

APLICAÇÃO PRÁTICA DOS PROCESSOS DE ANÁLISE DE CONTEÚDO E DE ANÁLISE DE CITAÇÕES EM ARTIGOS RELACIONADOS ÀS REDES NEURAIS ARTIFICIAIS

APLICACIÓN PRÁCTICA DE LOS PROCEDIMIENTOS DE ANÁLISIS DE CONTENIDO Y DE ANÁLISIS DE CITACIONES EN TRABAJOS RELACIONADOS CON LAS REDES NEURONALES ARTIFICIALES

Magali Rezende Gouvêa Meireles – magali@pucminas.br

Doutoranda em Ciência da Informação pela UFMG
Professora Adjunta do Instituto de Ciências Exatas e Informática da PUC

Beatriz Valadares Cendón – cendon@eci.ufmg.br

Doutora em *Library and Information Science* pela Universidade do Texas, EUA
Professora Associada da Escola de Ciência da Informação da UFMG

Resumo

Análises de conteúdo e de citações são metodologias que se valem da inferência e da interpretação para obter informações relevantes contidas em um conjunto de documentos. Estes métodos são amplamente discutidos na literatura, mas são poucos os artigos encontrados nos quais se descreve e exemplifica sua aplicação em estudos de caso específicos. Este artigo apresenta um estudo de caso que tem o duplo propósito de demonstrar o uso dos métodos citados para análise de um conjunto de documentos e de mostrar o potencial da sua utilização conjunta para a recuperação de informação. O artigo apresenta, inicialmente, as abordagens teóricas relativas aos dois métodos e a caracterização do material empírico da pesquisa, que utilizou artigos sobre Redes Neurais Artificiais. A seguir, são descritas as categorias semânticas utilizadas na análise de conteúdo e as variáveis selecionadas para a análise de citações. Por meio de análises quantitativas e qualitativas, exemplifica-se o uso da análise de citações e da análise de conteúdo e demonstra-se, ainda, que elas podem ser utilizadas de modo integrado para a obtenção de metadados a respeito de coleções de documentos.

Palavras-chave

Ciência da Informação. Análise de conteúdo. Análise de citações. Categorização. Redes Neurais Artificiais.

1 INTRODUÇÃO

As análises de conteúdo e de citações, por meio da sistematização de conteúdos existentes em documentos, são metodologias que valem-se da inferência e da interpretação para obter questões relevantes contidas em um conjunto de documentos. Os métodos são amplamente discutidos na literatura, mas são poucas as aplicações encontradas nas quais se descreve e exemplifica sua utilização em estudos de caso específicos.

Este artigo apresenta um estudo de caso que demonstra o uso dos métodos citados para análise de um conjunto de documentos sobre Redes Neurais Artificiais, mostrando, ainda, o potencial do uso conjunto dos métodos de análise de conteúdo e análise de citações, para a recuperação de informações que vão além daquelas explicitamente contidas nos documentos analisados.

Na análise de conteúdo são estabelecidas categorias semânticas relacionadas aos artigos e análises estatísticas da frequência de ocorrência destas categorias nos documentos em estudo. Com a análise de citações, são identificadas variáveis e disponibilizadas consultas de ocorrência de citações por cada uma das variáveis selecionadas. Com base na metodologia desenvolvida e nos resultados encontrados, o artigo demonstra a possibilidade de ampliação do uso das análises de conteúdo e de citações como uma ferramenta refinada de recuperação de informação. Enquanto, normalmente, os sistemas de recuperação de informação buscam informações e palavras-chave contidas no texto, as informações encontradas por meio das análises de citação e de conteúdo extrapolam o conteúdo disponível no nível dos textos, identificando tendências de desenvolvimento da área em estudo e dispersão e agrupamento da literatura científica no tempo e no espaço físico.

2 ABORDAGEM TEÓRICA

Antes de iniciar a apresentação do estudo de caso discutido neste trabalho, são apresentados conceitos relacionados à análise de conteúdo e à análise de citações.

2.1 Análise de conteúdo

Segundo Bardin (1977), a análise de conteúdo pode ser definida como um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores que permitam a inferência de conhecimentos relativos às condições de produção e recepção destas mensagens. Caracteriza-se, assim, como um método de tratamento da informação contida nas mensagens.

Para a utilização do método é necessária a criação de categorias relacionadas ao objeto de pesquisa. As deduções lógicas ou inferências que serão obtidas a partir das categorias serão responsáveis pela identificação das questões relevantes contidas no conteúdo das mensagens.

O analista trabalha com palavras que, isoladas, podem atribuir relações com a mensagem ou possibilitar que se faça inferência de conhecimento a partir da mensagem. São, dessa maneira, estabelecidas correspondências entre as estruturas linguísticas ou semânticas e as estruturas psicológicas ou sociológicas dos enunciados. A leitura do pesquisador responsável pela análise não é, portanto, uma leitura à letra, mas, o realçar de um sentido que se encontra em segundo plano (BARDIN, 1977).

Para que as inferências sejam realizadas, é necessário, como descrito na Figura 1, que a etapa anterior, a de descrição, tenha enumerado as características do texto após tratamento.

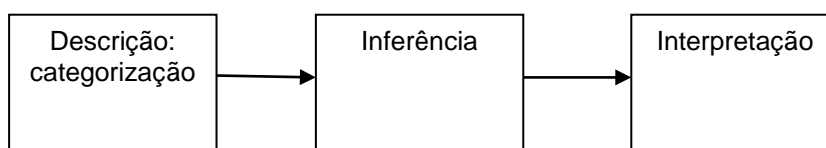


Figura 1 Etapas da análise de conteúdo

A análise categorial é uma das técnicas utilizadas para a realização da análise de conteúdo, e, em seu desenvolvimento, desmembra o discurso em categorias. Os critérios de escolha e de delimitação das categorias são determinados pelos temas relacionados aos objetos de pesquisa e identificados nos discursos dos sujeitos pesquisados (BARDIN *apud* VALENTIM, 2005). Um conjunto de categorias cuidadosamente selecionadas pode gerar indicações produtivas para o processo de inferência, contribuindo para que as interpretações possam espelhar resultados validados pelo método.

2.1.1 *Categorização e Classificação*

Historicamente, a noção de categoria tem sido abordada de modos diversos. Segundo Xavier (2008), o conceito de categorias, como discutido nos dias de hoje, nasceu com Aristóteles, que viveu entre os anos de 384 e 322 a.C.. “*Categorias*” é o primeiro dos cinco tratados que compõem “*Organon*”, a obra que expõe a lógica aristotélica.

Para Santos (2003), uma das consequências derivadas da importância que a matemática assume na ciência moderna, é a de que o método científico assenta-se na redução da complexidade. Se o mundo é complicado e a mente humana tem dificuldade em compreendê-lo, a primeira etapa para se alcançar o conhecimento é separar e, depois, classificar. Em uma etapa posterior, as relações entre o que se dividiu poderão ser encontradas.

Segundo Jacob (2004), categorização é o processo de dividir o mundo em grupos de entidades cujos membros têm similaridades entre eles dentro de um determinado contexto. Agregar as entidades em categorias leva o indivíduo a perceber ordem no mundo que o circunda. Segundo Barite (2000), as categorias são usadas como ferramentas para se descobrir certas regularidades do mundo material, concluindo-se que todos os objetos, pelo menos os que pertencem ao mundo material, possuem certas propriedades. Ainda afirma que as categorias

são abstrações simplificadas usadas pelos classificacionistas para investigar regularidades de objetos do mundo físico e ideal e para representar noções.

Utilizando-se o pensamento de Piaget, citado por Rojas (2005), o conhecimento é construído pelo sujeito com base na assimilação, na integração e na reorganização de estruturas que lhe permitem interpretar o mundo e interagir com ele. Para que o conhecimento possa ser disseminado, são realizados processos cognitivos conhecidos como classificação, ordenação e organização da informação.

A classificação é, portanto, uma capacidade conceitual humana, que utiliza as categorias como ferramentas. O processo de classificação envolve a associação ordenada e sistemática de cada entidade a uma única classe dentro de um conjunto de classes mutuamente exclusivas e que não se sobrepõem (JACOB, 2004).

2.2 Análise de citações

Um outro método que mostra relacionamento de um documento com outro independentemente da presença de termos iguais nos documentos avaliados é a análise de citações. As citações bibliográficas denotam relações particulares entre o artigo citante e o citado e, nesse sentido, a citação pode ser entendida como um processo social no momento em que considera toda a experiência anterior do autor, sua rede de conhecimento e suas próprias considerações (LEAL, 2005). Para Alvarenga (1998), o ato de citar é envolvido por um espectro de implicações psicológicas, sociológicas, políticas e históricas. Além disto, percebe-se a presença de características como o fenômeno de auto-citações, a influência entre autores e instituições e a adesão a paradigmas vigentes.

Cronin, citado por Leal (2005), afirma que o hábito de citar mostra conformidade e consistência no ato da produção intelectual, muitas vezes, governada por conhecimento tácito e normas internalizadas. Os fatos citados no texto ganham credibilidade quando se apresenta a literatura utilizada, interligando o leitor a outras fontes de informação sobre o assunto.

De acordo com Alvarenga (2009), cria-se um fluxo de informação e de comunicação entre aquele que cita e aquele que lê, proporcionando uma dinâmica circulante, que mantém viva a criação de um autor, concedendo-lhe significativo reconhecimento. Segundo Braga (1973), a inclusão de uma ou mais referências, em um documento, evidencia relações entre partes dos textos dos documentos citados e partes do texto do documento que as inclui.

Segundo Cronin (1984), a teoria das citações é difícil de ser construída por ser um ato privado e porque existem muitos modos de interpretação relacionados ao que se constitui uma citação correta ou aceitável. A citação deve perpassar todo o conhecimento prévio do autor até que sua idéia mental se transforme em uma marca ou um símbolo em seu texto.

Além de toda a importância dedicada ao ato de citar no processo de construção de um documento, durante o ciclo de suas atividades de pesquisa e divulgação de resultados, Borgman e Furner (2002) ressaltam que os autores tomam decisões relevantes nos processos que envolvem a elaboração e a

submissão de um trabalho e a escolha de profissionais que compartilharão o mesmo trabalho.

Alvarenga (1998) destaca, ainda, o fato de que os artigos relacionados por suas citações, geralmente, estabelecem relações semânticas entre si. A citação seria, também, uma maneira de se reportar aos fundamentos teórico-metodológicos presentes em uma área de pesquisa. As citações são, portanto, utilizadas para indicar leituras indispensáveis à compreensão do próprio documento, para evitar explicações teóricas detalhadas e para apresentar argumentos favoráveis à pesquisa.

Os resultados alcançados no desenvolvimento da análise de citações, segundo Alvarenga (1998), refletem aspectos quantitativos de campos de conhecimento, evidenciando variáveis, tais como produtividade de autores ou de fontes discursivas. O mapeamento da rede de relações estabelecidas entre autores e textos, por meio das citações, pode ser utilizado para se evidenciar ângulos específicos do processo de produção de conhecimento. São identificados, deste modo, os autores que constituem as frentes de pesquisa em determinado campo de conhecimento e algumas regularidades na evolução de uma disciplina, que podem ter sido influenciadas por fatos históricos.

Segundo Borgman e Furner (2002), a análise de citações é a abordagem bibliométrica mais conhecida. Como a bibliometria se encontra na fronteira das ciências sociais e das ciências físicas, Sengupta (1992) esclarece que suas técnicas têm aplicações na sociologia, na gestão da informação, na biblioteconomia, na história e na ciência política.

3 ESTUDO DE CASO SOBRE A APLICAÇÃO DE ANÁLISE DE CONTEÚDO E ANÁLISE DE CITAÇÕES EM ARTIGOS SOBRE REDES NEURAIS ARTIFICIAIS

A seguir, apresenta-se um estudo de caso que demonstra o uso dos métodos citados para análise e interpretação de um conjunto de documentos sobre Redes Neurais Artificiais, RNA. Na análise de conteúdo, são estabelecidas categorias semânticas que serão utilizadas para interpretar a informação contida nos conjuntos de artigos. Com a análise de citações, são identificados os autores citados nos trabalhos e outras variáveis. Estas informações são utilizadas para inferência de conhecimento relativo ao conjunto dos documentos analisados

O tema RNA foi selecionado por ser considerado atual e de interesse dentro do contexto de Ciência da Informação. Um grande número de pesquisadores tem utilizado RNA na recuperação de informação. É uma tentativa de simular a maneira pela qual são processadas informações no cérebro. Na aplicação de RNA relatada por Ferneda (2006), destaca-se o caráter dinâmico associado à recuperação de informação, uma vez que as representações dos documentos podem ser reavaliadas e alteradas de acordo com a especificação de relevância atribuída pelos usuários aos documentos recuperados. Entre os documentos recuperados podem surgir alguns que não estão diretamente relacionados aos termos de busca, mas que foram inferidos durante a pesquisa e possuem algum grau de relacionamento com a necessidade de informação do usuário.

3.1 Seleção do Material Empírico

Foram selecionados cinco textos nacionais para o desenvolvimento do trabalho, os quais abordam a utilização de Redes Neurais Artificiais em SRI. Foram escolhidos, para o experimento aqui relatado, apenas cinco artigos por se tratar de uma prova de conceitos para o desenvolvimento de um projeto de pesquisa maior em que, pretende-se demonstrar, utilizando-se um corpus de maior porte, o potencial da metodologia proposta para a recuperação da informação. Os cinco artigos selecionados discutem temáticas relacionadas ao processo de recuperação de informação utilizando-se Redes Neurais Artificiais.

Na fase de exploração do material, observou-se que quatro textos abordavam estudos de caso específicos para a aplicação de RNA e um texto apresentava as principais iniciativas de se aplicar os conceitos de RNA aos sistemas de recuperação de informação. Optou-se por mantê-lo dentre os documentos selecionados por se tratar de um documento importante de contextualização do tema e apresentação histórica do que podemos chamar de busca por um modelo computacional que simulasse o funcionamento das células do cérebro.

3.2 Análise de conteúdo: desenvolvimento do método

A análise de conteúdo foi realizada em duas etapas. A primeira delas foi realizada identificando-se, nos quatro textos que apresentam estudos de caso, três categorias, subdividas em subcategorias. Na segunda etapa, foi criada a tabela de frequência de ocorrência das categorias e subcategorias nos quatro artigos selecionados.

Para identificar as categorias e subcategorias, foram definidas unidades de registro e unidades de contexto. Foram consideradas como unidades de registro as palavras que contribuem para a criação das categorias. As unidades de contexto são os parágrafos ou segmentos da mensagem utilizados como unidades de compreensão para identificar as palavras que foram definidas como unidades de registro.

As categorias assim identificadas foram “característica funcional da RNA”, “arquitetura da RNA” e “algoritmo de aprendizado”.

Na categoria “característica funcional da RNA”, foram encontradas as subcategorias categorização de documentos, reconhecimento e recuperação de imagens e análise de similaridades de textos. A categoria “arquitetura da RNA” se subdividiu em redes MLP, *Multilayer Perceptron* (Redes Perceptron Multicamadas), e redes SOM, *Self-organizing Maps*. Na categoria “algoritmo de aprendizado”, foram selecionadas as subcategorias *Back Propagation* (Retro-propagação) e *Cascade Correlation* (Correlação em cascata). O Quadro 1 relaciona as categorias e as subcategorias encontradas nos artigos.

As RNAs são utilizadas, basicamente, em problemas de aproximação, predição, classificação, categorização e otimização. A grande maioria das aplicações reportadas na literatura se concentra na área industrial (MEIRELES; ALMEIDA; SIMÕES, 2003). Em um Sistema de Recuperação de Informação, SRI,

as RNA podem ser usadas para realizar o casamento de padrões entre as consultas e os documentos do acervo do sistema (SOUZA, 2006).

Categorias	Subcategorias
Característica funcional	Categorização de documentos
	Reconhecimento de imagens
	Análise de similaridade
Arquitetura	MLP
	SOM
Algoritmo de aprendizado	<i>Back Propagation</i>
	<i>Cascade Correlation</i>

Quadro 1 Categorias e subcategorias identificadas nos textos

Na primeira categoria, foram identificadas, nos artigos, três características funcionais. Dois dos artigos apresentam estudos relacionados à primeira subcategoria, denominada “categorização de documentos”. São experimentos que envolvem a descoberta de características estaticamente relevantes de um determinado conjunto de dados e como estes podem ser divididos em classes (REZENDE, 2003). Neste tipo de problema, somente os dados de entrada são conhecidos, cabendo à rede encontrar as categorias presentes neste conjunto de dados.

Um dos artigos discutiu uma abordagem em processamento de sinais e imagens, sendo classificado na segunda subcategoria “reconhecimento de imagens”. Esta subcategoria pode abranger aplicações de uma RNA responsável pelo reconhecimento de caracteres escritos a mão, pela compactação de dados, pela codificação, pelo reconhecimento da fala entre outras aplicações.

O último artigo, classificado na terceira subcategoria “análise de similaridade”, descreveu aplicações que envolvem análises de similaridade comuns em processos de classificação ou de reconhecimento de padrão.

A arquitetura de uma RNA, definida como a segunda categoria, pode ser caracterizada por parâmetros como número de camadas da rede, número de nós em cada camada, tipo de conexão entre os nós e topologia da rede. Duas subcategorias foram encontradas nos textos, “MLP” e “SOM”.

As redes MLP são redes que apresentam, pelo menos, uma camada intermediária e o processamento realizado por cada nó é definido pela combinação dos processamentos realizados pelos nós da camada anterior que estão conectados a ele.

As redes SOM foram desenvolvidas na década de 80. São estruturas inspiradas no mapa topológico presente no córtex cerebral. A rede SOM funciona, basicamente, construindo-se um mapa topológico onde os nós que estão topologicamente próximos respondem de modo semelhante a padrões de entrada semelhantes. Na literatura, existem exemplos de processos de categorização de dados usando as redes SOM (BRAGA; CARVALHO; LUDERMIR, 2000).

A terceira categoria encontrada foi “algoritmo de aprendizado”. As RNAs caracterizam-se pelo aprendizado por meio de exemplos. Para um determinado conjunto de dados, o algoritmo de aprendizado deve ser responsável pela

adaptação dos parâmetros da rede para que, após um número finito de iterações do algoritmo, a RNA apresente uma solução para o problema proposto. As duas subcategorias apresentadas, nesta categoria, foram “*Back Propagation*” e “*Cascade Correlation*”.

O algoritmo mais conhecido para treinamento da rede MLP é o algoritmo *Back Propagation*. A maioria dos métodos de aprendizado para RNAs do tipo MLP utiliza variações deste algoritmo (BRAGA; CARVALHO; LUDERMIR, 2000), sendo o seu desenvolvimento um dos principais motivos para o ressurgimento do interesse em RNAs.

As redes construtivas são aquelas em que a estrutura da rede é definida dinamicamente durante o seu treinamento. Esta mudança pode envolver a eliminação e ou o acréscimo de nós. A rede *Cascade Correlation* é um exemplo de rede construtiva.

A tabela contendo as categorias, as subcategorias e a frequência em que foram identificadas nos documentos será apresentada na seção de experimentos e resultados (Tabela 1, p. 87).

3.3 Análise de citações: desenvolvimento do método

Para a realização da análise de citações, foram identificadas, nas citações dos artigos selecionados, cinco variáveis, quais sejam: autor, ano de publicação, temática, idioma e país.

Os artigos e as suas referências bibliográficas foram cadastrados em uma base de dados, construída utilizando-se o *Microsoft Office Access 2003*. A partir do cadastro das variáveis citadas, foi possível disponibilizar consultas de ocorrência de citações por autor, de citações por referência, de referências por temática, de citações por ano, de citações por idioma e de citações por país.

Observando-se os autores citados, é possível relacioná-los às pesquisas realizadas em diferentes temas e identificar aqueles considerados referências em suas respectivas áreas. As consultas disponíveis podem identificar a produtividade de autores e países e a língua mais utilizada nas publicações da área.

Utilizando-se o ano de publicação, podem ser constatadas tendências nas publicações do passado e nas atuais e realizadas previsões de tendências para os próximos anos.

O agrupamento por temática identifica frentes de pesquisa e crescimento do conhecimento em determinada disciplina, destacando obsolescência e dispersão da literatura científica.

4 RESULTADOS

Esta seção apresenta os experimentos e resultados encontrados em cada uma das análises realizadas.

4.1 Análise de conteúdo: apresentação e análise dos resultados

Para a definição das unidades de contexto e das unidades de registro, foram utilizados os títulos, os resumos e as palavras-chave de cada um dos artigos.

Os quatro artigos que descrevem estudos de caso relacionados à aplicação de RNA estão numerados de acordo com o Quadro 2.

Número do Artigo	Título do artigo
1	Categorização Automática de Documentos: Estudo de Caso
2	O Uso de Técnicas de Aprendizado de Máquina na Categorização de Documentos
3	Reconhecimento e Recuperação de Imagens utilizando Redes Neurais Artificiais do Tipo MLP
4	Um mecanismo de recuperação de informação na <i>web</i> baseado em Redes Neurais

Quadro 2 Numeração dos artigos

O Quadro 3 apresenta a frequência de ocorrência das categorias e subcategorias nos artigos citados na figura anterior. A ocorrência é determinada pelo caractere “x”.

Categorias	Característica funcional			Arquitetura		Algoritmo de aprendizado	
	Categorização de documentos	Reconhecimento de imagens	Análise de similaridade	MLP	SOM	<i>Back Propagation</i>	<i>Cascade Correlation</i>
Artigos							
1	x	---	---	x	x	---	---
2	x	---	---	x	---	x	x
3	---	x	---	x	---	---	x
4	---	---	x	x		x	---
Totais	2	1	1	4	1	2	2

Quadro 3 Categorização dos artigos

Um tratamento estatístico simples dos dados contidos neste quadro pode gerar algumas interpretações.

Observou-se que o percentual de 50% dos documentos descreve estudos de categorização de documentos.

A arquitetura MLP foi encontrada em 100% dos documentos. Uma possível justificativa para esse resultado poderia ser o seu poder computacional muito maior do que aquele apresentado pelas redes sem camadas intermediárias. A arquitetura SOM foi identificada em apenas um dos artigos.

Em relação ao algoritmo de aprendizado, percebe-se, nessa amostragem, que há aplicação similar tanto do algoritmo *Back Propagation* quanto do algoritmo *Cascade Correlation*.

A análise dos dados contidos na tabela indica uma tendência de desenvolvimento de determinada área de conhecimento. São interpretações relevantes obtidas pela comparação dos dados contidos na amostragem e que poderão ser úteis ao usuário.

4.2 Análise de citações: apresentação e análise dos resultados

A primeira análise realizada foi a do número de ocorrências de citação por autor. A Figura 2 mostra que existem três autores com um número de citações em destaque e os comentários seguintes estão concentrados nos trabalhos destes três pesquisadores. Observando-se os currículos desses autores, Antonio de Pádua Braga, Teresa B. Ludermir e André P. de L. F. de Carvalho, constata-se que os três autores mais citados trabalharam juntos em algum instante de suas vidas. Foram contemporâneos durante os cursos de mestrado ou doutorado ou trabalharam nos mesmos departamentos ou nos mesmos grupos de pesquisa.

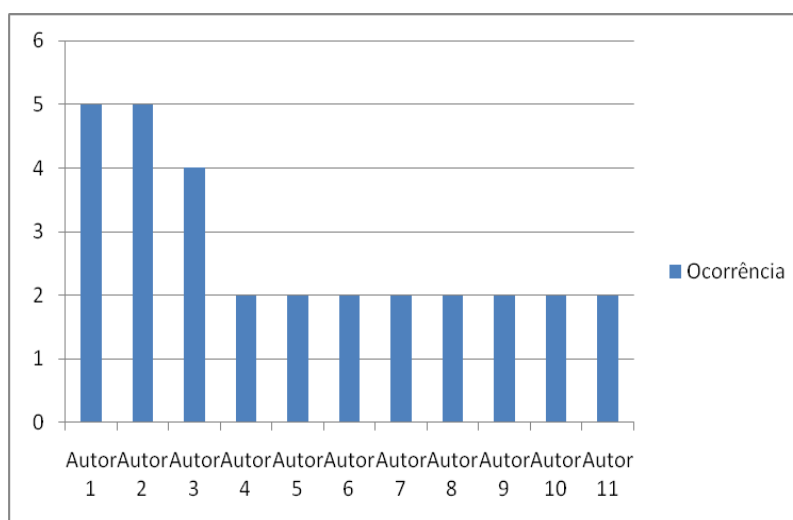


Figura 2 Ocorrência de citações por autor

O livro “Redes Neurais Artificiais: teorias e aplicações”, citado quatro vezes, conforme mostra a Figura 3, apresenta aspectos básicos e aplicações de RNAs. Utilizando-se de uma introdução gradual aos principais fundamentos de uma RNA, apresenta os modelos mais difundidos da tecnologia. É importante ressaltar que, à medida que avança em seu estudo, o leitor adquire a autonomia necessária para desenvolver suas próprias aplicações. As características mencionadas justificam a utilização do livro como bibliografia básica em qualquer pesquisa ou projeto da área.

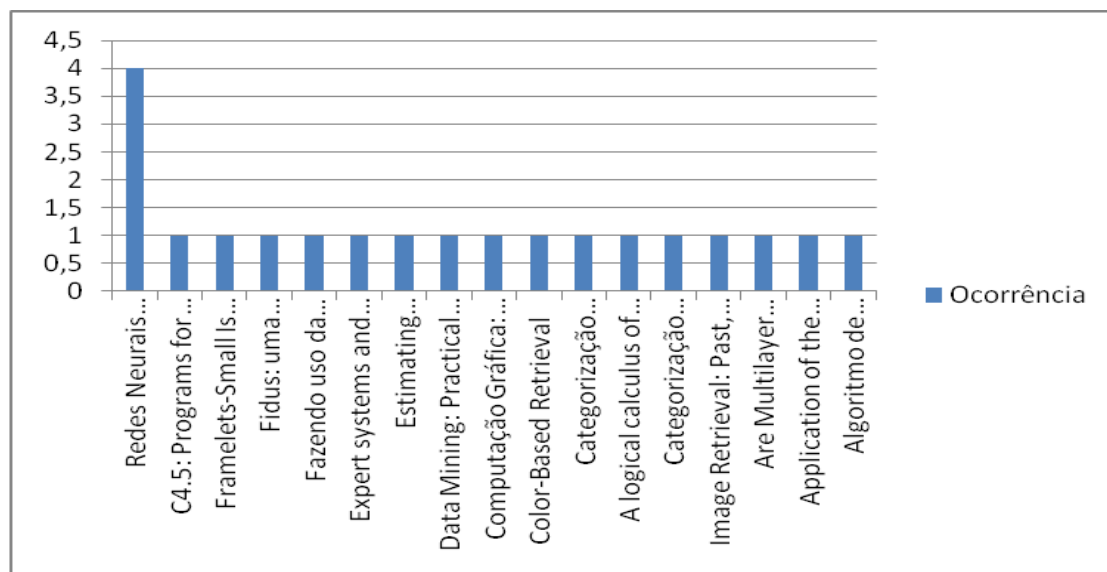


Figura 3 Ocorrência de citações por referência

A Figura 4 apresenta o número de citações por ano. Os artigos selecionados para a realização do trabalho foram publicados em 2006, 2005, 2002 e, dois deles, em 2003. Os artigos referenciados concentram-se entre os anos de 1998 e 2002.

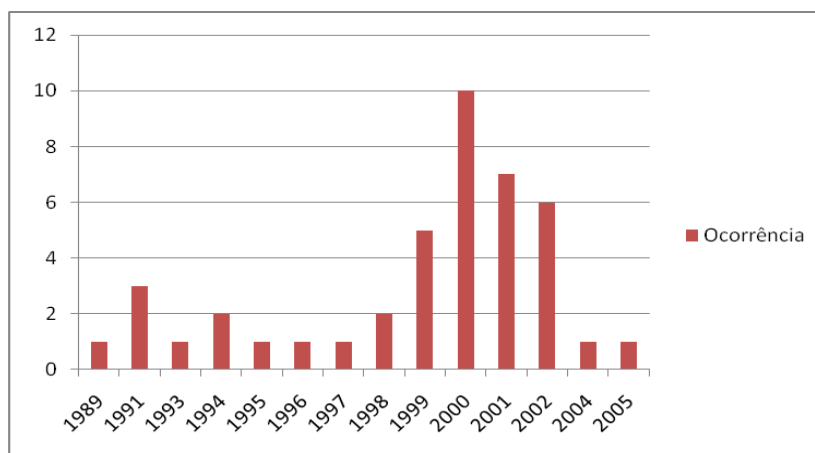


Figura 4 Ocorrência de citações por ano

Os artigos referenciados foram agrupados em dez categorias, descritas pela Tabela 1, e a frequência de ocorrência das categorias, em valores percentuais, está ilustrada pela Figura 5. Estas categorias foram definidas pela avaliação do título da publicação referenciada.

Os idiomas utilizados nas referências foram o português e o inglês. Das referências cadastradas, vinte publicações utilizaram o português e vinte e oito, o inglês. Os autores, em geral, submetem seus artigos em língua inglesa por possibilitar um maior nível de repercussão de seu trabalho, considerando-se ser o inglês uma das línguas mais faladas e entendidas no mundo. O número elevado

de citações em português pode ser justificado por terem sido selecionados, para o trabalho, apenas artigos nacionais.

Tabela 1 Categorização da Temática das Referências

Categorias	Número de artigos	Percentual
Recuperação de informação	14	29%
Redes Neurais Artificiais	10	21%
Categorização de documentos	7	15%
Aprendizado de máquina	6	13%
Processamento de informação	3	6%
Classificação de documentos	3	6%
Reconhecimento de Imagens	2	4%
Otimização	1	2%
Organização de documentos	1	2%
Inteligência Artificial	1	2%

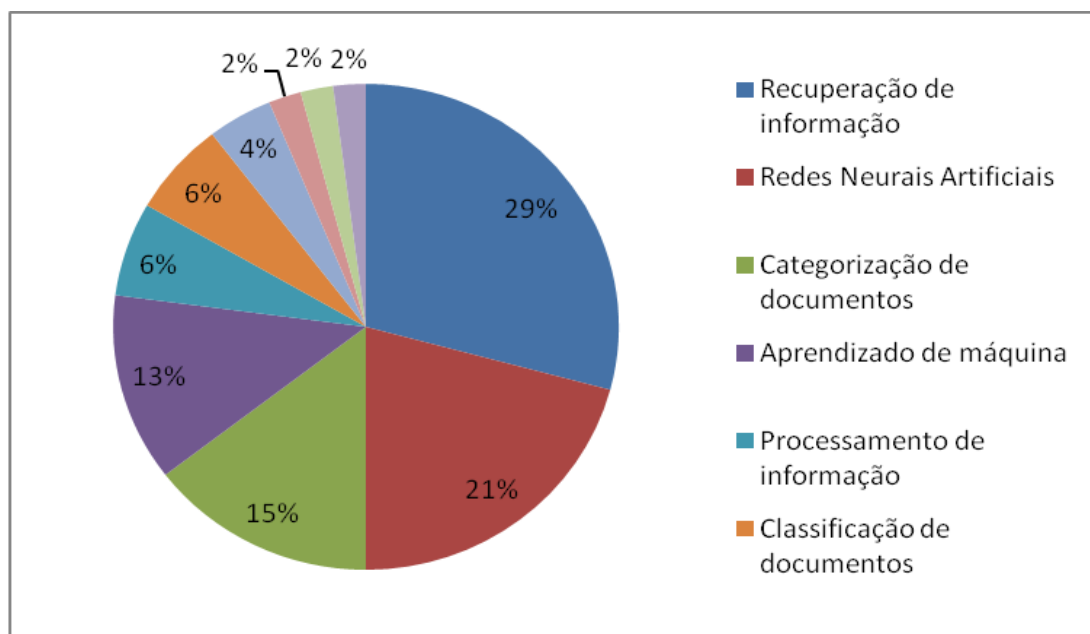


Figura 5 Referências agrupadas por temática

A Figura 6, apresentada a seguir, ressalta a participação dos Estados Unidos no número de publicações abordando o tema estudado. O grande número de citações realizada no Brasil, novamente, se justifica pelo fato dos artigos selecionados serem de autoria de brasileiros.

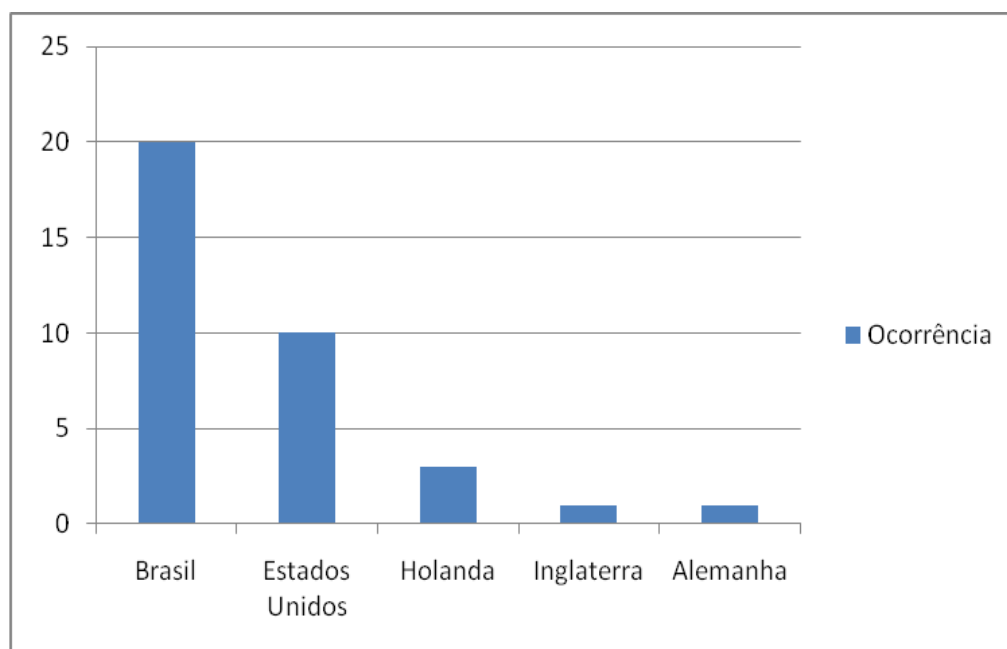


Figura 6 Ocorrência de citações por país

Os resultados apresentados identificam o crescimento de publicações em áreas específicas do conhecimento, apontando dispersão e agrupamento da literatura científica no tempo e no espaço físico. Observando-se os gráficos apresentados, é possível encontrar grupos de pesquisadores e países que vêm se destacando no desenvolvimento de uma disciplina, disponibilizando, para o usuário informações que vão além do conteúdo avaliado.

5 CONSIDERAÇÕES FINAIS

Este artigo teve o duplo propósito de apresentar uma demonstração da aplicação de análise de conteúdo e de análise de citações em um conjunto de documentos e de mostrar o potencial da utilização conjunta desses métodos para a recuperação de informação. O estudo de caso apresentado, por utilizar uma amostragem pouco significativa, não teve a pretensão de se tornar conclusivo em suas análises no que diz respeito às tendências de desenvolvimento da área avaliada, ou seja, Redes Neurais Artificiais. Além de ser pequeno o número de artigos analisados, por terem sido selecionados apenas textos nacionais, não se pode esperar que os resultados das análises reflitam de maneira acurada a utilização de RNAs nas aplicações industriais ou nas pesquisas realizadas na área acadêmica. A proposta central do trabalho aqui relatado foi criar uma prova de conceito para validar a utilização conjunta dos dois métodos para recuperação de informação, que será, futuramente, empregada em uma amostra com um número maior de artigos. Pretende-se, como continuidade desta proposta inicial, incluir outros padrões para análise de citações, que utilizem, por exemplo, métricas de centralidade ou de agrupamento e outros padrões inspirados nas metodologias de análises de redes sociais.

Na fase de pré-análise, foram selecionados cinco textos relacionados ao tema proposto e sistematizadas as idéias iniciais para que fosse gerado um

esquema do desenvolvimento das próximas etapas. Foram geradas categorias semânticas para a realização da análise de conteúdo. Para a análise de citações, foram identificados os países dos quais emanam as referências levantadas, o número de referências em cada uma das categorias determinadas pelo título da publicação referenciada, as referências mais citadas nos textos selecionados e a distribuição cronológica do número de citações. Os dados foram submetidos a análises qualitativas e quantitativas que trataram a informação contida nos artigos possibilitando conclusões que extrapolam o conteúdo dos artigos do corpus empírico. A análise de conteúdo complementou o estudo realizado pela análise de citações ao identificar os tópicos de uma determinada área de conhecimento sobre os quais os pesquisadores têm desenvolvido suas publicações. As conclusões apresentadas pela avaliação da tabela de frequência de ocorrência das categorias e subcategorias semânticas encontradas tornaram os resultados estatísticos obtidos na análise de citações ainda mais específicos, permitindo caracterizar não apenas o crescimento da literatura de uma área, mas, também, de suas especificidades, descritas pelas categorias encontradas na análise de conteúdo. Trata-se, portanto, da inferência de conhecimentos relativos aos artigos não necessariamente declarados no texto. Os dois métodos discutidos podem ser considerados ferramentas refinadas de recuperação de informação. Propõe-se, portanto, uma integração dos dois métodos e o uso de técnicas quantitativas e qualitativas na busca de informações relevantes que estão além do conteúdo avaliado.

O estudo apresentado mostrou, ainda, que ambos os métodos apresentam desafios para a sua aplicação na recuperação da informação do modo aqui proposto. A análise de conteúdo requer, para determinação das categorias semânticas aplicáveis aos documentos a serem analisados, o conhecimento do tópico específico em questão, que no caso aqui apresentado foram as RNAs. Outra dificuldade apresentada para a aplicação prática de ambos os métodos é a identificação, dentro da literatura da área escolhida, de documentos que, por espelharem os conceitos e os avanços tecnológicos presentes em um campo de estudo, sejam adequados para compor a coleção a ser analisada. Estas observações implicam na necessidade de se incorporar, na metodologia apresentada, mecanismos para lidar com estas dificuldades.

Conclui-se que utilizar análise de conteúdo e análise de citações, de modo integrado, mesclando metodologias qualitativas e quantitativas na análise dos resultados, é uma proposta relevante para a obtenção de meta-informações sobre coleções de documentos. A triangulação de técnicas, utilizando-se de inspiração qualitativa e quantitativa, vem de encontro ao caráter transdisciplinar da Ciência da Informação, que perpassa campos diferentes do conhecimento e reúne profissionais de diferentes formações.

AGRADECIMENTOS

As autoras gostariam de agradecer a Daniela Lucas da Silva pela elaboração e cessão da interface de gerenciamento da base de dados utilizada neste trabalho.

REFERÊNCIAS

ALVARENGA, Lídia. Bibliometria e arqueologia do saber de Michel Foucault: traços de identidade teórico-metodológica. *Ciência da Informação*, Brasília, v. 27, n. 3, set. 1998.

_____. *Citações como método de trabalho na bibliometria: notas de aula da disciplina Análises no Processo Classificatório para Sistemas de Recuperação de Informação*, PPGCI-UFMG, 24f, 2009.

BARDIN, Laurence. *Análise de conteúdo*. Lisboa: Edições 70, 1977. 226p.

BARITE, Mario Guido. The notion of "Category": its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge organization*, v. 27, n.1/n.2, p. 4-10, 2000.

BENDER, Túlio Cleber; OSÓRIO, Fernando Santos. *Reconhecimento e recuperação de imagens utilizando redes neurais artificiais do tipo MLP*. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL, 4. 2003, Campinas, SP. Anais... Campinas: SBC, 2003. v. 1. p. 1-10.

BORGMAN, Christine L.; FURNER, Jonathan. Scholarly Communication and Bibliometrics. *Annual Review of Information Science and Technology*, White Plains, v. 36, n. 1, p. 2-72, 2002.

BRAGA, Gilda Maria. Relações bibliométricas entre a frente de pesquisa e revisões de literatura: estudo aplicado à Ciência da Informação. *Ciência da Informação*, Rio de Janeiro, v. 2, n. 1, p. 9-26, 1973.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDERMIR, T. B. *Redes neurais artificiais: teoria e aplicações*. Rio de Janeiro: LTC, 2000.

CORRÊA, R. F.; LUDERMIR, T. B. *Categorização automática de documentos: estudo de caso*. In: BRAZILIAN SYMPOSIUM ON NEURAL NETWORKS, 7., 2002, Porto de Galinhas. *Proceedings...* Recife: SBC, 2002. v. 2.

CRONIN, B. *The citation process*. London: Taylor Graham, 1984.

FERNEDA, Edberto. Redes neurais e sua aplicação em sistemas de recuperação de informação. *Ciência da Informação*, Brasília, v.35, n.1, p.25-30, jan./abr. 2006.

JACOB, Elin. Classification and categorization: a difference that makes a difference. *Library trends*, v. 52, n.3, p. 515-540, 2004.

LEAL, Igor Campos. *Análise de citações da produção científica de uma comunidade: a construção de uma ferramenta e sua aplicação em um acervo de teses e dissertações do PPGCI-UFMG*. 2005. 94f. Dissertação (Mestrado em Ciência da Informação)- Universidade Federal de Minas Gerais. Belo Horizonte.2005.

MEIRELES, Magali Rezende Gouvêa; ALMEIDA, Paulo Eduardo Maciel; SIMÕES, Marcelo Godoy. A comprehensive review for industrial applicability of

artificial neural networks. *IEEE Transactions on Industrial Electronics*, EUA, v. 50, n. 3, p. 585-601, 2003.

PINHEIRO NETO, Hermes; VIEIRA, Sérgio Ricardo T. *Um mecanismo de recuperação de informação na web baseado em redes neurais*. Científico (Faculdade Rui Barbosa), v. 2, p. 4, 2005.

REZENDE, Solange Oliveira. *Sistemas Inteligentes: fundamentos e aplicações*. Barueri, SP: Manole, 2003.

ROJAS, Miguel Ángel Rendón. Relación entre los conceptos: información, conocimiento y valor. Semejanzas y diferencias. *Ciência da Informação*, Brasília, v. 34, n. 2, p. 52-61, maio/ago. 2005.

SANTOS, Boaventura de Souza. *Um discurso sobre as ciências*. São Paulo: Cortez, 2003.

SANTOS, Cássia T.; OSÓRIO, Fernando S. O Uso de Técnicas de Aprendizado de Máquina na Categorização de Documentos. In: CONFERÊNCIA LATINO AMERICANA DE INFORMÁTICA, 29., 2003, La Paz. *Proceendigs...* 2003.

SENGUPTA, I. N. Bibliometrics, Informetrics, Scientometrics and Librametrics: An Overview. *Libri*, Munksgaard, v. 42, n. 2, p. 75-98, 1992.

SOUZA, Renato R. Sistemas de Recuperação de Informações e Mecanismos de Busca na *web*: panorama atual e tendências, *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 11, n.2, p.161-173, mai./ago. 2006.

VALENTIM, M. L. P. *Métodos qualitativos de pesquisa em Ciência da Informação*. São Paulo: Polis, 2005. 176p.

XAVIER, Beatriz Rêgo. As categorias de Aristóteles e o conhecimento científico. *Pensar*, Fortaleza, v. 13, n. 1, p. 57-64, jan./jun. 2008.

Title

Practical application of content analysis and citation analysis processes on Artificial Neural Networks related articles

Abstract

Citation analysis and content analysis rely on inference and interpretation to obtain relevant information contained in a set of documents. These methods are broadly discussed in literature, but articles in which their practical use are described and exemplified are few. This paper presents a case study which aims to demonstrate the use of these methods in the analysis of a set of documents and to show the potential of their use for information retrieval. Initially, the paper presents the theoretical approaches concerning the two methods and the characterization of the empirical portion of the study based on Artificial Neural Networks. Next, the semantic categories applied in the content analysis are described, along with the variables chosen for the citation analysis. Based on quantitative and qualitative analysis, the application of the two methods is exemplified

showing that they can be used in an integrated way to elicit metadata on a document collection.

Keywords: Information Science. Content Analysis. Citation Analysis. Categorization. Artificial Neural Networks.

Titulo

Aplicación práctica de los procedimientos de análisis de contenido y de análisis de citas en trabajos relacionados con las Redes Neuronales Artificiales

Resumo

Los análisis del contenido y de citas son metodologías que se basan en la inferencia de la interpretación para obtener informaciones relevantes contenidas en un conjunto de documentos. Estos métodos son ampliamente discutidos en la literatura, pero solo en algunos artículos se describe e ilustra su aplicación en estudios de casos específicos. Este artículo presenta un estudio de caso que tiene el doble propósito: de demostrar el uso de métodos reportados para el análisis de un conjunto de documentos y mostrar el potencial de su uso combinado para recuperación de información. El artículo presenta, inicialmente, los planteamientos teóricos de los dos métodos y la caracterización del material empírico de la investigación, que utilizó artículos sobre Redes Neuronales Artificiales. A la continuación, se describen las categorías semánticas utilizadas en el análisis de contenido y las variables seleccionadas para el análisis de citas. Por medio del análisis cuantitativo y cualitativo, se tiene ejemplos de la utilización de análisis de citas y análisis de contenido y demuestran, además, que se pueden utilizar de manera integrada para obtener los metadatos de las colecciones de documentos.

Palabras clave

Ciencias de la Información. Análisis de contenido. Análisis de citas. Categorización. Redes Neuronales Artificiales.

Recebido em: 11/04/2010

Aceito em: 11/01/2011
