DOI: 10.5433/1981-8920.2020v25n4p477

ANÁLISE DA REPRESENTAÇÃO DA INFORMAÇÃO NA *WEB OF SCIENCE*: UM ESTUDO A PARTIR DO DOMÍNIO DE NUTRIÇÃO

ANALYSIS OF THE INFORMATION REPRESENTATION IN THE WEB OF SCIENCE: A STUDY OF NUTRITION FIELD

Elisângela Vilela dos Santos[a]
Fábio Mascarenhas Silva[b]

RESUMO
Introdução: A motivação do estudo é entender se as palavras-chave e as *keywords plus* dos artigos da área de Nutrição, indexados na *Web of Science*, recebem tratamento temático adequado que favoreça a visualização da informação. **Objetivo:** avaliar a qualidade da representação da informação na Base de dados *Web of Science* (WoS) por meio da análise dos campos palavras-chave de autor e *keywords plus* dos artigos da área de Nutrição em Saúde Pública, buscando observar o uso desses termos no contexto da visualização da informação. **Metodologia:** Trata-se de um estudo exploratório, que busca compreender a dinâmica da representação da informação do domínio de Nutrição na base de dados 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 termos da referida base e, na segunda, utilizou como parâmetro o vocabulário DeCS (Descritores de Ciências da Saúde) para encontrar pontos de intersecção entre esse vocabulário e a representação da informação na WoS. Tal análise teve como ponto central a garantia literária. **Resultados:** tanto as palavras-chave de autor quanto as *keywords plus* do *corpus* analisado conseguiram atender a critérios de indexação, todavia foram observados aspectos da representação temática que comprometem a visualização da informação. **Conclusões:** Quanto à garantia literária, o resultado mostrou que não existe uma linguagem comum entre o DeCS e os termos da WoS.

Descritores: Representação da Informação. Visualização da Informação. Indexação de assuntos. Nutrição em Saúde Pública. *Web of Science*.

[a] Mestre em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Pernambuco (PPGCI/UFPE). Bibliotecária - Documentalista da Universidade Federal de Alagoas (UFAL). E-mail: elisangela.vilela.s@gmail.com

[b] Doutor em Ciência da Informação pela Universidade de São Paulo (USP). Docente do Programa de Pós-graduação em Ciência da Informação da Universidade Federal de Pernambuco (PPGCI/UFPE). E-mail: fabiomascarenhas@gmail.com

1 INTRODUÇÃO

A Representação da Informação (RI), enquanto parte do processo de Organização da Informação (OI) que media conteúdos informacionais às pessoas, desempenha papel essencial na sociedade contemporânea. No cenário atual, em que a produção documental aumenta significativamente, e entendida a necessidade de seu acesso e uso, faz-se mister questionar e refletir metodologias de gestão, disseminação, acesso e uso da informação.

A comunidade científica tem discutido diferentes formas de RI, destaque para os estudos centrados nos aspectos extrínsecos dos documentos (como a catalogação descritiva) e nos aspectos intrínsecos dos documentos (como a representação temática) por meio da atividade de classificação e indexação de assuntos.

Este artigo trata dos aspectos intrínsecos da RI, precisamente da representação por meio de palavras-chave e descritores como forma de visualização do conhecimento publicado em artigos da área de Nutrição (contexto da Saúde Pública) indexados na base de dados *Web of Science* (WoS).

Salienta-se que palavras-chave e descritores, a depender da aplicabilidade, podem servir tanto como instrumento ou como produto de representação da informação. No caso das palavras-chave da WoS, estas devem ser entendidas como produtos do processo de indexação, já que permitem a recuperação por assuntos.

O uso de palavras-chave e descritores, enquanto produtos da representação da informação, deve permitir a comunicação entre usuário e sistema de informação. Ademais, por meio de palavras-chave e descritores, também é possível montar representações visuais dos principais assuntos de um domínio[1], de forma a acompanhar o progresso ou estagnação da produção científica de um campo a partir de diferentes temáticas.

[1] Domínio é definido por Dias (2015) como uma área de conhecimento ou um campo de especialidade. 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”.

A relação entre representação e Visualização da Informação (VI) tem suscitado interessantes debates na Ciência da Informação (CI). No entanto, há certa tendência em relacionar tais termos com a finalidade de recuperar informação. Há, então, uma necessidade de se observar a relação entre representação e visualização sob outra perspectiva, ou seja, a de compreender representações temáticas tendo como eixos principais a RI por meio de palavras-chave e/ou descritores e aVI, no sentido de representar visualmente tais representações.

Algumas bases de dados adotam vocabulários controlados para padronizar a indexação e, ao mesmo tempo, facilitar o processo de busca e recuperação do conhecimento científico. É o caso da 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 utilizam os Descritores em Ciências da Saúde (DeCS) como terminologia comum entre os pesquisadores da área de Saúde e Ciências da Vida. Nesse sentido, elegeu-se o DeCS para padronizar a busca na WoS e, ao mesmo tempo, utilizá-lo como parâmetro em uma das análises dessa pesquisa.

Motivada pela necessidade de compreender diferentes aspectos da indexação de assuntos no contexto da VI, a pesquisa elegeu a WoS para estudar o comportamento das palavras-chave de autor e *keywords plus* dos artigos da área de Nutrição. A escolha pela WoS se justifica por se tratar de uma base de renome mundial (PACKER, 2011) e que possibilita analisar a representação da informação no âmbito internacional.

Diante da problemática apresentada, elege-se a seguinte questão de pesquisa: *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?* Dessa forma, o objetivo geral deste trabalho é avaliar a qualidade da representação da informação na WoS através das palavras-chave de autor e das *keywords plus* dos artigos da área de Nutrição. Especificamente espera-se observar se tal representação favorece ou não a VI

obtida a partir desses termos. Para tal foram estabelecidos critérios de avaliação da indexação fundamentados na literatura.

2 REFERENCIAL TEÓRICO

Abril (2004) compreende que a Organização da Informação (OI), 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 e produtos de representação, por exemplo, que se consegue gerenciar a informação e comunicá-la a sociedade.

Svenonius (2000) ressalta que a OI requer sua descrição e está, 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. Para Taylor (2004) estes objetos são as informações registradas, 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 cuja finalidade é, geralmente, a recuperação da informação.

Para as autoras (2008) a Representação e Organização do Conhecimento (OC/RC), são aplicáveis às unidades do pensamento, isto é, aos conceitos e têm por objetivo a construção de modelos de mundo que se constituem em abstrações da realidade.

Enquanto campo disciplinar, Dahlberg (1978) compreende que a OC e RC têm por base o conceito e suas relações para representar diferentes domínios. Esses conceitos são enunciados verdadeiros a respeito de um objeto. Desse modo, entende-se que a OC e a RC não se limitam a conceitos individuais, mas buscam compreender conceitos construídos socialmente em diferentes contextos e domínios, já que nas diversas áreas existem muitos conceitos

concorrentes, os quais evoluem segundo a construção social do conhecimento (HJORLAND, 2009).

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. (CHAUMIER, 1988, p. 70-71). Acrescenta-se mais dois instrumentos importantes para representar informação e/ou conhecimento: as **taxonomias** e as **ontologias**.

Quanto aos produtos de representação da informação, os índices, os resumos e os catálogos de assunto são os principais. Geralmente o **índice** consiste em uma listagem alfabética ou sistemática de tópicos que indicam a localização de cada um deles em um documento ou em uma coleção de documentos (DIAS; NAVES, 2007; VIEIRA, 2014); enquanto que o **resumo** diz respeito a representação sucinta, mas exata do conteúdo de um documento (LANCASTER, 2004). Destaca-se que as **palavras-chave** também são produtos de representação temática, visto que proporcionam ao usuário recuperar documentos em um Sistema de Recuperação da Informação (SRI).

É importante não confundir palavra-chave (linguagem natural) com descritor (linguagem artificial). Nessa perspectiva, Migués e Neves (2013, p. 117) entendem que palavras-chave e descritores não são a mesma coisa. 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 provenientes de uma linguagem documentária.

Do ponto de vista da VI, a representação por meio de palavras-chave ou descritores pode contribuir tanto com a construção de interfaces amigáveis que permitam uma melhor interação entre usuário e SRI, quanto para o desenvolvimento de mapas de conhecimento, através da frequência de termos, de modo que seja possível analisar o comportamento das temáticas de um domínio.

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.

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).

3 PROCEDIMENTOS METODOLÓGICOS

Trata-se de uma pesquisa exploratória e de abordagem qualitativa que buscou avaliar a qualidade da representação da informação na WoS através das palavras-chave e *keywords plus* dos artigos da área de Nutrição, de modo a observar o comportamento desses termos diante da visualização da informação.

O corpus dessa pesquisa é composto pelos registros recuperados na WoS (vinte artigos mais citados da área de Nutrição), tendo como objeto de estudo as palavras-chave de autor e as *keywords plus* desses documentos - contabilizando um total de 253 termos, eliminadas as duplicações.

Utilizou-se o vocabulário DeCS – Descritores em Ciências da Saúde[2] como parâmetro para realizar a busca na WoS. Primeiro buscou-se identificar as categorias DeCS pertencentes a área de Nutrição e optou-se pela categoria SP6 – *Nutrition, Public Health*. Em seguida foram identificados 153 descritores dessa categoria, dos quais foram selecionados os vinte de maior ocorrência nas bases BVS apontadas pelo próprio DeCS. Este, por sua vez, é um dos vocabulários mais utilizados na área da saúde em nível mundial (ANDALIA; CHAPMAN, 2011).

[2] Disponível em: http://decs.bvs.br/

Para ter acesso ao *corpus* do trabalho fez-se uma busca na WoS, em modalidade de pesquisa básica, utilizando os vinte descritores DeCS com a seguinte expressão de busca: 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")*. O refinamento da busca foi feito por DT=Tipo de documento (artigo), WC=Categoria WoS (*Public Environmental Occupational Health*) e LA=Linguagem (inglês).

Foram coletados os seguintes metadados dos artigos recuperados: Título (TI), Resumo (AB), Ano de publicação (PY), palavras-chave de autor (DE), e *keywords plus* (ID). Utilizou-se a ferramenta *VOSviewer* para gerar gráficos, no intuito de ilustrar o comportamento do conjunto de termos nos clusters de coocorrência com os campos das palavras-chave e *keyword plus*.

Foram estabelecidos critérios de avaliação da indexação para serem aplicados ao objeto investigado, isto é, palavras-chave de autor (DE) e *keywords plus* (ID). Desse modo, 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).

Nessas condições, com base na literatura, elencaram-se os seguintes 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. É perceptível que grande parte dos critérios encontrados na literatura se concentra na recuperação da informação. Desse modo, para o contexto da VI foi necessário construir os seguintes:

a) **Controle de abrangência dos termos:** trata-se do controle do uso de termos abrangentes, que são aqueles que não conseguem expressar claramente um assunto.

b) **Uniformidade:** diz respeito ao controle de plural e singular, sinonímia, homonímia e outras variações de grafia do termo.

4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

A análise está descrita em duas etapas: uma centrada na indexação conforme os critérios estabelecidos, e a outra baseada nos descritores da área de Nutrição identificados no DeCS.

4.1 PRIMEIRA ANÁLISE: APLICAÇÃO DOS CRITÉRIOS ESTABELECIDOS

Partindo do pressuposto de que a VI resulta, a priori, de uma boa indexação, esta seção tem por objetivo aplicar os critérios de avaliação da indexação ao *corpus* investigado, assim foram realizadas as seguintes análises: **A partir da exaustividade; A partir da especificidade; A partir do controle da abrangência dos termos; A partir da uniformidade; e, A partir da consistência (coerência).**

a) Análise a partir da Exaustividade

Sob a ótica da VI, a exaustividade torna-se um aspecto negativo quando ocasiona a dispersão de termos ao redor dos mapas gerados a partir da coocorrência de palavras, desfavorecendo a visualização de domínios do conhecimento. A VI gerada a partir da coocorrência 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.

Com base no *corpus* analisado, identificou-se que a representação da informação na WoS é predominantemente exaustiva. Isso significa que tanto palavras-chave de autor (DE) quanto *keywords plus* (ID) aparecem em grande quantidade (Tabela 1).

Tabela 1 – Total 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 termos/artigo	**Variação 4/5/6**	**10**

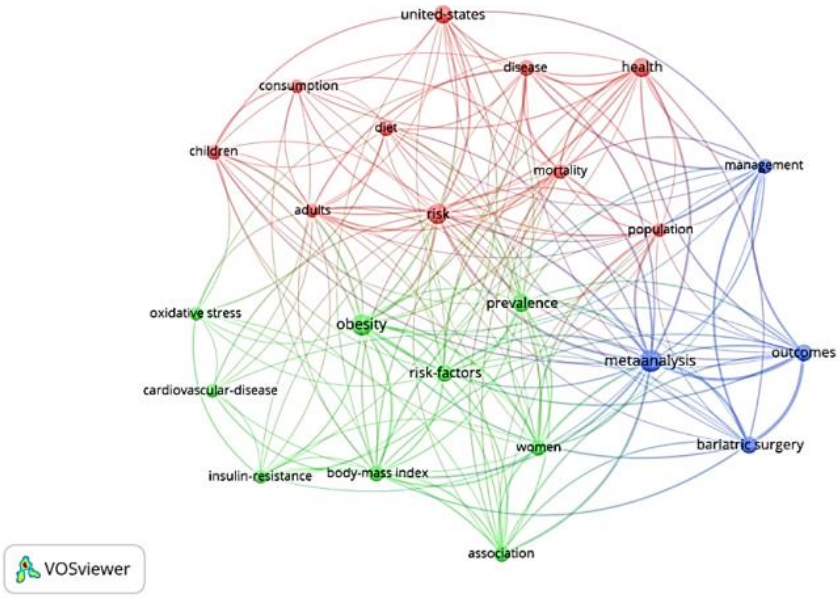
Fonte: Dados da pesquisa (2017).

As palavras-chave de autor alcançaram uma média de cinco termos por artigo, enquanto as *keywords plus* uma média de dez termos. Isso evidencia o critério de exaustividade aplicado a estes documentos. Percebeu-se ainda o quanto a exaustividade pode interferir na VI: de 253 termos (soma de palavras-chave de autor e *keywords plus*) apenas 25 (9,88% do total) possuem, no mínimo, duas ocorrências. Essa constatação remete ao argumento de Lancaster (2004) sobre a relação entre exaustividade e consistência da indexação: quanto maior a primeira, menor a segunda. Os dados do estudo apontam que quanto maior a exaustividade, menor o número de termos coincidentes entre si.

Ressalta-se que essa inconsistência em razão da exaustividade independe da quantidade de termos do *corpus*. Para tal constatação o *corpus* foi aumentado para um conjunto de 500 artigos, resultando em 4.257 termos, sendo 1.907 palavras-chave e 2.620 *keywords plus*. Em seguida foi gerado um mapa a partir dos termos com no mínimo 13 ocorrências (Figura 1). Assim, para ter um

mapa que represente melhor as relações entre os termos, necessita-se contemplar aqueles mais representativos do conjunto, neste caso, a melhor visualização pautou-se em 0,87% do total de termos.

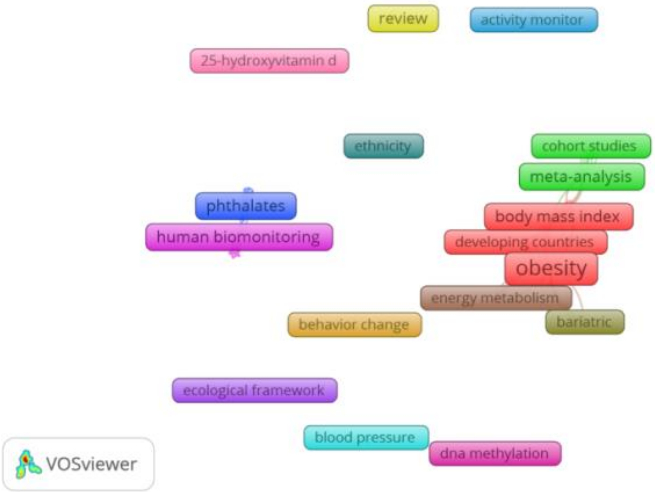
Figura 1 – Mapa de termos (Keyword Plus) com o mínimo 13 ocorrências



Fonte: Dados da pesquisa (2017).

Por outro lado, verificou-se o comportamento daquelas palavras-chave menos recorrentes; para tal utilizou-se o *corpus* inicial de 20 artigos. Nota-se (Figura 2) ampla dispersão e a quase imperceptível ligação entre as palavras-chave.

Figura 2 – Cluster das palavras-chave de autor com o mínimo de uma ocorrência



Fonte: Dados da pesquisa (2017).

Se por um lado a exaustividade favorece o processo de Recuperação da Informação, por outro acarreta a dispersão nos mapas de coocorrência de palavras, desfavorecendo, portanto, a VI.

b) Análise a partir da Especificidade

Esse critério verificou se a representação da informação na WoS privilegia mais termos abrangentes ou específicos. Trata de observar se as palavras de autor (DE) e *keywords plus* (ID) contemplam esse aspecto e, também, como estas reagem diante da VI, tendo como parâmetro o critério de especificidade. Segundo Lancaster (2004, p. 34), a especificidade é um dos princípios da indexação ao qual "um tópico deve ser indexado sob o termo mais específico que o abranja completamente".

É importante considerar que pode haver 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 a eleição de termos indevidamente específicos. Isso significa que, nem sempre, a eleição de vários termos específicos refletirá numa especificidade adequada.

Diante dos termos analisados, observou-se que a maioria é específico, contudo, há alguns 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. Em estudo similar, Zhang *et al*. (2016) analisaram a estrutura do conhecimento na área de adesão ao paciente.

No que concerne ao corpus dessa pesquisa, foram encontrados termos de especificidade elevada nos dois grupos de palavras – ID e DE. A especificidade elevada diz respeito aos termos específicos de pouca utilidade em um domínio. Ou seja, são palavras que provocam inconsistência e pulverização nos mapas de VI.

A partir dessa análise, percebeu-se que a especificidade das palavras-chave no *corpus* estudado é, em algum momento, consistente e em outro causa a dispersão de termos.

c) Análise a partir do Controle da Abrangência dos Termos

Este critério analisou a relação entre termos abrangentes e a VI. Consideram-se termos abrangentes aqueles que não são claros quanto ao que representam, como termos ambíguos, que causam dúvida ou que não apresentam a obviedade quanto ao pertencimento a um domínio. Na área da Nutrição, podem ser considerados termos óbvios aqueles utilizados na estratégia de busca e que, consequentemente, constam nos documentos recuperados, tais como "*nutrition*" e "*obesity*".

O inverso de termos óbvios são os amplos, abrangentes e às vezes irrelevantes. A abrangência dos termos 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, apresenta algum significado. O segundo nível de abrangência é aquele considerado elevado. Isto é, um termo indexado que, além de não ser óbvio o seu pertencimento a um domínio, não carrega significado relevante.

Dos artigos analisados, foram identificados termos considerados generalistas, tais como as DE com duas ocorrências ou menos, totalizando 24 palavras-chave generalistas. Quanto às ID, estas apresentaram uma variação maior de ocorrências (uma, duas e três) como pode ser observado na Tabela 2. O total de ID generalistas é 48.

Tabela 2 – *Keyword Plus* (ID) abrangentes

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

Fonte: Dados da pesquisa (2017).

Observados os dois tipos de termos de abrangência elevada (ID e DE),

nota-se que estes possuem ocorrência muito inferior. Do ponto de vista da VI, termos dessa natureza não são representativos em mapas de coocorrência devido a sua dispersão nesses mapas.

Para aferir melhor esses termos genéricos, fez-se buscas na WoS utilizando-os. Chamou atenção nos resultados a posição desses termos, seja no título, no resumo ou nas palavras-chave do próprio autor. Quando um termo consta simultaneamente nessas três instâncias ou ao menos em duas, o mesmo passa a ser considerado um descritor. Com relação à abrangência do termo, adotou-se como critério de análise sua posição nas palavras-chave de autor. Esse critério remete à garantia literária, que considera importante os termos adotados pelos especialistas do domínio.

A análise se procedeu da seguinte forma: foram selecionados cinco termos generalistas da Tabela 2 de *keywords plus* (***height, income, increasing fruit, indicators, in-house dust***), os quais foram pesquisados isoladamente na WoS. Do total recuperado, selecionaram-se 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.

Os termos apareceram com maior frequência nos resumos dos documentos do que em qualquer outro campo. Só o termo "i*ndicators*" 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 termos generalistas e que não devem ser utilizadas como descritores, pois quase não apareceram no campo de palavras-chave de autor, isto é, não fazem parte da linguagem dos especialistas da área.

d) Análise a partir da Uniformidade

Esta análise verificou a 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 buscando perceber se há uniformidade entre ambas. Esse tipo de verificação tem como objetivo compreender o grau de variação de grafia que

existe nos termos indexados.

Um problema comum devido à ausência de controle/padronização é a variância dos termos indexados. Para a recuperação da informação, diferenças de número (plural ou singular) e grafias diferentes do termo, já não representam um problema comprometedor, visto que a maioria dos SRIs possui soluções automáticas para tal. Contudo, sob a ótica da VI esse ainda é um problema que requer considerável esforço operacional para corrigir as inconsistências. Um exemplo é quando uma palavra aparece duas ou mais vezes com uma pequena diferença na grafia, ainda sim, será contabilizada tantas vezes aparecer. No corpus analisado, os termos a seguir (Tabela 3) apresentaram esse comportamento:

Tabela 3 – Termos com variações de grafia

Termos	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).

As normas de indexação da informação orientam para que as palavras-chave sejam escritas na sua forma mais natural (evitar uso de plural, hífen entre outros caracteres desnecessários) nota-se, na tabela 3, que a maioria das DE são escritas mais adequadamente em comparação as ID.

Sob este aspecto, 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.

e) Análise a partir da Consistência (Coerência)

Esse critério analisou o grau de consistência da representação temática dos artigos da WoS a partir dos termos com maior ocorrência no conjunto de DE e no conjunto de ID (Tabela 4). Buscou-se, nos dois grupos, identificar qual possui o maior índice de consistência, para tal, identificou-se o total de DE e ID, separadamente. Em seguida, calculou-se a proporção entre o número de descritores comuns e o número total de descritores.

Esse tipo de contagem tem por base a taxa de coerência mencionada por Slype (1991), na qual 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 (ARAÚJO JÚNIOR, 2007).

Para se adequar ao propósito desta análise, fez-se uma adaptação na taxa de coerência de Slype. O autor considera o índice de coerência entre os termos 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).

Tabela 4 – 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).

As *keywords plus* (ID) apresentaram maior índice de ocorrência em comparação às palavras-chave de autor (DE), o que implica 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 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%.

Quando a consistência dos termos é baixa, a elaboração de um mapa que represente um domínio torna-se inviável, comprometendo assim a VI. Isso ocorre porque esse tipo de visualização considera os termos de maior frequência.

4.2 Segunda Análise: Aplicação do Vocabulário DeCS

Esta etapa identificou os pontos de convergência entre o DeCS, as palavras-chave (DE) e as *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, aqui utilizado, busca identificar se a representação da informação do corpus selecionado na WoS corresponde à linguagem utilizada pelos profissionais da área de Nutrição do DeCS.

Desse modo, esta análise seguiu as seguintes etapas: a) verificar, separadamente, se os termos de maior ocorrência no grupo de DE e ID ocorrem no DeCS; b) verificar se os termos de especificidade elevada ocorrem ou não no DeCS, e; c) verificar, a partir do total de termos DE e do total de termos ID, quantos termos correspondem ao DeCS.

Essa verificação foi realizada de forma semiautomática utilizando o *software Microsoft Excel*. Primeiro foram listados os termos de maior ocorrência no conjunto de ID e DE, posteriormente foram pesquisados na planilha dos 153 descritores do DeCS.

Comparando o total de termos do corpus (253 sem repetição) com o DeCS, poucos termos coincidiram, revelando divergência entre eles.

Os termos da WoS que coincidiram no DeCS foram aqueles utilizados na expressão de busca. No entanto, estranhamente, nem todas as palavras da expressão de busca fazem parte do conjunto de termos da WoS.

Infere-se que não há 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 presente no DeCS são limitados por compor um conjunto exíguo de termos para representar um domínio. Além disso, este se restringe à Nutrição em Saúde Pública, logo, esta análise tem suas limitações e, portanto, deve-se ponderar tal fato, visto que há um vocabulário diversificado na indexação dos documentos da WoS.

Diante dos resultados, destaca-se:

a) A partir do **critério de exaustividade**, 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 dispersão de termos.

b) Sobre a **especificidade**, percebeu-se que tanto as DE quanto as ID contemplam esse aspecto, apesar de serem observados termos exaustivamente abrangentes. Nesse sentido, a exaustividade, somada à especificidade, não favoreceu a VI devido à dispersão dos termos.

c) Acerca do **critério de controle de abrangência dos termos**, estes de abrangência elevada apresentaram pouca ocorrência. Sendo assim, do ponto de vista da VI, termos muito abrangentes não são favoráveis para este propósito.

d) Com relação ao **critério de uniformidade**, há situações em que a mesma palavra foi escrita de forma variada, seja por uso indevido do hífen ou por uso de plural desnecessário.

e) Quanto à **consistência** da indexação, verificou-se que o índice de

consistência nos vinte artigos foi relativamente baixo (7,61% nas DE e 10,64% nas ID), levando em consideração a taxa de variação mencionada por Slype (1991) que deve ser entre 50 a 80%.

f) Já com relação ao **vocabulário DeCS**, verificou-se que não existe uma linguagem comum entre o DeCS e o conjunto de descritores dos artigos analisados.

5 CONSIDERAÇÕES FINAIS

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 pesquisas dedicadas a avaliar a representação da informação sob a ótica da visualização da informação, e uma das contribuições desta pesquisa foi acrescentar novas discussões para este contexto. A pesquisa também contribuiu com a reunião e elaboração de critérios para avaliar a qualidade da indexação.

Espera-se que este estudo possa ser ampliado para um *corpus* maior, de modo a fornecer uma visão macro sobre a dinâmica da representação da informação na WoS. Além disso, é possível estender a pesquisa para outros domínios, de modo a compreender a dinâmica da representação da informação sob à ótica da VI em diferentes áreas e, finalmente, o estudo pode se estender para outras bases de dados.

REFERÊNCIAS

ABRIL, G. Prólogo. *In*: GARCÍA GUTIÉRREZ, A. **Otra memoria es posible:** estrategias descolonizadoras del archivo mundial. Buenos Aires: La Crujía, 2004.

AGUILAR, A. G.; PINTO, A.; SEMELER, A.; SOARES, A. P. **Visualização de dados, informação e conhecimento**. Florianópolis: Ed. UFSC, 2017.

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

https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2003v8n15p18/5233. Acesso em: 30 set. 2016.

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 ACIMED**, La Habana, v. 22, n. 2, p. 142-154, 2011. Disponível em: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1024-94352011000200005. Acesso em: 30 set. 2016.

ANIZIO, J. L. A.; NASCIMENTO, G. F. C. L. Avaliação do processo de indexação na biblioteca da assessoria jurídica do banco do brasil. **Biblionline**, n. esp., 2012. Disponível em: http://hdl.handle.net/20.500.11959/brapci/100286. Acesso em: 18 set. 2017.

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

BRASCHER, 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. Disponível em: http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/809.

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. Disponível em: https://periodicos.ufpb.br/index.php/itec/article/view/33926. Acesso em: 15 nov. 2017.

BUFREM, L. S.; FREITAS, J. L. Interdomínios na literatura periódica científica da ciência da informação. **DataGramaZero**, [*S.l.*], v. 16, n. 6, 2015. Disponível em: http://hdl.handle.net/20.500.11959/brapci/8295. Acesso em: 30 set. 2016.

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. Disponível em: https://brapci.inf.br/index.php/article/download/19202.

CHEN, C. **Information visualization**: beyond the horizon. [*S.l.*]: Springer, 2006.

DAHLBERG, I. Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n. 2, p. 102 – 107, 1978. Disponível em: http://revista.ibict.br/ciinf/article/view/115/115. Acesso: 24 ago. 2017.

DeCS. **Descritores em Ciências da Saúde**, [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. 2017.

DIAS, E. W.; NAVES, M. L. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

FUJITA, M. S. L.; GIL-LEIVA, I. Avaliação da indexação por meio da recuperação da informação. **Ciência da Informação**, Brasília, v. 41, n. 1, p. 50-66, jan./abr. 2014. Disponível em: http://revista.ibict.br/ciinf/article/view/1418.

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

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

HJORLAND, B. Concept theory. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1519–1536, 2009. DOI: https://doi.org/10.1002/asi.21082.

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

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos/Livros, 1993.

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.

MIGUÉS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório científico 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.

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

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

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, 2000. 255 p.

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

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, R. F. 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. DOI: 10.5007/15182924.2011v16n32p73

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. Disponível em: https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-160/knowledge-organization-systems-kos-jahrgang-35-2008-heft-2-3. Acesso em: 10 de maio de 2017.

ZHANG, J.; YU, Q.; ZHENG, F.; LONG, C.; LU, Z.; DUAN, Z. Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. **Journal of the Association for Information Science and Technology**, [*S.l.*], v. 67, n. 4, p. 967–972, 2016. DOI: https://doi.org/10.1002/asi.23437. Acesso em: 15 abr. de 2017.

ANALYSIS OF THE INFORMATION REPRESENTATION IN THE WEB OF SCIENCE: A STUDY OF NUTRITION FIELD

ABSTRACT

Introduction: The motivation of the study is to understand whether the keywords and keywords plus of articles in the Nutrition area, indexed on the Web of Science, receive appropriate thematic treatment that favors Information Visualization. **Objective:** to evaluate the quality of information representation in the Web of Science (WoS) database by analyzing the fields author's keywords and keywords plus of articles in the field of Nutrition in Public Health, seeking to observe the use of these terms in the context of information display. **Methodology:** This is an exploratory study, which seeks to

understand the dynamics of information representation in the Nutrition domain in the WoS database. The corpus consists of a set of twenty articles with a high citation index, published in the period from 2006 to 2016. The analysis took place in two stages: in the first, it applied the criteria for evaluating the indexation to the set of terms of that base and, in the second, he used the DeCS vocabulary (Health Sciences Descriptors) as a parameter to find points of intersection between this vocabulary and the representation of information in WoS. Such analysis had as its central point the literary guarantee. **Results:** both the author's keywords and the keywords plus of the analyzed corpus were able to meet indexation criteria, however aspects of the thematic representation that compromise the visualization of the information were observed. **Conclusions:** As for the literary guarantee, the result showed that there is no common language between DeCS and WoS terms.

Descriptors: Information representation. Information visualization. Subject indexing. Public Health Nutrition. Web of Science.

REPRESENTACIÓN DE INFORMACIÓN EN LA WEB OF SCIENCE: UN ESTUDIO DEL DOMINIO DE LA NUTRICIÓN

RESUMEN
Introducción: La motivación del estudio es comprender si las palabras clave y las *keyword plus* de los artículos en el área de Nutrición, indexados en la *Web of Science*, reciben el tratamiento temático apropiado que favorece la visualización de datos. **Objetivo:** evaluar la calidad de la representación de la información en la base de datos de *Web of Science* (WoS) mediante el análisis de las palabras clave del autor y las *keywords plus* de los artículos en el campo de la Nutrición en Salud Pública, mirando el uso de estos términos en el contexto de visualización de datos. **Metodología:** Este es un estudio exploratorio, que busca comprender la dinámica de la representación de la información en el dominio de Nutrición en la base de datos de WoS. El *corpus* consiste en un conjunto de veinte artículos con un alto índice de citas, publicado en el período de 2006 a 2016. El análisis se realizó en dos etapas: en la primera, se aplicó los criterios para evaluar la indexación al conjunto de términos de esa base y, en el segundo, se utilizó el vocabulario DeCS (Descriptores de Ciencias de la Salud) como parámetro para encontrar puntos de intersección entre este vocabulario y la representación de información en WoS. Tal análisis tenía como punto central la garantía literaria. **Resultados:** tanto las palabras clave del autor como las *keyword plus* del *corpus* analizado pudieron cumplir con los criterios de indexación, sin embargo, se observaron aspectos de la representación temática que comprometen la visualización de datos. **Conclusiones:** En cuanto a la garantía literaria, el resultado mostró que no hay un lenguaje común entre los términos DeCS y WoS.

Descriptores: Representación de la información. Visualización de Datos. Indexación de asunto. Nutrición en salud pública. *Web of Science*.

Recebido em: 16.04.2020
Aceito em: 20.11.2020