UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO
MESTRADO EM CIÊNCIA DA INFORMAÇÃO

ELISÂNGELA VILELA DOS SANTOS

ANÁLISE DA REPRESENTAÇÃO DA INFORMAÇÃO NA WEB OF SCIENCE:
um estudo a partir do domínio de nutrição

Recife
2018

ELISÂNGELA VILELA DOS SANTOS

ANÁLISE DA REPRESENTAÇÃO DA INFORMAÇÃO NA WEB OF SCIENCE:
um estudo a partir do domínio de nutrição

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco como requisito final para a obtenção do título de Mestre em Ciência da Informação.

Área de concentração: Informação, Memória e Tecnologia.

Linha de Pesquisa: Comunicação e visualização da memória.

Orientador: Prof. Dr. Fábio Mascarenhas e Silva

Recife
2018

Catalogação na fonte
Bibliotecário Jonas Lucas Vieira, CRB4-1204

S237a Santos, Elisângela Vilela dos
Análise da representação da informação na Web of Science: um estudo a partir do domínio de Nutrição / Elisângela Vilela dos Santos. – Recife, 2018.
132 f.: il., fig.

Orientador: Fábio Mascarenhas e Silva.
Dissertação (Mestrado) – Universidade Federal de Pernambuco, Centro de Artes e Comunicação. Programa de Pós-Graduação em Ciência da Informação, 2018.

Inclui referências e apêndices.

1. Representação da informação. 2. Visualização da informação. 3. Nutrição e Saúde Pública. 4. Descritores de Ciências da Saúde. 5. *Web of Science*. I. Silva, Fábio Mascarenhas e (Orientador). II. Título.

020 CDD (22. ed.) UFPE (CAC 2018-159)

Serviço Público Federal
Universidade Federal de Pernambuco
Programa de Pós-graduação em Ciência da Informação - PPGCI

ELISÂNGELA VILELA DOS SANTOS

Análise da representação da informação na Web of Science: um estudo a partir do domínio de Nutrição

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de mestre em Ciência da Informação.

Aprovada em: 28/02/2018

BANCA EXAMINADORA

_____
Prof. Dr. Fábio Mascarenhas e Silva (Orientador)
Universidade Federal de Pernambuco

_____
Prof. Dr. Raimundo Nonato Macedo dos Santos (Examinador Interno)
Universidade Federal de Pernambuco

_____
Prof. Dr. Murilo Artur Araújo da Silveira (Examinador Externo)
Universidade Federal de Pernambuco

A Deus e ao Universo,
pelo dom da VIDA,

Dedico!

AGRADECIMENTOS

Agradeço, primeiramente, ao professor Fábio Mascarenhas, pelos ensinamentos, paciência, orientação, palavras de ânimo e, o mais importante, por confiar em mim. É, sem dúvida, um profissional muito humanizado, pelo qual sinto profunda admiração e respeito.

Agradeço aos membros da banca professor Raimundo Nonato e Professor Murilo Silveira pelas riquíssimas contribuições.

Agradeço ao Professor Fábio Pinho pelas preciosas sugestões durante a qualificação.

Agradeço ao PPGCI/UFPE por me proporcionar a viver essa experiência um tanto árdua, mas maravilhosa, que só me fez crescer como ser humano. Aprendi muito com os meus colegas, amigos e professores. Sou-lhes imensamente grata.

Agradeço aos meus colegas e amigos da turma PPGCI/UFPE 2016.1 pela experiência vivida, em especial ao meu amigo Alejandro, pela parceria nos trabalhos, pelas conversas enriquecedoras, pelas palavras de ânimo, por me ouvir e principalmente, por não me deixar desistir. És um ser de um coração grandioso e um profissional que admiro profundamente. Obrigada por estar ao meu lado nos momentos de dificuldade e medo. Acredito que construímos uma bela amizade.

Agradeço ao meu amigo Jorge por me fazer enxergar que este sonho seria possível e foi. Sem seus conselhos e algumas broncas, eu não teria chegado até aqui. Você é um amigo muito especial.

Agradeço as minhas amigas da UFAL, Lívia e Rejane, por todo o apoio e incentivo. Meninas, muito obrigada!!!

Agradeço a minha família por acatar minhas decisões e por me incentivar a seguir em frente.

Agradeço ao meu papaizito, a minha mamaezita (Salvador e Suely) e a minha querida tia Marisa por serem tão especiais comigo. Sempre me apoiando e me dando força para seguir e vencer o medo. Amo muito vocês!

Agradeço a pessoa mais especial desse mudo, meu amor, meu amigo e companheiro de vida, Rodrigo. Obrigada por viver tudo isso comigo, por aguentar as noites mal dormidas, por segurar a minha mão nos momentos de medo e insegurança, por sorrir comigo nas vitórias e conquistas... obrigada por ser meu Anjo da guarda. Amo você!

Por fim, agradeço a todos que, diretamente ou indiretamente, me ajudaram nessa caminhada.

Muito OBRIGADA!!!

“Tenho em mim todos
os sonhos do mundo”.

Fernando Pessoa

RESUMO

Objetiva avaliar a qualidade da representação da informação na WoS através das palavras-chave de autor e *keywords plus* dos artigos da área de Nutrição, a fim de observar como esses termos se comportam diante das visualizações geradas a partir da coocorrência de palavras. Quanto aos objetivos específicos, o estudo propõe: estabelecer critérios de avaliação da indexação fundamentados na literatura; aplicar os critérios de avaliação da indexação às palavras-chave de autor e às *keywords plus* atribuídas aos artigos da base de dados; apresentar os resultados das análises segundo os parâmetros da pesquisa sob a ótica da visualização da informação (VI). Trata-se de um estudo exploratório, que busca compreender a dinâmica da representação da informação do domínio de Nutrição, numa base de dados específica, isto é, na *Web of Science* (WoS). O *corpus* é composto por um conjunto de vinte artigos com alto índice de citação, publicados no período de 2006 a 2016. A análise se deu em duas etapas: na primeira, aplicou os critérios de avaliação da indexação ao conjunto de palavras da WoS e, na segunda, que teve como parâmetros os Descritores de Ciências da Saúde – DeCS, buscou-se encontrar pontos de intersecção entre o vocabulário DeCS e a representação da informação na WoS. Tal análise teve como ponto central a garantia literária. Os resultados mostraram que tanto as palavras-chave de autor quanto as *keywords plus* do *corpus* analisado conseguiram atender a alguns critérios de indexação, todavia foram observados alguns aspectos da representação temática que comprometem a VI. Quanto à garantia literária, tendo como parâmetro o DeCS, observou-se uma quantidade muito reduzida de palavras do *corpus* coincidentes no DeCS e na WoS.

Palavras-chave: Representação da informação. Visualização da informação. Nutrição e Saúde pública. Descritores de Ciências da Saúde. *Web of Science*.

ABSTRACT

This work aims to evaluate the quality of information representation at WoS using the author keywords and keywords plus of papers in the area of Nutrition in order to observe how these terms behave in front of the visualizations generated from the co-occurrence of words. Regarding the specific objectives, the study proposes: to establish an indexation evaluation criteria based on the literature; apply the evaluation criteria to the author's keywords and keywords plus attributed to the articles of the database; present the results of the analyzes according to the parameters of the research from the point of view of the information visualization (IV). This is an exploratory study, which aims to understand the dynamics of information representation of a given domain, in a given database (Web of Science – WoS). The corpus consists of a set of 20 articles with high citation rate, published between 2006 and 2016. The analysis took place in two stages: in the first one, the indexation evaluation criteria is applied to the set of WoS words and, in the second one, which had as parameters the Health Sciences Descriptors - DeCS, it is sought to find points of intersection between the DeCS vocabulary and the representation of information in WoS. This analysis had as its central point the literary guarantee. The results showed that both the author's keywords and the keywords plus of the Nutrition domain were able to meet some indexing criteria, however, some thematic representation failures were observed that compromise VI. As for the literary assurance, having as parameter the DeCS, a much reduced number of words of the corpus were observed in the DeCS and WoS at the same time.

Keywords: Information Representation. Information visualization. Public Health and Nutrition. Health Sciences Descriptors. Web of Science.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE SIGLAS

BDTDs	Bibliotecas Digitais de Teses e Dissertações
BIREME	Centro Latino-americano de Informação em Ciências da Saúde
BRAPCI	Base de dados de Periódicos em Ciência da Informação
BVS	Biblioteca Virtual em Saúde
CDD	Classificação Decimal de Dewey
CDU	Classificação Decimal Universal
CI	Ciência da Informação
CPCI-S	Conference Proceedings, Citation Index Science
CPCI-SSH	Conference Proceedings Citation Index Social Sciences and Humanites
CRG	Classification Research Group
DE	Palavrs-chave de autor
DeCSD	Descritores em Ciências da Saúde
ENANCIB	Encontro Nacional de Pesquisa em Ciência da Informação
HUPAA	Hospital Universitário Professor Alberto Antunes
IBICT	Instituto Brasileiro de Informação em Ciência e Tecnologia
ID	*Keywords Plus*
LDs	Linguagens Documentárias
LILACS	Base de Dados da Literatura Latino-Americana e do Caribe em Ciências da Saúde
MEDLINE	Base de Dados Especializada em Ciências Biomédicas e Ciências da Vida
MeSH	Medical Subject Headings da U.S. National Library of Medicine (NLM)
NCBI	National Center for Biotechnology Information
NIHE	U.S. National Institutes of Health
OC	Organização do Conhecimento
OI	Organização da Informação
ORC	Organização e representação do Conhecimento
ORI	organização e representação da informação
TTI	Tratamento Temático da Informação
UFAL	Universidade Federal de Alagoas
VI	Visualização da Informação
WoS	Web of Science

SUMÁRIO

1 INTRODUÇÃO

Atualmente há bases de dados nacionais e internacionais (de acesso aberto ou não), que indexam renomados periódicos científicos, com elevado índice de citação e fator de impacto. Estas bases são consideradas um dos principais canais de divulgação do conhecimento científico. Contudo, sabe-se que existem algumas deficiências na comunicação científica, principalmente no que tange à representação e recuperação da informação, carecendo de metodologias mais acuradas que minimizem as lacunas ainda existentes entre as formas de representação e o acesso à informação científica.

A comunidade científica da Ciência da Informação – CI tem discutido diferentes formas de representação da informação, dentre as quais se destacam aquelas dedicadas aos aspectos extrínsecos dos documentos (como a catalogação descritiva) e aos aspectos intrínsecos dos documentos, ou seja, a representação temática por meio da atividade de classificação e da indexação de assuntos. Esta última utiliza-se, principalmente, de palavras-chave e descritores para representar os conteúdos informacionais.

Neste trabalho, a discussão está direcionada aos aspectos intrínsecos da representação da informação, mais precisamente, da representação por meio de palavras-chave e descritores como forma de visualização do conhecimento científico da área de Nutrição no contexto da Saúde Pública.

Vale salientar que palvras-chave e descritores, dependendo da sua aplicabilidade, podem assumir tanto o papel de instrumento quanto de produto de representação da informação. Os descritores de um tesauro, por exemplo, são instrumentos de representação, uma vez que o tesauro fornece as diretrizes para a construção de produtos de representação da informação. Por outro lado, quando descritores e ou palavras-chave auxiliam o usuário a recuperar um documento, trata-se então de produtos do tratamento temático, tais como os índices e catálogos de assunto.

No caso das palavras-chave da WoS, as mesmas podem ser entendidas como produtos do processo de indexação, já que possibilitam ao usuário recuperar informações numa busca por assunto. Essa opção da base de dados permite que o sistema encontre documentos em que os termos das expressões de busca apareçam no título, no resumo, nas palavras-chave de autor e nas *keywords plus de* cada item. Portanto, nessa pesquisa, optou-se por utilizar o termo produto de representação da informação para o objeto aqui estudado, isto é, as palavras-chave do domínio de Nutrição da WoS.

O uso de palavras-chave e descritores, enquanto produtos da representação da informação, deve permitir a comunicação entre usuário e sistema de informação, isto é, possibilitar, através dos mecanismos de busca, que o sistema responda satisfatoriamente as indagações do usuário. Ademais, a partir das palavras-chave e descritores, também é possível montar representações visuais dos principais assuntos de domínios[1] específicos, de forma a acompanhar o progresso ou estagnação da produção científica de um campo a partir de diferentes temáticas.

Pode-se dizer que a visualização da informação estudada no âmbito da CI possui duas dimensões: uma voltada à construção de interfaces amigáveis em Sistemas de Recuperação da Informação – SRIs, de modo que os usuários possam obter melhores resultados de navegação, busca e recuperação de documentos úteis; e a outra dimensão está relacionada à construção de visualizações geradas por meio da coocorrência de palavras, de forma que seja possível identificar conjuntos de atores (palavras) com algum grau de semelhança. Geralmente esse tipo de visualização é gerado a partir de métodos bibliométricos e softwares específicos.

Para Fujita (2004, p. 258), "a palavra-chave é uma representação do conteúdo significativo do texto e também é utilizada para representar uma necessidade de informação na estratégia de busca". Percebe-se, então, que a palavra-chave se comporta como a tradução do conteúdo informacional em termos que o represente fielmente.

Assim como as palavras-chave, os descritores são termos utilizados para representar o conteúdo de um documento, só que de forma controlada. Geralmente, os descritores são elaborados por especialistas e de modo padronizado (LOPES, 2002).

É comum que as palavras-chave sejam atribuídas pelos próprios pesquisadores (linguagem natural) quando representam tematicamente suas publicações, em especial artigos de periódicos e trabalhos de eventos. Enquanto que os descritores são termos mais específicos, determinados por um grupo de especialistas, que fazem uso da linguagem artificial, dando origem às listas de vocabulários controlados, índices de assuntos e tesauros. Algumas bases de dados utilizam vocabulários controlados para padronizar a indexação e, ao mesmo tempo, facilitar o processo de busca e recuperação da informação, de modo que a linguagem utilizada pelos pesquisadores seja comum e a comunicação científica se torne mais eficiente.

---

[1]Domínio é definido por Dias (2015) como uma área de conhecimento ou um campo de especialidade. Para Bufrem e Freitas (2015, p. 5) entendem domínio científico como um "modo racional de delimitação de saberes dentro de um campo, na medida em que permite o aperfeiçoamento da produção do conhecimento".

Como exemplo de base de dados desse cunho é possível citar a Base de Dados da Literatura Latino-Americana e do Caribe em Ciências da Saúde – LILACS, e a Base de Dados Especializada em Ciências Biomédicas e Ciências da Vida – MEDLINE. Esta última desenvolvida pelo *U.S. National Institutes of Health* – NIH e administrada pelo *National Center for Biotechnology Information* – NCBI. Ambas as bases utilizam os Descritores em Ciências da Saúde – DeCS como terminologia comum entre os pesquisadores da área de Saúde e Ciências da Vida.

Voltando-se para as questões relacionadas à representação da informação, um dos principais problemas que reflete na qualidade da indexação é a abrangência dos termos selecionados para representar conceitos, isto é, uma representação temática sem nenhum grau de especificidade e sem controle terminológico adequado que, por seu turno, acaba acarretando uma recuperação exaustiva de documentos muitas vezes inutilizáveis.

Sobre a qualidade da indexação é importante esclarecer que tal termo diz respeito ao que Lancaster (2004) denominou de 'boa indexação'. Para o autor, a boa indexação é aquela "[...] que permite que se recuperem itens de uma base dados durante buscas para as quais sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas uteis" (LANCASTER, 2004, p. 83).

Já Gil-Leiva (2008) argumenta que a qualidade da indexação está relacionada aos elementos que caracterizam tanto o processo quanto o resultado da indexação, que são: a exaustividade, consistência, especificidade e a correção. Estes são alguns dos critérios para avaliar a qualidade da indexação e que aparecem com melhor esclarecimento na seção 3.3.1 da metodologia.

A especificidade, por sua vez, também pode se tornar um problema que reflete na qualidade da indexação. Isso porque nem toda especificidade é utilizada adequadamente. Nesse caso, entende-se por especificidade inadequada aquela que faz uso de termos extremamente específicos, que não condizem com a linguagem específica do domínio ao qual o assunto do documento que está sendo indexado pertence.

Outrossim, elaborar palavras-chave sem uma política específica ou um controle de vocabulário, implica na inconsistência da indexação, ambiguidade de termos, falta de padronização quanto ao uso de plural ou singular, além de termos irrelevantes (palavras

vazias[2]) que não conseguem representar claramente um assunto específico (STREHL, 1998; FUJITA, 2004).

A importância do uso de palavras-chave não se restringe aos estudos de representação e recuperação da informação. Estudos orientados à visualização da informação por meio da coocorrência de termos também se valem das palavras-chave ou descritores para representar as tendências temáticas de um determinado domínio. Esses estudos são importantes para acompanhar o desenvolvimento da ciência e as mudanças ocorridas ao longo do tempo sobre temas específicos.

Estudos dessa natureza só são possíveis quando existe uma representação fiel do conteúdo dos documentos, seguindo políticas de indexação específicas e controle terminológico. Do contrário, a ausência de controle terminológico pode acarretar a dispersão de termos que, por sua vez, não favorece a visualização da informação pela falta da coincidência e consistência de palavras.

A indexação em bases de dados, por meio de palavras-chave ou descritores, necessita de políticas e metodologias que possibilitem tanto a garantia da qualidade na recuperação da informação, quanto uma visualização fidedigna dos temas que estão sendo discutidos em áreas específicas. Quando as palavras-chave ou descritores são indexados adequadamente, aumenta-se a precisão no momento da busca, pois quanto mais específica for a busca e os termos indexados no sistema de informação corresponderem ao nível de especificidade do usuário, o índice de revocação diminui, garantindo que os documentos recuperados sejam o mais próximo possível da resposta que o usuário necessita.

Dito de outra forma, quando não se tem descritores que correspondam à linguagem utilizada pela comunidade discursiva (descritores amplos demais ou específicos demais), torna-se difícil identificar as principais temáticas de um domínio. Primeiro, por causa do caráter multidisciplinar que termos abrangentes possuem, o que torna a representação pouco clara e objetiva. Em segundo lugar, pela inconsistência das palavras-chave, quer seja por causa da abrangência ou da especificidade inadequada dos termos que acaba por refletir numa visualização pulverizada[3] e pouco representativa.

---

[2] Esse termo também é conhecido como *stopwords* e geralmente é muito utilizado em SRIs que fazem uso da indexação automática.

[3]Entende-se por pulverização da VI a quantidade de termos dispersos que não geram aglutinamento ou que geram pouco aglutinamento. Um exemplo disso é quando diversos termos são usados para designar um mesmo conceito. Se esses termos são pouco utilizados na literatura e caso sejam geradas visualizações a partir de suas ocorrências, se estas forem mínimas, ou seja, aparecem duas ou apenas uma vez, os clusters formados apresentarão resultados dispersos. (continua...)

Particularmente, o presente estudo se propõe a investigar e analisar se as palavras-chave e as *keywords plus* dos artigos científicos da área de Nutrição, indexados na *Web of Science* (WoS), atendem aos critérios de qualidade da indexação identificados na literatura e como esse conjunto de termos se comportam diante das visualizações que são geradas a partir da coocorrência de palavras.

A WoS permite o acesso aos resumos, às citações e às referências da produção científica de diversas áreas do conhecimento, incluindo a Ciência da Informação. Uma das razões que dão à WoS esse caráter de importância e reconhecimento é que a mesma permite acompanhar a produção de pesquisadores nacionais em periódicos internacionais, e isso possibilita verificar o comportamento das ciências, suas tendências e vanguardas, além de permitir detectar a possibilidade de construção de redes colaborativas, outro ponto que é fundamental no processo de comunicação da informação científica em termos internacionais.

Diante do que foi exposto, principalmente da problemática apresentada sobre alguns fatores que interferem na qualidade da indexação e, também, da relevância atribuída às palavras-chave e descritores, enquanto produtos de representação e comunicação da informação, surgiu a seguinte indagação: *As palavras-chave e as keywords plus dos artigos da área de Nutrição, indexados na WoS, recebem tratamento temático adequado que favoreça a visualização da informação?*

No intuito de responder a tal questionamento, apresentam-se os objetivos que nortearão o encaminhamento da pesquisa. Nesses termos, tem-se o seguinte objetivo geral: avaliar a qualidade da representação da informação na WoS através das palavras-chave dos autores e *keywords plus* dos artigos da área de Nutrição, a fim de observar como esses termos se comportam diante das visualizações geradas a partir da coocorrência de palavras. Com base no objetivo geral, foram elencados os seguintes objetivos específicos:

- Estabelecer critérios de avaliação da indexação fundamentados na literatura;

No contexto da Organização da Informação, a dispersão de termos na representação da informação é resultado da discordância entre indexadores e/ou da linguagem técnica utilizada pela comunidade discursiva de um domínio. Se não existe concordância na linguagem de uma área, consequentemente não haverá consistência na representação da informação. A partir da inconsistência dos termos é possível identificar duas situações: a primeira é que os termos que não possuem consistência não podem ser considerados como representativos para um domínio porque não são comumente utilizados. A segunda situação é que a visualização gerada a partir da coocorrência de palavras com termos pouco utilizados tende a ser pulverizada.

- Aplicar os critérios de avaliação da indexação às palavras-chave e às *keywords plus* atribuídas aos artigos da base de dados;
- Apresentar os resultados das análises segundo os parâmetros da pesquisa sob a ótica da visualização da informação.

Em um primeiro momento, esta pesquisa se justifica por compreender a necessidade de estudos mais aprofundados sobre a representação temática da informação, especialmente, no que tange ao uso de palavras-chave e descritores para a visualização da informação em domínios específicos (nesse caso, da área de Nutrição). Esse tipo de estudo é interessante porque permite visualizar a dinâmica da produção de conhecimentos em diferentes campos, sem que seja necessário fazer um levantamento exaustivo de produção científica. Ou seja, acredita-se que, a partir das palavras-chave e descritores, é possível construir representações visuais em diferentes áreas. Nesse sentido, a pesquisa dedicou-se a verificar se a representação temática da WoS, a partir das palavras-chave e *keywords plus* dos artigos da área de Nutrição, favorecem a visualização da informação.

Com relação à escolha da área, o interesse em investigar sobre informação em saúde, mais precisamente na área de Nutrição, tem como marco inicial a monografia pessoal no curso de graduação em Biblioteconomia, da Universidade Federal de Alagoas, intitulado "*Registro e uso de Informação em Saúde: um estudo do prontuário do paciente no apoio à tomada de decisão dos profissionais de Nutrição do HUPAA/UFAL*" que buscou averiguar como os profissionais de nutrição utilizavam o prontuário do paciente no processo de tomada de decisão quanto aos cuidados e serviços prestados aos pacientes. Investigaram-se, também, as formas de tratamento da informação nos prontuários e quais as outras fontes mais utilizadas por estes profissionais.

Outro ponto importante é que foi decidido trabalhar com os descritores da área de Nutrição, no contexto da Saúde Pública, pelo fato de que o DeCS – Descritores em Ciências da Saúde (que será utilizado como parâmetro para a busca na WoS) não contempla a área de Nutrição como uma categoria principal, mas como uma subcategoria da Saúde Pública (ver seção 3.2.2), o que poderia afetar a recuperação e análise dos termos se fosse realizada uma busca mais abrangente.

Quanto à escolha da base de dados, foi selecionada a WoS como o universo a ser investigado, por ser tratar de uma base de dados internacional e multidisciplinar que indexa periódicos renomados com alto índice de citação e fator de impacto elevado nas mais diversas áreas do conhecimento, inclusive na área de Nutrição. Essa base de dados é importante por ser

reconhecida internacionalmente e também porque divulga parte da produção de autores nacionais e internacionais da área em questão.

Sob essa perspectiva, Packer (2011) infere que a *Web of Science* é uma das maiores bases de dados multidisciplinar do mundo, sendo referência mundial para medir a produção científica dos países. Ou seja, busca investigar as revistas nucleares, as pesquisas que mais recebem investimento, as principais instituições de fomento à pesquisa, os autores mais producentes e mais citados, entre outros mapeamentos que acabam se tornando objeto essencial na construção de políticas públicas que favoreçam o investimento e incentivo à pesquisa científica.

Com relação ao tratamento temático da informação, outro ponto que merece destaque referente aos problemas de indexação é que a representação temática é uma atividade puramente subjetiva, tendo relação direta com os aspectos cognitivos do ser humano. O ato de ler e representar conteúdos acontece de forma diferente para cada indivíduo, pois as interpretações são diversas. Então, é natural existir termos mais coerentes e menos coerentes com o assunto que está sendo indexado.

No entanto, na literatura existem critérios de indexação que buscam facilitar o trabalho do indexador. Além disso é possível contar com instrumentos de representação da informação, como no caso dos tesauros e das listas de vocabulários controlados, para eleger os termos mais apropriados no momento da indexação.

A validação de termos considerados preferidos, para representar os conceitos de um dado domínio, deve levar em consideração as comunidades discursivas e as diversas garantias relacionadas à organização do conhecimento. Quando a indexação busca eleger termos a partir da frequência de pedidos em um sistema de recuperação da informação, tem-se a garantia de usuário e quando a indexação busca eleger termos a partir da frequência com que estes aparecem nas publicações, tem-se a garantia literária.

Vale salientar que não é objetivo dessa pesquisa contextualizar e conceituar os diferentes tipos de garantias, contudo é preciso esclarecer o que se entende por garantia de literatura ou garantia literária[4], uma vez que esta será mencionada com frequência no decorrer deste trabalho.

---

[4]Barité (et al., 2010, p. 124) compreende que "[...] a concepção original de garantia literária se sustenta na ideia nuclear de que a literatura de um domínio deve ser a fonte para extração e validação da terminologia a ser incorporada em um sistema de classificação, ou em qualquer outro sistema de organização do conhecimento". Nessa mesma linha, Dias (2015) acrescenta que a garantia literária é aquela oferecida pela própria literatura, através dos diversos termos adotados por determinado

Acredita-se que este trabalho, ainda, possa contribuir para a produção científica da CI sobre os métodos e técnicas de representação da informação em bases de dados internacionais, sobre a qualidade da representação da informação e sua contribuição para a visualização da informação em domínios específicos, de modo que seja possível maiores discussões entre os diversos profissionais da informação sobre o tema aqui investigado.

domínio e reconhecidos pelos pares. Trata-se, então da linguagem técnica de um domínio e que possui o aval da comunidade discursiva desse domínio.

2 QUADRO TEÓRICO E CONCEITUAL

Este capítulo se dedica a tratar dos assuntos pertinentes ao referencial teórico e conceitual que é indispensável ao entendimento e fundamentação de qualquer pesquisa científica. Trata-se então de abordar temas, conceitos, discussões e reflexões condizentes ao objeto de investigação aqui estudado, isto é, palavras-chave e *keywords plus*, observando sua qualidade representacional para fins de visualização da informação referente à área de Nutrição.

É interessante mencionar que a representação da informação possui diferentes perspectivas, objetivos e aplicações, dependendo do contexto em que esteja inserida e da finalidade à qual esteja destinada. Nessa pesquisa, o foco de discussão está pautado na representação da informação sob a perspectiva da visualização da informação. Sendo assim serão abordados assuntos sobre a representação de um modo geral, Representação da Informação e do Conhecimento, Linguagens de Representação da Informação, Instrumentos de Representação da Informação e do Conhecimento, Atividades, Processos e Qualidade da Indexação e algumas definições sobre Visualização da Informação, de modo específico.

A representação da informação por meio de palavras-chave e descritores é fundamental para comunicar o conhecimento científico. A indexação de assuntos permite que usuários e SRIs se comuniquem de modo que, através das estratégias de busca e da linguagem utilizada pelo sistema, seja possível recuperar documentos úteis.

Vale ressaltar que a indexação também permite a identificação de domínios específicos em diferentes áreas do conhecimento. Esses estudos voltados à visualização da informação são importantes para acompanhar o progresso e/ou estagnação das linhas de discussão da pesquisa científica. Nesse sentido, é necessário medir maiores esforços que possam contribuir com novas discussões sobre representação no contexto da visualização da informação, visto que, na literatura da CI existe uma concentração de estudos dedicados à representação da informação para fins de recuperação da informação.

Partindo dessas discussões, este capítulo está organizado da seguinte forma: a seção 2.1 trata das representações simbólicas e das técnicas de representação da informação científica. Nesse tópico estão pontuadas algumas concepções e definições do que é representação de um modo geral, algumas discussões sobre a representação através dos símbolos e mitos, buscando compreender seus significados, onde foram apontadas algumas intersecções com a representação na Ciência da Informação.

Na seção 2.2 foi traçada uma discussão da representação da informação numa perspectiva semiótica, já que a compreensão dos signos linguísticos é fundamental quando se trabalha com as Linguagens Documentárias (LDs). A seção 2.3, em conjunto com as seções 2.3.1 e 2.3.2, apresentarão abordagens e conceitos sobre Organização e Representação da Informação – ORI – e Organização e Representação do Conhecimento – ORC, apontando suas diferenças e similaridades.

Por conseguinte, nas seções 2.3.3, 2.3.4 são abordados temas sobre a comunicação científica e sua relação com a representação da informação, no sentido de mostrar a necessidade das práticas de ORI, tais como as técnicas de indexação e uso de LDs para a comunicação do conhecimento científico. Na seção 2.4 estão apresentadas algumas definições sobre atividade de indexação, em seguida, na seção 2.4.1 apresentam-se brevemente alguns instrumentos de representação tais como classificações, tesauros, taxonomias e ontologias e, também, alguns produtos do tratamento temático como o índice, resumo e catálogo de assunto. E finalizando o capítulo do referencial teórico, encontram-se na seção 2.4.2 as definições de palvras-chave, seu uso e aplicações, mostrando sua importância no tocante à visualização da informação.

2.1 Das Representações Simbólicas às técnicas de Representação da Informação

São comuns, na literatura da CI, estudos dedicados à representação e organização da informação. Nesse sentido, encontram-se esforços que se dedicam principalmente ao desenvolvimento de métodos e técnicas que sejam aplicados em bases de dados e diferentes SRIs para auxiliar na comunicação entre documentos e usuários. Essas concepções podem ser identificadas nos trabalhos de Lancaster (1993, 2004), Fujita (2004), Araújo Júnior (2007) e Fujita e Gil-Leiva (2014), por exemplo.

Estudos desse segmento ganham maior visibilidade à medida que serviços de informação são implementados para satisfazer as necessidades de seus usuários. A importância desses estudos pode ser notada não apenas no âmbito da Ciência da Informação, mas nas contribuições advindas desta área em conjunto com outros campos do conhecimento, utilizando-se de métodos e técnicas de representação.

O universo informacional, constituído pelos diferentes formatos e suportes em que as informações são registradas, necessita de tecnologias, de métodos e de técnicas mais precisas para a potencialização da comunicação do conhecimento. As atividades de representação e

visualização da informação científica é o ponto fundamental para que tais esforços sejam possíveis.

Cabe ressaltar que, apesar de o termo representação ser muito utilizada na Ciência da Informação, tal termo, assim como seus atributos, usos e aplicações, não são circunscritos a essa área. O universo representacional está relacionado a tudo o que existe para representar (dar significado de algo, a alguém), isso também inclui símbolos, objetos, sinais, códigos e signos.

A esse aspecto, Lima e Alvares (2012, p. 21) concordam que "desde o princípio das civilizações, o ser humano utiliza diversos recursos para simbolizar a realidade que o circunda. Ao produzir pinturas rupestres, o homem pré-histórico desenhava figuras que retratavam práticas do seu cotidiano". Ou seja, o homem utilizava (e ainda utiliza) símbolos para representar sua realidade. Representar o ato de colocar algo em lugar de, isto é, está relacionado ao conceito de substituição (ALVARENGA, 2003).

Entender as diferentes formas de representação, seus usos e aplicações são fundamentais para compreender de forma mais aprofundada o objeto de estudo dessa pesquisa. Isto é, em síntese, a qualidade da representação da informação como requisito para uma visualização mais confiável dos domínios do conhecimento científico. Para tanto, é necessário fazer uso da literatura e abordar os contextos em que se insere a representação numa concepção mais genérica, para posteriormente adentrar ao que se conhece por *representação da informação*, *representação do conhecimento* e/ou *representação da memória científica*.

Iniciando este discurso, é interessante mencionar a simbologia como área que se dedica aos estudos referentes à origem, criação e interpretação dos símbolos, e que por sua vez nos permite encontrar alguns aportes que contribuem para um melhor entendimento dos estudos sobre representação.

Aqui, faz-se necessário explicar que a contextualização da representação fará referência aos estudos dos símbolos, de forma resumida, aos estudos semióticos na intenção de apresentar o universo dos signos e dos objetos enquanto formas de representação de significados. Posteriormente serão abordados contextos e conceitos voltados à representação e organização da informação, visto que é pertinente ao objeto de investigação do estudo aqui realizado.

Sobre as considerações iniciais acerca da representação simbólica, ou melhor, dos símbolos enquanto elementos dotados de valores, significados e interpretações, entende-se que o símbolo pode adquirir valores tanto pessoais quanto coletivos. Isto é, ele vai significar

algo específico para alguém que possui determinados conhecimentos sobre o que está sendo representado. Do mesmo modo que ele também pode representar um significado universal perante a sociedade, diante consensos de uma cultura já estabelecida.

Isso, por sua vez, nos remete ao cerne da Semiótica desenvolvida por Saussure, Pierce e outros estudiosos dessa área, que compreendem o signo como um elemento que comporta um significado e um significante. Ou seja, podemos entender o signo como a coisa, o objeto representado; o significado pode ser compreendido pelas ações da coisa representada (o que faz, para que serve etc.); já o significante diz respeito à representação da coisa de forma singular na mente de cada um, isto é, a psique, o sujeito cognoscente interpreta a coisa como realmente ela é, individualmente.

Sob esse prisma, podemos enxergar o quão um objeto, uma imagem (símbolos) pode agir de forma diferente na mente de cada pessoa, ganhando interpretações que podem ser convergentes (por meio de consensos) ou divergentes.

Na obra de Jung (2008) “o homem e seus símbolos”, um exemplo vem a calhar muito bem: o autor relata um caso interessante de um “indiano” que, ao visitar a Inglaterra, chegou à conclusão de que os britânicos adoravam animais, comentando com seus amigos sobre sua percepção. O indiano pode ter chegado a tal conclusão, pelo fato de se deparar com vários monumentos de leões, bois e águias nas antigas igrejas da Inglaterra. Não sabia ele que estes animais são símbolos dos evangelistas, provenientes de uma visão do profeta Ezequiel que, por sua vez, faz referência ao deus egípcio do Sol (Horus) e seus quatro filhos. A Figura 1 representa os quatro evangelistas: o homem representa Mateus que, por sua vez, pode simbolizar a geração humana; Lucas é representado por um touro que está associado ao sacrifício de Jesus; O leão representa Marcos, simbolizando o clamor no deserto e a força da ressurreição, e João é representado por uma águia, simbolizando a elevação de Jesus aos céus, vide Figura 1.

Figura 1 – Representação dos quatro Evangelistas
(baixo-relevo da Catedral de Chartres)

Fonte: Jung (2008) em "*O homem e seus símbolos*".

Geralmente, animais em grupo de quatro são símbolos religiosos universais. Assim como na Figura 1, o humano e os outros três animais representam os evangelistas, símbolo do cristianismo, na mitologia egípcia existe uma representação similar, porém com significado diferente. Como no caso do mito do deus egípcio Horus[5] e seus quatro filhos, divindades representadas por um humano, um babuíno, um chacal e um gavião, vide Figura 2.

[5]Horus, filho de Isis e Osíris, é uma grande lenda da mitologia egípcia. Foi concebido para vingar a morte de seu pai Osíris, derrotado por seu tio Seth. Os quatro filhos de Horus, por seu turno, estão ligados ao ritual funerário, sendo cada um responsável por proteger um dos órgãos internos das múmias. Como de costume, no Antigo Egito, as entranhas dos finados eram retiradas durante o embalsamento da múmia e guardadas nos quatro vasos "Canopos" (nome retirado de uma cidade insular, no Baixo Egito, a oeste da foz do rio Nilo), cujas tampas eram representadas pela cabeça de cada um dos quatro deuses. (Imesti, com a cabeça humana) era responsável por guardar e proteger o fígado; (Hapi, com a cabeça de babuíno) guardava os pulmões; (Dua-motef, com a cabeça de cão ou chacal), guardava o estômago e (Kebeh-senuf, com cabeça de gavião) guardava e protegia os intestinos (CAMPBELL, J. 1999 p. 38).

Figura 2 –Filhos de *Horus*, deus egípcio do Sol

Fonte: Jung (2008) em *O homem e seus símbolos*

À luz do que foi supracitado, e retomando a história do indiano, percebe-se que os símbolos vistos na perspectiva do mesmo provocaram interpretações divergentes do seu verdadeiro significado. Ou seja, o símbolo adquiriu um significado pessoal, que vai de encontro ao significado coletivo, próprio de determinada cultura. A esse aspecto, acrescenta-se que as coisas ganham significados por meio da percepção humana, através dos modelos mentais que cada indivíduo carrega acerca do que está sendo representado. De imediato, o indiano, ao conhecer pouco sobre a cultura religiosa da Inglaterra, logo deu um significado pessoal aos monumentos que continham animais. O desconhecimento sobre uma cultura, interpretado a partir de uma vivência cultural diferente, possibilitou ao indiano considerar os britânicos como adoradores de animais, visto que, na Índia, alguns animais são considerados sagrados.

Compreender os significados dos símbolos é interessante para a presente pesquisa porque permite estabelecer relações entre a representação simbólica e as representações documentárias (sejam elas descritivas ou temáticas), pois da mesma forma que os símbolos carregam elementos significativos, as palavras são dotadas de elementos de sentido que dão significados às coisas. Por isso, entender o universo que rege a representação simbólica e sígnica, tona-se requisito para compreender com mais minúcia as representações no contexto informacional e documental.

Na obra de Jung (2008) pode-se compreender a questão dos símbolos sob duas perspectivas: 1) no que concerne à representação propriamente dita que, por sua vez, está

relacionada ao significado manifesto e imediato das coisas, e 2) no sentido do valor simbólico agregado, empregado ao que está sendo representado, ou seja, diz respeito a um aspecto que está relacionado ao "inconsciente" – que não consegue ser precisamente definido ou conceituado (JUNG, 2008).

Ambas as perspectivas se aproximam da mesma forma que se separam, pois, o símbolo (entendido por Jung como um objeto, uma imagem, um monumento, um signo linguístico, entre outros) representa informações sobre algo. Essas informações e significados que são empregados aos símbolos são provenientes da percepção humana. Contudo, ainda existe a interpretação que vai além da percepção do homem pelos seus sentidos físicos, isto é, a visão, o tato, e audição, por exemplo. Nesse caso, nos referimos ao valor simbólico que é único e particular de cada um (é o enxergar e interpretar com os olhos da alma).

Para Jung (2008, p. 19)

> [...] quando a mente humana explora um símbolo, é conduzida a ideias que estão fora da nossa razão. A imagem de uma roda pode levar nossos pensamentos ao conceito de um sol "**divino**", mas, neste ponto, nossa razão vai confessar sua incompetência: o homem é incapaz de descrever um ser **"divino".** Quando, com toda a nossa limitação intelectual, chamamos alguma coisa de "divina", estamos dando-lhe apenas um nome, que poderá estar baseado em uma crença, mas nunca em uma evidência concreta (JUNG, 2008, p. 19).

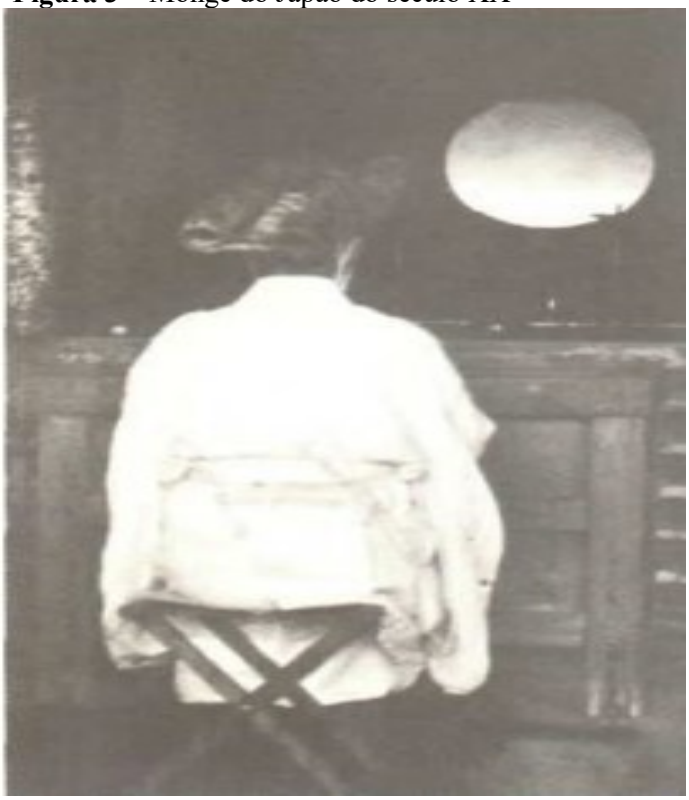
A esta perspectiva, Jung (2008) acrescenta que, pela incapacidade de a mente humana compreender inúmeras coisas, são utilizados os símbolos para representar conceitos que não são possíveis de definição ou de compreensão em sua plenitude. O autor compreende que o uso consciente dos símbolos está atrelado a um fato psicológico importante: que o homem produz símbolos inconscientemente através dos sonhos. Essa relação entre o consciente e o inconsciente é um dos pontos chave trabalhado pelo autor.

Contudo, vale salientar que o que nos interessa nessa obra são as observações que Jung faz a respeito dos símbolos, dos seus significados e representações. De antemão, serão poupados assuntos e/ou abordagens referentes à psicanálise e à interpretação dos sonhos. Este ponto não é de interesse da presente pesquisa.

Sobre a Figura 3, a seguir, pode-se fazer a seguinte reflexão: a princípio, vê-se apenas a Figura de uma pessoa sentada numa cadeira olhando para um espelho. No entanto, trata-se de um monge do Japão do século XX que está orando diante de um espelho, pois no xintoísmo, o espelho representa o Sol divino (JUNG, 2008). Para um leigo, além do manifesto imediato que a nossa razão consegue interpretar diante do que está sendo visto, nada de mais profundo pode significar ao ver a imagem de uma pessoa diante de um espelho. Nesse caso,

em especial, as vestimentas e o chapéu podem dar indícios de que se trata de uma cultura específica. Contudo, perceber de forma imediata que a imagem representa um culto do xintoísmo, não é tão simples para um leigo quanto para alguém que faz parte desta cultura.

Figura 3 – Monge do Japão do século XX



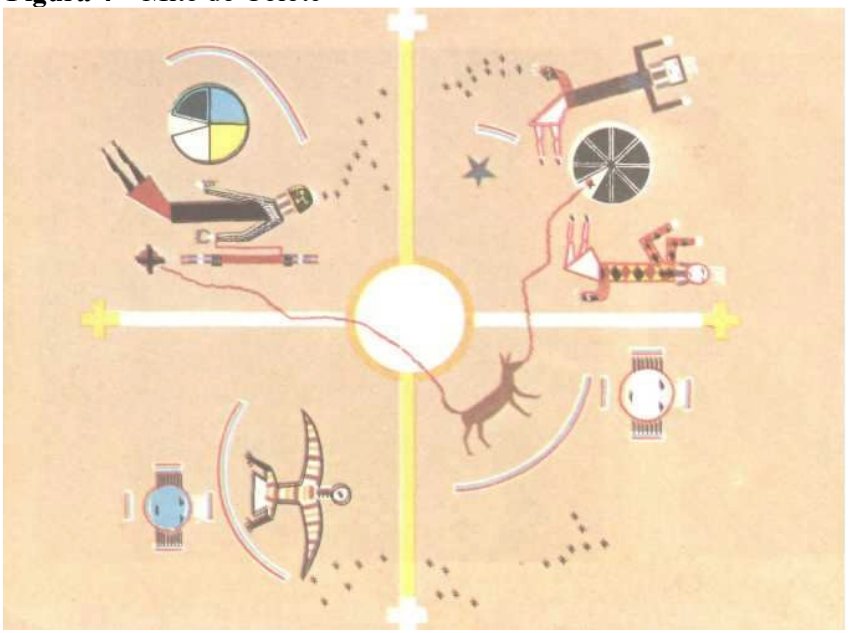
Fonte: Jung (2008)

A representação simbólica também pode ser percebida no âmbito dos grandes mitos antigos das diferentes culturas. Na mitologia, muitos símbolos são utilizados para representar acontecimentos importantes que, na grande maioria das vezes, estão relacionados aos fenômenos naturais e às origens do universo. O mito possibilita a humanidade a refletir sobre algumas questões mais profundas da vida e tentar compreender seu próprio mundo.

No livro de Jung (2008), o capítulo II é dedicado aos mitos antigos e sua relação com as questões da modernidade. Este capítulo, que tem as contribuições de Joseph L. Henderson, traz em sua composição as conexões que existem entre os símbolos mitológicos e a humanidade, mostrando que, no tocante à simbologia, não há distinção entre o homem primitivo e o homem moderno, pois, os mitos, assim como os demais símbolos, fazem parte do nosso cotidiano. São formas rebuscadas e aprofundadas de tentarmos compreender o que nem sempre a lógica e a razão nos permitem entender.

No mito do Coiote[6], por exemplo, é possível perceber uma representação simbólica interessante a respeito da evolução biológica do homem. O Coiote é uma das figuras emblemáticas da mitologia norte-americana. Um mito heróico da tribo dos "winnebagos", que nos permite observar quatro etapas distintas da evolução biológica do herói, desde a sua infância até adquirir a fase mais madura (HENDERSON, 2008).

Figura 4 – Mito do Coiote



Fonte: Henderson (2008)

A partir das representações simbólicas, vimos que o universo representacional é amplo e diversificado. Nessas condições, a linguagem também assume seu lugar dentro desse universo. A linguagem (seja ela escrita ou falada) é o meio mais comum que os homens utilizam para se comunicar uns com os outros. São, portanto, formas de representação de algum significado que se quer transmitir a alguém.

[6]Em 1948, Paul Radin escreveu e publicou algumas histórias sobre a tribo dos winnebagos, dentre elas, uma sob o título *"O ciclo heroico dos winnebagos"*. Através desta história é possível notar a progressão do mito desde o conceito mais primitivo do herói até o mais elaborado. Dentro desse mito foi possível constatar quatro ciclos distintos*: ciclo Trickster*,que corresponde ao primeiro período da vida, o mais primitivo, dominado por seus desejos, possuindo a mentalidade de uma criança.O ciclo *Hare*, onde este personagem ainda se apresenta sob a forma de um animal, não tendo ainda alcançado a plenitude da estatura humana. Continua sendo dominado pelas emoções (adolescência).
O terceiro ciclo é o *Red Horn*, que representa a idade adulta do homem e o quarto e último ciclo é Twin, que diz respeito à fase mais madura do indivíduo, ou seja, quando o homem atinge seu grau de sabedoria procurando o equilíbrio (HENDERSON, 2008).

Ao compreendermos esse universo simbólico e representacional, torna-se mais claro entender a representação no âmbito da CI. Assim como imagens e mitos são dotados de símbolos e significados, a palavra (signo linguístico) assume diferentes significados. No âmbito da Organização e Representação da Informação é comum o uso de signos linguísticos para representar conteúdos informacionais. Nesses termos, a Linguística e a Semiótica também são de suma importância para o presente estudo.

Na próxima seção será discutida a representação no contexto da semiótica, compreendendo o valor do signo, significado e significante, que juntos podem atribuir sentido ao que está sendo representado. Além disso, pretende-se com esta perspectiva enfatizar a relação entre os signos linguísticos e a domínio da Organização e Representação da Informação.

2.2. Uma perspectiva Semiótica

Esta seção se propõe a apresentar, sinteticamente, a representação no contexto semiótico, discorrendo sobre a teoria dos signos verbais e não verbais na perspectiva de Pierce e Saussure, a fim de identificar pontos de intersecção entre os estudos de representação no contexto da Ciência da Informação. Para tanto, serão utilizados conceitos e abordagens sobre linguagem e linguística documentária, de forma que seja possível estreitar os laços entre tais áreas de estudo. Para compor esta seção, serão apresentadas também concepções utilizadas por Santaella, autora reconhecida na literatura brasileira sobre os estudos dedicados à semiótica.

Segundo Santaella (1983, p. 1) "O nome Semiótica vem da raiz grega *semeion*, que quer dizer signo. Semiótica é a ciência dos signos. [...] contudo, não se trata dos signos do zodíaco, mas dos signos da linguagem. Ou seja, a Semiótica é a ciência geral de todas as linguagens".

Para Noth (2003), algumas escolas não concordam que a semiótica esteja voltada à teoria dos signos, tentam, portanto, dar definições mais restritivas, como no caso, por exemplo, da escola de Greimas, que entende a semiótica como uma ciência que se ocupa apenas da comunicação humana, definindo-a como uma teoria da significação e não teoria dos signos.

Quanto ao percurso histórico da semiótica, Noth (2003) aponta que um dos primeiros estudos voltado à Semiótica pode ser encontrado na história da Medicina, cujo médico grego Galeno de Pérgamo (139-199) trouxe à tona o diagnóstico dos signos das doenças, colocando a diagnóstica como a parte semiótica da Medicina. O autor acrescenta também os estudos filosóficos de John Locke (1632-1704) sobre a doutrina dos signos, e o tratado *Semiotik* de Johann Heinrich Lambert (1728-1777) como um dos primeiros estudos acerca do tema (NOTH, 2003).

Com relação à terminologia da Semiótica, várias denominações foram utilizadas ao longo da história para designar o estudo dos signos, dentre elas temos como exemplo: a Semântica, Sematologia, Semasiologia e a Semiologia. Esta última ainda continua sendo confundida com o termo semiótica, embora ambos possuam alguma relação. Mas, para evitar confusão ao longo desse texto, é necessário apontarmos algumas diferenciações.

Na história da Semiótica, esta pode ser designada como "uma ciência mais geral dos signos, incluindo os signos animais e da natureza, enquanto que a Semiologia passou a referir-se unicamente à teoria dos signos humanos, culturais e, especialmente, textuais" (NOTH, 2003, p. 23). Geralmente, a Semiologia está atrelada à Semiótica fundada no contexto da Linguística, campo muito conhecido pelos estudos de Saussure.

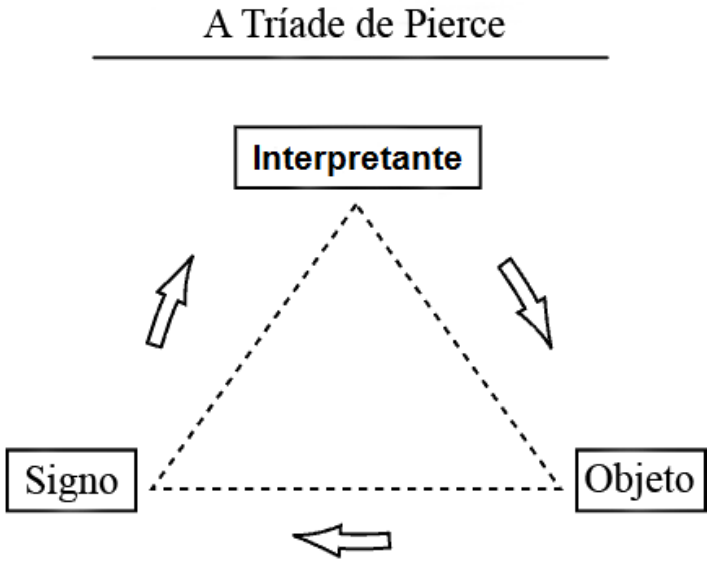
A Semiologia defendida por Saussure trata de uma disciplina pela qual os signos são estudados no meio da vida social. Para ele, é uma ciência que faz parte da Psicologia Social, que estuda a teoria geral de todos os sistemas de signos, que estabelece a comunicação entre os homens (SAUSSURE, 2006).

Já a Semiótica de Peirce trabalha a ideia do signo com base numa relação triádica, ou seja, envolvendo três entidades: signo, objeto e interpretante (COELHO-NETTO, 2007; NOTH, 2003; PEIRCE, 1975; SANTAELLA, 1983).

> Um signo, ou *representamem*, é algo que, sob certo aspecto ou de algum modo, representa alguma coisa para alguém. Dirige-se a alguém, isto é, cria na mente dessa pessoa um signo equivalente ou talvez um signo melhor desenvolvido. Ao signo, assim criado, denomino *interpretante* do primeiro signo. O signo representa alguma coisa, seu *objeto* (PEIRCE, 1975, p. 94).

Essas três entidades abordadas por Peirce formam a relação triádica de signo que, com base numa proposta de Ogden e Richards, pode ser graficamente representada (COELHO-NETTO, 2007). A Figura 5 mostra essa proposta.

Figura 5– Relação triádica de Peirce



Fonte: Ogden e Richerd (1972) citado por Coelho-Netto (2007)

É pertinente acrescentarmos as considerações de Saussure (2006) a respeito do signo, que compreende o mesmo como a *união entre um conceito (significado) e uma imagem acústica (significante)*. Para Saussure o signo deve ser entendido como

> [...] *signo linguístico*, especificamente. Este é arbitrário – isto é, não há uma relação necessária entre ele e o objeto representado – e diferente do símbolo que, segundo Saussure nunca é completamente arbitrário. Saussure dá o exemplo do símbolo da justiça (uma balança) que não poderia ser substituída por outro (uma luva, uma caneta etc.) [...] (COELHO-NETTO, 2007, p.21).

Compreende-se então que a Semiologia saussuriana trabalha com o signo linguístico. Este, por sua vez é arbitrário, porque na realidade ele é apenas uma convenção humana da língua falada. De forma simplificada, o **signo** é composto de significado e significante. Entende-se o **significado** como o conceito que permite a formação da imagem na mente humana quando esta entra em contato com o significante. Ou seja, significado é o plano do *conteúdo*, a parte inteligível, o mundo das ideias. Já o **significante**, por sua vez, diz respeito à *imagem acústica* do signo, representa a impressão acústica da palavra falada (é o plano da expressão, a parte física da palavra – escrita ou som). Para um melhor esclarecimento, seguem alguns exemplos no Quadro 1.

Quadro 1 – exemplo de signo significado e significante

SITUAÇÃO	SIGNIFICANTE	SIGNIFICADO	SIGNO (REFERENTE)
No trânsito	Sinal vermelho	Pare	Ato de parar
No corpo humano	Febre	Está doente	A doença
Na natureza	Pôr do sol	Vai anoitecer	O fenômeno de anoitecer

Fonte: elaborado pela autora, 2017.

Nesta seção não se pretende aprofundar a discussão sobre as relações epistemológicas entre a Semiótica, a Linguística e a Semiologia, contudo, indicar pontos de intersecção entre estes estudos com a representação da informação por meio das linguagens documentárias, uma vez que a linguagem é de interesse comum dessas áreas.

Para firmar a importância dos estudos de Semiótica e Linguística para a Ciência da Informação, vale atentarmos aos achados de Smiraglia (2014) que compreende que, entre os domínios do conhecimento e da linguagem, existe uma afinidade natural, já que a linguagem é o primeiro e o principal meio em que o conhecimento é comunicado. O conhecimento, tal qual a informação, são objetos de estudo da Ciência da Informação, e, as suas representações, por meio do uso das Linguagens Documentárias (LDs), são fundamentais para comunicá-los à sociedade.

Sendo assim, as representações no âmbito da CI são realizadas a partir de palavras-chave, descritores, vocabulários controlados, tesauros, ontologias, entre outros produtos e instrumentos. Estas formas de representação, tendo como característica comum a palavra, nos permitem inferir que a Linguística contribui diretamente com a área de Organização e Representação da Informação, visto que a palavra é um signo linguístico e este, por sua vez, faz parte tanto do universo de estudo da Linguística quanto da Semiótica.

Conforme Kobashi (2007, [p. 3]), "dentre os signos peirceanos, o símbolo, o signo estabelecido por convenção, é certamente o mais importante para as Linguagens Documentárias. [...] O paradigma da comunicabilidade, da Semiótica peirceana, introduz o sujeito em uma comunidade de linguagem".

A este aspecto Vogel (2007) acrescenta que a Documentação, enquanto disciplina que se dedica ao tratamento da informação para fins de recuperação, transformando estoques de informação em fluxos informacionais, reconhece que as linguagens documentárias utilizam os parâmetros da Linguística para sua elaboração, ou seja, o tratamento da informação envolve um processo de natureza linguística.

Tálamo (2001, p. 142) reconhece que [...] “os processos informacionais se realizam em universos simbólicos, cuja constituição procede por mecanismos lógico-linguísticos e terminológicos, materializando-se em processos comunicacionais socialmente constituídos”.

A esta perspectiva, acrescenta-se que a linguagem é um fenômeno de cultura, que só funciona culturalmente porque é um fenômeno de comunicação, e estes fenômenos só existem porque se estruturam como linguagem. Nesse sentido, as práticas sociais se constituem como práticas de produção de linguagem e de sentidos (SANTAELLA, 1983, p. 2).

Concorda-se com Cintra et al. (2002, p. 34) que “no amplo universo da linguagem, as LDs possuem um status muito particular: por meio delas pode-se representar, de maneira sintética as informações materializadas nos textos”. Sendo assim, compreende-se que a Linguística, bem como a Semiótica, podem ser participantes ativas no desenvolvimento de linguagens de representação na Ciência da Informação. A seção 2.3.4, abordará, de forma mais aprofundada, os tipos de linguagens de representação da informação.

2.3 Informação e Conhecimento: relações e distinções

É fato que a informação técnica e científica é a mola propulsora para o desenvolvimento de um país. Os investimentos do Estado em Ciência, Tecnologia & Inovação contribuem para o seu desenvolvimento, visto que são os resultados dessas pesquisas que proporcionam a criação de produtos e serviços que atendem as demandas sociais. As informações oriundas dessas pesquisas precisam ser comunicadas à sociedade para que novos conhecimentos venham a surgir. Nesse sentido, vale atentarmos aos achados de Cintra et al. (2002, p. 10) quando afirmam que

> [...] a informação cumpre o papel decisivo na mudança dos destinos da humanidade, uma vez que ela está diretamente ligada ao conhecimento e ao desenvolvimento de cada uma das áreas do saber, já que todo conhecimento começa por algum tipo de informação e se constitui em informação [...].

Cabe reforçar que Kobashi e Tálamo (2003, p. 9**)** entendem a informação como um alimento e, se “a carência de alimento provoca a fome, a carência de informação provoca a carência de conhecimento”. Entendendo aqui, carência diferente de escassez, isto é, a produção de informação é constante e, portanto, não existe sua escassez. Contudo, sua comunicação e recuperação é um problema interminável.

Para as autoras, o problema da fome não está relacionado à escassez de alimentos, mas à má distribuição desse bem. Do mesmo modo que não existe escassez de informação, pelo contrário, o que existe é muita informação e pouca solução dos problemas relacionados ao seu fluxo, isto é, problemas que inviabilizam sua recuperação e acesso, dificultando a ação do indivíduo conhecer (KOBASHI; TALAMO, 2003).

Outro ponto importante é que, diante da produção excessiva de informação, fica difícil encontrar algo que seja útil e relevante para se gerar conhecimento. Desse modo, não seria falácia concordar com Zygmunt Bauman (2016), quando ele diz que "estamos nos afogando em informações e famintos por sabedoria". Nesse caso específico, estamos famintos por conhecimento.

De fato, nem tudo o que é "informado" é válido para se gerar conhecimento e desenvolvimento, pois quantidade não implica em qualidade. Na sociedade atual, entendida como Sociedade da Informação, acreditamos que se informa muito e ao mesmo tempo não se informa nada. Essa analogia é tida pelo fato de que informações relevantes podem estar perdidas em meio ao emaranhado de informações inutilizáveis.

A esta concepção, Ribeiro (2009), em sua dissertação intitulada *Visualização de dados na Internet,* traz algumas indagações e reflexões acerca da atual condição da informação no ciberespaço, questionando sobre até que ponto o excesso de informação ocasionado pelo desenvolvimento das tecnologias de informação e comunicação pode contribuir no desenvolvimento de novos conhecimentos.

O autor indaga se "haveria uma ampliação do conhecimento proporcional ao aumento do fluxo de informação em escala global" (RIBEIRO, 2009, p. 12). E "como o volume de informação exponencialmente crescente pode ser representado, a fim de proporcionar um ambiente de navegação que favoreça a emergência de conhecimento?" (RIBEIRO, 2009, p. 17). Estes questionamentos são pontos centrais da pesquisa de Ribeiro (2009) e para responder tais indagações o autor faz uma analogia da metáfora da "Biblioteca de Babel", de Jorge Luís Borges, traçando algumas conexões com a atual sociedade movida pelo informalismo tecnológico.

A analogia que Ribeiro (2009) faz à Biblioteca de Babel com os dias atuais é que, a princípio, a ideia de uma biblioteca que abarque todo o conhecimento produzido pelo homem é de fato tentadora. No entanto, no decorrer do conto, existe a frustração por parte do narrador com tamanha quantidade de material, uma vez que este apresenta referências redundantes, incompletas e que, muitas vezes, não fazem o menor sentido. Tal realidade não parece ser tão

diferente da nossa, visto que, em meio à produção em massa de informações e sua disponibilização na internet, encontra-se muito conteúdo inutilizável.

Nesse contexto, Ribeiro (2009) argumenta que o excesso de dados não é condição para geração de novos conhecimentos. Mesmo com as facilidades das bases de dados, com o acesso rápido através de links, nem sempre é possível encontrar o conhecimento que é necessário, isto é, as pessoas se perdem em novos labirintos, não mais das bibliotecas físicas, mas nos labirintos das referências do ciberespaço.

Da mesma forma que as tecnologias contribuíram para a produção e disseminação da informação em larga escala, o seu excesso se tornou um problema para aqueles que buscam por conhecimentos específicos e que não logram de muito tempo para encontrá-los, surgindo então, a necessidade de se desenvolver mecanismos que facilitem a recuperação e acesso ao conhecimento desejado em tempo hábil.

Conforme Ribeiro (2009) a solução para resolver o problema do excesso de informação não implica em impor limites de publicação no ciberespaço, pois isso seria inútil. Na realidade, possíveis soluções poderiam ser encontradas a partir de pesquisas para o desenvolvimento de novas cartografias que facilitem a interação e representação visual dos conteúdos informacionais no ciberespaço. É dentro da linha de pesquisa de Organização e Representação da Informação que se podem encontrar os aportes necessários ao desenvolvimento e aplicação de técnicas que facilitem a recuperação e acesso ao conhecimento.

A Organização e Representação da Informação, assim como a Organização e Representação do Conhecimento são temas bastante discutidos na CI. Tais discussões, na maioria das vezes, são direcionadas à organização e representação dos conteúdos informacionais de forma que seja possível sua visualização, recuperação e acesso.

Todavia, é interessante compreender melhor o uso desses termos na CI, de forma que não causemos confusão entre um e outro. Para tanto, também é necessário refletirmos brevemente sobre os conceitos de informação e conhecimento, visto que ambos possuem características distintas.

Conforme Bräscher e Café (2008, p. 3), para compreendermos os termos informação e conhecimento, dois aspectos importantes são necessários: “a) relacionar seus conceitos às funções que damos a eles nos contextos em que se inserem; e b) diferenciá-los de conceitos próximos a eles incluídos no sistema referencial”. Para as autoras, algumas contribuições já são possíveis de diferenciar os conceitos de *dados*, *informação* e *conhecimento*. Segundo Fernández-Molina (1994, p. 328), citado por Brascher e Café (2008, p. 3): os *dados* são

informação potencial, que somente são percebidos por um receptor se forem convertidos em informação e esta passa a converter-se em *conhecimento* no momento em que produz uma modificação na estrutura do conhecimento do receptor.

Fogl (1979, p. 21), em seu trabalho *Relations of the concepts 'information' and 'knowledge'* considera que a informação é uma unidade composta por três elementos: "1) Conhecimento (conteúdo da informação), 2) Linguagem (um instrumento de expressão de itens de informação) e 3) Suporte (objetos materiais ou energia)". Todavia, não existe uma relação direta entre o suporte (objeto) e a informação, pois para o autor, o conhecimento advindo da percepção humana é a única fonte de origem da informação (FOGL, 1979). Sendo assim, a informação ultrapassa os limites do suporte físico.

Nesse sentido, diferentemente da Teoria Matemática de Shannon e Weaver, em que o conceito de informação é limitado à eficácia do processo de comunicação através da lógica computacional, ou seja, ao processo de emissão, transmissão e recepção de informação, na CI, aquela ganha uma nova dimensão, deixa de ser considerada apenas como coisa (objeto) e passa a ser compreendida como um processo construído socialmente (ARAÚJO, 2009).

Com base nos conceitos apontados por Fogl (1979), Brascher e Café (2008, p. 4), no Quadro 2 estão apresentadas as principais características acerca da informação e do conhecimento:

Quadro 2 – Características da informação e do conhecimento

INFORMAÇÃO	CONHECIMENTO
1 – é uma forma material da existência do conhecimento; 2 – é um item definitivo do conhecimento expresso por meio da linguagem natural ou outros sistemas de signos percebidos pelos órgãos e sentidos; 3 – existe e exerce sua função social por meio de um suporte físico; 4 – existe objetivamente fora da consciência individual e independente dela, desde o momento de sua origem.	1 – é o resultado da cognição (processo de reflexão das leis e das propriedades de objetos e fenômenos da realidade objetiva na consciência humana); 2 – é o conteúdo ideal da consciência humana;

Fonte: Adaptado de Brascher e Café (2008, p. 4).

De acordo com as características apontadas pelas autoras, a informação adquire um caráter materialista, direcionada ao mundo físico, ao universo dos registros, e que Buckland

(1991), por sua vez, denominou de informação como coisa. Já o conhecimento adquire um caráter cognitivo, ou seja, está no mundo das ideias, onde conhecer é um processo mental.

Para Ribeiro (2009), *dados* são meros registros que, depois de interpretados, se transformam em *informação*, esta, por sua vez, se transforma em *conhecimento* depois de compreendida a partir de experiências prévias pelos indivíduos.

Desta forma, podemos dizer que dado é qualquer registro sem estrutura lógica de sentido, ou seja, é algo extremamente "cru". A informação, por sua vez, é uma mensagem estruturalmente organizada, que apresenta algum significado, nesse caso, é o "cozimento" dos dados. Por fim, o conhecimento é algo produzido pela percepção e consciência do homem por meio do que foi informado, ou seja, é a "ingestão" e "digestão" do próprio conhecimento materializado em informação.

Na concepção de Currás (2010) a informação e o conhecimento estão estritamente interligados, de forma que um é condição para o outro. Nesse sentido, a autora compreende que:

> A informação é causa primeira para produzir conhecimento, quando chega ao cérebro e impacta os neurônios. Então começam a acontecer, de forma sucessiva ou simultânea processos de percepção, apreensão, análise, classificação, arquivo em memória, avaliação que constituem o conhecimento pessoal, subjetivo e condicionado pelo substrato individual e cultural de cada indivíduo. Numa elaboração mental posterior, mais complexa, o conhecimento passa a constituir as ideias, linhas de pensamento. Essas são as que voltam a se converter em informação útil, quando surge a ocasião. (CURRÁS, 2010, p. 24-25).

Compreender os conceitos de informação e conhecimento adotados nesta pesquisa é necessário para entendermos de forma mais clara algumas concepções a respeito da Organização e Representação da Informação e Organização e Representação do Conhecimento, de forma que seja possível identificar suas diferenças e similaridades. Pois, tais termos são comuns na literatura da CI e muitas vezes se confundem. Contudo, para não causar dúvidas quanto ao uso do termo mais apropriado para o presente estudo, foram abordadas brevemente algumas concepções e definições sobre o assunto.

2.3.1 Organização e Representação do Conhecimento

A Organização do Conhecimento (OC) pode ser entendida como uma atividade inerente ao ser humano desde o início da sua existência, pois o homem, enquanto ser social,

sente a necessidade de compartilhar conhecimentos entre seus semelhantes. A este aspecto, Lima e Alvares (2012, p.27) compreendem que "a organização social do conhecimento é a prática cotidiana na organização dos seres, na divisão social do trabalho, na sociologia do conhecimento, na sociologia das profissões, das inovações e de tudo mais que nos cerca".

Este anseio por informação e conhecimento é o que torna possível a evolução da humanidade. No entanto, para que o conhecimento seja comunicado de forma satisfatória é necessário que o mesmo esteja registrado em algum suporte e organizado de modo a permitir sua disseminação, acesso e uso.

É importante frisar que esta organização não deve se restringir apenas ao processo meramente comunicativo, pois se assim fosse, isto colocaria a recuperação da informação como sendo o objeto final da atividade de OC, o que seria uma falácia. A necessidade de OC vai além de uma breve troca de documentos recuperados, isto é, possui uma implicação maior no sentido de contribuir para a propagação do saber, geração de novos conhecimentos que irão contribuir para o desenvolvimento político, econômico, social e cultural da humanidade.

Voltando um pouco há tempos mais remotos, as bibliotecas podem ser consideradas como o primeiro intento institucional das práticas de organização do conhecimento registrado, onde tal esforço é advindo das necessidades de conservação, ordenação e catalogação dos registros de eventos religiosos, políticos, econômicos e administrativos. Primeiramente por meio de tabletes de argila e posteriormente pelo papiro. Mais tarde, a invenção da imprensa também influenciou a organização dos livros, bem como outro tipo de documentos gráficos.

Historicamente, o desenvolvimento das bibliotecas serviu como base para a criação de esquemas e sistemas de OC os quais, essencialmente, possuem como objetivo a classificação e representação do conhecimento registrado, sendo a classificação a primeira tentativa de OC.

Seu desenvolvimento vem acontecendo desde os tempos antigos por meio de diversas contribuições, entre as quais é possível mencionar Platão (427-234 a.C.) que classificou as Ciências em Física, Ética e Lógica; Aristóteles (384-322 a.C.) e sua classificação dicotômica; Calímaco que foi o primeiro a registrar o número de linhas das obras, as palavras iniciais e os dados bibliográficos por meio de tabulas (250 a.C.); von Gesner, que produziu a *Bibliotheca Universalis* e seu índice de assunto (1545); Naudé e seu esquema de classificação apresentado na sua obra *Bibliotheca Cordesianae Catalogus* (1643); Panizzi, quem pulicou as Regras para a Compilação de um Catálogo (1841); Dewey e sua Classificação Decimal (1876); Cutter, quem publicou as Regras para um Catálogo Dicionário (1876); Otlet e La Fontaine, com sua Classificação Decimal Universal de (1905); o Sistema de Classificação Facetada de Ranganathan (1933), entre outros (PINHO, 2009).

Desde esta perspectiva a OC foi ganhando espaço, não somente pela necessidade da sua aplicação prática para o universo documental, mas também como campo de reflexão e produção teórica, que por sua vez foi reforçado com a criação da *International Society for Knowledge Organization* (ISKO) em 1989 (ISKO, 2016). Acrescenta-se também que a OC enquanto campo de pesquisa busca desenvolver metodologias para a construção de instrumentos e produtos que favoreçam a busca e recuperação da informação.

Vale salientar que, em termos de áreas do conhecimento e/ou campos científicos, um dos pilares que sustentam a OC são os seus paradigmas. A este aspecto, Hjorland (1998; 2003) e Smiraglia (2002) concordam que a OC, enquanto área e disciplina científica, recebeu a contribuição do Empirismo, Racionalismo, Historicismo e Pragmatismo como paradigmas para compor suas bases epistemológicas.

Enquanto disciplina científica, a OC busca o "[...] desenvolvimento de técnicas para a construção, a gestão, o uso e a avaliação de classificações científicas, taxonomias, nomenclaturas e linguagens documentárias" (BARITÉ, 2001, p. 41).

A atividade de OC é focalizada no âmbito da Ciência da Informação (CI), uma vez que esta lida com a descrição, indexação e classificação dos artefatos (mensagens, textos, documentos) pelos quais o conhecimento (incluindo sentimentos, emoções, desejos) é representado e compartilhado com outras pessoas (HJØRLAND, 2003).

Autores como Saracevic (1995), Le Coadic (1996) e Robredo (2011) veem a CI como uma ciência emergente do Pós-Segunda Guerra Mundial, com a incumbência de organizar o enorme volume de informações que crescia exponencialmente devido, principalmente, aos avanços técnicos e científicos e à propagação das grandes redes de computadores em termos globais. No entanto, essa vertente da CI é pautada, principalmente, na disseminação e recuperação da informação, tendo um enfoque puramente tecnicista, o que, de certa forma, foge um pouco da proposta atual da OC enquanto campo disciplinar.

A OC não se limita à recuperação da informação, pois é uma área de estudo que se dedica a investigar e trabalhar as questões epistemológicas de OC, ultrapassando, desta forma, a visão pragmática e tecnicista da recuperação de documentos. Na OC, a recuperação da informação deixa de ser um fim e passa a ser um meio. Ou seja, a técnica não mais se sobressai ao processo de organização, representação e disseminação do conhecimento. O conceito passa a ser base fundamental na OC e, por este motivo, necessita ser compreendido em diferentes contextos sociais e culturais.

Para Dahlberg (1995) embora o conhecimento implique na certeza conclusiva, tanto subjetiva quanto objetiva, da existência de um fato, ele é intransferível, ou seja, só pode ser

adquirido por uma pessoa por meio de sua própria reflexão. Tal afirmação nos faz refletir sobre como organizar algo que é intangível por natureza. Porém, o conhecimento aqui discutido é o conhecimento registrado, ou seja, aquele que é transformado em informação por meio do registro em um suporte.

Desse ponto de vista, Guimarães (2001) compreende a OC como "[...] o estudo das possibilidades de organização do conhecimento registrado sob a perspectiva de geração de novos conhecimentos que, uma vez registrado, transforma-se em informação (conhecimento ação) para gerar novo conhecimento".

Tanto a OC quanto sua representação são de suma importância para comunicar o conhecimento através de metodologias e produtos que possibilitem sua recuperação e acesso. A Representação do Conhecimento (RC) é entendida por Barité (1997 *apud* PINHO, 2006, p. 28) como:

> O conjunto dos processos de simbolização notacional ou conceitual do saber humano no âmbito de qualquer disciplina. Na representação do conhecimento se compreende a classificação, a indexação e o conjunto de aspectos informáticos e linguísticos, relacionados com a tradução simbólica do conhecimento.

Compreende-se que a Organização e a Representação do Conhecimento, enquanto campo disciplinar, tem por base o conceito e suas relações para representar diferentes domínios. Esses conceitos são enunciados verdadeiros a respeito de um objeto (DAHLBERG, 1978).

Hjorland (2009) acredita que os conceitos não devem ser compreendidos isoladamente, pois, para o autor, existem muitos conceitos concorrentes em diferentes domínios. Além disso, os conceitos evoluem segundo a construção social do conhecimento. Sendo o conceito a base da OC, a Teoria do Conceito de Dahlberg é de suma importância para os estudos da área.

A seguir serão apresentadas algumas considerações acerca da organização e representação da informação de forma que seja possível identificar as similaridades e diferenças básicas com a OC/RC.

2.3.2 Organização e Representação da Informação

Para Taylor (2004), organizar é uma necessidade natural do ser humano. A este aspecto o autor comenta que segundo os psicólogos o cérebro de crianças muito pequenas

organiza imagens em categorias. Alguns indivíduos sentem maior necessidade de organizar do que outros. Tais só conseguem trabalhar se cada coisa estiver no seu devido lugar ou só conseguem realizar algum novo projeto quando se tudo estiver em ordem.

Porém, independentemente dos tipos de pessoas, o processo de aprendizagem de todo ser humano, baseia-se na capacidade de analisar dados, informação e conhecimento. O autor acrescenta que a necessidade de organizar é naturalmente advinda da necessidade de encontrar algo que necessitamos (TAYLOR, 2004).

Abril (2004) compreende que a Organização da Informação, enquanto campo disciplinar se preocupa em propor princípios e métodos para representar conhecimento materializado em informação. A representação é parte fundamental do processo de organização, visualização e comunicação da informação, visto que é a partir dos instrumentos de representação, como as tabelas de classificação, os códigos de catalogação, as listas de cabeçalhos de assunto, entre outros instrumentos, que se consegue dar ordem aos documentos e a construir produtos de representação desses documentos para comunicá-los à sociedade.

De acordo com Lima e Alvares (2012, p.21), "representar é o ato de utilizar elementos simbólicos – palavras, figuras, imagens, desenhos, mímicas, esquemas, entre outros – para substituir um objeto, uma ideia ou um fato". Nesta mesma perspectiva, Novellino (1998, p. 137) compreende que a

> representação da informação é a substituição de uma entidade linguística longa e complexa - o texto de um documento - por sua descrição abreviada. Sua função é demonstrar a essência do documento. A representação da informação é um processo primeiro da transferência da informação e necessário para enfatizar o que é essencial no documento, considerando sua recuperação.

A representação da informação pode ser realizada sob o ponto de vista de dois aspectos: a) aspectos extrínsecos do documento (diz respeito aos aspectos mais objetivos, tais como: o título, o autor, a quantidade de páginas, etc.) o que chamamos de representação descritiva; b) aspectos intrínsecos do documento (diz respeito aos aspectos mais subjetivos, ou seja, a representação do conteúdo do documento) o que chamamos de representação temática, realizada a partir do processo de análise documentária (DIAS; NAVES, 2007).

Na concepção de Guimarães (2003) a análise documentária retrata dois aspectos distintos: a forma (análise formal) e o conteúdo (análise de conteúdo). A primeira diz respeito ao processo de catalogação para fins de identificação e localização dos documentos. E a segunda diz respeito "aos processos de condensação e de representação por meio de

linguagens documentárias com objetivo específico de produzir resumos e índices de assunto [...]" (GUIMARÃES, 2003, p. 102).

Compreende-se que a representação da informação se dá tanto por meio de detalhes mais explícitos, como as questões relativas à catalogação, quanto por meio de detalhes mais implícitos, como é o caso da representação de assunto por meio das linguagens documentárias (naturais e/ou artificiais). Nesse sentido, as linguagens de representação da informação são fundamentais no contexto da comunicação científica.

Na concepção de Brascher e Café (2008, p. 5), "o objetivo do processo de organização da informação é possibilitar o acesso ao conhecimento contido na informação". Ou seja, é um processo que envolve a descrição física dos documentos, de modo que o produto desse processo descritivo representa atributos de um objeto informacional específico por meio de linguagens também específicas para fim de recuperação da informação (BRASCHER; CAFÉ, 2008).

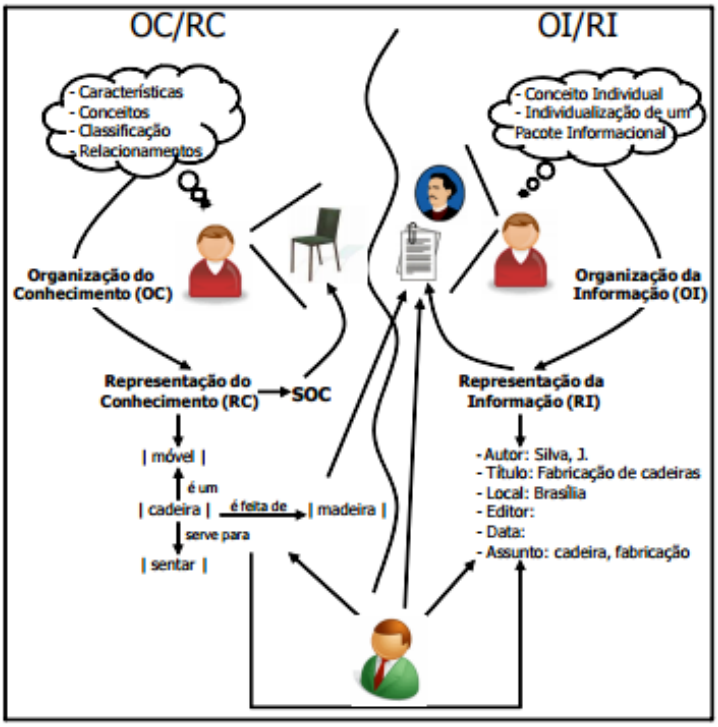
Svenonius (2000) ressalta que a organização da informação requer sua descrição e esta, por sua vez, diz respeito a um enunciado de um objeto informacional e das relações que se estabelecem entre esse objeto com outros que o identifica.

Taylor (2004) compreende que estes objetos são as informações registradas, e que o autor denominou de (pacotes informacionais), e que se constituem em unidades de informação organizáveis para fins de recuperação.

Brascher e Café (2008) delimitam o conceito de Representação da Informação (RI) ao proposto por Alvarenga (2003, 2006) como *representação secundária*, que é uma prática essencialmente dos sistemas de informação documentais. Este tipo de representação está relacionado à identificação dos assuntos constantes nos documentos, de modo que sejam representados descritivamente e tematicamente para fins de recuperação.

Na Figura 6, é possível compreender, de forma sintetizada, as principais características da Organização e Recuperação da Informação e da Organização e Representação do Conhecimento, conforme o trabalho de Brascher e Café (2008).

Figura 6 – OC/RC, OI/RI



Fonte: Brascher e Café (2008, p. 7).

A partir do estudo de Brascher e Café (2008) sobre os termos aqui discutidos, foi possível elencar alguns apontamentos:

- Organização e Representação da Informação é um processo que envolve a descrição física e temática dos documentos;

- O propósito da OI e RI, geralmente, tem como finalidade a recuperação da informação;

- A RI utiliza linguagens documentárias parar representar tanto as características físicas do documento quanto o assunto que nele é tratado;

- A RI se atém a conceitos individuais de objetos individuais, no sentindo de representar a partir das ideias que o autor expõe no texto. Ou seja, quando se classifica algo, está-se atribuindo um conceito individual a um objeto individual (exemplo: a indexação de um documento);

- A OC é entendida como processo de modelagem do conhecimento que visa a construção de representações do conhecimento;

- A OC se ocupa dos conceitos e das relações entre os mesmos para compreender um determinado domínio;

- "A organização do conhecimento, por sua vez, visa à construção de modelos de mundo que se constituem em abstrações da realidade[7]".
- A OC e a RC não se limita a conceitos individuais, mas busca entender conceitos construídos socialmente em diferentes contextos e domínios.
- "A representação do conhecimento é feita por meio de diferentes tipos de sistemas de organização do conhecimento (SOC) que são sistemas conceituais que representam determinado domínio".
- Os conceitos são as bases da OC;

Brascher e Café (2008, p. 6) concluem que existem "dois tipos distintos de processos de organização, um que se aplica às ocorrências individuais de objetos informacionais - o processo de organização da informação, e outro que se aplica a unidades do pensamento (conceitos) - o processo de organização do conhecimento".

No entanto, mesmo que muitas vezes a representação da informação seja vista apenas como um meio para a recuperação da informação reconhece-se que o uso de palavras-chave e descritores também podem ser utilizados para identificar os domínios das áreas do conhecimento. A título de exemplo, temos a dissertação de Ferreira (2012), intitulado *A representação da memória científica da Ciência da Informação brasileira: um estudo com as palavras-chave do ENANCIB*, que investiga as palavras-chave do ENANCIB, a fim de constatar se tais palavras conseguem representar a memória científica da CI. A autora busca montar um panorama visual da construção do conhecimento, por meio de elementos de representação da informação a partir das palavras-chave.

Verifica-se então que, tanto a OI/RI quanto a OC/RC utilizam linguagens documentárias para representar a informação e o conhecimento. Essencialmente o que distingue a RC da RI é que a primeira tem como base os conceitos e as relações que se estabelecem entre os mesmos para representar o conhecimento. Diferentemente da RI que também representa conhecimentos, mas do ponto de vista dos objetos informacionais, isto é, o conhecimento materializado em informação registrada.

Taylor (2004) acredita que não organizamos conhecimento, mas sim informação. Para ele o conhecimento só acontece na mente humana. Dessa forma, seria correto afirmar que um

[7] Trata-se de uma representação abstrata do mundo real construída para determinada finalidade, de modo a tornar possível a melhor compreensão e comunicação entre o usuário e o sistema de recuperação da informação. Um exemplo desse modelo de representação é a ontologia.

objeto informacional contém registros informacionais e não o conhecimento propriamente dito. O conhecimento só existirá a partir do momento em que o receptor consegue decifrar a mensagem inscrita, interpretá-la e compreendê-la.

Em parte, concorda-se com Taylor (2004) que o conhecimento é algo que só acontece na mente humana, todavia há autores que defendem que o conhecimento tratado na CI concerne ao materializado em informação, isto é, o conhecimento registrado. O dado é matéria prima para a informação, esta, por sua vez, é a matéria prima para o conhecimento. Nesse sentido, um precisa do outro para existir.

Adotamos o uso do termo Representação da Informação, uma vez que esta pesquisa se dedica essencialmente à análise da qualidade da representação da informação através das palavras-chave de autor e *keywords plus* de um conjunto de documentos da área de Nutrição indexados na WoS, de modo a observar o comportamento dessas palavras quando aplicadas à Visualização da Informação.

2.3.3 Contribuições da ORI no processo de Comunicação Científica

Não se faz ciência isoladamente. Pois, tomando como exemplo Thomas Kuhn (1998) em sua obra "as estruturas das revoluções científicas", o autor pontua que o progresso da ciência se dá com as crises de paradigmas e a construção de suas teorias. No entanto, os paradigmas e as teorias não surgem ao acaso, eles são construídos, estudados, compreendidos e aceitos pela comunidade científica. Entende-se, então, o conhecimento como uma construção social.

A comunicação da ciência é essencial para se gerar conhecimento e desenvolvimento. O conhecimento é o melhor caminho que homem pode percorrer para realizar grandes feitos e transformar o mundo. As comunidades científicas se comunicam por diversos canais, quer sejam estes formais ou informais. Contudo, a comunicação do conhecimento envolve várias variáveis, desde as formas de representação da informação até os canais propriamente ditos.

São as descobertas científicas que proporcionam a construção de novos conhecimentos que resultará no desenvolvimento da ciência, tecnologia e inovação. E sua comunicação é o que proporciona novas investigações, dando continuidade ao ciclo (produção, disseminação, assimilação e geração de novos conhecimentos).

Comunicação científica é a troca de informações entre os membros da comunidade científica, permitindo que esses possam somar esforços individuais para trocarem

continuamente informações com os seus pares. Informações que são adquiridas dos seus predecessores e passadas aos seus sucessores. Sendo, portanto, indispensável ao fazer científico (TARGINO, 1998).

Conforme Mueller (2006) existe uma íntima relação entre comunicação científica e comunidade científica, de modo que a primeira é a infraestrutura da segunda. A esta concepção podemos ressaltar que é de interesse da comunidade científica a divulgação de suas pesquisas, pois é a única forma de divulgar as tendências de uma área, e dar início a novas investigações (interesse coletivo). Do ponto de vista individual, é de interesse do pesquisador se tornar conhecido e reconhecido pelos pares.

"A posição de prestígio dos cientistas e dos periódicos é mantida e sustentada por um sistema de avaliação baseado em vários indicadores, tais como quantidade de publicações, índices de citação e visibilidade internacional" (MUELLER, 2006, p. 30). Daí o caráter produtivista e, ao mesmo tempo, meritocrático da ciência.

Informação e conhecimento são os pilares que sustentam a sociedade (científica ou não) no sentido de proporcionar grandes descobertas, acontecimentos e transformações. Ademais, o conhecimento é a cura da ignorância, é o que possibilita tornar uma sociedade mais esclarecida, mais justa e mais próspera. Desse modo, não faz sentido a existência do conhecimento sem a sua disseminação.

Para que o conhecimento científico chegue à sociedade, torna-se necessário a aplicação de métodos e técnicas que possibilitem sua comunicação.

No âmbito da CI, a organização e representação da informação é um dos métodos de fundamental importância para dar visibilidade ao conhecimento científico. Os sistemas de informação carecem de sua utilização para possibilitar a busca, recuperação e acesso ao conhecimento.

Geralmente, a prática de representação da informação está associada ao uso de Linguagens Documentárias (LDs) como no caso dos vocabulários controlados, listas de cabeçalhos de assuntos, tesauros, ontologias, sistemas de classificação e também o uso da linguagem natural, como no caso das palavras-chave, que são retiradas do próprio documento. Desse modo, concorda-se que uma linguagem documentária é "simultaneamente, um modo de representação e uma forma de comunicação da informação" (TÁLAMO, 1997).

Boccato e Fujita (2006, p. 28) consideram que "a linguagem documentária, enquanto veículo de comunicação deve representar os campos conceituais respeitando a cultura da comunidade à qual a linguagem serve".

Nesse sentido, acredita-se que a representação da informação não deve ser executada de forma meramente tecnicista. É necessário levar em consideração os aspectos sociais e culturais que permeiam os princípios de organização e representação da informação e do conhecimento. Sobre esta concepção, vale-se das palavras de Lima e Alvares (2012) que consideram a prática de organização do conhecimento válida apenas para o conhecimento socializado, aquele que possui uma dimensão cíclica e que, quando compartilhado, consegue gerar novos conhecimentos.

Portanto, por estes e outros motivos, as palavras-chave e descritores merecem estudos mais aprofundados que evitem sua ambiguidade e inconsistência nos diversos sistemas de recuperação da informação. Nesse sentido, o sucesso da comunicação científica dependerá da forma com a qual a informação está sendo representada em diferentes bases de dados.

As LDs, assim como a linguagem natural, desempenham papel importante na comunicação da informação, desta forma, a próxima seção discutirá um pouco mais as questões relacionadas aos tipos de LDs e o seu papel na representação e visualização da informação. (VOGEL, 2007, p. 30)

2.3.4 Linguagens de Representação da Informação

A linguagem é entendida como a forma mais comum de o homem se comunicar e conviver em sociedade. Para Cintra et al. (2002, p. 26) "[...] todas as práticas humanas são tipos de linguagens, já que elas têm a função de demarcar, significar e comunicar". Ou seja, é o que possibilita dar significado as coisas e compreender as estruturas da comunicação.

> A prática da linguagem é marcada por uma tendência natural do homem: compreender, governar e modificar o mundo. Com efeito, o homem busca, incansavelmente, encontrar uma ordem para as coisas, já que um mundo caótico seria incompreensível, insuportável; por isso ele busca encontrar, em meio à aparência caótica, uma ordem, mesmo que subjacente, uma estrutura capaz de explicar as coisas (CINTRA et al., 2002, p. 27).

Nesse sentido, compreende-se que sem a linguagem seria impossível a constituição da sociedade. O homem enquanto ser social precisa se comunicar com seus semelhantes; o homem enquanto ser racional precisa dar ordem às coisas, fazer com que estas tenham sentido para que assim possa compreender os mistérios da vida. Isso só é possível através do uso da linguagem, seja ela na sua forma oral ou escrita.

As comunidades científicas se comunicam entre si através dos canais informais e formais de comunicação. Geralmente, os canais formais utilizam instrumentos e produtos de ORI para comunicar suas informações. Nesse contexto, estão presentes as linguagens documentárias e linguagens naturais.

Conforme Dodebei (2002), o conhecimento tratado na Ciência da Informação é discutido no âmbito representacional. Isto é, os estoques de informação precisam ser organizados e representados adequadamente para serem recuperados e acessados da forma mais eficiente possível. Nesse contexto, as linguagens documentárias desempenham papel fundamental na representação da informação.

No contexto da representação documentária, temos dois tipos de linguagens: 1) linguagem natural (linguagem simples, sem nenhum tipo de controle) e 2) linguagem artificial ou documentária (geralmente são elaboradas por especialistas e possuem mais especificidade e controle).

Para Lancaster (1993, p. 200) a linguagem natural é:

> [...] sinônimo de 'discurso comum', isto é, a linguagem utilizada habitualmente na escrita e na fala, e que é o contrário de 'vocabulário controlado'. No contexto da recuperação da informação, a expressão normalmente se refere às palavras que ocorrem em textos impressos e, por isso, considera-se como seu sinônimo a expressão 'texto livre'. Um texto livre consiste em: 1) o título, 2) um resumo, ou 3) o texto integral de uma publicação.

As linguagens naturais são, por assim dizer, a forma mais simples de representar o conhecimento, enquanto que as linguagens artificiais são representações mais complexas e específicas, advindas, normalmente, de um vocabulário controlado.

De acordo com Dodebei (2002, p. 56),

> [...] as LD atuam nos sistemas de recuperação da informação em dois níveis: orientando o analista sobre os melhores termos para representar o assunto de um documento, e orientando o pesquisador sobre a escolha dos termos que corresponderiam à representação do assunto por ele procurado.

Já para Cintra et al. (2002) o uso de linguagem natural nesse processo não seria muito adequado, pois devido a fenômenos oriundos de sua natureza como ambiguidade e polissemia, por exemplo, a linguagem natural poderia ser incompreensível e duvidosa para o pesquisador.

Em contrapartida, Lancaster (1993; 2004) compreende que tanto a linguagem natural quanto à linguagem artificial possui valor importante na representação do conhecimento. Isso significa que, nem sempre, numa busca controlada, o usuário irá obter tudo o que deseja, pois, muitas vezes, o controle de vocabulário exclui informações relevantes indexadas com linguagem natural. Do mesmo modo que uma indexação mais genérica, que utiliza linguagem natural, pode refletir de forma negativa no resultado de uma busca, quando é desejável encontrar informações específicas, mas o sistema retorna informações desnecessárias devido ao alto índice de revocação oriundo da falta de controle de vocabulário

É preciso compreender que para obter sucesso numa busca de informação, é necessário que "[...] a pergunta e a resposta sejam formuladas no mesmo sistema. Assim é necessário converter uma pergunta feita em Linguagem Natural (LN) para o sistema em que foi traduzido o conteúdo do documento, isto é, para uma LD" (CINTRA et al., 2002, p. 39).

Conforme Van Slype (1983) a LD é entendida como um sistema de representação dos documentos, tendo como finalidade a recuperação dos mesmos. A LD possui uma estrutura própria, ou seja, controle, padronização, às vezes, estruturas hierárquicas como no caso dos tesauros.

Para Svenonius (2000) as linguagens de representação da informação podem se subdividir em linguagens orientadas a descrever a informação e linguagens dedicadas a descrever o documento (suporte físico). Nesse sentido, trata-se de linguagens de representação descritiva e linguagens de representação temática (esta última, por sua vez, diz respeito aos instrumentos de representação de conteúdo, por meio da atividade de classificação ou do processo de indexação).

Dentro do contexto da representação documentária, Dodebei (2002) entende que o tratamento documentário ou tratamento da informação compreende dois aspectos básicos: a) o que diz respeito ao *processo* em que o objeto é transformado em item documentário para ser recuperado e b) o que diz respeito aos *produtos* gerados por esses processos ou construídos para melhorar a comunicação entre informação e usuário.

2.4 A atividade de Indexação: contextos e definições

A atividade de indexação, também conhecida por catalogação de assunto, constitui as bases do Tratamento Temático da Informação (TTI). Tem como objetivo sintetizar os

assuntos dos documentos em linguagens de representação, de modo a facilitar sua recuperação e acesso.

Geralmente a indexação é utilizada para fins de recuperação da informação, mas também é essencial na representação com vistas à visualização da informação. A visualização aqui discutida diz respeito ao uso de palavras-chave, termos e/ou descritores para representar as unidades constituídas nos domínios do conhecimento, de forma que seja possível identificar as temáticas e tendências em diferentes áreas de estudo e como estas evoluem ao longo do tempo.

Dias e Naves (2007) comentam que o TTI é constituído de **processos, instrumentos e produtos**. Quanto aos **processos**, os autores consideram a *descrição física* e a *descrição temática*. Dentro do processo de descrição também existe outro processo que é a *análise de assunto*– esta permite ao indexador escolher os termos que irão representar tematicamente um documento. Nesse sentido, o **processo** do tratamento temático diz respeito à análise de assunto, sua síntese e representação.

Naves (2001, p. 192) comenta que o processo de análise de assunto compreende duas etapas distintas: "a análise de assunto, quando ocorre a extração de conceitos que possam representar o conteúdo de um documento, expressos em linguagem natural, e a tradução desses termos para termos contemplados nos instrumentos de indexação, que são as chamadas linguagens de indexação [...]".

Quanto aos **instrumentos**, os autores consideram que os principais instrumentos relacionados ao tratamento descritivo são os códigos de catalogação e os formatos de metadados, enquanto que no tratamento temático, as linguagens de indexação são os principais instrumentos (DIAS, NAVES, 2007).

Já os produtos do tratamento da informação que, geralmente, são construídos a partir dos instrumentos de representação, permitem a comunicação entre a informação e o usuário.

Para Dias e Naves (2007, p. 24) os principais produtos do tratamento da informação são os registros bibliográficos e catalográficos, os resumos, os metadados, os pontos de acesso de catálogos, os pontos de acesso de bibliografias, e os arranjo sistemático de coleções e documentos. É pertinente observar que tal abordagem, caracteriza a indexação como um processo do tratamento da informação mais direcionado às bibliotecas e centros de informação.

Sousa e Fujita (2014, p. 22) também entendem a indexação como um processo e argumentam que:

> O processo de indexação, além de ter foco no que é abordado no documento, também deve ser direcionado para a necessidade de informação do usuário, materializada por ele na forma de pergunta. É um processo com duas direções: um lado os dos documentos e de outro, as necessidades de informação dos usuários.

Segundo Van Slype (1977 apud CHAUMIER, 1988, p. 64), a indexação comporta quatro operações distintas, a saber:

- ✓ Conhecimento do conteúdo do documento;
- ✓ Escolha dos conceitos a serem representados, baseando-se na aplicação da regra da seletividade e exaustividade;
- ✓ Tradução dos conceitos selecionados da forma em que aparecem impressos no documento, para os descritores do *thesaurus* aplicando a regra da especificidade;
- ✓ Incorporação dos elementos sintáticos.

Na perspectiva de Strehl (1998) a indexação é composta de duas etapas principais: a análise conceitual e a tradução. A análise conceitual se dedica a definir os assuntos abordados no documento, e a tradução diz respeito à conversão dos conceitos identificados na análise para uma linguagem de indexação seja ela natural ou artificial.

Cunha e Cavalcanti (2008, p. 193) definem a indexação como uma "representação do conteúdo temático de um documento por meio dos elementos de uma linguagem documentária ou de termos extraídos do próprio documento (palavras-chave, frases-chave) ".

Dependendo da natureza e dos objetivos do sistema de informação, a linguagem natural pode ser a mais indicada para representar os conteúdos informacionais, isto é, uma linguagem não tão específica e não tão controlada, geralmente, extraída do próprio texto sob o ponto de vista do indexador. Esse tipo de indexação deve ser aplicado quando o sistema de busca e recuperação da informação possui demandas mais gerais e menos específicas.

Enquanto isso, o vocabulário controlado tende a evitar a dispersão de palavras, eliminar a ambiguidade dos termos, controlar sinônimos, diferenciar homógrafos e ligar termos que possuam uma relação estreita entre si (STREHL, 1998).

Nesse sentido, "o vocabulário controlado torna-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadores – premissa fundamental da comunicação de informações dentro de um sistema" (STREHL, 1998, p. 331). Este tipo de vocabulário deve ser utilizado em sistemas de informação que exigem uma linguagem mais rebuscada, mais específica de um domínio ou de diversos domínios.

Segundo Mai (2000), a representação de assunto pode conter duas, três ou quatro etapas, sendo que no procedimento de duas etapas, uma delas é dedicada à determinação do assunto, a outra é a tradução do assunto em uma linguagem documentária. No procedimento de três etapas acrescenta-se às duas primeiras mais uma etapa que trata da formulação do assunto explicita ou implicitamente. Já no procedimento de quatro etapas, a tradução do assunto se subdivide em duas etapas, uma em que o indexador traduz o assunto da linguagem natural para uma linguagem de indexação e outra em que é construída a entrada de assunto utilizando as linguagens de indexação.

Na literatura da CI existem diversas abordagens que compreendem as etapas da indexação de forma mais abrangente ou mais simplificada. Alguns autores consideram necessárias apenas duas etapas, enquanto que outros compreendem o universo da indexação composto por até cinco ou mais etapas. Independentemente da quantidade, todas são necessárias e importantes para que a representação da informação seja realizada da maneira mais clara e objetiva possível. Em seguida serão mostradas algumas das principais etapas de indexação sob diferentes perspectivas.

Quadro 3 – As etapas da Indexação

	ETAPAS	AUTOR/ES
Duas etapas	a) Determinação do assunto; b) Tradução dos conceitos nos termos da linguagem de indexação.	Unisist (1981)
	a) Reconhecimento e extração dos conceitos dos conceitos informativos; b) Tradução desses conceitos na linguagem documental.	Chaumier (1988)
	a) Análise dos conceitos e das perguntas para a seleção dos conceitos explícitos ou implícitos; b) Armazenamento das palavras chave tal como estão ou sua normalização por meio de um vocabulário controlado.	Gil Leiva (1997)
	a) Análise conceitual; b) Tradução.	Lancaster (2004)
	a) Análise de assunto; b) Tradução.	Fujita (2013)
Três etapas	a) Exame do documento e estabelecimento do assunto de seu conteúdo; b) Identificação dos conceitos presentes no assunto; c) Tradução desses conceitos nos termos de uma linguagem de indexação.	Norma 12675 (ABNT, 1972)
	a) Análise conceitual do conteúdo do documento; b) Expressão dessa análise por meio de códigos, palavras ou frases representativas do assunto; c) Tradução das descrições dos assuntos para a linguagem de indexação.	Robredo (2005)
	a) Análise do documento; b) Descrição do assunto; c) Entrada de assunto.	Mai (2001)
Quatro etapas	a) Contato com o documento; b) Identificação dos conceitos explícitos e implícitos do documento; c) Tradução dos conceitos expressados em linguagem natural por descritores; d) Estabelecimento de ligações sintáticas entre os descritores.	Van Slype (1977)
Cinco ou mais etapas	a) Lembrar os objetivos da operação, se necessário; b) Tomar conhecimento prévio do documento; c) Determinar o assunto principal do documento; d) Identificar os elementos do conteúdo que devem ser descritos e extrair os termos correspondentes; e) Verificar a pertinência dos termos selecionados; f) Traduzir os termos da linguagem natural nos termos correspondentes da linguagem documental, se for o caso; g) Verificar a pertinência da descrição; e h) Formalizar a descrição se o sistema prevê regras especiais de apresentação ou de escrita.	Guinchat e Menou (1994, p.177)
	a) Registro dos dados bibliográficos; b) Análise do conteúdo dos documentos a partir do título, resumo e texto completo; c) Determinação do assunto; d) Conversão dos conceitos extraídos em linguagem de indexação; e) Reexaminar a indexação.	Cleveland e Cleveland (1990 p. 104)

Fonte: Lapa (2014, p 68)

A partir do quadro anterior é possível perceber duas etapas essenciais da indexação que se repetem gradativamente em diferentes períodos e sob a ótica de diversos autores. Trata-se então da análise de assunto e a tradução deste em termos que o representem, isto é, palavras-chave e/ou descritores. As demais etapas complementam e reforçam a análise e tradução de conteúdo, sendo também importantes para garantir que um documento seja

representado satisfatoriamente, tanto no sentido de ser recuperável quanto no sentido de ser representado por termos que favoreçam a visualização da informação de determinado domínio do conhecimento.

Na literatura da CI existe uma série de conhecimentos acerca do Tratamento Temático da Informação que compreendem diferentes abordagens. Nesse contexto, Guimarães (2008 apud PINHO, 2010, p. 36, grifo do autor) identifica que as discussões teóricas da área seguem três correntes distintas: "a da catalogação de assunto (*subject cataloguing*), de influência norte-americana, a da indexação (*indexing*), de influência inglesa e a da análise documental (*analyse documentaire*), de influência francesa".

Com relação à primeira linha (a da catalogação), as atividades são desenvolvidas no âmbito das bibliotecas, tendo influência da Escola de Chicago, que tradicionalmente decorrem da catalogação alfabética de *Cutter* e dos cabeçalhos de assunto da *Library of Congress*. Na segunda linha (a da indexação) essas atividades abrangem tanto as bibliotecas quanto os centros de documentação especializados e também as editoras que sofreram influências do CRG (*Classification Research Group*) tendo os índices como produtos finais da atividade documental. E na terceira e última linha (a da análise documental) tem-se a explicação do próprio processo de tratamento temático (GUIMARÃES, 2008 apud PINHO, 2010).

Com o desenvolvimento das tecnologias de informação e computação, a indexação passou a ser adotada no campo tecnológico, ou seja, vários esforços contribuíram para o desenvolvimento de metodologias e tecnologias que favorecessem a aplicação das técnicas de indexação por máquinas (computadores).

Narukawa, Leiva e Fujita (2009) acrescentam que foi a partir do desenvolvimento tecnológico em meados do século XX que começaram a surgir discussões sobre a automatização da indexação para fins de resultados mais precisos.

Sem dúvidas, a necessidade de automatizar a indexação está muito direcionada a eficiência dos sistemas de recuperação da informação. Desse ponto de vista, Gil Leiva (2008, p. 320-321) argumenta acerca dos pontos favoráveis da indexação realizada por computador que são: a) a rapidez da indexação automática, ao contrário da indexação manual que é demorada, subjetiva e de alto custo; b) a diminuição de erros e a consequente elevação da eficiência na recuperação da informação; c) indexação mais precisa e recuperação mais rica.

Concorda-se com Correa (2011, p. 43), vide figura 7 que demonstra de forma visual, a importância da indexação no contexto da organização, representação e visualização da informação. Para orientar a interpretação da figura, o autor explica que:

> As setas tracejadas mostram algumas teorias, modelos e campos de estudos que influenciam o desenvolvimento de técnicas de indexação. As setas contínuas mostram a influência da indexação por meio de processos específicos (Análise da informação e Medidas de similaridade), nos estudos sobre visualização. As setas cheias indicam sua interação com campos de conhecimentos associados tanto à Ciência da Informação quanto à Ciência da Computação (CORRÊA, 2011, p. 42).

Figura 7 – Importância da indexação: influencias e interações

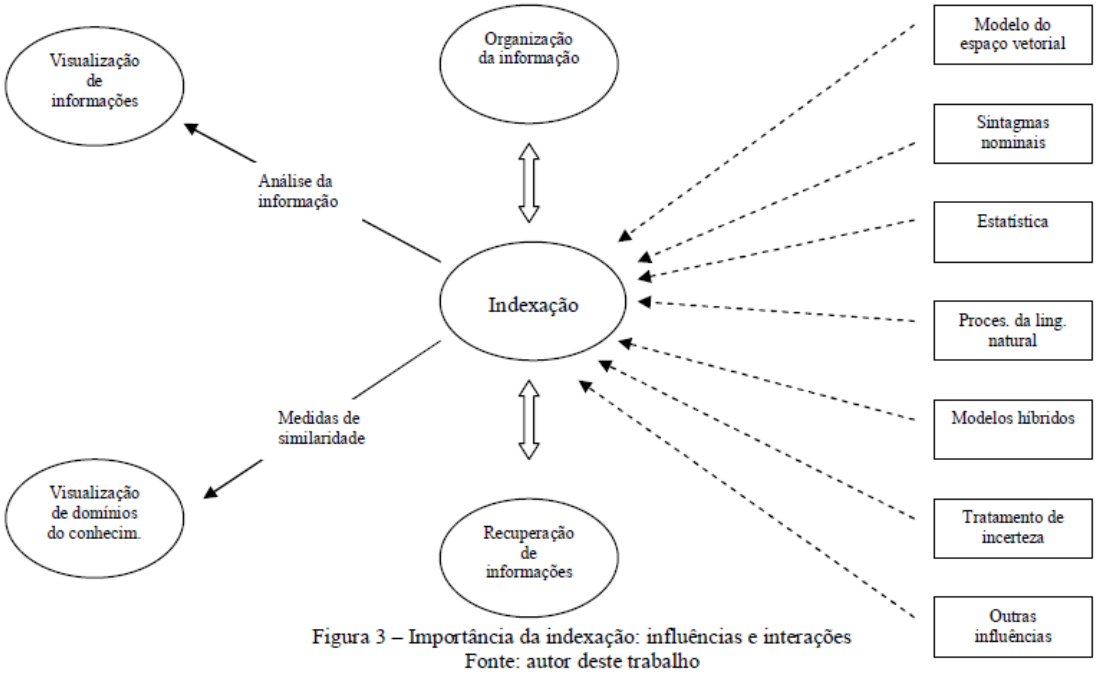


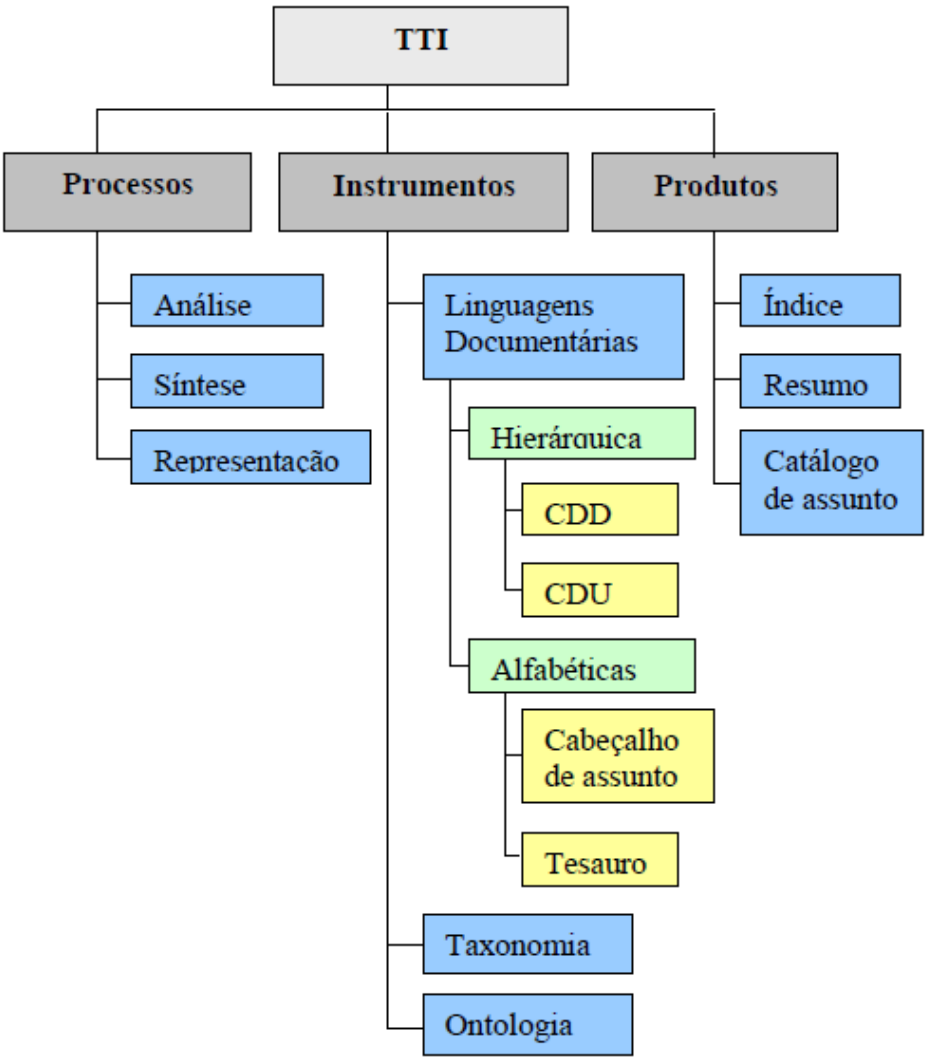
Figura 3 – Importância da indexação: influências e interações
Fonte: autor deste trabalho

Fonte: Corrêa (2011, p. 43)

Para Chaumier (1988) a indexação é a operação central de um sistema para a armazenagem e pesquisa das informações. Ou seja, é a partir das representações decorrentes do processo de indexação que a comunicação da informação acontece de forma mais fluida. Especialmente em bases de dados internacionais que se utilizam de palavras-chave e descritores para representar os conteúdos dos trabalhos que ali existem, para que assim eles possam ser encontrados e acessados e visualizados internacionalmente.

Como visto no decorrer desse estudo, o tratamento temático da informação é composto por processos, produtos e instrumentos. Os **processos** dizem respeito às etapas do tratamento da informação, os **produtos** são resultados dos processos e os **instrumentos** tendem a fornecer as diretrizes para que o processo seja cumprido da melhor forma possível. Na Figura 8 é possível observar com mais clareza a sistematização dos processos, produtos e instrumentos do tratamento temático da informação.

Figura 8 – Aspectos inerentes ao TTI



Fonte: Vieira (2014, p. 93).

Evidencia-se uma estreita relação entre processos, produtos e instrumentos do TTI, visto que a indexação, enquanto processo de representação temática, é realizada a partir de instrumentos que fornecem as diretrizes necessárias para a construção de produtos que possam auxiliar o usuário a encontrar o que deseja em um sistema de busca e recuperação da informação.

Na próxima seção serão comentados de forma sintetizada alguns instrumentos e produtos de representação da informação. Contudo, o objeto de estudo dessa pesquisa são as palavras-chave e *keywords plus* de um conjunto de documentos específico. Logo, a discussão seguinte não se aprofundará em todos os tipos de instrumentos e produtos dantes mencionados.

Muitos autores defendem que as palavras-chave são instrumentos de representação da informação. Contudo, é necessário esclarecer que as palavras-chave, por si só, não são de fato

um instrumento do tratamento temático. No entanto, quando controladas, seguindo políticas e diretrizes de indexação, transformam-se em vocabulário controlado e, assim, adquirem o caráter de instrumento de representação.

Outros autores acreditam que as palavras-chave são produtos do processo de indexação, que tem como finalidade a recuperação da informação. Mas é importante salientar que todo instrumento pode ser também um produto de representação e vice-versa. No caso das palavras-chave e *keywords plus* da WoS, as mesmas são consideradas produtos de representação uma vez que a partir delas é possível acessar diversos conteúdos informacionais.

2.4.1 Instrumentos e Produtos de Representação da Informação

Nos estudos sobre representação documentária é perceptível que as linguagens de representação da informação, bem como os seus instrumentos e produtos são essenciais para a comunicação dos documentos informacionais em qualquer sistema de informação. Nesse sentido, compreende-se que a comunicação científica só é possível pela existência dos canais de comunicação que, em conjunto com os instrumentos e produtos de representação da informação, possibilitam a busca, recuperação e acesso aos recursos informacionais.

Além disso, é necessário acentuar que os instrumentos e os produtos de representação da possibilitam a visualização de domínios do conhecimento. A partir dessa visualização é possível detectar os temas mais discutidos nesses domínios e, assim, compreender a evolução dos campos científicos.

Chaumier (1988) considera que existem dois importantes instrumentos de representação da informação: os *sistemas de classificação*, que são linguagens de estruturas hierárquicas e os *tesauros*, que são linguagens de estruturas combinatórias ou alfabéticas.

Na concepção do autor supracitado (1988), os sistemas de classificação são considerados o primeiro tipo de instrumento de indexação voltado à representação documentária. Baseia-se na pré-coordenação para extrair os conceitos e no encaixe das classes de conceitos, partindo sempre do geral para o mais específico, seguindo uma estrutura hierárquica.

Já o tesauro é definido como "[...] uma linguagem controlada, constituída de descritores (palavras ou expressões) passíveis de combinação entre si, no momento da indexação, para exprimir noções complexas" (CHAUMIER, 1988, p. 70-71).

Vogel (2007, p. 91) argumenta que o tesauro é "[...] um objeto cultural que registra e representa o conhecimento segundo parâmetros estáveis, previamente determinados e manifestos sob a forma de redes de relações entre descritores [...]". Ou seja, os tesauros são linguagens documentárias padronizadas que seguem determinadas diretrizes e normas para sua elaboração, de modo que as relações estabelecidas entre os termos utilizados, consigam cobrir determinados domínios do conhecimento.

Enquanto produto documentário, o tesauro tem suas origens no século XIX, quando em 1852, quando Peter Mark Rogel publicou seu *The Roget's Thesaurus*, uma coleção de palavras organizada não em ordem alfabética, como nos dicionários, mas de acordo com as ideias que expressam (VICKERY, 1960).

Segundo Vickery (1960) o *Thesaurus* de Roget tem duas características básicas: seu *propósito* e sua *forma*. Seu *propósito* é ajudar o usuário a passar de uma ideia para a palavra que ele pode usar para expressar essa ideia em um texto escrito. Quanto a sua forma esse tesauro é dividido em duas partes. A primeira, é uma estrutura classificatória de ideias com diferentes categorias subdividas em tópicos e, a segunda parte, é composta por um índice alfabético, que faz associação entre os cabeçalhos (sob os quais ocorrem as palavras e frases) e os números, que representam as ideias na parte sistemática (VICKERY, 1960; CAMPOS, 2001).

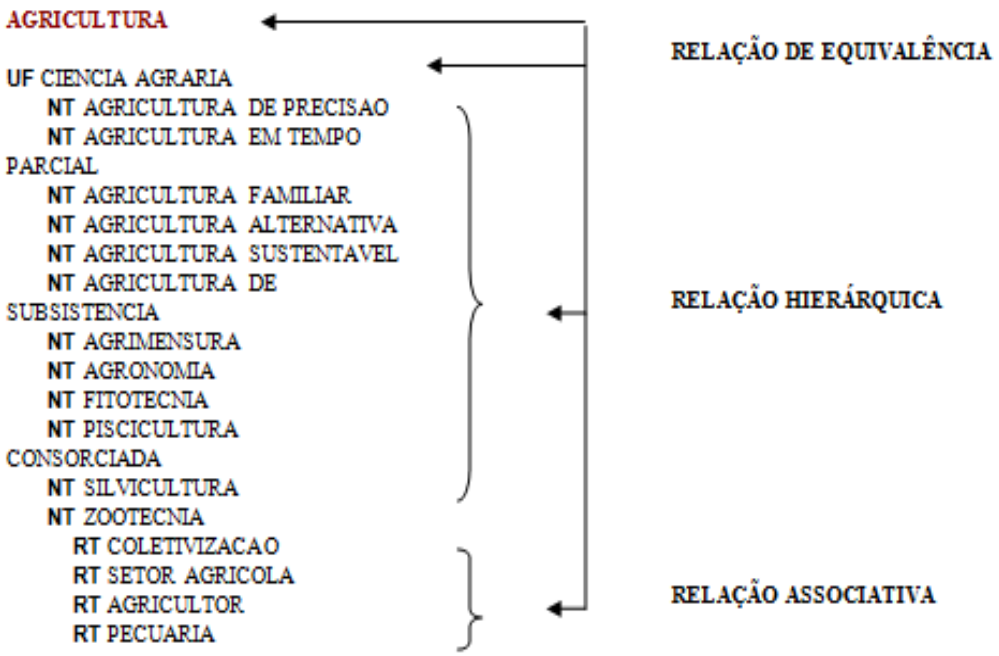
Contudo, para Chaumier (1988) o tesauro, enquanto instrumento de representação documentária, vai além de uma simples lista alfabética de descritores. Pois, para o autor, o tesauro permite *relações semânticas* entre os termos de forma que guiem o indexador no momento da indexação. Tais relações são classificadas em três tipos: de **equivalência**, **hierárquica** e **de associação**. A relação de equivalência permite remeter os termos sinônimos e quase sinônimos ao termo descritor; a relação hierárquica permite relacionar um termo ao outro de modo hierarquizado, ou seja, sempre do geral para o particular; a relação associativa permite unir dois termos que possuem conotações entre si (geralmente, foi usada por muitos anos no sentido de indicar a expressão "ver também").

Os tesauros também utilizam alguns operadores de sentido para indicar as relações entre suas unidades, isto é:

- ✓ **TG** – Termo genérico (de maior abrangência);
- ✓ **TE** – Termo específico (de maior especificidade);
- ✓ **TR** – Termo relacionado (relações associativas, não hierarquizadas);
- ✓ **UP** – Usado para (indica os termos sinônimos e quase sinônimos);
- ✓ **USE** – Indica preferência do termo (o termo mais adequado);
- ✓ **NE** ou **NA** – Nota Explicativa ou Nota de Aplicação (definições ou aplicações para o uso do termo) (CHAUMIER, 1988; DODEBEI, 2002; NARUKAWA, 2011; VOGEL, 2007).

A figura a seguir mostrar um exemplo de tesauro contendo os três tipos de relações: de **equivalência**, **hierárquica** e de **associação**:

Figura 9– Relacionamentos entre termos estabelecidos nos tesauros



Fonte: Narukawa (2011, p. 33) adaptado do THESAGRO (*Thesaurus* Agrícola Nacional)

Além dos sistemas de classificação bibliográfica e dos tesauros, existem outros instrumentos de representação da informação e do conhecimento, tais como as **taxonomias** e **ontologias**.

As **taxonomias** são um instrumento de representação da informação que utiliza procedimento classificatório para agrupar e categorizar assuntos. Isto é, através de um assunto, é possível criar categorias que se subdividem em classes e subclasses, de modo a construir uma lista de assuntos estruturada hierarquicamente (TRISTÃO; FACHIN, ALARCON, 2004).

Segundo Aquino, Carlan e Brascher (2009) o termo taxonomia é de origem grega *táxis* (ordem) e *onoma* (*nombre*) e foi derivado do ramo da Biologia que estuda a classificação lógica e científica dos seres vivos. Ou seja, trata-se da Biologia Sistemática, fruto do trabalho do médico e botânico Carolus Linnaeus. Contudo, devido o desenvolvimento das tecnologias de informação, e a necessidade de organizar e recuperar informação de forma mais eficiente, as taxonomias foram inseridas e implementadas nos ambientes digitais, tornando-se alvo de estudo da Ciência da Informação (AQUINO; CARLAN; BRASCHER, 2009).

Além disso, as taxonomias têm sido empregadas em portais corporativos de empresas, bibliotecas digitais e sítios de instituições governamentais com o objetivo de organizar e recuperar informações (AQUINO; CARLAN; BRASCHER, 2009). Ou seja, a taxonomia é

um vocabulário controlado que organiza logicamente os conteúdos informacionais para fins de recuperação da informação.

As **ontologias**, por seu turno, que também são uma espécie de vocabulário controlado, que busca representar, de forma estrutural e relacional, os conceitos de um domínio.

Brascher e Carlan (2010, p. 160) comentam que as ontologias "definem conceitos e relações de alguma área do conhecimento, de forma compartilhada e consensual e promovem e facilitam a interoperabilidade entre sistemas de informação, em um processo 'inteligente' dos agentes (computadores)".

Segundo Almeida e Bax (2003, p. 7) "uma ontologia é criada por especialistas e define as regras que regulam a combinação entre termos e relações em um domínio do conhecimento. Os usuários formulam consultas usando conceitos definidos pela ontologia". Nesse sentido, a ontologia tem por objetivo promover melhorias no processo de recuperação da informação.

Tanto os tesauros quanto as taxonomias e as ontologias são vocabulários controlados com características comuns e construídos seguindo uma classificação lógica, hierárquica e relacional. Tratam da representação e estruturação de conceitos e seus relacionamentos em um determinado domínio e contribuem para a construção dos produtos de representação e recuperação da informação.

Os principais produtos de representação da informação são os índices, os resumos e os catálogos de assunto. Geralmente o índice consiste em uma listagem alfabética ou sistemática de tópicos que indicam a localização de cada um deles num documento ou em uma coleção de documentos. Mas também pode ser um conjunto ordenado de códigos de representação de assuntos que podem servir de critério de busca para localizar documentos (DIAS; NAVES, 2007; VIEIRA, 2014.)

O resumo, enquanto produto do tratamento temático, diz respeito a representação sucinta, mas exata do conteúdo de um documento, tendo como finalidade a posterior recuperação desse documento e também a complementar os pontos de acesso gerados pelos termos de indexação (LANCASTER, 2004).

Já o catálogo de assunto é um produto da representação temática que é organizado "mediante determinação de cabeçalhos de assunto que funcionam como enunciados de assuntos formados a partir da composição ordenada de palavras" (SILVA; FUJITA, 2004). O principal objetivo do catálogo de assunto é possibilitar ao usuário a encontrar um item desejado ou a identificar o que existe sobre determinada temática em uma unidade de informação.

Tanto os instrumentos quanto os produtos do tratamento temático são essenciais para a recuperação da informação. Como mencionado no início dessa pesquisa, as palavras-chave também são produtos de representação e recuperação da informação. Nesse sentido, tendo as palavras-chave e as *keywords plus,* dos documentos da área de Nutrição da WoS, como objeto de investigação dessa pesquisa, a seguinte seção abordará alguns conceitos e definições acerca desse produto do tratamento temático, buscando mostrar sua utilização e aplicação em sistemas de recuperação da informação e como a mesma pode contribuir para a construção de representações visuais.

2.4.2 Palavras-chave e Visualização da Informação

O acesso ao conhecimento científico depende de um conjunto de variáveis que podem estar relacionadas a fatores políticos, ideológicos e metodológicos. Nesse contexto se inserem as linguagens de representação da informação que também estão atreladas a tais fatores.

É comum, no campo científico, o uso de palavras-chave e descritores para representar e comunicar o conhecimento. Dahlberg (1978, p. 101) aponta que "desde que o homem foi capaz de falar, empregou palavras (conjunto de símbolos) para designar os objetos de suas circunstâncias assim como para traduzir os pensamentos formulados sobre os mesmos". Para a autora (1978, p. 101), "o conhecimento fixou-se através dos elementos de linguagem" e esta, por sua vez, permite que o homem se comunique com seus semelhantes (DAHLBERG, 1978).

Na Teoria do Conceito proposta por Dahlberg (1978), evidencia-se que o conceito é a base para a organização do conhecimento. Isso porque o conceito permite a representação de enunciados verdadeiros sobre um determinado objeto, ou seja, do conhecimento registrado sob a forma de documento (impresso ou eletrônico). Essa representação é fixada por um símbolo linguístico verbal ou não verbal (DAHLBERG, 1978) e as palavras-chave fazem parte desse universo representacional.

Segundo Cunha e Cavalcanti (2008, p. 274), em seu Dicionário de Biblioteconomia e Arquivologia, a palavra-chave é definida como "palavra significativa encontrada no título de um documento, no resumo ou no texto. Essa palavra (ou grupo de palavras) caracteriza o conteúdo temático do item e é usada em catálogos e índices de assuntos".

Fujita (2004, p. 258) afirma que "a palavra-chave é uma representação do conteúdo significativo do texto e também é utilizada para representar uma necessidade de informação

na estratégia de busca". Percebemos, então, que a palavra-chave se comporta como a tradução do conteúdo informacional em termos que o represente fielmente.

Sua elaboração é resultado do processo de indexação, em que é feito, primeiro, a análise de conteúdo (análise documentária), momento em que são selecionados os conceitos que representam o assunto, depois é realizada a tradução, isto é, a transformação do assunto em termos, que pode ser feita por extração (uso da linguagem natural) ou por atribuição (uso da linguagem artificial).

A palavra-chave - enquanto uma forma de representação da informação - deve permitir a comunicação entre usuário e sistema de informação, possibilitando, através dos mecanismos de busca, que o sistema responda satisfatoriamente as indagações do usuário, retornando para este, informações das quais necessita.

É importante não confundir **palavra-chave** com **descritor**. Nessa perspectiva, Migués e Neves (2013, p. 117) compreendem que "as palavras-chave não são o mesmo que descritores. Enquanto àquelas são retiradas do repertório linguístico do autor do texto, estes são frutos de análise profissional para a escolha dos termos mais representativos, os quais são traduzidos em linguagem documentária".

"[...] Para uma palavra-chave tornar-se um descritor, ela tem que passar por um rígido controle de sinônimos, significado e importância na árvore de um determinado assunto" (BRANDAU; MONTEIRO; BRAILE, 2005, p. 8).

É possível que as palavras-chave sejam elaboradas seguindo um controle vocabular. Isso vai depender da política adotada pela equipe da instituição que as elaboram, quer sejam periódicos científicos, bibliotecas tradicionais ou digitais, ou mesmo em bases de dados. Assim, as palavras-chave desempenham a mesma função que os descritores (representar conteúdos informacionais e comunicá-los), o que releva a sua utilidade para o processo da comunicação científica e, por isso, devem ser elaboradas com mais minúcia.

Diferentes estudos podem ser realizados a partir do uso das palavras-chave, como os estudos bibliométricos, os estudos de análise de domínio, os estudos de avaliação de sistemas de recuperação da informação, entre outros. Além disso, as palavras-chave servem de insumo para a construção de vocabulários controlados e estruturados, o que as tornam de fundamental importância na representação dos campos do conhecimento.

Quando indexadas corretamente, as palavras-chave contribuem para a melhor precisão no processo de recuperação da informação, pois quanto mais específica for a busca e, os termos indexados no sistema de informação corresponderem ao nível de especificidade do

usuário, o índice de revocação diminui, garantindo que os documentos recuperados sejam o mais próximo possível da resposta que o usuário necessita.

Alguns autores como Fujita (2004), Araújo Júnior (2007), Dias e Naves (2007) e Borges e Lima (2012) concordam que as palavras-chave, sem um controle terminológico, não são tão eficientes nos sistemas de recuperação da informação (geralmente, esse controle diz respeito à especificidade dos termos que irão representar determinado conhecimento). Porém, é necessário ressalvar que dependendo do propósito do uso das palavras-chave, nem sempre a utilização da especificidade vai ser a melhor escolha. Do ponto de vista da recuperação da informação, a especificidade se torna um ponto positivo quando o SRI tem por característica assuntos específicos. Logo, seria incoerente fazer uso de termos mais abrangentes.

Do ponto de vista da Visualização da Informação (VI), principalmente daquela advinda dos métodos bibliométricos (que utilizam a contagem de palavras-chave para mapear as tendências de um domínio), a especificidade só será favorável quando os termos coincidirem, isto é, quando houver elevada ocorrência de palavras. Pois, do contrário, a especificidade inconsistente, acarretará na pulverização de termos, o que acaba não favorecendo a VI. Para uma melhor compreensão do que se trata a VI e quais suas aplicações, buscou-se alguns conceitos na literatura para explicá-la brevemente.

A VI pode ser entendida como uma área da ciência que tem por objetivo o estudo e aplicação das técnicas de representações gráficas para facilitar a compreensão de informações complexas. Trata-se, então, de todo tipo de representação imagética para apresentar dados, informações e conhecimentos, de forma que estes sejam percebidos e compreendidos com mais clareza.

Segundo Freitas (et al., 2001, p. [143]) a VI "é uma área de aplicação de técnicas de computação gráfica, geralmente interativas, visando auxiliar o processo de análise e compreensão de um conjunto de dados, através de representações gráficas manipuláveis. Essa concepção de visualização tem como ponto central a satisfação do usuário em um sistema de busca e recuperação da informação.

Aguilar (et al., 2017) compreendem que a VI tem como foco central a simplificação de conteúdos para a compreensão da ideia geral, de modo a facilitar sua percepção. Os autores concordam que a visualização pode ser aplicada em todos os âmbitos (sociais, econômicos, políticos e científicos), pois busca facilitar o entendimento de necessidades básicas para uma sociedade, a compreensão de tomadas de decisão, utilizando gráficos e outros tipos de representações visuais, ao invés de utilizar informações extensas e complexas. Enfim, a VI possui vários tipos de aplicações.

Vieira e Correa (2011) acrescentam que com a sobrecarga de informações que aumenta continuamente, fica muito difícil encontrar tudo o que se deseja de forma rápida e eficiente. Desse modo, a visualização de informações apresenta potencial para ajudar as pessoas a encontrarem o que precisam de forma mais efetiva e intuitiva. Sendo assim, a visualização proporciona ao usuário consultar coleções de informação sem demandar muito esforço, auxilia na identificação de padrões quando se está diante de uma situação que envolve a manipulação de grandes quantidades de dados e é um modo de enxergar o "invisível" ou de descobrir o "desconhecido" (CHEN 2006).

Diante do que foi exposto sobre VI, entende-se que esta possui duas dimensões, a primeira que tem por objetivo o uso de interfaces gráficas e interativas, implementadas em SRI para facilitar a busca e a recuperação da informação de forma mais simples e objetiva. Já a segunda dimensão está relacionada à apresentação de dados, informações e conhecimentos através de imagens, gráficos e cartografias, com o intuito de facilitar o seu entendimento (AGUILAR, et al., 2017). Nessa última, se encaixa a visualização de dados bibliométricos, como no caso dos mapas gerados pela coocorrência de palavras-chave. É esse tipo de visualização que interessa a presente pesquisa.

Acerca da primeira dimensão da VI, é possível aferir, a partir de Chen (2006) que a visualização permite que os conteúdos informacionais sejam identificados de forma mais rápida, de modo que os usuários de qualquer sistema de informação encontrem o que necessitam com mais precisão, além disso, a VI possibilita a descoberta de novos conhecimentos.

Já a segunda dimensão da VI possibilita, através dos dados bibliométricos, gerar mapas de representação da produção científica em diferentes áreas do conhecimento (KOBASHI; SANTOS, 2008). Esse tipo de visualização consegue capturar os principais conceitos de um campo e demonstrá-los através de representações cartográficas.

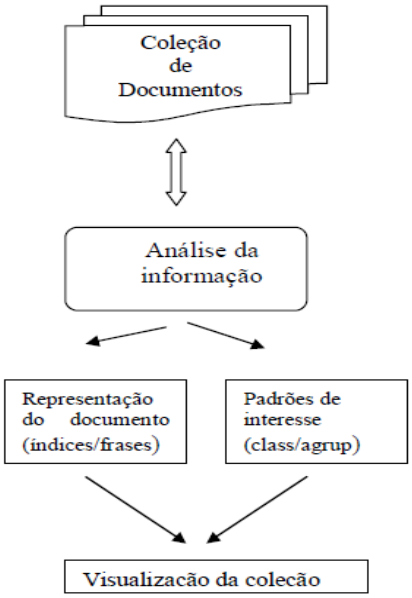
Para representar de forma visual os assuntos de uma área de estudo, é necessário que os termos utilizados na indexação sejam elaborados adequadamente, isto é, não devem ser amplos e ambíguos, mas claros e objetivos. Outro ponto importante é que os termos também não podem ser tão específicos, ao ponto de serem desconhecidos na literatura. Desse modo, quando são construídas representações gráficas para identificar assuntos, as palavras de maior frequência representarão os domínios, enquanto que as palavras de menos frequência ficam dispersas na visualização e não são consideradas representativas por serem pouco utilizadas.

As técnicas de visualização da informação buscam representar graficamente dados de diferentes domínios, de modo que a representação visual aumente a capacidade da percepção

humana em interpretar as informações apresentadas e deduzir novos conhecimentos (FREITAS, et al., 2001). Na literatura existem várias técnicas de VI. As principais são: mapas das ciências, mapas conceituais, mapas geográficos, desenho de grafos, desenho de árvores, browser hiperbólico, *clusters*, nuvem de *tags*, entre outras (VIEIRA; CORREA, 2011; VIEIRA, 2014; AGUILAR, et al., 2017).

A representação visual também necessita da análise da informação, isto é, uma das etapas da Representação Documentária. Nesse sentido, Correa (2011, p. 87) apresenta um esquema da influência da análise da informação no desenvolvimento de sistemas de visualização, vide Figura 10.

Figura 10 – Análise da informação e visualização



Fonte: Correa (2011, p. 87)

A representação da informação bem elaborada e de forma consistente é fundamental para que exista uma comunicação mais efetiva entre o usuário e o sistema. Contudo, algumas medidas devem ser tomadas para que a busca e a recuperação da informação, bem como os estudos voltados à visualização de domínios não sejam prejudicados. É interessante que exista equilíbrio, uma combinação entre termos específicos e controlados com termos mais gerais e não controlados. A indexação não é uma tarefa fácil, sendo importante medir prioridades no momento de tomar decisões ao representar determinado assunto. Com isso, se o objetivo for uma recuperação mais eficiente, os termos controlados são a opção mais adequada. Enquanto isso, se o objetivo for uma representação para a visualização de domínios, talvez a especificidade não seja a saída.

Os termos recuperação da informação e visualização da informação caminham juntos. Muitos sistemas de informação desenvolvem formas de visualização de domínios do conhecimento para fins de recuperação da informação. Nesse sentido, uma indexação de qualidade compreende tanto os aspectos que envolvem a RI quanto à visualização da informação.

Contudo, existem algumas particularidades que acabam por dar rumos diferentes aos estudos de representação voltados à visualização da informação e a recuperação da informação. Para este último, essencialmente, a qualidade da indexação é verificada pelo grau de exaustividade, especificidade e/ou consistência que o assunto de um documento é representado, dependendo da finalidade e da política adota por determinado SRI. Quanto aos aspectos dedicados à visualização da informação, a indexação torna-se mais apropriada quando não contém termos amplos ou específicos demais ao ponto de não serem utilizados comumente pelos pesquisadores da área investigada. Quando a indexação acarreta a dispersão de termos, devido ao rico aspecto semântico das palavras, fica difícil identificar as tendências de um domínio.

3 PERCURSO E PROCEDIMENTOS METODOLÓGICOS

Esta seção se dedica a abordar os caminhos que foram percorridos para alcançar os objetos propostos por este estudo. Trata-se então de definir o tipo de pesquisa quanto aos objetivos e às técnicas de análise dos dados; o tipo de abordagem utilizada, o *corpus* da pesquisa, os procedimentos de coleta e organização dos dados e os critérios utilizados para análise e avaliação dos dados.

3.1 Tipo e abordagem da pesquisa

O percurso metodológico dessa pesquisa foi dividido em duas partes: uma teórica que diz respeito à busca por literatura que fundamentasse o presente estudo e, uma prática, que se dedica à coleta e análise das palavras-chave e descritores da área de Nutrição no contexto da saúde pública, de artigos científicos disponíveis na base de dados *Web of Science* (WoS).

No que concerne à primeira etapa, foram utilizados diversos documentos em diferentes formatos como (artigos científicos, livros, anais de eventos, teses e dissertações, conteúdo de sites, entre outros). As fontes utilizadas para recuperar tais documentos foram bases de dados nacionais e internacionais, BDTDs, repositórios, bibliotecas físicas e digitais, entre outros meios. As principais fontes de busca foram: *Web of Science*, *Scopus*, Scielo, Brapci, BVS, Lilacs, BIREME, site do Ministério da Saúde, BDTD-IBICT, Google Acadêmico e anais do ENANCIB.

Esta pesquisa pode ser definida como exploratória, visto que busca: avaliar a qualidade da representação da informação na WoS através das palavras-chave e *keywords plus* dos artigos da área de Nutrição, a fim de observar como esses termos se comportam diante das visualizações geradas a partir de sua coocorrência. Gil (2002) acrescenta que este tipo de pesquisa busca proporcionar maior familiaridade com o problema a fim de explicitá-lo ou a construir hipóteses.

O foco desse trabalho é analisar a dinâmica dos dados coletados e, a partir desse ponto, encontrar respostas para a indagação inicial, refletir sobre o objeto e abrir novos questionamentos e posições que contribuam para a dinâmica da pesquisa científica nesse contexto.

3.2 *Corpus* da pesquisa

O *corpus* dessa pesquisa é composto pelos registros recuperados na base de dados WoS, isto é, pela representação temática dos artigos científicos relacionados à área de Nutrição no contexto da Saúde Pública, tendo como objeto de estudo as palavras-chave dos autores e as *keywords plus* da base de dados. Entretanto, os Descritores das Ciências da Saúde – DeCS também fazem parte desse *corpus*, uma vez que foram utilizados como parâmetro para realizar a busca na WoS. Nessa seção serão apresentadas as principais características da base de dados, seus objetivos e os recursos oferecidos para pesquisar diferentes tipos de documentos. Além disso, também serão mostradas as principais características do DeCS, a fim de exemplificar como foram pesquisados os descritores em Nutrição.

3.2.1 Características da *Web of Science*

A *Web of Science*, antes conhecida como *Web of Knowlege*, é um serviço de citações científicas on-line, criado pelo ISI – *Institute for Scientific Information*, mantido anteriormente pela *Thompson Reuters* e atualmente pela *Clarivate Analytics*. Tem como propósito fornecer uma pesquisa abrangente de citações, dando acesso a vários bancos de dados por meio de assinatura.

A ideia de recuperar a informação através de citações foi cunhada por Eugene Garfield (fundador do ISI), conhecido como o "pai da indexação de citações da literatura acadêmica" e também considerado pioneiro nas áreas de Bibliometria e Cientometria moderna. Para Garfield a teoria da indexação por citação diz que se existe um artigo relevante que cita determinados autores, e outros artigos que citam esses mesmos autores, certamente estes artigos serão de mesma relevância. Ou seja, as referências citadas por um autor identificam o relacionamento entre documentos que tratam do mesmo assunto de forma mais precisa (GARFIELD, 1995).

A WoS é considerada uma das maiores bases de dados do mundo que reúne desde 1997 até o presente momento, a literatura acadêmica em Ciências, Ciências Sociais, Artes e Humanidades, publicada nos principais periódicos de acesso aberto da América Latina, Portugal, Espanha e África do Sul (WEB OF ..., 2017).

Esta base ou banco de dados contempla a literatura científica de diversos países, proporcionando maior visibilidade à produção do conhecimento científico em termos internacionais.

Segundo Nunes (2014, 43), em consulta ao site da *Thomson Reuters*, no ano de 2013, a *Web of Science* cobria 256 disciplinas, dividindo-se em *Conference Proceedings, Citation Index Science* – CPCI-S e *Conference Proceedings Citation Index Social Sciences and Humanites* – CPCI-SSH, que juntas possuem mais de 148.000 conferências desde 1990, além de referências e contagens de citações, cumulativas desde 1999. Todos os anos são adicionados cerca de 400.000 anais, com link direto para acessar o texto completo dos trabalhos apresentados nos congressos.

Muito utilizada por pesquisadores, empresas e governos de vários países, esta base garante sua importância e credibilidade no mercado. No entanto, sua alta valorização não diz respeito apenas à consulta a documentos, mas também porque a WoS possui indicadores bibliométricos que possibilita acompanhar o fator de impacto dos periódicos por meio do número de suas publicações, além de permitir acompanhar o número de citações que determinado autor recebeu.

A busca na WoS pode ser por pesquisa básica, busca por autor, por referência citada e por pesquisa avançada. A base possui os operadores booleanos (AND, OR e NOT), os caracteres coringas (*), (?) e ($) que são usados para fazer truncamento e obter mais controle sobre a recuperação de plurais e variantes ortográficas; entre outras ferramentas que facilitam a busca e a recuperação de documentos. A seguir, por meio do Quadro 4, são mostrados os campos de pesquisa que a WoS oferece.

Quadro 4 – Campos dos registros bibliográficos da WoS

CAMPOS DE PESQUISA WoS
TS=Tópico: assunto TI=Título AU=Autor AI=Identificadores de autor ED=Editor SO=Nome da publicação DO=DOI DY=Ano de publicação AD=Endereço OG=Organizações – Aprimorada

Continua...

Continuação

CF=Conferência LA=Idioma DT=Tipo de documento FO=Agência financiadora FG=Número do subsídio UT=Número de acesso PMID=ID PubMed

Fonte: *Web of Science*, 2017.

Um detalhe importante do Quadro 4 é que a pesquisa por tópico (TS) permite recuperar documentos em que os termos de busca aparecem no título, no resumo, nas palavras-chave de autor e nas *keywords plus* dos documentos.

Keywords plus são descritores elaborados pela própria base de dados, através dos títulos das referências citadas e são capazes de capturar o conteúdo de um artigo com maior profundidade e variedade (GARFIELD, 1990). Compreende-se, então, que as *keywords plus* conseguem ampliar e complementar o resultado de uma busca, retornado, muitas vezes, documentos relevantes que somente com as palavras-chave de autor não seriam recuperados.

3.2.2 Descritores em Ciências da Saúde – DeCS

O DeCS – Descritores em Ciências da Saúde é um vocabulário estruturado e trilíngue que foi criado pela BIREME para servir como linguagem única na indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos, e outros tipos de materiais. Também é utilizado na pesquisa e recuperação de assuntos da literatura científica nas fontes de informação disponíveis na Biblioteca Virtual em Saúde – BVS como LILACS, MEDLINE e outras (DECS, 2017).

Foi desenvolvido a partir do MeSH - *Medical Subject Headings da U.S. National Library of Medicine* (NLM) com o objetivo de permitir o uso de terminologia comum para pesquisa em três idiomas, proporcionando um meio consistente e único para a recuperação da informação independentemente do idioma. Nunes (2014, p. 68) acrescenta que o DeCS é a versão latino-americana do MeSH, sendo um dos tesauros mais utilizados na área da saúde em nível mundial (ANDALIA; CHAPMAN, 2011).

Quanto a atualização dos descritores, esta é feita anualmente pelo MeSH. Devido à dinamicidade do conhecimento, ou seja, das mudanças que ocorrem na ciência, os termos passam por atualização de forma que a cada ano um mínimo de 1000 interações na base de dados dentre alterações, substituições e criações de novos termos ou áreas (DeCS, 2017).

A forma de organização dos conceitos e termos que compõem o DeCS é em uma estrutura hierárquica, possibilitando uma pesquisa por termos mais amplos ou mais específicos ou por todos os termos que pertençam à mesma estrutura hierárquica. A seguir serão mostradas as categorias hierárquicas do DeCS (Figura 11).

Figura 11–Categorias Hierárquicas do DeCS

Fonte: Consulta DeCS, 2017.

A princípio, percebe-se que só existem vinte categorias hierárquicas, mas outras tantas estão inclusas nessas categorias principais. A área de Nutrição, como pode ser observada, não é uma categoria principal, mas uma subcategoria que faz parte da categoria de Saúde Pública. Nessas condições, ficou decidido pesquisar por palavras-chave e descritores da área de Nutrição no contexto da Saúde Pública.

3.3 Procedimentos de Coleta e Organização dos Dados

Esta seção se dedica a explicar como os dados foram coletados e organizados para posterior análise e interpretação. Tais procedimentos seguiram algumas etapas: na **primeira etapa** foi feito um levantamento dos descritores mais utilizados na área de Nutrição a partir do DeCS. A busca foi realizada em inglês, uma vez que o DeCS, apesar de ser uma versão latino-americana do MeSH, a maior parte dos seus descritores permanece em inglês. Ademais, como o *corpus* de análise dessa pesquisa foi obtido a partir de uma base de dados

internacional, optamos por adotar uma língua comum. Nessa primeira etapa, para identificar em qual categoria se encontrava a área de Nutrição, utilizamos a expressão de busca "nutrition" através do índice hierárquico, que nos remeteu para a categoria SP1 – *Public Health* e à subcategoria SP6 – *Nutrition, Public Health*. Após identificar o código das categorias, utilizou-se um dos serviços do DeCS denominado "Ocorrência de conceitos DeCS na BVS", que permitiu identificar todos os termos pertencentes à categoria SP6. Nesse caso, foram identificados 153 descritores da área de Nutrição, dos quais, foram selecionados, para realizar a busca na WoS, os vinte descritores mais recorrentes nas bases de dados apontadas pelo próprio DeCS.

A escolha por um número menor de descritores, justifica-se, em primeiro lugar, por não se tratar de uma pesquisa quantitativa. Portanto, a quantidade de dados não é o seu objetivo. Em segundo lugar, elegeu-se os vinte primeiros descritores de maior ocorrência nas bases da BVS porque acredita-se que estes possuem mais prestígio, mais aceitação e usabilidade pela comunidade discursiva.

Todavia, é necessário ressalvar que pelo fato de a área de Nutrição pertencer à categoria Saúde Pública dentro do DeCS, além de utilizar os descritores dantes mencionados, foi preciso adotar determinados critérios de busca na WoS, delimitando a pesquisa ao contexto da Saúde Pública. A seguir, na Tabela 1, são mostrados os vinte descritores do DeCS de maior ocorrência nas bases da BVS.

Tabela1 – Os dez descritores do DeCS de maior ocorrência na BVS

Conceito	LILACS	MEDLINE	EQUIDAD	BBO	BDENF	HOMEOINDEX	DESASTRES	MEDCARIB	PAHO	WHOLIS	IBECS	REPIDISCA	Ocorrência
Body Weight	1761	167926	5	22	10	4	0	206	21	35	311	19	170320
Obesity	5097	141145	137	60	125	36	1	174	66	70	2507	149	149567
Diet	2277	119220	15	66	16	18	7	134	56	96	467	87	122459
Body Mass Index	2711	97169	10	21	26	0	0	59	3	6	1274	12	101291
Feeding Behavior	3153	67132	0	9	32	9	1	41	16	76	315	36	70820
Dietary Fats	371	42895	0	4	1	1	0	18	4	11	180	3	43488
Food Contamination	1420	33986	1	2	3	2	17	35	219	763	107	1074	37629
Anthropometry	2619	33277	2	46	30	0	4	108	55	55	913	33	37142
Birth Weight	1194	34477	4	3	11	0	0	89	47	23	130	13	35991
Breast Feeding	3366	28437	16	162	441	5	1	203	275	391	412	71	33780

Continua...

Continuação

Dietary Proteins	**402**	**32779**	**0**	**4**	**2**	**0**	**0**	**89**	**19**	**10**	**112**	**1**	**33418**
Body Height	815	32310	1	8	1	0	0	139	13	28	73	10	33398
Nutritional Status	3659	27154	9	20	56	0	22	295	208	134	903	265	32725
Weight Loss	529	28380	4	3	3	1	0	20	2	7	364	3	29316
Cephalometry	1687	24058	0	1059	0	0	0	12	1	0	93	0	26910
Weight Gain	689	25805	3	2	3	2	0	35	6	5	162	4	26716
Food	1172	23037	24	12	11	10	27	58	86	201	107	286	25031
Dietary Carbohydrates	234	22644	0	10	0	0	0	21	1	5	63	2	22980
Food Handling	1079	19647	0	2	0	1	9	51	67	183	56	19	21114
Nutrition Disorders	1530	16341	1	5	14	3	29	303	280	117	484	124	19231
Dietary Proteins	402	32779	0	4	2	0	0	89	19	10	112	1	33418

Fonte: dados da pesquisa, 2017.

A segunda etapa dos procedimentos diz respeito à busca e coleta dos dados na WoS. Para este tipo de coleta foram utilizados os descritores do quadro anterior, fazendo uma busca com os mesmos de uma única vez e acrescentando mais um descritor que delimitasse a pesquisa ao contexto da saúde pública. Ou seja, o descritor "public health". A expressão de busca utilizada na base de dados será mostrada mais adiante.

Adotou-se a modalidade de pesquisa básica na WoS, ressaltando que esse tipo de busca não compromete a qualidade da pesquisa, visto que existem vários recursos que podem ser utilizados para filtrar a busca tanto na modalidade de pesquisa básica quanto na modalidade de pesquisa avançada. Desta forma, a seguir temos os delimitadores utilizados na referida base de dados:

a) Delimitadores WoS

TS=TÓPICO[8]

DT=DOCUMENT TYPE

LA=LANGUAGE

WC=CATEGORIA WoS

SO=SOURCE TITLE

b) Expressão de busca na WoS

TS = ("body weight"OR "obesity" OR "diet" OR "body mass index" OR "feeding behavior" OR "dietary fats" OR "food contamination" OR "anthropometry" OR "berth weight" OR "breast feeding" OR "dietary proteins"OR "body height" OR "nutritional status"OR "weight loss" OR "cephalometry" OR "weight gain"OR "food" OR "dietary carbohydrates"OR "food handling" OR "nutrition disorders" AND TS: ("public health")

Refinamentos da busca

DT: ARTIGO

WC: PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH[9];

LA: INGLÊS

[8] Trata-se da busca por assunto.

[9]Por não existir uma categoria de saúde pública na WoS, foi tomada a decisão de refinar a busca pela categoria Public Environmental Occupational Health, pois essa foi a única categoria encontrada na base de dados, que possui alguma relação com saúde pública. Este refinamento tem como objetivo recuperar documentos da área de Nutrição no contexto da saúde pública, visto que os descritores do DeCS possuem essa característica.

Timespan: 2006-2016.
Indexes: SCI-EXPANDED

Além da expressão de busca acima explicitada, foram aplicados alguns delimitadores, já mencionados anteriormente. Os artigos recuperados a partir dessa busca foram numerosos, contudo, decidiu-se analisar **os vinte artigos mais citados na WoS,** lembrando que a escolha por uma quantidade mínima de documentos, justifica-se, primeiro, por não se tratar de uma pesquisa quantitativa; em segundo lugar, seria necessário um período de tempo maior para realizar análise mais minuciosas com um *corpus* elevado.

Com relação à data de publicação dos artigos, vale salientar que se considerou o período de 2006 a 2016 (uma década) como um intervalo considerável para que fossem observadas as mudanças científicas decorrentes da dinâmica do conhecimento. Desse modo, acredita-se que documentos recuperados em um intervalo maior possibilitam detectar o progresso de determinada área a partir dos termos representados em intervalos distintos. Detectar os assuntos mais discutidos por uma área em diferentes períodos possibilita identificar se houve progresso ou estagnação de determinadas temáticas.

A terceira etapa desse procedimento foi dedicada a construir critérios de análise da indexação, embasados na literatura, para avaliar a qualidade das palavras-chave e *keywords plus* identificados nos artigos da base de dados aqui explicitada. Para tanto, foram utilizados os textos de Slype (1991), Strehl (1998), Lancaster (1993, 2004), Dias e Naves (2007), Gil-Leiva (2008, 2012), Anízio e Nascimento (2012), Fujita (2012), Fujita e Gil-Leiva (2014), Lapa (2014), Borges e Lima (2015), e Dias (2015). Esses critérios têm como finalidade analisar a qualidade da representação da informação na WoS, especificamente no domínio de Nutrição, com vistas à visualização da informação.

3.3.1 Definição dos Critérios de Qualidade da Indexação

Essa etapa tem como objetivo definir os critérios de avaliação da qualidade da indexação, identificados na literatura da CI. A partir da análise de literatura, notou-se que existem diversas discussões acerca da qualidade da indexação ou da representação da informação, entre outros termos adotados pelos pesquisadores da área. É importante deixar claro que, nem todos os critérios identificados atenderam a proposta de análise dessa pesquisa.

Nessas condições, elencaram-se apenas aqueles que se adequaram ao estudo. Seguem os critérios:

a) **Exaustividade:** diz respeito à quantidade suficiente de termos indexados para representar o conteúdo de um documento de forma abrangente, aumentando a revocação em sistemas de recuperação da informação – SRIs.
b) **Especificidade:** compreende a elaboração de termos mais específicos para representar o conteúdo documental, de forma que sejam evitados termos genéricos. A especificidade diminui a revocação e aumenta a precisão em SRIs.
c) **Consistência:** diz respeito à similaridade dos termos que são atribuídos a determinado documento por diferentes indexadores ou por um mesmo indexador em momentos distintos numa unidade de informação. Mas também pode ser entendida como o grau de semelhança entre termos que representam um conjunto de documentos que tratam do mesmo assunto.

Os critérios anteriormente definidos estão mais direcionados à qualidade da indexação para a Recuperação da Informação – RI. Nesse sentido, tais critérios não atendem de imediato ao que foi proposto para esta pesquisa. Todavia, esses critérios também podem se adequar à avaliação da indexação para fins de visualização da informação, isto é, avaliar, a partir de cada critério, como que as palavras-chave e *keywords plus* se comportam diante das visualizações geradas a partir da coocorrência de palavras.

Desta forma, resolveu-se elaborar e adaptar outros critérios que possam atender os objetivos propostos.

d) **Evitar termos ambíguos:** entende-se por termo ambíguo aquele que possui mais de um significado. Palavras com diversos significados podem comprometer a qualidade da representação de um domínio. Em caso de adotar seu uso, é necessário especificar usando um qualificador de forma que fique claro o contexto e o significado do termo.
e) **Uniformidade:** diz respeito ao controle de plural e singular, sinonímia, homonímia e outras variações de grafia do termo. Para estudos voltados à identificação de domínios do conhecimento é importante que palavras-chave e descritores sejam utilizados de forma padronizada. Quando isso não acontece,

torna-se difícil analisar grupos de palavras ou termos com vistas a identificar as tendências de um domínio.

3.3.2 Tratamento e Organização dos Dados

Essa seção busca explicar como os metadados do *corpus* de estudo foram coletados e organizados para posterior análise.

Após definida as estratégias de busca, realizou-se a coleta propriamente dita. Assim foram coletados o seguinte conjunto de metadados dos artigos da WoS: Título (TI), Resumo (AB), Ano de publicação (PY), Palavras-chave de autor (DE), e *keywords plus* (ID) (estes últimos são os descritores elaborados pela própria WoS).

A WoS permite que os resultados sejam coletados em diferentes formatos de arquivo. Nessa pesquisa optou-se pelo download direto no formato "separado por tabulação (Win, UTF-8)" e posteriormente convertido para o software *Excel* de forma que viabilizasse a organização em tabelas. No Apêndice A se encontra um exemplo de registro completo de um artigo, no formato HTML, a fim de mostrar como os metadados dos documentos da WoS são organizados.

A partir do *Excel* foi utilizada a formula de contagem para calcular o número de palavras de autor e de *keywords plus* separadamente e também para contar a ocorrência desses termos e gerar os quadros que estão dispostos na seção das análises.

Ressalta-se que se utilizou a ferramenta *VOSviewer* para gerar gráficos no intuito de ilustrar o comportamento do conjunto de palavras nos clusters formados a partir da coocorrência de termos nos campos das palavras-chave e *keyword plus*. Essas imagens ajudaram a esclarecer apontamentos referente às características dos termos e suas respectivas associações em clusters diante dos critérios aplicados.

Salienta-se que as análises apresentadas na seção 4 almejam responder à questão primordial deste estudo: *as palavras-chave e as keywords plus dos artigos da área de Nutrição, indexados na WoS, possuem tratamento temático adequado que favoreça a visualização da informação?*

Vale destacar que em alguns momentos, utilizou-se um *corpus* de 500 documentos para ilustrar como as palavras se comportam quando expandidas para um universo maior. Estes documentos foram recuperados a partir da expressão de busca utilizando os dez

descritores do DeCS de maior ocorrência nas bases da BVS, os quais foram utilizados para gerar mapas no VOSviewer e confirmar algumas impressões observadas nas análises.

Essa escolha se justifica por ser necessário esclarecer que independentemente da extensão do *corpus*, as palavras-chave da base de dados e do domínio analisado possuem características e um comportamento particular. Nesse sentido, é possível que exista consistência ou inconsistência nesse conjunto de termos, tanto em um universo maior quanto em um universo menor de dados. Essa questão pode ser observada na seção 4.1.1 que trata da análise a partir da exaustividade.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A presente seção se dedica a apresentar os resultados da pesquisa e as discussões acerca dos dados analisados tendo como objetivo avaliar a qualidade da representação da informação na WoS através das palavras-chave e *keywords plus* dos artigos da área de Nutrição, a fim de observar como esses termos se comportam diante das visualizações geradas a partir da coocorrência de palavras.

A análise segue duas etapas importantes: uma voltada para a análise da indexação conforme os critérios estabelecidos e a outra tendo como parâmetro os descritores da área de Nutrição identificados no DeCS. Esta última se dedica, essencialmente, a estabelecer pontos de intersecção entre o DeCS e a representação da informação na WoS referente à área de Nutrição. Ou seja, identificar se existe semelhança entre a linguagem utilizada por ambos, objetivando verificar e comparar se os termos de maior ou menor frequência na WoS acontecem no DeCS ou se são significativamente diferentes.

A proposta desse tipo de análise é focar na garantia literária, pois se o objetivo central dessa pesquisa é avaliar a qualidade da representação da informação do domínio de Nutrição, a garantia literária torna-se uma ferramenta fundamental para este propósito. Dias (2015) acredita que a garantia de literatura é essencial para validar os conceitos e os termos utilizados na representação de um domínio.

Com relação à qualidade da indexação, cabe ressaltar que não existe uma verdade absoluta sobre esse assunto, pois qualidade é um termo subjetivo que alterna de acordo com a percepção de cada indivíduo. Contudo, existem alguns critérios ou normas de indexação encontrados na literatura que possibilitam avaliar a representação temática de documentos, buscando identificar suas falhas e corrigi-las e, ao mesmo tempo, para que tais critérios permitam construir princípios básicos de indexação.

A qualidade da indexação, segundo Lancaster (2004), está relacionada à eficiência da recuperação da informação, ou seja, quando os termos indexados permitem que o sistema recupere documentos úteis e evite documentos inúteis para uma necessidade de informação.

Não cabe aqui afirmar, tampouco concluir que a indexação da WoS está correta ou incorreta. Mas vale esclarecer que é possível desenvolver uma indexação que seja mais clara, consistente e objetiva. Uma indexação em que os termos sejam compatíveis com as necessidades dos usuários expressas nas suas formas de buscas e também que seja uma indexação compatível com o conteúdo dos documentos e com as linguagens controladas já existentes nos diversos domínios.

A visualização da informação aqui discutida não diz respeito à construção de interfaces amigáveis para fins de recuperação da informação. Na realidade, está relacionada à elaboração de representações gráficas de domínios do conhecimento por meio de palavras-chave e descritores. A proposta dessa dissertação, como já mencionada no início da pesquisa é analisar a representação da informação na WoS por meio das palavras-chave e *keywords plus*, a fim de identificar se tais termos favorecem ou não a construção de representações visuais do domínio de Nutrição. Para tanto, foram identificados na literatura da CI alguns critérios de avaliação da qualidade da indexação, os quais foram moldados para se adequarem à proposta da pesquisa.

Entende-se que a VI trata de uma área da ciência que tem por objetivo estudar e desenvolver diferentes formas de representações gráficas para apresentar informações, de modo que a compreensão das mesmas, se torne mais fácil (VIEIRA, 2014). Num SRI, a VI, que é construída a partir de interfaces gráficas, contribui para uma melhor comunicação entre sistema de informação e usuário. Ou seja, atua no intuito de facilitar o acesso à informação.

Sob a ótica da representação da informação, é possível aferir que a VI terá excelência quando existir relevância na escolha dos termos que irão representar determinada realidade. Desse modo, a qualidade da indexação é fundamental para que o resultado da VI seja satisfatório. Nesse sentido, não seria um equívoco afirmar que a maneira como os conteúdos informacionais são indexados contribui para uma melhor ou pior visualização da informação.

Lembrando que assim como o próprio nome já diz representar, do latim *representare*, que significa “1. Ser a imagem ou a reprodução de. 2. Ser um exemplo ou um caso concreto de [...]” (DICIO, [2017?]). Ou seja, a representação temática deve ser o mais fiel possível do conteúdo do documento, não acrescentando nada a mais ou a menos do que o documento trata. Também precisa ser clara na sua forma de apresentação-tradução como, por exemplo, evitar termos ambíguos e de abrangência elevada.

Geralmente, a VI utilizada em SRIs, dedica-se a representar domínios específicos do conhecimento de modo a facilitar a recuperação de informações úteis por meio do uso de interface gráfica e de navegação. Ou seja, esse tipo de interface permite o usuário interagir com o sistema de forma a encontrar a informação de que necessita, numa busca mais efetiva. A seguir serão apresentadas as análises propriamente ditas.

4.1 Primeira Análise: aplicação dos critérios estabelecidos

Partindo do pressuposto de que uma boa visualização da informação é resultado, *a priori*, de uma boa indexação, faz-se necessário não somente apontar os critérios que qualificam essa indexação, mas também aplicá-los ao *corpus* investigado, de modo que se possa aferir algo sobre o mesmo. Nessas condições, seguem as análises a partir dos critérios de avaliação da indexação.

4.1.1 Análise a partir da Exaustividade

A exaustividade diz respeito à quantidade ilimitada de termos livres que são indexados para representar o conteúdo de um documento, sem que exista necessariamente o uso da especificidade. No caso do critério de exaustividade, este será positivo quando, em um sistema de busca e recuperação da informação, os termos indexados permitirem a revocação de documentos úteis sobre determinado assunto.

Por outro lado, a exaustividade será negativa quando a resposta do sistema para com o usuário revocar documentos inutilizáveis. Aspecto este que Lancaster (1993, p. 75) denominou de "boa indexação" que é quando a indexação permite que uma base de dados retorne respostas úteis e impeça a recuperação de respostas inúteis.

Segundo Slype (1991) a qualidade da indexação sob o ponto de vista da exaustividade está relacionada à seleção de conceitos de fatos-significativos que contém informação pertinente para os usuários. Ou seja, uma exaustividade muito reduzida resulta na não recuperação de documentos pertinentes, da mesma forma que uma exaustividade muito elevada ocasiona a recuperação de documentos que não contém informações pertinentes.

Do ponto de vista da VI, a exaustividade não é considerada como um aspecto positivo porque ocasiona a pulverização de termos, não favorecendo a visualização de domínios do conhecimento.

A visualização da informação gerada a partir da coocorrência de palavras depende do grau de coincidência sintática entre os termos (ou semântica se o sistema for habilitado a tais associações), ou seja, quando se repetem no mínimo duas vezes. A partir dessa coincidência mapas de conhecimento podem ser gerados considerando não somente a coocorrência, mas também outros atributos dos conceitos, que indiquem aderência e relação semântica em um delimitado domínio.

Lancaster (2004) denomina a coincidência de termos de coerência da indexação, também conhecida na CI por consistência da indexação. E, para o autor, quanto maior a exaustividade, menor a coerência entre os termos. Nessas condições, a exaustividade é um critério de qualidade da indexação que não favorece a visualização de domínios do conhecimento, embora, em alguns casos, ela favoreça a recuperação da informação como já mencionado no decorrer desta pesquisa.

Com base no *corpus* analisado, identificou-se que a representação da informação na WoS é predominantemente exaustiva. Isso significa que tanto palavras-chave (DE) quanto *keywords plus* (ID) são termos que aparecem em grande quantidade. Na Tabela 2, a seguir, é possível observar com mais clareza esse detalhe.

Tabela2 – Quantitativo de termos por artigo

Artigo	Quant. DE	Quant. ID
1	6	10
2	6	10
3	5	10
4	5	10
5	4	10
6	3	10
7	5	10
8	5	10
9	4	10
10	4	10
11	4	10
12	6	10
13	5	10
14	7	10
15	4	10
16	6	10
17	5	9
18	6	10
19	6	10
20	8	7
Média palavras/artigo	**Variação 4/5/6**	**10**

Fonte: dados da pesquisa, 2017.

Diante do quadro aqui exposto, observa-se que a as palavras-chave de autor possuem uma média de cinco termos por artigo e que as *keywords plus* possuem uma média de dez termos. Isso evidencia o critério de exaustividade aplicado a estes documentos.

A partir desse *corpus* também foi observado que mesmo a indexação sendo exaustiva, principalmente no caso das *keywords plus*, existe uma quantidade limite de termos atribuídos por documento. No contexto específico das *keywords plus*, estas não ultrapassam o limite de

dez termos. Já nas palavras-chave de autor não se percebeu com clareza esse limite, pois é possível que também a exigência seja de dez palavras por artigo, mas no *corpus* de análise, notou-se que o número maior de palavras de autor foi de oito termos por documento e este número só acontece uma vez.

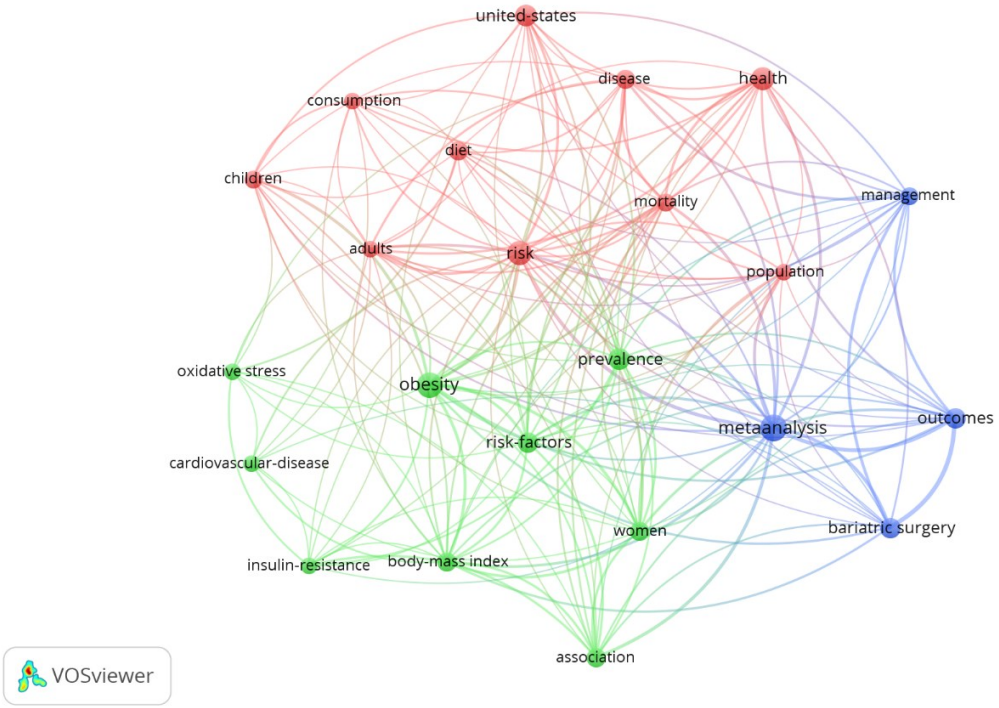
Foi possível identificar, no *corpus*, como a exaustividade interfere na VI. Primeiro observou-se que do total de 253 descritores (soma de palavras-chave de autor e *keywords plus*) apenas 25 termos possuem, no mínimo duas ocorrências. Ou seja, correspondem a 9, 88% do total de palavras.

O que se pode aferir acerca da exaustividade nesse estudo, é que a mesma acarreta na inconsistência das palavras. Sem uma quantidade considerável de palavras consistentes, isto é, que coocorrem, o número dos *clusters* formados diminui e a visualização desses dados, acaba por mostrar um universo muito inferior ao conjunto total de palavras.

É importante ressaltar que a inconsistência das palavras não está relacionada à quantidade de termos que existem no *corpus* de análise, mas se trata de uma característica específica das palavras-chave e *keywords plus* dos documentos da WoS, que são indexados exaustivamente e, consequentemente, acarreta a dispersão de termos.

A fim de constatar que a inconsistência das palavras independe da quantidade de termos existente no *corpus*, este foi aumentado para um conjunto de 500 documentos (*corpus* ilustrativo), onde foi possível contabilizar 4.257 palavras, das quais (1.907 são palavras-chave de autor e 2.620 são *keywords plus*). Utilizando as *keywords plus* como objeto de análise, as mesmas foram aplicadas no software *VOSviewer*, com a finalidade de criar um mapa com as palavras de maior ocorrência, vide Figura 12. Nesse sentido, elegeu-se aquelas com no mínimo treze ocorrências, que foi o maior nível de ocorrência mínima identificado pelo software para esse grupo de palavras. Diante desses dados, percebeu-se que quando são gerados mapas com palavras de maior frequência, o *corpus* diminui significativamente. No caso das *keywords plus*, em especial, aquelas que possuem frequência mínima treze, representam apenas 0,87% do total de palavras.

Figura 12 – *Keywords Plus* com o mínimo de treze ocorrências



Fonte: Dados da pesquisa, 2017.

O que é interessante no mapa da Figura 12, é que, por mais que seja uma quantidade pequena de palavras com ocorrência elevada, nota-se que os cluster formados estão interligados, o que proporciona uma melhor visualização e interpretação da rede que é formada entre esses cluster, diferente de quando os dados são dispersos e sem nenhuma relação entre si.

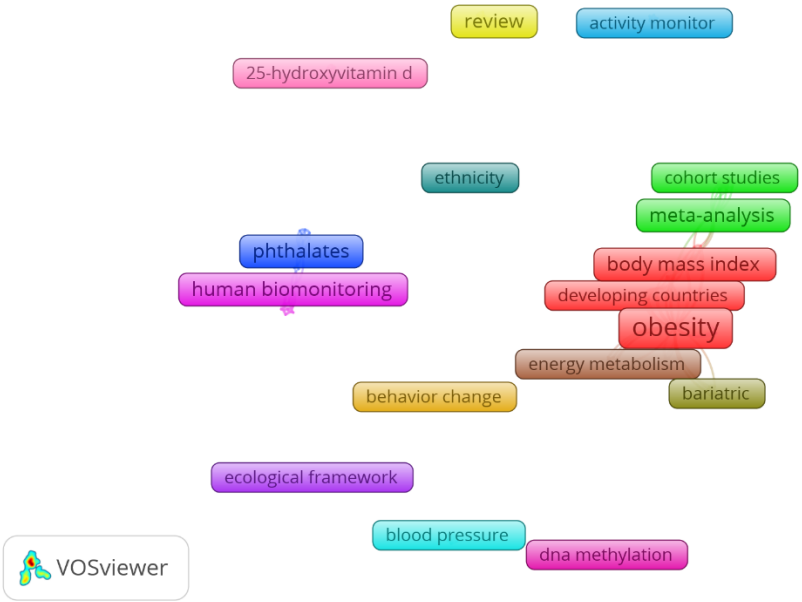
Para demonstrar como os termos menos recorrentes se comportam dentro da VI, foi gerada uma visualização, a partir dos 20 artigos iniciais da pesquisa, no *VOSviewer*, (Vide Figura 13). Percebeu-se, a partir desse mapa, que os clusters formados ficaram dispersos e sem nenhuma ligação entre si. Nesse caso, para gerar o mapa, utilizou-se a opção de contagem mínima de uma ocorrência, visto que o software permite também essa opção.

Convém salientar que visualizar palavras com ocorrência tão baixa não é algo interessante para estudar como um domínio se comporta. Até porque esses estudos, que também estão diretamente relacionados aos métodos bibliométricos[10], buscam contar palavras

[10]O método bibliométrico, segundo Zhang et al. (2016, p. [967], tradução nossa), "é uma ferramenta de pesquisa proeminente que pode representar sistematicamente a natureza de disciplinas científicas específicas, destacando *hotspots* de pesquisa e tendências de pesquisa".

que coincidem, para que assim se possa identificar as tendências que um domínio possui, pela frequência dos termos mais utilizados.

Figura 13 – Cluster das palavras-chave de autor com o mínimo de uma ocorrência



Fonte: dados da pesquisa, 2017.

Os *clusters* são formados por grupos de palavras que mais coocorrem conjuntamente no mesmo contexto. Ou seja, não necessariamente as palavras que apresentam quantidade de ocorrência semelhante, vão aparecer no mesmo cluster, pois elas precisam, essencialmente, ocorrer juntas. Na figura acima, por exemplo, o *cluster* de cor verde, onde aparecem as palavras "meta-analysis" e "cohort studies", cada uma possui número de ocorrência diferente (a primeira, contem duas ocorrências e, a segunda, apenas uma). No entanto, essas palavras possuem uma proximidade maior do que com as demais. Essa aproximação se dá pelo fato de que as mesmas ocorrem no mesmo documento e porque possuem força de ligação semelhante.

Acredita-se que, por mais que a exaustividade seja, em alguns momentos, um aspecto positivo na recuperação da informação, ela acarreta na dispersão de palavras, não favorecendo, portanto, a VI. A este aspecto, Ferreira (2012, p. 72) argumenta que

> se por um lado a disparidade de palavras-chave cria um rico aspecto semântico útil aos processos de recuperação da informação, por outro, essa

---

Conforme Santos e Kobashi (2008, p. 109) "a bibliometria é uma metodologia de recenseamento das atividades científicas e correlatas, por meio de análise de dados que apresentem as mesmas particularidades".

> "pulverização" de termos dificulta o processo de agrupamento por similitude, tornando a construção de mapas e gráficos um processo complexo e custoso.

Nessa perspectiva, concorda-se com Lancaster (2004) que a exaustividade diminui a coerência da indexação que, por sua vez, na VI não é considerado um ponto favorável para investigar as tendências temáticas de um domínio. Mapear as tendências de um domínio requer, essencialmente, que os termos ali presentes ocorram em diferentes documentos que tratem do mesmo assunto. Quando existe essa pulverização de palavras, torna-se mais difícil aferir que um termo é ou não preferível para representar as temáticas de uma área do conhecimento.

4.1.2 Análise a partir da Especificidade

Nessa análise, busca-se esclarecer algumas concepções sobre especificidade enquanto critério de avaliação da indexação. É sabido que quando se trata de indexação, vêm à tona dois conceitos importantes: exaustividade e especificidade. Como já foi traçada uma análise na perspectiva do primeiro termo, cabe aqui direcionar o foco de discussão para o segundo termo, ou seja, a especificidade. Contudo, é necessário esclarecer que por mais que os dois conceitos tenham dimensões diferentes, ambos possuem alguma relação, pois muitas vezes são utilizados conjuntamente para uma finalidade comum.

O propósito desse critério é identificar se a representação da informação na WoS especialmente, da área de Nutrição, possui mais termos abrangentes ou específicos. Trata-se de observar se as palavras de autor (DE) e *keywords plus* (ID) contemplam esse aspecto e também como estas se comportam diante da visualização da informação tendo como parâmetro o critério de especificidade.

É sabido que ao contrário da exaustividade, que utiliza quantos termos forem necessários para representar um conteúdo, a especificidade, por seu turno, é um dos princípios da indexação ao qual "um tópico deve ser indexado sob o termo mais específico que o abranja completamente" (LANCASTER, 2004, p. 34).

A especificidade diz respeito ao nível de profundidade da indexação, ou seja, aumenta o índice de precisão, buscando evitar documentos inúteis num SRI. Sobre a especificidade, Piedade (1983, p. 12) a define como "a exatidão com que os descritores ou símbolos de classificação utilizados permitem representar o assunto dos documentos".

Foi percebido, na seção anterior, que a indexação da WoS é predominantemente exaustiva. Isso implica numa quantidade de termos elevada para representar o assunto ou assuntos de que tratam os artigos científicos da área em estudo (Nutrição). A partir dessa observação, busca-se saber se a exaustividade, no *corpus* investigado, interfere ou não na especificidade. Será que a elevada quantidade de termos diminui o grau de especificidade das palavras-chave e *keywords plus*? Pretende-se, então, responder essa indagação no decorrer do presente discurso.

Acerca deste aspecto, é importante considerar que podem existir duas possibilidades de termos com uma indexação predominantemente exaustiva: a primeira situação é que o resultado seja de uma indexação abrangente com uma queda gradativa de termos específicos; e a segunda situação é que a exaustividade pode acarretar na eleição de termos indevidamente específicos. Isso significa que, nem sempre, a eleição de vários termos específicos refletirá numa especificidade adequada. Nem sempre um documento irá conter vários assuntos específicos para serem traduzidos como descritores, acarretando na eleição de termos mais genéricos, que contemplem o critério de exaustividade. Também é possível que um indexador não tenha conhecimento suficiente da linguagem técnica da área à qual um assunto está sendo indexado, elegendo, assim, termos específicos que não condizem com a linguagem técnica da comunidade discursiva do domínio.

Outro ponto importante é que tanto do ponto de vista da recuperação quanto do ponto de vista da visualização da informação, não é muito viável que a indexação seja absolutamente específica. Pois nem sempre a especificidade utilizada pelo indexador corresponde aos termos preferidos utilizados pela comunidade discursiva de um domínio.

Embora a especificidade seja um aspecto importante para tratar de assuntos específicos, é necessário que o indexador profissional, assim como o autor do texto adotem o máximo de cautela no momento de eleger os termos para representar os assuntos de que tratam o documento. É fundamental que o autor do trabalho, ao assumir a função de indexador do seu próprio texto, tome conhecimento dos conceitos e dos termos essenciais de sua área, pois isso ajuda a reduzir o índice de dispersão de palavras desnecessárias.

Em um sistema de recuperação da informação, a especificidade é considerada adequada e eficaz quando a demanda dos usuários do sistema é essencialmente específica, quando isso não ocorre, a especificidade pode comprometer o bom funcionamento do sistema. Já do ponto de vista da VI, a especificidade se torna adequada quando os termos específicos coincidem, ou seja, quando existe o consenso pela comunidade discursiva ao usar termos específicos. Não adianta tentar ser específico sem que o termo seja conhecido na literatura. É

preciso o olhar atento do indexador, quer seja em uma base de dados, quer seja o do próprio autor do trabalho, no momento de eleger os termos mais adequados para representar conceitos.

No *corpus* de análise, observou-se que a maioria dos termos são específicos, contudo, existem alguns termos considerados abrangentes. Com relação às palavras de maior ocorrência, verificou-se que grande parte também é específica, tanto no conjunto de DE quanto no conjunto de ID. Vale lembrar que, o *corpus* aqui estudado, diz respeito ao conjunto de 20 artigos o que, de certa forma, é uma quantidade muito reduzida para aferir se a indexação da área de Nutrição na base de dados como um todo é mais específica ou mais abrangente. Todavia, apontando o que foi observado no *corpus*, acredita-se que as palavras-chave de autor tendem a ser mais específicas do que as *keywords plus*.

A fim de enfatizar tal afirmação, vale-se do trabalho de Zhang et al. (2016) onde os autores realizaram um estudo que buscou investigar a estrutura do conhecimento na área de adesão ao paciente. Nesse estudo fez-se uma análise comparativa entre as palavras-chave de autor e *keywords plus*, de modo a investigar a eficácia destas últimas enquanto parâmetro para capturar o conteúdo dos conceitos científicos presentes nos documentos. Os autores chegaram à conclusão de que as *keywords plus* são mais abrangentes e genéricas do que as palavras-chave de autor.

Em outro estudo, realizado por Garfield & Sher (1993), citado por Zhang (2016), os autores fizeram uma pesquisa de literatura utilizando palavras-chave de autor isoladamente, o resultado foi de 100 documentos relevantes. Em seguida utilizaram uma combinação de palavras-chave de autor e *keywords plus*. O resultado foi de 63 documentos adicionais. Desses 63 documentos, 13 dos 63 itens foram recuperados por palavras-chave de autor e *keywords plus* juntas, enquanto que 45 itens foram recuperados apenas por *keywords plus* e 5 itens apenas por palavras de autor. Isso significa que as *keywords plus* tem um caráter mais abrangente na recuperação da informação, permitindo maior revocação de documentos.

Direcionando o foco para o *corpus* analisado, foram encontradas algumas palavras de especificidade elevada, tanto no grupo de DE quanto no conjunto de ID. A especificidade elevada diz respeito aos termos específicos que não são comumente utilizados. Ou seja, ocorrem apenas uma vez, e mesmo unindo palavras-chave de autor com *keywords plus*, estas palavras continuaram ocorrendo sozinhas. A seguir, na Tabela 3, serão mostradas as palavras consideradas de especificidade elevada.

Tabela 3 – Palavras de especificidade elevada

DE	Ocorrência	ID	Ocorrência
pfos	1	di(2-ethylhexyl)phthalate dehp	1
25-hydroxyvitamin d	1	1,25-dihydroxyvitamin-d	1
pfoa	1	25-hydroxyvitamin-d	1
pfc	1	di-isononyl phthalate	1
bmi	1	hypovitaminosis-d	1
		inc.	1
		n-butyl phthalate	1
		trout oncorhynchus-mykiss	1

Fonte: Dados da pesquisa, 2017.

Considerando os estudos que apontaram as ID como palavras mais abrangentes do que as DE, é curioso o resultado do Tabela 3, que apresenta um número maior de palavras de especificidade elevada no grupo de ID e não no grupo de DE. Uma possível justificativa para este fenômeno é que as DE possuem uma quantidade menor de palavras em comparação as ID. Pois, no geral, os artigos recuperados possuem mais *keywords plus*.

Outra observação que pode ser possível acerca da quantidade maior de termos "indevidamente" específicos ocorrer nas ID, é o fato de que as mesmas são eleitas automaticamente por um programa de computador que contabiliza a frequência das palavras nos títulos das referências citadas. Isso implica na eleição de termos não necessariamente adequados como descritor, pois sem o aval dos especialistas da área é improvável garantir a qualidade na eleição desse tipo de termo. Desse modo, seria natural uma ocorrência menor de eventuais "erros" de indexação por especialistas de um domínio, em comparação a uma indexação que é realizada automaticamente, tendo como critério, apenas, a frequência com que o termo aparece no texto. É o que se observa no grupo de ID.

A partir dessa análise, percebeu-se que a especificidade das palavras-chave no *corpus* estudado, em algum momento, é consistente e em outro, causa a pulverização de termos. Essa pulverização foi identificada tanto pelo software *Excel*, que separou os termos sem repetições e depois contabilizou a ocorrência dos mesmos (nesse caso foram identificadas palavras de especificidade elevada que só ocorreram uma vez), quanto pelo software *VOSviewer* que, ao contar mutuamente palavras-chave de autor e *keywords plus* com apenas uma ocorrência, identificou que várias palavras-chave específicas fazem parte desse grupo.

Aqui não cabe tomar como verdade absoluta que a indexação do domínio de Nutrição na WoS é muito específica ou muito abrangente. Para se chegar a tal conclusão seria necessário um estudo mais minucioso e aprofundado, de forma a captar mais detalhadamente os diversos aspectos dos termos.

Para o presente momento, o que se pode aferir é que, no *corpus* de 20 documentos, a indexação contempla o critério de especificidade. Contudo, nas *keywords plus* perceberam-se termos mais amplos em comparação as palavras-chave de autor. Além disso, acredita-se que a natureza exaustiva da indexação no domínio de Nutrição, não diminui o grau de especificidade das palavras, no entanto, notou-se a existência de algumas palavras específicas e inconsistentes, isto é, sem possuir mais de uma ocorrência. Esse aspecto denota que a exaustividade somada à especificidade, especialmente nesse grupo de palavras, não colaborou para a consistência dos termos, ocasionando, por exemplo, a pulverização de palavras na visualização da informação, como visto na Figura 13 da seção anterior.

4.1.3 Análise a partir do Controle da Abrangência dos Termos

Consideram-se termos abrangentes aqueles que não possuem clareza quanto ao assunto que está sendo representado. São termos que apresentam ambiguidade, que causam dúvida ou que não apresentam a obviedade de que fazem parte de um domínio. Na área de Nutrição, os termos óbvios podem ser considerados aqueles utilizados na estratégia de busca e que, consequentemente, apareceram nos documentos recuperados, como, por exemplo, as palavras “nutrition”, “obesity” e “public health”. Esta última não é necessariamente um termo óbvio do domínio em questão, mas é um descritor que foi selecionado para complementar a busca na WoS, unindo e, ao mesmo tempo, restringindo o *corpus* a dois domínios específicos (nutrição e saúde pública).

Ao contrário dos termos óbvios, existem aqueles considerados amplos, abrangentes e às vezes irrelevantes. Cabe salientar que a abrangência das palavras pode ser de dois níveis. O primeiro nível trata de uma abrangência leve, ou seja, mesmo que o termo não seja óbvio para o domínio, ele apresenta algum significado, como no caso da palavra “Estados Unidos” que no *corpus* aparece três vezes no conjunto de *keywords plus* e uma vez no conjunto das palavras de autor. Por mais que o referido termo seja considerado abrangente, ele possui algum significado. Em uma análise de domínio, por exemplo, talvez pudesse aferir que a palavra “obesity” possui forte relação com “United Stated”. Nesse sentido, seria possível

conjecturar que existe um grande índice de pesquisas relacionadas à obesidade nos Estados Unidos. Esta é apenas uma situação hipotética.

O segundo nível de abrangência é aquele considerado elevado. Isto é, um termo indexado que, além de não ser óbvio que pertence a um dado domínio, não possui um significado muito relevante. Trata-se então de stopwords que não favorecem nem a recuperação nem a visualização da informação.

Geralmente as *stopwords* (palavras vazias) são bastante utilizadas na indexação automática. Pois nesse tipo de indexação é gerada previamente uma lista de palavras vazias para que o programa de mineração de texto compare as palavras que aparecem com maior frequência no texto com essa lista de *stopwords* e, posteriormente, as elimine do conjunto de termos selecionados para serem descritores.

Um exemplo de *stopword* encontrada no *corpus* de análise é a palavra "risk". Nesse caso é possível aferir que dificilmente uma busca seria feita por esse termo. Não significa que esse tipo de busca seja impossível de acontecer, no entanto, é muito difícil ocorrer numa base de dados em que a maioria dos usuários são pesquisadores e profissionais que têm necessidades específicas. Pessoas com necessidades específicas não costumam utilizar expressões de buscas com termos tão amplos.

Do ponto de vista da representação da informação, não é comum eleger um termo amplamente genérico como descritor para representar um domínio. Se forem consideradas a garantia literária e as comunidades discursivas, termos com essa característica não devem ser considerados como candidatos a descritores.

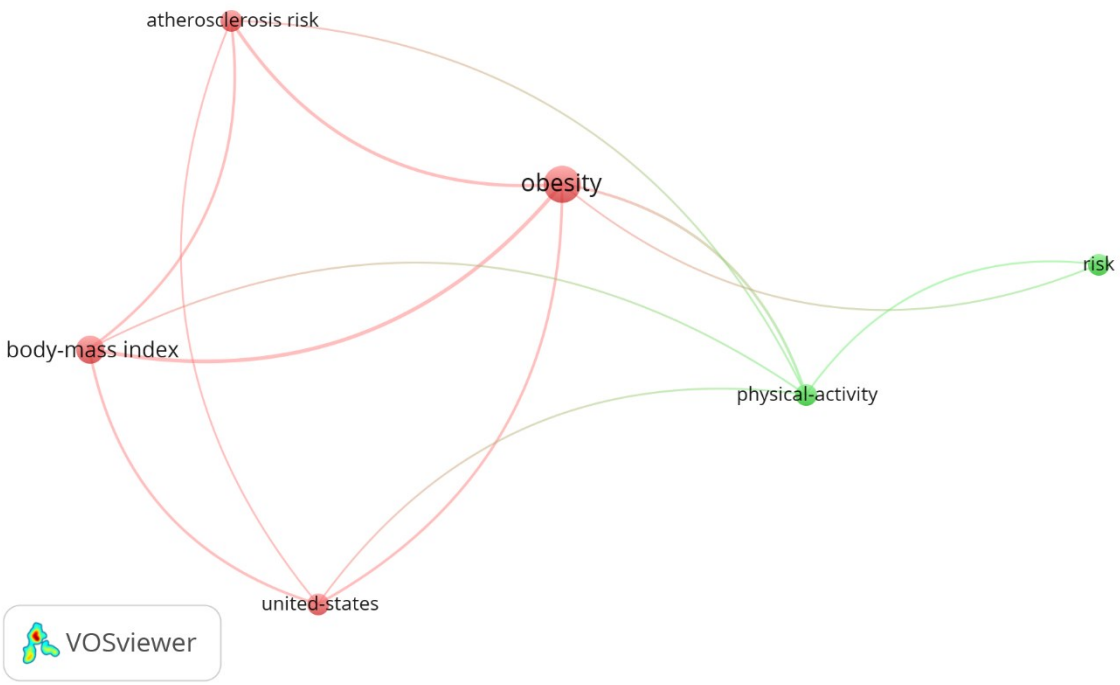
Sob a ótica da visualização da informação, termos abrangentes podem interferir de dois modos: primeiramente, quando as visualizações construídas são prejudicadas por não haver mais de uma ocorrência do termo e isso acaba gerando pulverização de palavras ao redor do mapa construído. Em segundo lugar, quando a visualização apresentada adquire um caráter abrangente e pouco representativo por causa desses termos.

Acrescenta-se também que se a representação do *corpus* de análise apresentar termos com abrangência elevada e estes só aparecerem uma única vez, a visualização acaba sendo prejudicada por diminuir os atores que irão formar o mapa de cluster. Exemplo: se num conjunto de cinquenta palavras-chave existem apenas dez que ocorrem mais de uma vez, a visualização gerada a partir desse número será pouco representativa, no sentido de que haverá uma disparidade elevada entre a quantidade de termos mais recorrentes e menos recorrentes.

Para melhor ilustrar essa situação, na Figura 14, é possível identificar diante da visualização gerada pelo software *VOSviewer* que os termos de maior ocorrência e força de

ligação não são generalistas, contudo a quantidade desses termos é bem inferior a quantidade do total de termos do *corpus*. Sendo assim, a visualização gerada pelo software, tendo como critério o mínimo de três ocorrências por descritor, é considerada pouco representativa. Ou seja, do total de 253 termos (palavras-chave e *keyword plus* juntas) apenas seis correspondem ao limite mínimo de três ocorrências.

Figura 14 – Cluster total das palavras com o mínimo de três ocorrências



Fonte: dados da pesquisa, 2017.

Para um melhor esclarecimento da figura anterior, apresentam-se, na Tabela 4, as palavras com os *clusters* formados por suas ocorrências e força de ligação. Consegue-se identificar a qual cluster cada palavra pertence. É notório que a palavra que forma o maior nó é “obesity”, pois, na Figura 14 assim como na Tabela 4, está clara a grande influência que essa palavra possui entre as demais. Certamente, isso acontece por se tratar de um termo óbvio utilizado na expressão de busca. Portanto, é natural que os documentos recuperados contenham essa palavra em maior quantidade do que as que não foram utilizadas nas expressões de busca, isto é, os demais descritores que fazem parte do documento recuperado. Já o termo “risk” é o mais abrangente e o que possui menos ligação com os demais.

Tabela 4 – Cluster e força dos nós das palavras com o mínimo de três ocorrências

Palavras-chave	Cluster	Ocorrências	Força do nó
Obesity	1	9	12
body-mass index	1	5	9
astherosclerosis risk	1	3	7
united-states	1	3	6
physical-activity	2	3	6
Risk	2	3	2

Fonte: Dados da pesquisa, 2017.

Com relação à abrangência da indexação, foram identificados, no *corpus* de análise, alguns termos considerados generalistas. Esses dados podem ser visualizados nas Tabelas 5 e 6, onde se apresentam com suas respectivas ocorrências. O total de palavras-chave de autor consideradas generalistas é de 24 e o total de *keywords plus* é 48.

Tabela 5 – Palavras-chave de autor (DE) consideradas abrangentes

DE abrangentes	ocorrência	DE abrangentes	ocorrência
social class	2	data interpretation	1
review	2	prevalence	1
monitoring	1	exposure assessment	1
policy	1	residence characteristics	1
prevention	1	statistical	1
sex	1	race	1
adult	1	drug	1
middle age	1	ethnicity	1
sleep	1	log book	1
incidence	1	indoor air	1
cellular phone	1	plasticizers	1
consumer exposure	1	house dust	1

Fonte: dados da pesquisa, 2017.

Tabela 6 – *Keywords Plus* (ID) consideradas abrangentes

ID abrangentes	ocorrência	ID abrangentes	ocorrência
risk	3	internal exposure	1
weight	2	internet	1
disease	2	intervention	1
population	2	la carte	1
adults	1	location	1
alcohol	1	mice	1
assay	1	mobile phone	1
association	1	models	1
birth	1	neighborhood characteristics	1
cellular phone	1	nonresponse	1
children	1	participation	1
communities	1	patterns	1
costs	1	policy	1
disorders	1	populations	1
disparities	1	poverty	1
environments	1	prevalence	1
exposure	1	purchase fresh fruit	1
global burden	1	recommendation	1
health	1	reliability	1
height	1	service	1
income	1	supermarkets	1
increasing fruit	1	transportation	1
indicators	1	urban form	1
in-house dust	1	women	1

Fonte: dados da pesquisa, 2017.

Para esta análise, considerou-se como termos generalistas, aqueles de nível de abrangência elevado. No entanto, sabe-se que os julgamentos acabam sendo inexoravelmente tendenciosos. Para uma melhor credibilidade desta análise, foram feitas algumas buscas, na própria WoS, pelos termos designados genéricos. A partir do resultado da busca, foram observadas as localizações em que o suposto termo genérico aparece, lembrando que um dos critérios de julgar se um termo pode ser considerado ou não como um descritor, é sua presença no título, no resumo e nas palavras-chave do próprio autor. Quando o termo aparece simultaneamente nessas três instâncias ou pelo menos em duas delas, o mesmo pode ser utilizado como um descritor.

Todavia, cabe ressalvar que não é qualquer palavra que aparece no título e no resumo do documento que pode ser considerada uma candidata a descritor, pois, algumas palavras, só

fazem sentido se estiverem em uma frase ou em conjunto com outra palavra. Pode-se observar isso na seguinte sentença "O **cultivo** de laranja lima, no estado de Alagoas, tem se mostrado crescente nos últimos anos". A palavra "cultivo" por si só não consegue expressar nenhum significado. Ela necessita estar relacionada com outra ou fazer parte de alguma sentença para que, assim, tenha sentido completo.

Para aferir se o termo é ou não abrangente, considerou-se, como critério de análise, sua posição nas palvras-chave de autor. Esse critério remete novamente à garantia literária. Considera-se também que o autor do documento é um tipo particular de especialista no assunto que está sendo tratado. Nesse caso, em cada busca, dentre os vinte artigos selecionados para essa verificação, se os termos não aparecerem como palavra-chave de autor, considera-se o mesmo abrangente e de pouca utilidade pela comunidade discursiva.

Essa análise se procedeu da seguinte forma: foram selecionadas cinco palavras generalistas da Tabela 6, de *keywords plus*. Cada palavra foi pesquisada isoladamente na WoS. Do total de resultados recuperados, foram selecionados os vinte primeiros artigos das cinco buscas e, a partir desse ponto, foram feitas as análises tendo como critério a posição dos termos nos documentos. Vejamos a seguir estas palavras.

- ✓ **Termo "height"** – a pesquisa por essa palavra retornou 335.447 documentos dos quais foram selecionados os vinte primeiros para a análise. Observou-se que essa palavra aparece apenas em um artigo como palavra-chave de autor (DE) e, mesmo assim, não aparece sozinha, mas em conjunto com outra palavra, formando "Faltings Height". Somente em um artigo, "height" aparece no campo de *keyword plus* (ID), mas como um descritor composto. Trata-se, então, da palavra "monmonotocity height". Nos demais documentos, "height" aparece apenas nos seus respectivos resumos.

- ✓ **Termo "income"** – foram recuperados 190.545 documentos, a partir da busca realizada com esse termo. Na análise dos vinte primeiros, observou-se que, em dois artigos, a palavra "income" aparece no campo das DE como uma palavra-chave composta: em um é "stochastic income" e no outro é "income risk". Num terceiro artigo, aparece como ID sozinha: "income"; num quarto artigo aparece como ID acompanhando outra palavra, ou seja, "income inequality" e nos 16 documentos restantes, aparece somente nos resumos dos mesmos.

- ✓ **Termo "increasing fruit"** – esse termo recuperou 622 documentos. Dos vinte analisados, a palavra aparece em dois artigos como ID; em um terceiro artigo aparece no título do mesmo e nos dezessete restantes, aparece em seus resumos.

- ✓ **Termo "indicators"** – essa palavra recuperou 244.797 documentos. No *corpus* analisado, observou-se que a palavra aparece em dois artigos como DE, mas em conjunto com outra palavra. Trata-se do termo "freshness indicators", no terceiro e quarto artigo, encontra-se no título do documento; num quinto artigo aparece como ID sozinha e num sexto artigo aparece como DE, também, sozinha.

- ✓ **Termo "in-house dust"** – o termo recuperou 1.546 documentos. No *corpus* selecionado, foi identificado que, em onze artigos, essa palavra aparece como ID sozinha. Aparece também no título de uma conferência e no título de dois artigos. Nos demais documentos, aparece apenas nos resumos.

A partir do exposto, observou-se que os termos apareceram com maior frequência nos resumos dos documentos do que em qualquer outro campo. Só o termo "indicators" que apareceu uma vez sozinho como palavra-chave de autor. Os termos "income" e "increasing fruit" também foram identificados sozinhos como *keywords plus*. O primeiro em um artigo e o segundo em dois artigos. Logo, considera-se este conjunto de palavras generalistas e que não devem ser utilizadas como descritores, pois quase não apareceram no campo de palavras-chave de autor. Se não fazem parte da linguagem dos especialistas da área, não deveriam ser utilizadas como descritores.

4.1.4 Análise a partir da Uniformidade

Para esta pesquisa, este critério averiguou se há uniformidade na representação da informação considerando o domínio de Nutrição da WoS. Isso se fez a partir de análises das palavras-chave de autor (DE) e *keywords plus* (ID) isoladamente, como também, comparou-as visando perceber se há uniformidade entre ambas, isto é, se os termos incidentes nos grupos de ID e DE estão representados, quanto a grafia, de forma idêntica. Esse tipo de verificação busca compreender o grau de variação que existe nas palavras indexadas.

A uniformidade aqui discutida diz respeito ao controle de sinonímia, homonímia e quantidade (plural e singular), além de outras formas variantes dos termos, como, por

exemplo, os que utilizam caracteres especiais como hífen, aspas, parênteses ou quaisquer outros.

A sinonímia diz respeito à relação de equivalência entre, pelo menos, duas palavras. Num vocabulário controlado, quando um conceito pode ser representado por mais de dois termos, escolhe-se aquele que é mais utilizado na literatura do domínio, fazendo-se remissivas para os demais (AFFONSO, 1997; STREHL, 1998; LIMA e ANÍZIO, 2012). Em um sistema de busca e recuperação da informação, interessa o termo que é mais comum nas expressões de buscas. Ou seja, o termo preferido pelo usuário do sistema. Diante do controle de sinonímia, podemos perceber duas garantias: no primeiro caso, tem-se a garantia literária e no segundo, tem-se a garantia de usuário.

Já a homonímia é o fenômeno pelo qual as palavras de diferentes significados têm a mesma grafia (homônimos homógrafos) ou a mesma pronúncia (homônimos homófonos). Há também os homônimos perfeitos que possuem a mesma grafia e pronúncia. Ex. de homônimo homógrafo: gosto (verbo) e gosto (substantivo); Ex. de homônimo homófono: sessão (de cinema) e seção (departamento); Ex. de homônimo perfeito: nutrição (substantivo) e nutrição (disciplina).

O uso do plural ou do singular para designar um termo é um dos problemas mais corriqueiros encontrado na indexação de assuntos. Segundo Strehl (1998), Lancaster (2004), Lima e Anízio (2012) as palvras-chave ou os termos de um vocabulário controlado devem ser usados no singular, exceto quando a palavra só pode ser empregada no plural, como no caso de Estados Unidos.

Outras formas variantes de escrita diz respeito às palavras compostas que, em alguns momentos, são utilizadas com hífen e, em outros, sem hífen. Nesse caso, é necessária uma padronização para que a pulverização de termos diminua.

Um dos problemas mais comuns decorrentes da ausência de controle/padronização é a variância dos termos indexados. Do ponto de vista da recuperação da informação, diferenças como o uso do plural ou singular e de variantes formas de grafia do termo, já não representam um problema comprometedor, visto que a maioria dos SRIs têm adotado recursos capazes de solucionar essas diferenças.

No caso da WoS, a base dispõe do recurso de truncamento, que permite recuperar palavras com variações de plural e singular e, também possui o recurso de lematização automática nas pesquisas por tópico e título, ou seja, permite que os documentos sejam recuperados com variações dos termos de busca, mesmo que não tenham sido digitadas. Por exemplo: o termo color recupera colour; mouse recupera mice, entre outros (WEB OF

SCIENCE, 2017). Contudo, para fins de visualização esse ainda é um problema que tem exigido soluções que demandam um grande esforço operacional para corrigir as inconsistências. Neste caso, torna-se necessário aplicar técnicas de uniformização de termos quando, por exemplo, busca-se gerar redes a partir de co-relações temáticas.

No caso da sinonímia, essa questão torna-se ainda mais complexa, pois identificar o termo sinônimo preterido em uma base de dados nem sempre é uma tarefa simples. Um dos problemas (se é que assim podemos chamá-lo) é a riqueza semântica. Em razão dela, documentos que tratam de temas afins (ou mesmo idênticos) são representados de variadas formas. Isto pode impedir a recuperação de um documento que foi indexado por determinado termo enquanto outro foi utilizado na estratégia de busca. Uma alternativa para atenuar a questão da sinonímia é a adoção de um vocabulário controlado, de modo a controlar esses termos, elegendo aquele de maior usabilidade no sistema de busca e estipular remissivas para os demais. É necessário mencionar que o critério de uniformidade, aqui empregado, não levará em consideração o controle de sinonímia, pois isso foge ao escopo desta análise.

Sob a ótica da visualização da informação, a variedade de termos também não é um aspecto favorável, pois se torna mais difícil compreender as estruturas de um domínio. Termos semanticamente variados, grafias diferentes e ausência de controle de vocabulário, tende a proporcionar uma VI “pulverizada”. Quando uma palavra aparece duas vezes escrita de forma diferente, a mesma será contabilizada duas vezes. Nesse caso, ocorre a duplicidade do termo, que é algo irrelevante. Foi o que aconteceu com algumas palavras do *corpus* analisado. Na tabela a seguir serão apresentadas essas palavras.

Tabela 7 – Palavras com variações de grafia

Palavras	Ocorrências
body mass index (DE)	2
body-mass index (ID)	5
ethinic groups (DE)	1
ethinic-groups (ID)	1
united states (DE)	1
united-states (ID)	3
public health (DE)	1
public-health (ID)	2
physical activity (DE)	1
physical-activity (ID)	3
adult (DE)	1
adults (ID)	1
risk-factor (ID)	1
risk-factors (ID)	1

Fonte: dados da pesquisa, 2017.

Percebe-se na Tabela 7, que a maioria das DE são escritas de forma mais adequada, isto é, sem o uso do hífen e no singular, todavia, não se pode afirmar o mesmo para o grupo de ID, que foge de algumas regras básicas de indexação, inclusive o uso de termos no plural, mesmo quando desnecessário. Conforme visto na literatura, as palavras-chave devem ser escritas na sua forma mais natural. O uso de hífen só deve ser usado nas palavras compostas que, naturalmente, necessitam do hífen, porém, se se trata de uma questão de estilo, o ideal é optar pelo termo escrito de forma simples.

Observa-se também que, por mais que as ID não se adequem às recomendações de indexação, os termos desse grupo apresentam maior índice de ocorrência quando comparadas ao grupo de palavras-chave de autor (DE). Mas é importante frisar que a ocorrência elevada das ID, não faz delas melhores que as DE. Pois não se sabe, ao certo, se as ID são geradas apenas pelo índice de ocorrência nas referências citadas ou se existe um outro critério as avalie antes de serem selecionadas como um termo para a busca e recuperação da informação na base de dados.

Considerando as normas de indexação da informação, as ID não correspondem aos padrões de indexação na sua forma mais apropriada, no entanto, se forem consideradas suas ocorrências nos documentos, essas palavras são mais favoráveis para construir representações visuais a partir da contagem de palavras.

Acredita-se que a variedade de termos para representar um conceito, seja pela diversidade semântica ou devido à falta de padronização das palavras, de certa forma, pulveriza os catálogos dos SRIs, prejudicando, consequentemente, a VI. Na visualização, quando isso acontece, o tratamento bibliométrico é laborioso, pois há a necessidade de padronização dos termos, seja de forma manual ou com a ajuda de algum software. O que se pode inferir sobre o conjunto de palavras do *corpus*, é que o critério de uniformidade não é atendido, pois as palavras possuem variação de grafia.

Sob o aspecto da uniformidade, Zeng (2008) infere que os sistemas de organização do conhecimento precisam de uma estrutura multidimensional, de modo a transpor fronteiras culturais e geográficas de acesso e representação. Para tanto, necessitam privilegiar as suas funções primordiais: eliminação da ambiguidade, controle de sinonímia, homonímia e relacionamentos semânticos entre os termos.

4.1.5 Análise a partir da Consistência (Coerência)

Esse critério buscou averiguar o grau de consistência da representação temática da WoS e, para tanto, considerou-se as palavras de maior ocorrência no conjunto de DE e no conjunto de ID. Então, os dois grupos de palavras foram analisados para identificar qual possui o maior índice de consistência. Para calcular tal índice, identificou-se a quantidade de DE e ID, separadamente, com pelo menos duas ocorrências, pois nesse universo, pouquíssimos foram os casos de palavras acima de duas ocorrências. Após essa contagem, calculou-se a proporção entre o número de descritores comuns e o número total de descritores. O cálculo foi aplicado nos dois grupos de palavras, individualmente.

Esse tipo de contagem tem por base a taxa de coerência mencionada por Slype (1991), em que a coerência é calculada a partir do número de descritores comuns (que ocorrem em um documento por indexadores diferentes) e o número total de descritores comuns ou incomuns variando entre 50 a 80%, considerando a qualidade do manual de indexação, formação dos indexadores e a cautela com que estes desenvolvem seu trabalho (ARAUJO JÚNIOR, .2007).

De modo semelhante, Lancaster (2004) menciona alguns trabalhos que calculam a coerência da indexação. Dentre eles, o de Leonard (1975), que considera a medida mais comum de calcular a consistência da indexação pela simples relação "$AB / (A+B)$, onde A representa os termos atribuídos pelo indexador a, B representa os termos atribuídos pelo

indexador *b*, e *AB* representa os termos com os quais *a* e *b* concordam" (LANCASTER, 2004, p. [68]).Outro trabalho importante mencionado por Lancaster(2004) é o de Cooper (1997), em que a coerência interindexadores é definida como a "proporção de indexadores que atribuem o termo menos a proporção daqueles que não atribuem" (LANCASTER, 2004, p. 69).

A coerência da indexação é entendida como o grau de concordância na representação da informação de um documento por meio de um conjunto de termos de indexação. Essa consistência pode ser verificada comparando a concordância dos termos por dois ou mais indexadores (consistência interindexadores) e comparando a concordância dos termos de um único indexador para com um documento em momentos distintos (consistência intraindexador) (LANCASTER, 2004; GIL-LEIVA, 2008).

A taxa de coerência de Slype sofreu uma adaptação para se adequar ao propósito desta análise. O autor considera o índice de coerência entre os termos de um documento por diferentes indexadores ou grupos de indexadores, ou seja, a consistência interindexadores. A análise em questão propõe que o índice de consistência seja obtido pela proporção de termos coincidentes de um conjunto de documentos que tratem do mesmo assunto e o total de termos desse conjunto (coincidentes ou não). Considera-se, para esta análise, que uma indexação coerente é aquela em que os termos que representam os conceitos de documentos que tratam do mesmo assunto devem ser idênticos, de modo a evitar sua dispersão, ambiguidade e imprecisão.

Seguindo o objetivo inicial dessa análise, mostra-se na Tabela 8, as DE e ID de maior ocorrência nos documentos da WoS. Em seguida, calcula-se o índice de consistência para cada grupo de palavras.

Tabela 8 – Conjunto de ID e DE de Maior Ocorrência

DE	ocorrência	ID	ocorrência
obesity	7	body-mass index	5
body mass index	2	obesity	4
human biomonitoring	2	atherosclerosis risk	3
meta-analysis	2	physical-activity	3
phthalates	2	risk	3
review	2	united-states	3
social class	2	accelerometer	2
		di(2-ethylhexyl)phthalate dehp	2
		disease	2
		metabolic syndrome	2
		nursery-school children	2
		nutrition examination survey	2
		population	2
		public-health	2
		randomized controlled-trial	2
		us adults	2
		waist circumference	2
		weight	2

Fonte: dados da pesquisa, 2017.

Observando a quantidade de palavras que possuem, no mínimo duas ocorrências, verifica-se que as *keywords plus* (ID) apresentam maior índice de ocorrência em comparação às palavras-chave de autor (DE), o que implica também em maior índice de consistência. O índice de consistência (IC) foi obtido da seguinte forma: $IC=DC/DT$ x 100, onde DC são os descritores comuns (que ocorrem ao menos duas vezes) e *DT* é o total de descritores comuns ou incomuns. Para as DE, o *IC* corresponde, aproximadamente, a 7,61%, enquanto que para as ID, o *IC* foi de 10,64%. Nota-se que o índice de consistência das DE e das ID é baixo, considerando a variação mencionada por Slype (1991) que deve ser entre 50% a 80%.

Lancaster (2004), a partir de um estudo de Hooper (1965), afirma que é muito difícil de se alcançar um alto índice de coerência interindexadores, já que a análise realizada por Hooper, em 14 estudos diferentes, permitiu ao autor chegar a valores muito dispersos, valores entre 10 a 80%. E nos estudos que foi possível recalcular a taxa de coerência, os resultados tiveram uma variação de 24 a 80%.

O principal problema da inconsistência de termos em um SRI é a pulverização do catálogo do sistema, pois a discordância entre os termos empregados para representar um conceito implica na ambiguidade e na imprecisão dos descritores. Isso acaba comprometendo a melhor resposta do sistema para determinada busca.

Com relação à visualização da informação construída a partir da ocorrência de palavras, quando a consistência de termos é muito baixa, torna-se difícil gerar um mapa adequado para visualizar um domínio. Isso ocorre porque esse tipo de visualização considera os termos de maior frequência como descritores para representar os assuntos mais discutidos em um domínio. Quando se tem uma alta frequência de termos, os *clusters* gerados são mais ricos, consistentes e interligados e, com isso, é possível fazer análises mais detalhadas sobre as tendências de uma área em determinado período. Do contrário, uma VI com elevado índice de termos inconsistentes, proporciona a construção de mapas com palavras de pouco aglutinamento e de maior dispersão ao redor do mapa gerado.

4.2 Segunda Análise: aplicação do vocabulário DeCS

Como já mencionado, em alguns momentos, no decorrer do trabalho, esta etapa da análise objetiva identificar os pontos de convergência entre o DeCS e as palvras-chave (DE) e *keywords plus* (ID) dos artigos de Nutrição indexados na WoS. Tal análise considerou a garantia literária como parâmetro para justificar as descobertas e afirmativas sobre o *corpus* analisado, mediante o critério aqui proposto.

O conceito de garantia literária, aplicado a esta pesquisa, busca identificar se a representação da informação do *corpus* selecionado na WoS corresponde à linguagem utilizada pelos profissionais da área de Nutrição do DeCS. Entende-se que, para estudos voltados à visualização da informação, em um dado domínio do conhecimento, é importante considerar que a representação temática deve utilizar uma linguagem comum entre a comunidade discursiva de determinado campo científico.

Contudo, explicando de antemão, não foi possível identificar, na WoS, listas de cabeçalhos de assunto ou outro tipo de vocabulário controlado que pudessem auxiliar os usuários da base (de diferentes domínios) a utilizar os termos mais apropriados nas suas expressões de busca. Também não foi realizado nenhum tipo de consulta aos editores da base de dados para saber se existe alguma política de indexação, que contenha alguma diretriz, que determine a indexação com base em algum vocabulário controlado. Acredita-se até que isso seria algo utópico, levando em consideração a complexidade dessa base de dados.

Na representação da informação da WoS, percebeu-se duas situações. A primeira é que os trabalhos são indexados e também recuperados pelas palvras-chave do próprio autor do texto; em segundo lugar, a WoS complementa a indexação desses documentos com as

keywords plus. Sendo estas, resultado de uma indexação automática que contabiliza palavras de maior frequência nos títulos das referências citadas, de modo a selecionar as mais recorrentes como descritor para ampliar a recuperação dos documentos. Observa-se, então, uma indexação que parte dos especialistas do domínio, isto é, do próprio autor do trabalho; e outra indexação que não tem necessariamente como garantir que o termo é adequado para o domínio.

Seria interessante verificar, de forma mais aprofundada, quais os critérios utilizados para elaborar as *keywords plus*, ou seja, quais as classes de palavras que são eliminadas dessa indexação, quais palavras são consideradas *stopwords*? Será que apenas artigos definidos e indefinidos, conjunções, preposições? E os termos que não se encaixam nessas classes gramaticais, mas que também não são representativos como, por exemplo, os termos extremamente generalistas, será que por trás dessa indexação automática, não existe um indexador especialista que possa verificar esse tipo de termo irrelevante? Descobrir esses detalhes daria um outro tipo de pesquisa.

Outro ponto importante para justificar essa análise é que algumas bases de dados utilizam vocabulários controlados para indexar seus documentos e também para auxiliar na busca e na recuperação da informação. Como no caso das bases da BVS que utilizam o MeSH e o DeCS como vocabulários controlados aplicados para as finalidades anteriormente mencionadas. Sendo assim, é coerente analisar se o domínio de Nutrição da WoS possui alguma coincidência com o vocabulário do DeCS.

Desse modo, cabe destacar que o propósito desta análise está organizado do seguinte modo: a) verifica-se, separadamente, se as palavras de maior ocorrência no grupo de DE e ID ocorrem no DeCS; b) verifica-se se os termos ditos mais específicos ocorrem ou não no DeCS e c) verifica-se, a partir do total de termos DE e do total de termos ID, quantas palavras correspondem ao DeCS.

Essa verificação foi realizada de forma semiautomática no software *Excel*. Primeiro foram listadas as palavras de maior ocorrência no conjunto de ID e DE, posteriormente foram pesquisadas na planilha dos 153 descritores do DeCS. Este processo de verificação se repete para as palavras consideradas de especificidade elevada e para o conjunto total de termos (incluindo ID e DE).

Entende-se, por vocabulário controlado, uma lista de termos autorizados, utilizada na indexação de documentos. No entanto, não se trata de uma mera lista de termos, pois, geralmente, possui uma estrutura lógica e semântica, tendo como principais objetivos, o

controle de sinônimos, diferenciar homógrafos e ligar termos que possuem uma relação estreita entre si (LANCASTER, 2004).

Fazer uso de vocabulário controlado em um SRI traz muitos benefícios para a busca e recuperação de documentos. Quando se tem um controle dos termos que podem ser utilizados como descritores e, ao mesmo tempo, quando o sistema fornece ao usuário informações sobre os termos preferíveis para uma expressão de busca, supõe-se que há maior eficiência na resposta do mesmo. Sobre este aspecto, Strehl (1998, p. 331) acrescenta que "o vocabulário controlado torna-se o ponto de convergência entre as linguagens utilizadas por autores, indexadores e pesquisadores – premissa fundamental para comunicação de informações dentro de um sistema".

Sobre as DE e ID de maior ocorrência, acrescenta-se que estas foram pesquisadas de forma semiautomática, no conjunto de 153 descritores do DeCS, para verificar quantas coincidem com este vocabulário. As palavras de maior ocorrência são as mesmas que foram elencadas no Tabela 8, da seção 4.1.5. Percebeu-se, então, que poucas palavras desse conjunto, ocorrem no DeCS. As DE que ocorrem no DeCS são "obesity", "body mass index" e "review". Nota-se que as duas primeiras são palavras óbvias, utilizadas na expressão de busca da WoS.

O termo "*review*", como mencionado na seção 4.1.3, é um termo abrangente. É curiosa sua presença no vocabulário DeCS, pois trata-se de uma palavra que não é clara quanto ao domínio de Nutrição e, ao mesmo tempo, não possui um especificador ou delimitador que a torne mais objetiva.

Os termos de especificidade elevada, apresentados na Tabela 3, da seção 4.1.2, foram consultados um a um no DeCS. A partir dessa investigação, percebeu-se que nenhuma dessas palavras de especificidade elevada fazem parte do DeCS. Essa constatação reforça, ainda mais, a ideia de que os termos de especificidade elevada, identificados no *corpus* de análise, não são termos preferidos pela comunidade discursiva para representar o domínio em estudo.

Comparando o total de termos do *corpus* (253 palavras sem repetição) com o DeCS, verificou-se que poucas palavras coincidiram. Isso insinua alto grau de divergência entre a linguagem da base e a linguagem do DeCS. No Quadro 9 seguem as palavras da WoS coincidentes no DeCS.

Tabela 9 – Palavras-chave da WoS coincidentes no DeCS

Palavras-chave (soma DE e ID)	Ocorrência DE	Ocorrência ID	Ocorrência total
Obesity	7	4	11
Diet	0	1	1
body mass index	2	0	2
Anthropometry	1	0	1

Fonte: Dados da pesquisa, 2017.

Ressalta-se que as palavras da WoS que coincidiram no DeCS foram exatamente algumas das quais utilizou-se na expressão de busca. Nesse sentido, é natural que tais palavras coincidam nos dois conjuntos de termos, isto é, nos documentos da base de dados e no vocabulário controlado mencionado. Mas é curioso que nem todos os termos utilizados na expressão de busca fazem parte do conjunto de palavras da WoS.

Talvez isso seja reflexo da redução do *corpus* para somente os vinte artigos mais citados na base de dados, pois acredita-se que, num SRI, os documentos são recuperados seguindo uma ordem de *score* numérico. Um exemplo de *score* é o número de coincidência de termos no documento. Nesse caso, considera-se como prioridade um documento em que coincidem três termos de uma expressão de busca ao invés de um documento que coincidem dois termos (LANCASTER, 2004). Esse *score* vai diminuindo conforme diminui a quantidade de palavras utilizadas na busca e que aparecem no documento. Então, uma possível explicação para que os vinte artigos desse trabalho não apresentem todas as palavras utilizadas na busca, é que, os documentos que continham o restante das palavras receberam um *score* inferior ao artigo de menor *score* dentro dos vinte documentos considerados, uma vez que o critério de ordenação utilizado foi o número de citações e não a quantidade de palavras-chave que aparecem nos artigos.

Concorda-se, então, que não existe uma linguagem comum entre o DeCS, as palavras-chave de autor e as *keywords plus* dos artigos de Nutrição da WoS. Contudo, é importante esclarecer que os descritores de Nutrição encontrados no DeCS, possuem uma certa limitação, pois trata-se de um conjunto muito exíguo de termos para representar um domínio. Além disso, este se restringe à Nutrição em Saúde Pública. Sendo assim, acredita-se que esta análise tende a ser um pouco limitada e, portanto, não deve ser considerada como uma verdade irrefutável, já que nos deparamos com um vocabulário diversificado na indexação dos documentos da WoS.

5 CONSIDERAÇÕES FINAIS

Inicialmente foi traçada, para esta pesquisa, uma estrutura textual convergente com o objetivo proposto, o qual, buscou investigar a qualidade da representação da informação em um domínio específico, sob a ótica da Visualização da Informação (VI). Tal estudo foi motivado pela problemática decorrente da natureza subjetiva da indexação e, também, pela necessidade de reunir aportes teóricos e técnicos que pudessem contribuir na construção de critérios de avaliação e também como ferramenta de apoio no desenvolvimento da atividade de indexação.

Numa sociedade movida pelo informacionalismo Castells (1999), a informação permeia diversas esferas de atividades, como a econômica, política, social e cultural. Nesse contexto, a informação técnica e científica, também orientada a atender as demandas sociais, torna-se a mola propulsora para o desenvolvimento de um país. No entanto, para que o conhecimento técnico e científico tenha um alcance local ou global, são necessários canais que proporcionem sua comunicação, como também métodos e técnicas que promovam sua melhor organização e representação para posterior recuperação e acesso.

A informação científica só faz sentido existir se for para ser comunicada. Dentre os principais canais de comunicação científica, existem as bases de dados, sejam nacionais ou internacionais. Todavia, esses canais não são passíveis de apresentarem problemas e limitações quanto ao tratamento temático da informação. A partir dessa premissa, o estudo elegeu a base de dados *Web of Science* como canal de comunicação científica para investigar o comportamento da representação da informação no domínio de Nutrição.

A escolha da base justificou-se por se tratar de uma base de dados internacionalmente renomada e que também indexa periódicos com fator de impacto elevado. Quanto ao domínio escolhido, vale destacar que existem alguns vocabulários controlados da área de Nutrição e, por este motivo, considerou-se mais viável investigar esse domínio, utilizando como um dos parâmetros de análise os Descritores de Ciências da Saúde (DeCS) que, por seu turno, possui uma subcategoria de Nutrição aplicada ao contexto da Saúde Pública.

Escolheu-se trabalhar com as palavras-chave de autor (DE) e *keywords plus* (ID) dos artigos da área em estudo, por acreditar que palavras-chave tendem a representar a informação de forma sucinta e coerente e, ao mesmo tempo, podem contribuir para a construção de mapas temáticos, de modo que seja possível visualizar as tendências de um domínio. Buscou-se, desta forma, compreender o comportamento das DE e ID diante das visualizações geradas a partir de um total de vinte artigos. A análise foi dividida em duas partes: uma dedicada à

aplicação de critérios de qualidade de indexação e, outra, tendo o DeCS como parâmetro para avaliar se existe algum índice de similaridade entre a representação da informação da WoS e o vocabulário controlado DeCS. Este tipo de análise teve como foco a garantia literária, pois acredita-se que estudos voltados à visualização da informação (quer seja para construir mapas de conhecimento voltados à recuperação da informação ou para a visualizar temáticas de domínios do conhecimento) devem levar em consideração a garantia de literatura, isto é, a linguagem comumente utilizada por especialistas de uma área.

Diante dos resultados, foi possível fazer algumas observações acerca das palavras-chave de autor e *keywords plus*. Segue-se uma síntese dessas observações:

- ✓ A partir do critério de **exaustividade**, observou-se que a representação da informação na WoS é predominantemente exaustiva. Tanto ID quanto DE são termos que aparecem em grande quantidade nos artigos analisados. Para a recuperação da informação, este é um critério favorável, todavia, sob a ótica da VI, tal critério não favorece a visualização, por apresentar inconsistência e pulverização de termos. Sabe-se que, para gerar visualizações a partir da coocorrência de palavras, é necessário que os termos coincidam. Mas a exaustividade identificada no conjunto dos artigos da WoS não atende a este aspecto.

- ✓ Sobre a **especificidade**, percebeu-se que tanto as DE quanto as ID contemplam esse aspecto, apesar de serem observadas algumas palavras abrangentes. A especificidade é considerada adequada e eficaz quando a demanda dos usuários do sistema é essencialmente específica. Já do ponto de vista da VI, a especificidade se torna adequada quando os termos específicos coincidem, ou seja, quando existe o consenso pela comunidade discursiva ao usar termos específicos. Concorda-se que, no *corpus* de vinte documentos, a indexação contempla o critério de especificidade, mas observou-se que, devido à característica exaustiva da representação temática desses documentos, poucas palavras específicas foram coincidentes. Nesse sentido, a exaustividade, somada à especificidade, não favoreceu a VI devido à pulverização dos termos.

- ✓ Acerca do critério de **controle de abrangência** dos termos, percebeu-se algumas palavras abrangentes. Consideram-se como termos abrangentes,

aqueles que não possuem clareza quanto ao assunto que está sendo representado. A abrangência pode ser do tipo leve, ou seja, mesmo os termos não sendo óbvios para um domínio, ele apresenta algum significado, como no caso de "*United States"*. A abrangência também pode ser do tipo elevada, como no caso das *stopwords*. No *corpus* analisado as palavras de abrangência elevada apresentaram pouca ocorrência. Sendo assim, do ponto de vista da VI, termos muito abrangentes não favorecem a visualização.

- ✓ Com relação ao critério de **uniformidade**, entende-se que esse critério objetiva o controle de sinonímia, homonímia e quantidade (plural e singular), além de outras formas variantes dos termos, como, por exemplo, os que utilizam caracteres especiais como hífen, aspas, parênteses ou quaisquer outros. Um dos problemas mais comuns decorrentes da ausência de controle/padronização é a variação dos termos indexados. Do ponto de vista da recuperação da informação esse problema já não é tão grave, pois a maioria dos SRIs possuem recursos capazes de solucionar essas diferenças. Contudo, para a VI, a variedade de termos ocasionada pela falta de padronização, não é um aspecto favorável, pois, assim, torna-se mais difícil compreender as estruturas de um domínio. Termos com uma diversidade de variações, tendem a proporcionar uma VI pulverizada. Na representação temática da WoS, notou-se algumas situações em que a mesma palavra foi escrita de forma diferente, seja por uso indevido do hífen, seja por uso de plural desnecessário. Nesse sentido, compreende-se que a indexação da WoS não contempla o critério de uniformidade.

- ✓ Quanto à **consistência** da indexação, entende-se, por este critério, o grau de concordância na representação da informação de um documento por meio de um conjunto de termos indexados. Buscou-se averiguar o grau de consistência da representação temática da WoS, tendo como parâmetro as palavras de maior ocorrência no conjunto de DE e no conjunto de ID. Obteve-se um índice de consistência de, aproximadamente, 7,61% nas DE e 10,64% nas ID. Verificou-se que o índice de consistência nos vinte artigos foi relativamente baixo, levando em consideração a taxa de variação mencionada por Slype (1991) que deve ser entre 50 a 80%.

- ✓ Já com relação ao **vocabulário DeCS**, verificou-se que não existe uma linguagem comum entre o DeCS e o conjunto dedescritores dos artigos analisados. Das palavras de maior ocorrência no conjunto total de ID e DE, apenas 3 delas aparecem no DeCS, são as palavras "obseity", "body mass index" e "review". Com relação às palavras-chave de especificidade elevada, constatou-se que nenhuma delas aparecem no DeCS. Comparando o total de termos do *corpus* com o vocabulário de Nutrição, notou-se que, das 253 palavras, apenas 4 ocorreram no DeCS. Ou seja, existe inconsistência entre o DeCS e a linguagem da WoS, especificamente, no domínio de Nutrição.

Acredita-se que a pesquisa alcançou o objetivo proposto, pois foi possível avaliar a qualidade da representação da informação na WoS por meio dos critérios estabelecidos. Percebeu-se que tanto as palavras-chave quanto as *keywords plus* do domínio de Nutrição, conseguem atender alguns critérios de indexação, todavia observaram-se alguns aspectos da representação temática que comprometem a visualização da informação (VI):

a) a exaustividade elevada, sem termos coincidentes, resultou na pulverização da VI; b) algumas palavras de especificidade elevada, que não são comumente utilizadas, implicaram na inconsistência de termos e na pulverização da VI; c) foram encontrados termos abrangentes e considerados como não representativos para o domínio, ou seja, são termos com apenas uma ocorrência que também não favorecem a VI; d) identificaram-se palavras iguais com variação de grafia, ou seja, isso significa que não há uniformidade na representação da informação. Um exemplo é a variação de grafia nas palavras "public health" e "public-health"; e) observou-se baixo índice de consistência no *corpus* analisado, isto é, poucas palavras ocorreram, no mínimo, duas vezes. Fator, este, que afeta a VI; f) quanto à garantia literária, tendo como parâmetro o DeCS, observou-se uma quantidade muito reduzida de palavras do *corpus* coincidentes no DeCS e na WoS; g) observou-se que as *keywords plus* são mais abrangentes que as palavras-chave de autor, mas do ponto de vista da VI, as ID favorecem uma melhor visualização por apresentar mais elementos coincidentes, ou seja, as ID, do *corpus* analisado, possuem maior ocorrência do que as DE.

É fato que os problemas de organização e recuperação da informação são tão antigos quanto atuais. A CI, enquanto área que se dedica a estudar e compreender os problemas de informação, compõe-se de métodos e técnicas para tratar a informação registrada e torná-la disponível. Ainda assim, sabe-se que estes estudos merecem maior enfoque para que se

possam minimizar as lacunas ainda existentes entre organização da informação e acesso ao conhecimento científico.

Destaca-se que os estudos sobre representação temática possuem maior dedicação à indexação voltada à recuperação da informação, o que, de certa forma, limita a aplicação das técnicas de tratamento temático em outros contextos. Devido à carência de estudos dedicados a avaliar a representação da informação para fins de visualização da informação, acredita-se que esta pesquisa teve como uma de suas contribuições acrescentar novas discussões para este contexto.

O estudo também contribuiu para a reunião e elaboração de critérios para avaliar a qualidade da indexação. Sendo esta entendida como o grau de eficiência de um sistema de recuperação da informação em promover respostas úteis e o grau de satisfação de um usuário para com os documentos recuperados durante uma busca. Ou sob a ótica da VI, uma boa indexação é aquela que garante a eleição de termos claros, objetivos e coincidentes para representar um domínio, de modo que as visualizações geradas não sejam pulverizadas.

Espera-se que esta pesquisa possa contribuir para estudos futuros relativos ao desenvolvimento de novas técnicas e metodologias de avaliação da qualidade da representação da informação no contexto da VI.

Salienta-se que novos estudos a respeito da temática contribuem para dar maior visibilidade ao tratamento temático da informação, no que diz respeito às práticas de organização, representação e visualização do conhecimento científico. Ademais, o estudo também pode contribuir para que profissionais, pesquisadores, estudantes tenham conhecimento da base de dados que foi utilizada nessa pesquisa, de forma que as mesmas possam servir de fontes de pesquisa para trabalhos futuros.

REFERÊNCIAS

ABRIL, G. Prólogo. In: GARCÍA GUTIÉRREZ, A. **Otra memoria es posible: estrategias descolonizadoras del archivo mundial**. Sevilla: Un. de Sevilla, 2004.

AGUILAR, A. G. et al. **Visualização de dados, informação e conhecimento**. Florianópolis: Ed. UFSC, 2017.

ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definição, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, Brasília, v. 32, n. 3, p. 7-20, set./dez. 2003. Disponível em: <http://www.scielo.br/pdf/ci/v32n3/19019.pdf>. Acesso em: 14 mar. 2017.

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Enc. Bibli**: R. Eletr. Bibliotecon. Ci. Inf., Florinópolis, n. 15, 2003.

ALVARENGA, L. Organização da informação nas bibliotecas digitais. In: NAVES, M. M. L.; KURAMOTO, H. (Org.). **Organização da informação: princípios e tendências**. Brasília: Briquet de Lemos, 2006.

ANDALIA, R. C.; CHAPMAN, M. C. S. Elementos sobre indización y búsqueda de la información por medio de vocabularios controlados en bases de datos biomédicas. **Revista Cubana de Ciéncias de la Salud**, La Habana, v. 22, n. 2, p. 142-154, 2011.

AQUINO, I. J.; CARLAN, E.; BRASCHER, M. B. Princípios classificatórios para a construção de taxonomias. **PontodeAcesso**, Salvador, v. 3, n. 3, p. 196-215, 2009.

ARAÚJO, C. A. A. Correntes teóricas da ciência da informação. **Ciência da Informação**, 2009, vol.38, n.3, p. 192-204.

ARAUJO JUNIOR, R. H. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.

BARITÉ, M. G. R. et al. Garantia literária: elementos para uma revisão crítica após um século. **Transinformação**, v. 22, n. 2, p. 123-138, 2010. Disponível em: <http://www.scielo.br/pdf/tinf/v22n2/a03v22n2.pdf>. Acesso em: 01 out. 2016.

BARITÉ, M. **Glosario sobre organización y representación del conocimiento, clasificación, indización, terminología**. Montevideo: Comisión Sectorial de Investigación Científica, 1997.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org**.). Educação, Universidade e Pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p.35-60.

BOCCATO, V. R. C.; FUJITA, M. S. L. Avaliação da linguagem documentária Decs na área de fonoaudiologia na perspectiva do usuário: estudo de observação da recuperação da informação como protocolo verbal. **Encontros Bibli**, Florianópolis, n. 21, p. 16-33, 1º sem. 2006.

BORGES, G. S. B.; LIMA, G. A. Desenvolvimento de software de indexação automática: breve avaliação dos principais critérios. **Informação & Tecnologia (ITEC)**, Marília/João Pessoa, v. 2, n. 2, p. 49-70, jul./dec., 2015.

BRANDAU, R.; MONTEIRO, R.; BRAILE, D. M. Importância do uso correto dos descritores nos artigos científicos. **Rev Bras Cir Cardiovasc**, São José do Rio Preto, v. 20, n. 1, mar. 2005. Disponível em: <http://www.scielo.br/pdf/rbccv/v20n1/v20n1a04.pdf>. Acesso em: 16 jul. 2016

BRÄSCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo, **Anais**... São Paulo: ANCIB, 2008.

BRÄSCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRASCHER, M. (Org.). **Passeios pelos bosques da informação**: estudos sobre representação e organização da informação e do conhecimento. Brasília: IBICT, 2010.

BUFREM, L. S.; FREITAS, J. L. Interdomínios na literatura periódica científica da ciência da informação. **DataGramaZero,** v. 16, n. 6, p. A02, 2015. Disponível em: <http://www.brapci.inf.br/v/a/18491>. Acesso em: 30 set. 2016.

CAMPOS, M. L. A. **Linguagens documentárias:** teorias que fundamentam sua elaboração. Niterói, RJ: EDUFF, 2001. 133 p.

CASTELLS, Manuel. **A sociedade em rede**. 10. ed. São Paulo: Paz e Terra, 1999.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. Trad. José Augusto Chaves Guimarães. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.21, n.1/2, p. 63-79, jan./jun. 1988.

CINTRA, A. M. M. et al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002. 92 p.

COELHO-NETTO, T. J. **Semiótica, informação e comunicação**. 7. ed. São Paulo: Perspectiva, 2007. 217 p.

CONSULTA AO DeCS. [2017?]. Disponível em: <http://decs.bvs.br/cgi-bin/wxis1660.exe/decsserver/> Acesso em: 19 jun. 2017.

CORREA, C. A. **Indexação automática e visualização de informações**: um estudo baseado em Lógica paraconsistente. 2011. 152 f. Tese (Doutorado em Ciência da Informação) – ECA–USPE, São Paulo, 2011.

CUNHA, M. B.; CAVALCANTI, C. R. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos; Lauro de Freitas: Livros, 2008. 451 p.

CURRÁS, E. **Ontologias, taxonomia e tesauro em teoria de sistemas e sistemática**. Brasília: Thesaurus, 2010, 182 p.

DAHLBERG, I. Current trends in Knowledge Organization. In: GARCÍA MARCO, F. J. (Ed.). **Organización del conocimiento en sistemas de información y documentación**. Zaragoza: Librería General, 1995. p. 7-25.

DAHLBERG, I. Teoria do Conceito. **Ciência da Informação,** Rio de Janeiro, 7(2), 102 – 107, 1978. Disponível em: <http://revista.ibict.br/ciinf/article/view/115/115 >. Acesso: 24 ago. 2016.

Descritores em Ciências da Saúde: DeCS. [2017?]. Disponível em: < http://decs.bvsalud.org >. Acesso em 22 de abr. 2017.

DIAS, C. L. C. O. A análise de domínio, as comunidades discursivas, a garantia de literatura e outras garantias. **Informação & Sociedade: Estudos**, João Pessoa, v. 25, n. 2, p. 7-17, 2015. Disponível em: <http://www.brapci.inf.br/v/a/18418>. Acesso em: 30 set. 2016.

DICIONÁRIO ONLINE DE PORTUGÊS: DICIO. [2017?]. Disponível em: <https://www.dicio.com.br/representar/>. Acesso em: 30 out. 2017.

DODEBEI, V. L. D. **Tesauro:** linguagem de representação da memória documentária. Niterói: Intertexto, 2002. 119 p.

FERREIRA, M. S. **A representação da memória científica da Ciência da Informação brasileira**: um estudo com as palavras-chave do ENANCIB. 2012, 179f. Dissertação (mestrado em Ciência da Informação) – Universidade Federal de Pernambuco, Centro de Artes e Comunicação. Recife, 2012. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/10440/1/Marilucy-PGCI%20Mestr..pdf>. Aceso em: 20 abr. 2017.

FOGL, J. Relations of the concepts 'information' and 'knowledge'. International Fórum on Information and Documentation, **The Hague**, v.4, n.1, p. 21-24, 1979.

FREITAS, C. M. D.S. et al. Introdução à Visualização de Informações. **RITA**: revista de Informática Teórica e Aplicada, Porto Alegre, v. 8, n. 2, p. 143-158, out. 2001.

FUJITA, M. S. L. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Rev. Bras**. Ed. Esp., Marília, v.10, n.3, p.257-272, set./dez. 2004.Disponível em: <http://www.abpee.net/homepageabpee04_06/artigos_em_pdf/revista10numero3pdf/1fujita.pdf> Acesso em: 15 out. 2016.

FUJITA, M. S. L.; GIL-LEIVA, I. Avaliação da indexação por meio da recuperação da informação. **Ci. Inf.**, Brasília, v. 41, n. 1, p. 50-66, jan./abr. 2014.

GARFIELD, E. Citation indexes for science: a new dimension in documentation through association of ideas. **Science**, Washington, v. 122, n. 3159, p. 108-111, 1955.

GARFIELD, E. Historiographs, librarianship, and the history of science. In: **Toward a theory of librarianship**: papers in honor of Jesse Hauk Shera, ed. by Conrad H. Rawski (Metuchen, N. J.: Scarecrow Press, 1973), p. 380-402.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Editora Atlas, 2002.

GIL LEIVA, I. **Manual de indización**: Teoría y práctica. Gijón: Trea, 2008. 429 p

GUIMARÃES, J.A.C. A análise documentária no âmbito do tratamento temático da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L. (Orgs.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003. p. 100-117.

GUIMARÃES, J. A. C. Perspectivas de ensino e pesquisa em organização do conhecimento em cursos de Biblioteconomia: uma reflexão. In: CARRARA, K. (org.). **Educação, universidade e pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p. 61-74.

HENDERSON, J. L. Os mitos antigos e o homem moderno. In: JUNG, C. G. (Org.) **O homem e seus símbolos**. 2. ed. Rio de Janeiro: Nova Fronteira, 2011. 429 p.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2., p. 87-111, 2003. Disponível em: <http://ppggoc.eci.ufmg.br/downloads/bibliografia/Hjorland2003.pdf> Acesso em: 10 ago. 2016.

HJØRLAND, B. Theory and metatheory of information science: a new interpretation. **Journal of Documentation**. London, v. 54, n. 5 p. 606-621, 1998.

ISKO – INTERNATIONAL SOCIETY OF KNOWLEDGE ORGANIZATION. **ISKO`s mission**, 2016. Disponível em: < http://www.isko.org/about.html>. Acesso: 29 mar. 2016.

JUNG, C. G. (Org.) **O homem e seus símbolos**. 2. ed. Rio de Janeiro: Nova Fronteira, 2011. 429 p.

KOBASHI, N. Y. Fundamentos semânticos e pragmáticos da construção de instrumentos de representação de informação. **Datagrammazero**: Revista de Ciência da Informação, Rio de Janeiro, v.8, n.6, dez. 2007.

KOBASHI, N. Y.; SANTOS, R. N. M. Arqueologia do trabalho imaterial: uma aplicação bibliométrica à análise de dissertações e teses. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, v. 13, n. esp., p. 106-115, 2008. Disponível em: <http://www.brapci.inf.br/v/a/5004>. Acesso em: 29 set. 2017.

KOBASHI, N. Y.; TÁLAMO, M. F G. M. (2003). Informação: fenômeno e objeto de estudo da sociedade contemporânea. **Transinformação**, Campinas, v. 15, n. esp., p. 7- 21, set./dez. 2003.

KUHN, Thomas S. **A estrutura das revoluções científicas**. 5. ed. São Paulo: Editora Perspectiva, 1998.
LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos/Livros, 1993.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos/Livros, 2004.

LAPA, R. C.. Indexação Automática no Brasil no âmbito da Ciência da Informação (1973-2012). 2014. 287 f. Dissertação (Mestrado) - Curso de Ciência da Informação, Departamento de Ciência da Informação, Universidade Federal de Pernambuco, Recife, 2014.

LE COADIC, Y. F. **A ciência da informação**. Brasília: Briquet de Lemos, 1996

LIMA, José Leonardo Oliveira; ALVARES, Lillian. Organização e representação da informação e do conhecimento. In.: ALVARES, Lillian (Org.). **Organização da informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações**. São Paulo: B4 Editores, 2012. p. 21-47.

LOPES, I. L. Uso das linguagens controlada e natural em bases de dados: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 1, p. 41-52, jan./abr. 2002. Disponível em: <http://revista.ibict.br/ciinf/article/view/976> Acesso em: 20 nov. 2016.

MAI, J-E. Deconstructing the indexing process. Advances in **Librarianship,** v. 23, p. 269-298, 2000.

MELO, M. A. F.; BRÄSCHER, M. Termo, conceito e relações conceituais: um estudo das propostas de Dahlberg e Hjorland. **Ciência da Informação**, Brasília, DF, v. 41 n. 1, p.67-80, jan./abr., 2014. Disponível em: <http://revista.ibict.br/ciinf/article/view/1419/1597>. Acesso: 24 set. 2016.

MIGUÉS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório cientifico da Universidade de Coimbra. **PontodeAcesso**. Salvador, v. 7, n. 1, p.116-131, abr. 2013. Disponível em: <https://portalseer.ufba.br/index.php/revistaici/article/view/8045/5810>. Acesso em: 20 abr. 2017.

NARUKAWA, C. M. **Estudo de vocabulário controlado na indexação automática**: aplicação no processo de indexação do sistema de indización semiautomatica (SISA). 2011. 222 f. Dissertação (Mestrado em Ciência da Informação) -Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2011.

NARUKAWA, C. M.; GIL LEIVA, I.; FUJITA, M. S. L. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de odontologia. **Inf. & Soc**.: Est., João Pessoa, v. 19, n. 2, p. 99-118, maio/ago. 2009.

NUNEZ, Z. A. G. **Produção científica brasileira em medicina topical indexada nas bases de dados web of cience e scopus entre os anos de 2005 a 2012**. 2014. 143 f. Dissertação (Mestrado em Comunicação e Informação) -Faculdade de Biblioteconomia e Documentação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

NÖTH, W. **Panorama da semiótica de Platão a Peirce**. 3. ed. São Paulo: Annablume, 2003. 149 p.

PACKER, A.L. Os periódicos brasileiros e a comunicação da pesquisa nacional. **Revista USP**, n.89, p.26-61, 2011. Disponível em: <https://www.revistas.usp.br/revusp/article/view/13868> Acesso em: 20 nov. 2016.

PEIRCER, C. S. **Semiótica e filosofia**. 2. ed. São Paulo: Editora Cultrix, 1975. 164 p.

PINHO, Fábio Assis. **Aspectos éticos em representação do conhecimento**: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol. 2006. 123 f. Dissertação (Mestrado em Ciência da Informação). – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2006. Disponível em: <http://www.enancib.ppgci.ufba.br/premio/UNESP_Pinho.pdf>. Acesso em: 11 nov. 2016.

PINHO, F. A. **Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina**: uma análise da precisão em linguagens de indexação brasileiras. 2010. 149 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

PINHO, F. A. **Fundamentos da organização do conhecimento**. Recife: Editora Universitária da UFPE, 2009.

RIBEIRO, D. M. **Visualização de dados na Internet**. 2009. 124 f. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Pontifícia Universidade Católica de São Paulo, São Paulo, 2009.

ROBREDO, J. Filosofia e informação? Reflexões. **RICI**: Rev. Ibero-amer. Ci. Inf., Brasília, v. 4, n. 2, p. 2-39, ago./dez. 2011. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/6207/5100>. Acesso em: 20 nov. 2016.

SANTAELLA, L. **O que é semiótica. São Paulo**: Brasiliense, 1983. (Coleção Primeiros Passos).

SAUSSURE, F. **Curso de linguística geral**. São Paulo: Cultrix, 2006.

SARACEVIC, T. Interdisciplinary nature of Information Science. Ciência da Informação, v. 24, n. 1, 1995. Disponível em: <http://revista.ibict.br/ciinf/article/download/608/610>. Acesso: 29 ago. 2016.

SILVA, M.R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, Campinas, v.16, n. 2, p. 133-161, maio/ago. 2004. Disponível em: <http://www.scielo.br/pdf/tinf/v16n2/03.pdf> Acesso em: 19 out. 2017.

SMIRAGLIA, R. P. The progress of theory in knowledge organization. **Library Trends**, Champaing, v. 50, n. 3, p. 519-537, 2001. Disponível em: <https://www.ideals.illinois.edu/bitstream/handle/2142/8414/librarytrendsv50i3d_opt.pdf?sequence=1&isAllowed=y> Acesso em: 30 maio 2016.

SMIRAGLIA, R. P. **The Elements of Knowledge Organization**. Dordrecht: Springer, 2014.

SOUZA, B. P.; FUJITA, M. S. L. Análise de assunto no processo de indexação: um percurso entre teoria e norma. **Inf. & Soc. Est**., João Pessoa, v.24, n.1, p. 19-34, jan./abr. 2014.

STRHEL, L. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. **Ciência da Informação**, Brasília, v. 27, n. 3, p. 329-35, set./dez. 1998. Disponível em: < http://revista.ibict.br/ciinf/article/view/787/816> Acesso em: 20 nov. 2016.

SVENONIUS, E. **The intellectual foundations of information organization**. Cambridge: The MIT Press, c2000. 255 p.

TÁLAMO, M.F.G.M. **Linguagem documentária**. São Paulo: APB, 1997. (Ensaios APB, n.45).

TÁLAMO, M. F. G. M. Terminologia e documentação. **TradTerm**: Revista do Centro Interdepartamental de Tradução e Terminologia, São Paulo, n. 7, p.141-152, 2001.

TAYLOR, A.G. **The organization of information**. 2.ed. London: Westport Connecticut, 2004.

TARGINO, M. G. **Comunicação científica**: o artigo de periódico nas atividadesde ensino e pesquisa do docente universitário brasileiro na pós-graduação. 1998. 387 f. Tese (Doutorado em Ciência da Informação) – Departamento de Ciência da Informação da Universidade de Brasília, Brasília, 1998.

TRISTÃO, A. M. D.; FACHIN, G. R. B.; ALARCON, O. E. Sistema de classificação facetada e tesauros: instrumentos para organização do conhecimento. **Ciência da Informação**. Brasília, v. 33, n. 2, p.161/171, maio/ago. 2004.

VAN SLYPE, G. **Linguagem documentária e linguística**. Trad. Cordélia R. Cavalcanti. Brasília: UnB; Departamento de Biblioteconomia, 1983.

VAN SLYPE, Georges. **Los lenguages de indización**: concepción, construcción y utalización en los sistemas documentales. Madrid: Fundación German Sánchez Ruipérez, 1991. (Biblioteca del livro).

VICKERY, Brian C. Thesaurus: a new word in documentation. **Journal of Documentation**, v. 16, n. 4, p. 181-189, dec. 1960.

VIEIRA, J. M. L. **A contribuição da organização e da visualização da informação para os sistemas de recuperação de informação**. 2014. 227 f. Dissertação (Mestrado em Ciência da Informação) – Centro de Artes e Comunicação, Universidade Federal de Pernambuco, Recife, 2014.

VIEIRA, J. M. L.; CORRÊA, Renato Fernandes. Visualização da Informação na construção de interfaces amigáveis para Sistemas de Recuperação de Informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 16, n. 32, p. 73-93, jul./dez.
2011.

VOGEL, M. J. M. (2007). **A noção de estrutura linguística e de processo de estruturação e sua influência no conceito e na elaboração de linguagens documentárias**. 2007, 124 f. Dissertação (Mestrado em Ciência da Informação) –ECA-USP, São Paulo, 2007.

WEB OF SCIENCE: **Coleção Principal (Thomson Reuters)**. [2017?]. Disponível em: <http://buscador-periodicos-capes-gov-br.ez16.periodicos.capes.gov.br/V/6JQYMHX4RTQ4R3NFYDQPYVP8BR9M2SM6QS2JBBC9F4UH6IJ1MN-01255?func=find-db-info&doc_num=000002653> Acesso em: 27 jun. 2017.

ZENG, M. L. Knowledge organization systems (KOS). **Knowledge Organization**: international journal devoted to concept theory, classification, indexing, and knowledge representation, Frankfurt, v. 35, n. 2-3, p. 160-182, 2008.

APÊNDICE A – REFERÊNCIAS DO *CORPUS* ANALISADO

BROWN, I. et al. Salt intakes around the world: implications for public health. **International Journal of Epidemiology**, v. 38, n. 3, p. 791-813, 2009.

COLE-LEWIS, H.; KERSHAW, T. Text Messaging as a Tool for Behavior Change in Disease Prevention and Management. **Epidemiologic Reviews**, v. 32, n. 1, p. 56-69, 2010.

DOLINOY, D. et al. Maternal Genistein Alters Coat Color and Protects Avy Mouse Offspring from Obesity by Modifying the Fetal Epigenome. **Environmental Health Perspectives**, v. 114, n. 4, p. 567-572, 2006.

FROMME, H. et al. Perfluorinated compounds – Exposure assessment for the general population in western countries. **International Journal of Hygiene and Environmental Health**, v. 212, n. 3, p. 239-270, 2009.

HAGSTRÖMER, M.; OJA, P.; SJÖSTRÖM, M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. **Public Health Nutrition**, v. 9, n. 06, 2006.

HEUDORF, U.; MERSCH-SUNDERMANN, V.; ANGERER, J. Phthalates: Toxicology and exposure. **International Journal of Hygiene and Environmental Health**, v. 210, n. 5, p. 623-634, 2007.

HOLICK, M. Vitamin D Status: Measurement, Interpretation, and Clinical Application. **Annals of Epidemiology**, v. 19, n. 2, p. 73-78, 2009.

LEE, C. et al. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: a meta-analysis. **Journal of Clinical Epidemiology**, v. 61, n. 7, p. 646-653, 2008.

MATTHEWS, C. et al. Amount of Time Spent in Sedentary Behaviors in the United States, 2003-2004. **American Journal of Epidemiology**, v. 167, n. 7, p. 875-881, 2008.

MCGRATH, J. et al. Schizophrenia: A Concise Overview of Incidence, Prevalence, and Mortality. **Epidemiologic Reviews**, v. 30, n. 1, p. 67-76, 2008.

MCLAREN, L. Socioeconomic Status and Obesity. **Epidemiologic Reviews**, v. 29, n. 1, p. 29-48, 2007.

ORSINI, N. et al. Meta-Analysis for Linear and Nonlinear Dose-Response Relations: Examples, an Evaluation of Approximations, and Software. **American Journal of Epidemiology**, v. 175, n. 1, p. 66-73, 2011.

PAPAS, M. et al. The Built Environment and Obesity. **Epidemiologic Reviews**, v. 29, n. 1, p. 129-143, 2007.

PEPPARD, P. et al. Increased Prevalence of Sleep-Disordered Breathing in Adults. **American Journal of Epidemiology**, v. 177, n. 9, p. 1006-1014, 2013.

POWELL, L. et al. Food store availability and neighborhood characteristics in the United States.**Preventive Medicine**, v. 44, n. 3, p. 189-195, 2007.

SALLIS, J. et al. AN ECOLOGICAL APPROACH TO CREATING ACTIVE LIVING COMMUNITIES. **Annual Review of Public Health**, v. 27, n. 1, p. 297-322, 2006.

STORY, M. et al. Creating Healthy Food and Eating Environments: Policy and Environmental Approaches. **Annual Review of Public Health**, v. 29, n. 1, p. 253-272, 2008.

STURM, R. Increases in morbid obesity in the USA: 2000–2005. **Public Health**, v. 121, n. 7, p. 492-496, 2007.

WANG, Y.; BEYDOUN, M. The Obesity Epidemic in the United States Gender, Age, Socioeconomic, Racial/Ethnic, and Geographic Characteristics: A Systematic Review and Meta-Regression Analysis. **Epidemiologic Reviews**, v. 29, n. 1, p. 6-28, 2007.

WORMUTH, M. et al. What Are the Sources of Exposure to Eight Frequently Used Phthalic Acid Esters in Europeans?. **Risk Analysis**, v. 26, n. 3, p. 803-824, 2006.

APÊNDICE B – APRESENTAÇÃO DOS METADADOS DA WoS

Quadro 5 – Organização dos metadados na WoS

AU	Roberfroid, M
AF	Roberfroid, Marcel
TI	Prebiotics: The concept revisited
SO	JOURNAL OF NUTRITION
LA	English
DT	Article; Proceedings Paper
CT	World Dairy Summit of the International-Dairy-Federation
CY	SEP, 2003
CL	Brugge, BELGIUM
SP	Int Dairy Federat
ID	16S RIBOSOMAL-RNA; IN-SITU HYBRIDIZATION; INCREASES FECAL BIFIDOBACTERIA; HEALTHY HUMANS; DIETARY FIBER; HUMAN COLON; FRUCTO-OLIGOSACCHARIDES; OLIGONUCLEOTIDE PROBES; INTESTINAL MICROFLORA; MICROBIAL ECOLOGY
AB	A prebiotic is "a selectively fermented ingredient that allows specific changes, both in the composition and/or activity in the gastrointestinal microflora that confers benefits upon host well-being and health." Today, only 2 dietary nondigestible oligosaccharides fulfill all the criteria for prebiotic classification. The daily dose of the prebiotic is not a determinant of the prebiotic effect, which is mainly influenced by the number of bifidobacteria/g in feces before supplementation of the diet with the prebiotic begins. The ingested prebiotic stimulates the whole indigenous population of bifidobacteria to growth, and the larger that population, the larger is the number of new bacterial cells appearing in feces.

Continua

Continuação

The "dose argument" is thus not supported by the scientific data: it is misleading for consumers and should not be allowed. A prebiotic index is proposed, defined as "the increase in the absolute number of bifidobacteria expressed divided by the daily dose of prebiotic ingested."	
C1	Univ Catholique Louvain, B-1348 Louvain, Belgium.
RP	Roberfroid, M (reprint author), Univ Catholique Louvain, B-1348 Louvain, Belgium.
EM	marcel@fefem.com
NR	68
TC	410
Z9	435
U1	9
U2	122
PU	AMER SOCIETY NUTRITIONAL SCIENCE
PI	BETHESDA
PA	9650 ROCKVILLE PIKE, RM L-2407A, BETHESDA, MD 20814 USA
SN	0022-3166
J9	J NUTR
JI	J. Nutr.
PD	MAR
PY	2007
VL	137
IS	3
SU	S
BP	830S
EP	837S

Continua...

Continuação

PG	8
WC	Nutrition & Dietetics
SC	Nutrition & Dietetics
GA	140WU
UT	WoS:000244538300015
PM	17311983
OA	No
DA	2017-08-10
ER	
EF	

Fonte: Dados da pesquisa, 2017.