

ALEXANDRA MOREIRA¹

LÍDIA ALVARENGA²

**UMA PROPOSTA DE METODOLOGIA PARA ANÁLISE DE CONTEÚDO;
MÉTODO ANALÍTICO-SINTÉTICO COM O AUXÍLIO DO SISTEMA KWIC**

Grupo temático: Representação do conhecimento/Indexação/Teoria da Classificação

RESUMO

A análise de conteúdo é uma técnica utilizada na investigação do conteúdo das comunicações que se baseia no enquadramento dos elementos da comunicação em categorias seguida de uma análise quantitativa e qualitativa. Os usuários desta técnica podem se beneficiar de ferramentas automáticas que permitam uma extração dos elementos lingüísticos e sua manipulação dentro do contexto de ocorrência, uma vez que deste modo é possível uma melhor visualização do papel de cada elemento. Ferramentas desse tipo são denominadas de *concordância* e dentre elas se destaca o KWIC (*Keywords in context*), muito utilizada na área de lingüística de *corpus*³. Este trabalho apresenta uma metodologia de análise de conteúdo que faz uso de um sistema KWIC e do método analítico-sintético, proposto por Dahlberg. A metodologia foi empregada em uma pesquisa onde se buscava um esclarecimento terminológico e a caracterização de dois instrumentos utilizados na organização do conhecimento: tesaurus e ontologias. A metodologia mostrou-se bastante útil na análise das definições sobre os dois instrumentos.

¹ Alexandra Moreira. Mestranda da Escola de Ciência da Informação da UFMG.
xandramoreira@yahoo.com.br.

² Lídia Alvarenga. Mestre em Ciência da Informação e Doutora em Educação. Professora do Programa de Pós-Graduação em Ciência da Informação da UFMG.
lidiaalvarenga@eci.ufmg.br.

³ A lingüística de corpus é a área da lingüística que se ocupa da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais, em formato legível por computador, que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

1 INTRODUÇÃO

A análise de conteúdo é uma técnica oriunda das ciências sociais e é utilizada na investigação do conteúdo das comunicações que se baseia no enquadramento dos elementos da comunicação em categorias seguida de uma análise quantitativa e qualitativa. Segundo MARCONI E LAKATOS (1982), a vertente quantitativa foi um acréscimo recente à técnica e que tornou a análise mais precisa, sem descartar a necessidade de existir uma análise qualitativa. Os usuários desta técnica podem se beneficiar de ferramentas automáticas que permitam uma extração dos elementos lingüísticos e sua manipulação dentro do contexto de ocorrência, uma vez que deste modo é possível uma melhor visualização do papel de cada elemento. Ferramentas desse tipo são denominadas de *concordância* e dentre elas se destaca o KWIC (*Keywords In Context*), muito utilizada nas áreas lingüística de *corpus* e na ciência da informação para a confecção de índices.

Este trabalho apresenta uma metodologia de análise de conteúdo que faz uso de um sistema KWIC e do método analítico-sintético, proposto por Dahlberg (1978). A metodologia foi empregada em uma pesquisa onde se buscava um esclarecimento terminológico e a caracterização de dois instrumentos utilizados na organização do conhecimento: tesouros e ontologias. Para tal, utilizou-se subsídios teóricos e materiais de fontes bibliográficas relacionadas às temáticas estudadas, englobando as áreas ciência da informação e ciência da computação, sendo que nesta última o enfoque foi a subárea da inteligência artificial. A partir do material coletado foram extraídas definições que foram analisadas de acordo com o método analítico-sintético, agrupando termos por categorias, por meio de um processo indutivo.

O objetivo deste artigo é descrever a metodologia utilizada e mostrar que o uso do método analítico-sintético auxiliado por um sistema de concordância produz uma ferramenta útil onde a técnica de análise de conteúdo faz-se necessária.

2 A ANÁLISE DE CONTEÚDO

As técnicas adotadas na pesquisa foram análise quantitativa e qualitativa do conteúdo. O uso da análise qualitativa, em conjunto com a quantitativa enriquece a análise dos dados que se pretende compreender. Os aspectos quantitativos são pontos de partida e de apoio para a análise e as interpretações qualitativas permitem uma melhor compreensão dos dados obtidos.

A técnica aqui descrita também pode ser vista como uma técnica de *análise de conteúdo*. Análise de conteúdo pode ser entendida como "uma técnica de pesquisa para descrição objetiva, sistemática e quantitativa do conteúdo evidente da comunicação" (BERELSON *apud* MARCONI & LAKATOS, 1982, p. 99). Já para ANDER-EGG (1978, p.178) é "a técnica mais difundida para investigar o conteúdo das comunicações de massas, mediante a classificação, em categorias, dos elementos da comunicação". Assim, em relação a este trabalho o conteúdo das definições, tanto de ontologias e de tesouros, seriam analisados utilizando-se de categorias sistemáticas de termos emergentes das definições estudadas, possibilitando uma abordagem quantitativa subsequente.

Berelson apresenta uma série de princípios específicos em relação às questões que podem ser estudadas, ao aplicar esta técnica, tais como: Questões a respeito das características do conteúdo; questões relativas ao produtor ou motivos de conteúdo; e questões em relação à audiência ou efeitos de conteúdo. Neste trabalho se justificaria o

primeiro item: questões a respeito das características do conteúdo tornando possível comparar níveis dos dois conceitos: tesouros e ontologias.

Para tanto, a análise de conteúdo considerou algumas restrições, segundo SELLTIZ *et al.* (1965, p. 392):

1. As categorias da análise usadas para classificar o conteúdo são definidas clara e explicitamente para que outros indivíduos possam aplicá-las ao mesmo conteúdo, a fim de verificar as conclusões;
2. analista não é livre para selecionar e registrar simplesmente aquilo que chama sua atenção, por ser interessante, mas deve classificar metodicamente todos os assuntos importantes, em sua amostra;
3. Certo processo quantitativo é usado para proporcionar a média da importância e ênfase da matéria de várias idéias verificadas e para permitir confrontos com outras amostras do material.

Esta etapa seria a fase quantitativa compreendendo uma análise objetiva, que permitirá uma comparação dos resultados alcançados. A etapa seguinte seria uma discussão qualitativa dos resultados obtidos na fase quantitativa sobre os objetos estudados, utilizando-se da interpretação e discussão dos resultados alcançados tendo como subsídio o conhecimento prévio sobre a temática estudada, tal como expresso em capítulo específico.

3 MÉTODO ANALÍTICO-SINTÉTICO

O método proposto por DAHLBERG (1978) para análise e estruturação de conceitos baseia-se na análise das proposições verdadeiras sobre o conceito, gerando um

conjunto de características, que tem como característica mais genérica uma categoria, e na síntese destas características na forma de um termo ou nome cujo significado está estabelecido precisamente por meio de uma definição. Por envolver tanto uma etapa de análise como uma etapa de síntese o método é denominado de analítico-sintético. Este método provê uma forma segura para compressão da intensão de um conceito para sua inserção em uma estrutura de conceitos de um dado domínio.

Para cumprir o objetivo geral de determinar o significado dos termos “ontologia” e “tesauro” dentro nas áreas descritas é preciso usou-se um método que permitisse o registro dos significados subjacentes a cada definição corrente e sua posterior comparação; o método escolhido para essa tarefa foi o método analítico-sintético

O método analítico-sintético tem sido empregado dentro da ciência da informação em situações que envolvem a construção de estruturas conceituais, como a construção de tesouros, como em Gomes (1990), ou em situações que envolvam a compressão e definição adequada de um conceito como em Alvarenga (1993).

Em seu trabalho Alvarenga utiliza-se do método analítico-sintético para analisar e propor uma definição para o conceito de “publicações oficiais”. Devido a analogia com o presente trabalho, os passos utilizados por Alvarenga no desenvolvimento de seu trabalho foram adaptados para esta pesquisa, incluindo-se a aplicação do KWIC para separar os termos constantes das definições. A seguir apresenta-se os passos da pesquisa de forma detalhada.

- Levantamento em separado das definições e conceitos sobre ontologias e tesouros. Nesta etapa foram extraídas da literatura as definições e conceitos sobre ontologia na área da ciência da computação(CC) e de tesouros na ciência da informação(CI).

- Submissão das definições a um programa para produção de uma ordenação alfabética de palavras de acordo com sistema de indexação KWIC (*Keywords in context*) que é uma forma de análise de texto onde as palavras são destacadas e listadas alfabeticamente, mantendo toda a frase e preservando o contexto original, ou seja, as palavras que antecedem e sucedem a palavra destacada. Nessa fase foram eliminadas palavras não significativas (*stopwords*). No resultado do KWIC também aparece o número total de ocorrência da palavra.
- Agrupamento dos termos em categorias ou classes amplas. As categorias foram determinadas pelas semelhanças ou equivalências semânticas entre os termos. Esta fase demandou uma tentativa de juntar num mesmo grupo termos que possuíssem conteúdos semânticos de uma mesma categoria.
- Discussão sobre as categorias. Nesta etapa discute-se as categorias escolhidas bem como os atributos e características a elas relacionadas.
- Transcrição em uma matriz de ocorrência dos conceitos ocorridos em cada categorias tanto sobre tesouro como sobre ontologias. Esta matriz torna-se de fundamental importância para que possa ser feita uma análise comparativa do conteúdo proposto;
- Análise das definições e conceitos segundo a ocorrência das categorias determinadas na fase anterior. Desta forma é possível verificar diferenças e semelhanças entre os dois conceitos.

- Uma discussão ampla dos resultados à luz do conhecimento prévio, compreende a abordagem qualitativa prevista nessa pesquisa levando a conclusões.

4 PROCESSO DE SELEÇÃO DOS TEXTOS

Devido a grande quantidade de documentos não seria possível analisar toda a bibliografia relacionada com os temas durante o período de desenvolvimento da presente pesquisa de mestrado. Portanto, foi necessário o estabelecimento de critérios norteadores para a seleção dos documentos, optando-se por selecionar os textos mais relevantes encontrados na literatura.

A atribuição de relevância a documento é uma tarefa difícil e envolve critérios polêmicos e/ou subjetivos. Acrescente-se a isso a necessidade de estabelecer critérios suficientemente sólidos para tornar o trabalho coerente e aceito pelas duas comunidades acadêmicas envolvidas.

Na fase de elaboração do projeto de pesquisa foram contatados alguns especialistas que, em consultas informais, indicaram os autores que mais se destacavam sobre o tema. Na área da ciência da computação contou-se com o auxílio de professores do Departamento de Informática da Universidade Federal de Viçosa. No caso dos tesouros, como se trata de um instrumento oriundo do contexto da ciência da informação, não foi necessário recorrer a especialistas de outras unidades acadêmicas, fora da Escola de Ciência da Informação da UFMG.

Pode-se observar que existe uma diferença na natureza do material colhido nas duas áreas. Na ciência da informação o termo tesouro possui um significado mais consolidado

em razão dos anos de pesquisa e de utilização. Existem inúmeros documentos, artigos científicos, normas e manuais que tratam do assunto sob diversos ângulos.

Já na área da ciência da computação pode-se observar que o termo ontologia encontra-se em fase de amadurecimento. O termo possui várias definições e ainda não existe uma metodologia amplamente aceita que guie a elaboração de ontologias. Logo, no caso da pesquisa sobre ontologia, tornou-se óbvia a necessidade do estabelecimento de critérios claros para a seleção dos documentos e definições sobre o assunto.

O critério adotado para a seleção de artigos da ciência da computação relacionados com ontologias, para a fase de desenvolvimento da pesquisa, foi o número de citações recebidas. A frequência de citações foi obtida por meio de consultas ao site da CiteSeer. A CiteSeer (também conhecida por ResearchIndex) é uma biblioteca digital de literatura científica que tem por objetivo auxiliar na disseminação de literatura científica relacionada principalmente com a ciência da computação. A biblioteca é mantida pelo NEC Research Institute e pode ser acessada no endereço eletrônico *citeseer.org*. Dentre os serviços fornecidos pelo CiteSeer destacam-se:

- Indexação autônoma de citações – o CiteSeer analisa de forma, por meio de um sistema computacional autônomo, documentos em formato eletrônico, retirando citações e armazenando-as em um índice para consultas posteriores. Os documentos podem ser fornecidos pelos próprios autores ou buscas automaticamente pelo sistema, por meio da Internet. O sistema opera de forma completamente autônoma, ou seja, sem nenhuma intervenção humana. O sistema é capaz de analisar diversas formas de citação. Lawrence (Lawrence et al., 1999) (Lawrence, 2001) apresenta maiores detalhes sobre o funcionamento do mecanismo de indexação autônoma de citações do CiteSeer;

- Análises estatísticas de citações para todos os documentos citados;
- Busca por palavras chaves com resultados ordenados por data, citações ou citações esperadas;

Em uma busca no CiteSeer utilizando a palavra “ontologies” foram retornados documentos que totalizaram 2674 citações. Os artigos foram ordenados por uma função que leva em conta o número de citações e a data de publicação do artigo, sendo que artigos mais recentes recebem um peso maior. A FIG. 1 mostra um segmento do resultado da consulta contendo os três artigos melhor classificados. A listagem dos resultados tal como mostra a FIG. 2 apresenta a distribuição das citações em função do ano de publicação dos artigos .

The diagram shows three callout boxes at the top: 'ranque' (rank), 'citações' (citations), and 'auto citações' (self-citations). Lines connect these boxes to a larger rectangular box containing a list of three articles. The first article is highlighted in blue text. The second article is in black text. The third article is in black text.

Context Doc 203.4 152 (1): Gruber, T.R. (1995). *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*, Int. Journal of Human-Computer Studies, Vol. 43, pp.907-928.

Context Doc 92.8 83 (4): Gruber, T.R. (1992). *ONTOLINGUA: A Mechanism to Support Portable Ontologies*, technical report, Knowledge Systems Laboratory, Stanford University, Stanford, United States.

Context Doc 63 43 (4): M. Uschold and Gruninger M. **Ontologies: Principles, methods and applications.** *Knowledge Engineering Review*, 11(2), 1996.

Figura 1 - Trecho do resultado da consulta usando a palavra “ontologies”

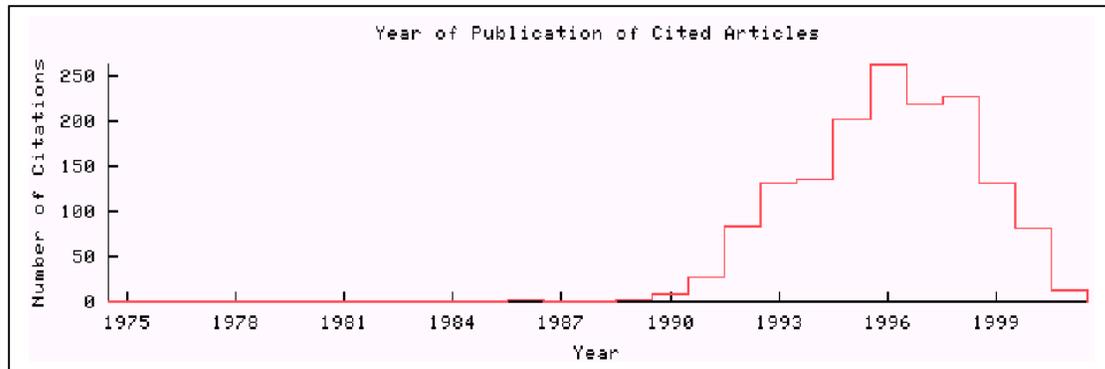


Figura 2 - Número de citações em função do ano de publicação dos artigos

A listagem retornada pela pesquisa realizada no CiteSeer revelou que após o artigo de posição 196 da listagem se seguiam artigos com apenas uma citação, demonstrando uma concentração de citações em poucos artigos. Além disso, alguns dos artigos mais citados pertencem a um mesmo autor, como pode ser observado na FIG. 1, o que parece indicar que poucos autores tratam o tema com profundidade. Nota-se também, um aumento do número de citações a partir da década de 90, mostrando que foi nesta época que surgiu o interesse pelo tema na ciência da computação. Pode-se observar também que, devido a erros de formatação, alguns artigos aparecem na listagem mais de uma vez, sendo interpretado pelo sistema como artigos diferentes. No entanto, este fato não teve qualquer influência sobre os resultados da pesquisa, exigindo apenas uma maior atenção na seleção dos artigos.

Em uma análise do conteúdo dos artigos pode-se notar que as pesquisas podiam ser classificadas, em sua maioria, em: pesquisas para desenvolvimento de ferramentas para criação de ontologias; metodologias para o desenvolvimento de ontologias; e aplicações baseadas em ontologias. Na primeira categoria enquadram-se produtos como Ontolingua (Gruber, 1992), ODE (Blazquez et al., 1998) Tadzebao e WebOnto (Domingue, 1998) que

podem ser usados para criar, manter e compartilhar ontologias. Na segunda categoria tem-se metodologias como a proposta por Guarino (1997), Gruninger e Fox (1995) e Uschold (1996). Na terceira categoria tem-se o uso de ontologias na gerência de conhecimento (Benjamins et al., 1998), no desenvolvimento de sistemas baseados em conhecimento (Van Heijst et al., 1997) e na integração de empresas (Fox e Gruninger, 1994). Os textos foram separados segundo esta classificação e selecionados aqueles que propunham definições originais para o termo ontologia.

Nota-se que o termo “ontologia” na área da filosofia não foi analisado por ser um termo bem consolidado pela comunidade de filósofos. Porém o entendimento do termo na filosofia foi utilizado nas discussões sobre as definições do termo na ciência da computação, visando verificar se existe uma preocupação, por parte da comunidade da ciência da computação, de fundamentar o termo filosoficamente.

5 MATERIAL

Neste trabalho o material empírico são as definições contidas nos textos selecionados sobre tesouros na CI e ontologias na CC. A análise das definições utilizadas neste estudo podem ser vistas como uma tentativa dos pesquisadores de estabelecer os elementos ou características, entendidos por eles como essenciais para a compreensão do objeto, sintetizando-os em um enunciado, ou em um conjunto de enunciados considerados uma definição.

O processo de seleção de definições já foi descrito anteriormente neste texto. No caso da temática ontologia ressalta-se nesta pesquisa a dificuldade o fato de que todos os

textos analisados são em inglês, havendo portanto, necessidade de se traduzir as definições. No caso dos tesouros esta dificuldade não existiu, devido a existência de autores consagrados que escrevem em português. Além disso, existem publicações em português de normas com definições sobre tesouros e traduções de textos de autores clássicos.

Para o termo tesouros foram encontradas seis (06) definições, sendo que duas definições constam do mesmo documento da UNESCO, que apresenta uma definição envolvendo um sentido estrutural e um sentido funcional ambas utilizadas neste trabalho. A quantidade de definições encontradas para tesouros foi a metade da quantidade de definições encontradas para o termo ontologia na ciência da computação. Provavelmente, isto se deve ao fato do termo tesouro estar mais consolidado na ciência da informação, fazendo com que exista uma menor quantidade de variações.

Para o termo ontologias foram utilizadas treze (13) definições extraídas de vinte e nove artigos retornados da pesquisa realizada no CiteSeer. A pesquisa foi realizada com a palavra “*ontologies*” e os artigos foram ordenados por frequência de citação. A razão de terem sido extraídas apenas treze definições de vinte e nove artigos deve-se ao fato de que nem todos os artigos apresentam definições, sendo que os que não apresentam definições adotam a proposta por Gruber. Esta é uma evidência da aceitação que a definição proposta por Gruber está obtendo dentro da ciência da computação. Além disso, alguns dos artigos são de um mesmo autor que, portanto, usa uma mesma definição em todos artigos.

6 O KWIC

As definições selecionadas foram submetidas ao programa ptx para produção do KWIC. As listagens obtidas por meio do ptx foram trabalhadas, gerando duas novas listas contendo conjuntos de termos agrupados em categorias para uma posterior análise quantitativa e qualitativa.

KWIC (*KeyWord In Context*) é uma ferramenta de indexação, enquadrada dentro da classe de ferramentas denominadas de *concordância*. Nas palavras de GILCHRIST (1971), uma concordância é "um índice para palavras ou frases de um livro. ... A palavra é, usualmente, exibida com o seu contexto". SARDINHA (2000) reforça a necessidade da apresentação do contexto de ocorrência da palavra quando diz que uma concordância "consiste numa listagem dos co-textos (palavras ao redor) nos quais um dado item (palavra isolada, composta, estrutura, etc.) ocorre." Segundo Sardinha, os tipos mais comuns de concordância são o KWIC e o KWOC (*KeyWord Out of Context*). No caso da concordância KWIC cada palavra relevante é mostrada no centro da listagem, ordenada alfabeticamente, ladeada pelas palavras antes e depois dela no texto. Já a concordância KWOC é uma listagem de títulos ordenados pelas palavras relevantes que neles ocorrem. O exemplo abaixo mostra o uso KWIC no destaque das palavras dentro de seu contexto.

No. da definição	No. de ocorrências		
		CONCEITUAL.....2	
4		Ontologia como um sistema	conceitual relacionado a uma base de conhecimento.
5		Ontologia como uma representação de um sistema	conceitual

CONCEITUALIZAÇÃO.....1			
6	Especificação formal e explícita de uma	conceitualização	
CONCEITUALIZAÇÕES....1			
12	Ontologias são acordos sobre	conceitualizações	compartilhadas.

Figura 3 - Exemplo do uso KWIC

Os instrumentos de concordância não são recentes. GILCHRIST (1971) menciona o uso de concordância feito por Cruden em 1738 sobre o texto da Bíblia. Ele menciona também que em 1856, Crestadoro sugeriu o de KWIC sobre títulos em um trabalho denominado de *The Art of Making Catalogues*. Atualmente, os sistemas de concordância, principalmente o KWIC, são utilizados para análise de texto, em estudos de lexicometria e no auxílio a buscas.

A concordância KWIC foi a selecionada para ser usada no presente estudo devido a sua facilidade para evidenciar as palavras dentro de seu contexto (neste caso as definições). O uso desta ferramenta, permitiu uma visualização mais rápida das palavras que ocorrem nas definições. O programa utilizado para a criação do KWIC foi o ptx (permuted indexer), versão beta 0.3, disponibilizado livremente sob a licença da Fundação de Software Livre, *Inc.* e distribuído junto com o sistema operacional Linux.

7 ESCOLHA DOS TERMOS

A seleção de termos obedeceu a abordagem onomasiológica⁴, ou seja, selecionou-se os termos que denotassem conceitos do domínio. Como consequência desta abordagem, alguns termos compostos não foram desmembrados em seus termos constituintes, uma vez que denotam conceitos relevantes no domínio ou unidades lexicais indissociáveis. Por outro lado, outros termos compostos foram desmembrados em seus termos constituintes uma vez que estes denotavam conceitos existentes no contexto. Por exemplo, o termo “base de conhecimento” não foi desmembrado em “base” e “conhecimento”, uma vez que o termo composto denota uma entidade conhecida pela comunidade de inteligência artificial. Já o termo “linguagem de informação” foi desmembrado nos termos “linguagem” e “informação” uma vez que o termo composto não representa algo mais significativo do que seus termos constituintes.

Utilizou-se definições de dicionário para descrever parte dos termos privilegiados da rotação pelo sistema KWIC. Os termos restantes foram descritos por meio de definições de autores da área ou por meio de descrições que capturam o significado do termo no contexto. Esta ação de ir buscar a noção do objeto a ser categorizado é compreendida como uma abordagem pragmática onde se registra interpretações adequadas ao contrário da abordagem epistemológica que são definidos no nível abstrato. No manual de tesouro monolíngues é indicado que apenas os termos básicos sejam definidos por causa do gasto de tempo nesta atividade, mas aqui tentou-se caracterizar todos os elementos utilizados.

Uma vez que o programa para geração do KWIC gera uma saída para palavras e não para termos, foi necessário uma edição da saída para incluir os termos que formassem uma unidade semântica. Isto foi feito, tomando-se cada palavra emitida pelo programa e

⁴ O estudo das expressões de que dispõe uma língua para traduzir determinada noção, e que parte, pois, do significado para estudar o significante. (Ferreira, 1975)

examinando se denotava um conceito do domínio. Caso contrário, era feita uma composição com as palavras ocorrendo em sua vizinhança e uma nova verificação, em relação ao sentido, era realizada.

8 CATEGORIAS

Após a seleção das definições e extração dos termos, estes identificados pelas unidades semânticas presentes, iniciou-se a etapa de elaboração das categorias. O objetivo deste processo é juntar, sob um mesmo grupo, conceitos semelhantes representados por termos distintos. As categorias foram determinadas por meio das semelhanças entre as características inerentes aos termos em um processo de indução. Há de se considerar que as definições foram retiradas de áreas distintas, com terminologias próprias e, portanto, em determinadas situações, apesar de usarem termos distintos podem estar se referenciando a um mesmo conceito. O uso de categorias visa equalizar os termos, uma vez que os termos abrigados sob a mesma categoria compartilham do significado apontado por ela.

Lista-se a seguir um trecho da matriz de ocorrências dos termos nas categorias levantada durante a pesquisa:

CATEGORIAS /TERMOS	ONTOLOGIA	TESAURO
CATEGORIA "OBJETO"		
BASE DE CONHECIMENTO	X	
DOCUMENTOS / DOCUMENTÁRIOS		X
ENTIDADES	X	
INFORMAÇÃO		X
CATEGORIA "CONTEÚDO SEMÂNTICO"		
CONCEITO (S)	X	X
CONCEITUALIZAÇÕES	X	
CONHECIMENTO	X	X
DEFINIÇÕES	X	
SEMÂNTICA / SEMANTICAMENTE		X
SIGNIFICADOS	X	
CATEGORIA "PROCESSOS"		
CONTROLE		X

DESCRIÇÃO / DESCRIVER / DESCRREVENDO	X	
DECLARATIVO	X	
ESPECIFICAÇÃO	X	
INDEXAÇÃO		X
INTERPRETAÇÕES	X	
REPRESENTAÇÃO	X	
TRADUÇÃO		X
CATEGORIA "AGENTES"		
USUÁRIO / USUÁRIOS		X
INDEXADORES		X

Tabela 1 - Trecho da distribuição dos termos por categoria

A partir deste processo realizou o processo quantitativo e qualitativo do estudo, que não será detalhado neste artigo.

9 RESULTADOS E CONCLUSÕES

A metodologia mostrou-se bastante útil na análise das definições sobre as duas temáticas. O sistema KWIC permitiu uma análise rápida e automatizada das palavras que ocorrem nas definições. Esta manipulação permitiu uma transposição mais tranqüila da análise de palavras para termos. Uma vez obtidos os termos, o método analítico-sintético mostrou-se uma ferramenta útil tanto para a elaboração das categorias quanto enquadramentos dos termos nas categorias. O uso da metodologias descrita no presente artigo possibilitou uma análise segura e sistemática sobre o conteúdo das definições sobre tesouros e ontologias. Em decorrência desta sistematização pode-se chegar a algumas conclusões válidas sobre as diferenças e semelhanças entre esses dois instrumentos usados na representação de conhecimento.

Sugere-se a aplicação da metodologia apresentada como forma de análise de conteúdo em outros temas da área.

REFERÊNCIA BIBLIOGRAFIA

1. ALVARENGA, Lídia. Definição de publicações oficiais. *R. Esc. Bibliotecon. UFMG*, Belo Horizonte, v. 22, n. 2, p. 213-238, Jul/Dez.1993.
2. ANDER-EGG, Ezequiel. *Introducción a las técnicas de investigación social: para trabajadores sociales*. 7.ed. Buenos Aires : Hmanitas, 1978.
3. BENJAMINS, V. R., FENSEL, D., GÓMEZ PÉREZ, A. Knowledge management through ontologies. In: PAKM-98, p. 5.1-5.12, 1998.
4. BLAZQUEZ, M., FERNANDEZ, M., GARCIA-PINAR, J.M., *et al.* Building Ontologies at the Knowledge Level using the Ontology Design Environment. In: KNOWLEDGE ACQUISITION MODELING AND MANAGEMENT WORKSHOP, KAW'98, 11, 1998, Banff. *Proceedings of...* Banff, 1998.
URL: <http://ksi.cpsc.ucalgary.ca/KAW/KAW98/blazquez/>
5. DAHLBERG, Ingetraut. A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT. *International Classification*, v. 5, n. 3, p. 142-151, 1978.
6. FOX, M. S., GRUNINGER, M. Ontologies for Enterprise Integration. In: CONF. ON COOPERATIVE INFORMATION SYSTEMS, 2, 1994, Toronto Ont. *Proceedings...* Toronto Ont., 1994.
7. GILCHRIST, Alan. *The thesaurus in retrieval*. London . ASLIB, 1971.
8. GOMES, H. E., *et al.* *Manual de elaboração de tesauros monolíngües*. Brasília : CNPq / PNB, 1990. 78 p.
9. GRUBER, T. R. *Ontolingua: a mechanism to support portable ontologies*. Stanford : Knowledge Systems Laboratory Stanford University, 1992. (Relatório KSL91-66).
10. GRUNINGER, M, FOX, M. S. Methodology for the design and evaluation of ontologies. In : INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 1995, Montreal, Quebec,Canada. *Proceedings...* Montreal, Quebec,Canada, 1995.
11. GUARINO, N. 1997b. Understanding, building, and using ontologies: a commentary to "*Using explicit ontologies in KBS development*", by Van Heijst, Schreiber, and Wielinga. *International Journal of Human and Computer Studies*, 46.
12. DOMINGUE, J. Tadzebao and WebOnto: discussing, browsing, and editing

ontologies on the Web. In: KNOWLEDGE ACQUISITION, MODELLING AND MANAGEMENT WORKSHOP, KAW'98, 11, April 1998, Banff, Canada. *Proceedings...*Banff, Canada, April 1998.

13. LANCASTER, F. W. *Vocabulary control for information retrieval*. 2. ed. Virginia : IRP, 1986. 270 p.
14. MARCONI, Marina de Andrade, LAKATOS, Eva Maria. *Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados*. São Paulo : Atlas, 1982.
15. SELLTIZ, C. et al. *Métodos de pesquisa nas relações sociais*. São Paulo : Heder, 1965.
16. LAWRENCE, Steve S., BOLLACKER K., LEE GILES, C. *Digital Libraries and Autonomous Citation Indexing*. *IEEE Computer*, v. 32, n. 6, p. 67-71, 1999.
17. LAWRENCE, S. Access to Scientific Literature.. *The Nature Yearbook of Science and Technology*. London, Declan Butler (Ed.) England : Macmillan, 2001. p. 86–88.
18. USCHOLD, M. Building Ontologies: Towards a Unified Methodology. proceedings of Expert Systems. In: ANNUAL CONFERENCE OF THE BRITISH COMPUTER SOCIETY SPECIALIST GROUP ON EXPERT SYSTEMS, 16, p. 16-18 December, 1996, Cambridge. *Anais...*, Cambridge, 1996.
19. VAN HEIST, G., SCHREIBER A. Th., WIELINGA B. J. Using explicit ontologies in KBS development. *International Journal of Human-Computer Studies*, 45, p. 183-292, 1997.