P., and Zimek, A. (2009). Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. ACM Trans. Knowl. Discov. Data. 3, 1, Article 1, 58 pages.

[20] Knorr, E.M.; Ng, R.T. (1998). Algorithms for Mining Distance-Based Outliers in Large Datasets. Proceedings of the 24th International Conference on Very Large Data Bases, VLDB.

[21] Schons, C. H. (2007). O volume de informações na Internet e sua desorganização: reflexões e perspectivas. I n f . I n f . , L o n d r i n a , v . 1 2 , n . 1.

[22] Marcondes, C. H.; Sayão, L. F (2002). Documentos digitais e novas formas de cooperação entre sistemas de informação em C&t. Ciência da Informação, Brasília, v. 31, n. 3, p. 42-54.

[23] Ferneda, Edberto (2006). Redes neurais e sua aplicação em sistemas de recuperação de informação. Ci. Inf., Brasília, v. 35, n. 1, p. 25-30.

[24] Amo, Sandra (2004). Técnicas de Mineração de Dados . XXIV Congresso da Sociedade Brasileira de Computação. Jornada de Atualização em Informatica, 31 de Julho a 6 de Agosto, Salvador Brazil.

[25] Lancaster, F. W (2004). Indexação e Resumos: Teoria e Prática. Brasília: Briquet de Lemos.

[26] Ferneda, E.; Pinheiro, C.B.F (2005). Representação dinâmica de documentos em bibliotecas digitais. São Paulo. v. 2, p. 145-166.

[27] MOZER, M.C. (1984) Inductive information retrieval using parallel distributed computation. San Diego: University of California. (ICS Technical Report 8406).

[28] BELEW, R. K. (1989). Adaptive information retrieval. In: ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 12. , Cambridge. Proceedings... Cambridge: ACM, p.11-20.

[29] O W3C e a Web Semântica. CPqD - abril/2009 – Workshop Rede IP do Futuro. <http://www.w3c.br/palestras/cpqd-campinas-2009/CPqD20090416.pdf>. Acesso em 13/01/2010.