

UMA ABORDAGEM À LINGUAGEM DE INDEXAÇÃO DOS ARTIGOS CIENTÍFICOS DEPOSITADOS NO REPOSITÓRIO CIENTÍFICO DA UNIVERSIDADE DE COIMBRA

Resumo - Este estudo tem como objetivo geral analisar as palavras-chave usadas pelos autores da Universidade de Coimbra nos artigos da área das Ciências da Saúde, comparando-as com os termos da lista de encabeçamentos de matérias Medical Subject Headings (MeSH), a que recorre a National Library of Medicine, na análise de conteúdo desses artigos. O corpus é constituído por 207 documentos, depositados no repositório científico da Universidade de Coimbra – Estudo Geral – e que foram publicados em revistas internacionais e indexados na base de dados MEDLINE. Apresenta como objetivos específicos: (i) identificar e aferir o número de palavras-chave colocadas na versão depositada no Estudo Geral e na versão publicada pela editora; e o número de descritores atribuídos pela MEDLINE; (ii) verificar as palavras-chave, usadas pelos autores, que coincidem ou se relacionam com os termos do MeSH. Os resultados obtidos indicam que o número de palavras-chave registadas no Estudo Geral inclui apenas cerca de metade de todas as palavras-chave registadas nas versões publicadas pelas editoras e é significativamente inferior relativamente aos termos empregues pela MEDLINE. Das palavras-chave identificadas, cerca de dois terços coincidem, ou estabelecem um tipo de relação associativa com os descritores aplicados do MeSH.

Palavras-chave: Recuperação da informação; Linguagens de indexação; Palavras-chave; Artigos científicos; Medical Subject Headings

AN APPROACH TO INDEXING LANGUAGE OF JOURNAL ARTICLES DEPOSITED IN THE UNIVERSITY OF COIMBRA SCIENTIFIC REPOSITORY

Abstract - In this article, the main goal is to analyze the keywords given by authors from articles in Health Science and related fields, available in *Estudo Geral*, the University of Coimbra institutional repository, and compare them with the Medical Subject Headings (MeSH), the controlled vocabulary published by the National Library of Medicine. Two-hundred and seven articles deposited in *Estudo Geral*, published in scientific international journals and indexed in MEDLINE database, were considered. A specific objective is: (i) to identify and count the number of keywords in the version deposited in the institutional repository, in the published version, and the number of controlled vocabulary of MeSH; (ii) to verify the number of keywords that are identical or related to MeSH terms. The results indicate that the number of keywords recorded in *Estudo Geral* includes only about half of all keywords in the published version recorded by publishers, and is significantly lower than the terms used by MEDLINE. About two-thirds of those keywords agree or establish some kind of associative relationship with MeSH terms applied.

Keywords: Information retrieval; Indexing language; Keywords; Scientific articles; Medical Subject Headings

Ana Miguéis

Coordenadora do Serviço Integrado
das Bibliotecas da Universidade de
Coimbra
amigueis@sib.uc.pt

Bruno Neves

Investigador do Serviço Integrado das
Bibliotecas da Universidade de
Coimbra
bneves@sib.uc.pt

1. INTRODUÇÃO

O modo como a comunicação do conhecimento científico se organiza e representa levou ao aparecimento de instrumentos de acesso e recuperação de informação. Os diversos modelos de representação e descoberta coexistem num ambiente competitivo que se caracteriza por uma dinâmica permanente e em constante mutação. A organização da informação e a sua representação são preocupações centrais da ciência da informação, que tem como uma das suas finalidades a partilha e comunicação da informação, considerada um recurso estratégico e significativo na cadeia que gera o conhecimento (Tonello, Lunardelli e Almeida Júnior, 2012).

Um desses instrumentos é a indexação por assuntos, que se entende como “uma operação documental que consiste na análise do conteúdo de um documento com o propósito de o representar através de uma linguagem documental e cujo objetivo final é a sua recuperação” (Simões, 2008, 235). A análise dos documentos e da informação que veiculam, faz parte do processo de indexação por assuntos que, através de princípios orientadores e de normas¹ de apoio aos indexadores, lhes permite a objetividade necessária na análise dos documentos, e na determinação, com precisão, da opção mais correta para a representação dos conceitos, tendo bem presente o contexto em que se observa, de modo a garantir a sua coerência e a conferir uniformidade aos instrumentos adotados (Mendes e Simões, 2002). Este papel de mediador da informação, na análise documental, permite ao profissional assumir uma atitude ativa junto de utilizadores mais exigentes e com necessidades mais complexas, apreendendo a realidade informacional e social da atualidade e participando nos processos de disponibilização da informação (Tonello, Lunardelli e Almeida Júnior, 2012).

A representação do conteúdo dos documentos pode realizar-se através de conjuntos de termos, extraídos diretamente da linguagem natural, e que permitem a recuperação da informação, traduzindo-se através de diferentes tipos de linguagens. Uma delas é a linguagem livre, que se designa deste modo por não apresentar qualquer controlo dos termos, quer formal, quer semântico e cujos pontos de acesso são as palavras ou expressões utilizados pelos próprios autores no título, resumo e texto integral dos documentos. Esta linguagem afirma-

¹ As normas, como a NP 3715 (1989) e a NP 4036 (1992) centram-se, a primeira, nos métodos para análise de documentos e determinação do seu conteúdo; a segunda na construção de tesouros monolíngues

se e consolida-se em áreas específicas e restritas, que particularizam e identificam um código de comunicação num determinado campo científico (Simões, 2008 235), e que se materializam em listas de palavras-chave. Estas listas, constituídas pelos termos vocabulares, são retiradas dos documentos e correspondem, habitualmente, a uma escolha dos autores.

1.1 As palavras-chave

As palavras-chave podem ser definidas como sendo “termos significativos de um documento, que representam os conceitos que caracterizam, de forma inequívoca, o conteúdo de um documento, geralmente extraído do título, resumo ou do próprio texto” (Simões, 2008, 237). Diferenciam-se dos termos de indexação, propriamente ditos, que resultam dos processos de análise conceptual e sua tradução num vocabulário controlado. A atribuição das palavras-chave é livre, podendo ser definida pelos autores dos documentos, pelos editores ou, até, ser gerada de forma automática.

No âmbito dos artigos científicos, as palavras-chave expõem a abrangência de um assunto e os seus conceitos principais, úteis para a indexação em mecanismos de pesquisa ou categorização do texto (Ercan e Cicekli apud Gonçalves, 2008). Os estudos sobre o uso das palavras-chave, pela falta de padronização que existe nesta terminologia, assumem um papel diferente dos estudos sobre indexação e terminologia, embora ambos abordem as linguagens vocabulares (Gonçalves, 2008).

No processo de submissão de artigos, os editores de revistas científicas solicitam usualmente aos autores que coloquem as palavras-chave nos seus textos; mais raramente, são as próprias editoras que, recorrendo a profissionais, introduzem esses termos, mais adequados para a representação dos conceitos (Gonçalves, 2008). O uso das palavras-chave tem algumas vantagens, como potenciar o acesso ao conteúdo dos documentos, para além da informação que é representada pelo título e resumo; de traduzir o pensamento dos autores; e de manter o contacto com a realidade da prática quotidiana, acompanhando a evolução científica e tecnológica refletida pelos documentos.

A investigação sobre a importância e características das palavras-chave tem incidido sobre vários aspetos (Gil-Leiva e Alonso-Arroyo, 2007; Gonçalves, 2008), como seja o da eficiência na recuperação da informação; o uso para a extração automática, a partir de diferentes metodologias e algoritmos; o uso por parte dos autores e editores; mais recentemente, a sua utilização nos comportamentos de etiquetagem (metatags); e a comparação com os títulos, resumos e textos integrais, o que confirma do interesse que despertam. No entanto, na revisão da literatura, identificaram-se poucos trabalhos que abordem a comparação das palavras-chave fornecidas por autores de artigos científicos com os descritores empregues pelos indexadores e o contributo que podem dar para a escolha dos termos de indexação. Entre eles, encontram-se dois artigos de Gil-Leiva e Alonso-Arroyo (2005 e 2007) que analisaram as palavras-chave dos autores e as compararam com os descritores usados em várias bases de dados, concluindo que existe uma relação direta, ou indireta, na seleção dos descritores.

1.2 O Estudo Geral

Foi com base neste pressuposto, o de conhecer a relação conceptual entre as palavras-chave e os descritores empregues, que se realizou este trabalho, analisando, comparativamente, as palavras-chave empregues pelos autores nos seus artigos, e os termos da linguagem de indexação que lhe são associados. Para tal, definiu-se como ponto de partida, o repositório científico da Universidade de Coimbra, designado Estudo Geral², seleccionando um conjunto de artigos depositados, da área das Ciências da Saúde,

O Estudo Geral tem a missão de divulgar e preservar a produção científica de autores ligados à Universidade de Coimbra, afirmando-se como um instrumento que preserva e maximiza o impacto da investigação realizada no seio da instituição, com a possibilidade de introduzir novos modelos de investigação e de ensino. Os repositórios institucionais surgiram no âmbito do movimento de Acesso Aberto à literatura científica (*Open Access*) – ao qual o

² O Estudo Geral foi criado, em 2008, para preservar e divulgar, de uma forma sistemática, a produção académica e científica da Universidade de Coimbra, no âmbito do movimento de Acesso Aberto à literatura científica. Está disponível na WWW: <https://estudogeral.sib.uc.pt/>

Conselho de Reitores das Universidades Portuguesas aderiu em 2006 – como uma forma de reação às barreiras existentes na obtenção do acesso à informação científica e na divulgação da investigação produzida, fundamental para os investigadores poderem prosseguir nos seus trabalhos de investigação (Borges, 2009).

A organização do Estudo Geral obedece à estrutura orgânica da Universidade de Coimbra, em que, num primeiro nível, as comunidades representam as diversas Faculdades, que por sua vez se dividem em subcomunidades correspondentes a unidades orgânicas menores (os departamentos ou centros de investigação). A cada uma das comunidades ou subcomunidades estão associadas as coleções, que refletem a tipologia documental. Os “Artigos em Revistas Internacionais” representam uma dessas tipologias que assume particular relevância, pelo facto de os resultados da investigação científica realizada pelos seus autores, serem divulgados através deste modelo de publicação, estando esta prática fortemente implementada e apoiada por critérios internacionalmente dominantes de avaliação. O acesso a estes artigos científicos, armazenados em repositórios de acesso aberto, faz-se a partir dos metadados e do conteúdo de texto integral, aplicando técnicas que representam e expõem as relações semânticas entre documentos, em que os dados são relacionados.

Para este estudo, a escolha do conjunto de documentos a analisar incidiu na coleção “Artigos em Revistas Internacionais” da comunidade Faculdade de Medicina, que se justifica, como se esclarecerá no ponto seguinte, pela disponibilidade de informação em volume significativo simultaneamente em três níveis de análise: o do autor (no Estudo Geral), do editor (na revista científica) e no indexador (na base de dados MEDLINE).

2. METODOLOGIA

Definiu-se, como ponto de partida, o repositório científico da Universidade de Coimbra, o Estudo Geral, selecionando um conjunto de artigos depositados, da área das Ciências da Saúde, identificando todas as palavras-chave introduzidas pelos respetivos autores e comparando esses termos com os descritores de indexação atribuídos pela National Library of Medicine (NLM) dos Estados Unidos aos artigos da base de dados MEDLINE³.

Após a revisão bibliográfica exploratória na literatura especializada, delimitou-se o corpus, constituído por 207 artigos científicos publicados em revistas internacionais, da comunidade da Faculdade de Medicina, depositados no Estudo Geral, que compreendia, no seu conjunto, um total de 335 artigos científicos, editados entre 1990 e 2012, e depositados até final de dezembro de 2012.

Para o cumprimento dos objetivos a atingir, optou-se por uma metodologia de natureza exploratório-descritiva, circunscrita a um estudo de campo, completada por uma análise qualitativa.

Definiram-se algumas condições prévias, no início da recolha da informação, e que consistiram em incluir, apenas, artigos com acesso ao texto integral, com palavras-chave em língua inglesa e editados em publicações periódicas referenciadas na base de dados MEDLINE.

Foi pela importância atribuída à MEDLINE que a escolha recaiu sobre esta base de dados. A MEDLINE, como uma das principais bases de dados bibliográficos de artigos científicos da área das Ciências da Saúde, com particular destaque no campo da Biomedicina, é pesquisável através da PubMed⁴, plataforma Web que permite o acesso a vários recursos de informação biomédica e ciências da vida, e inclui hiperligações para o conteúdo do texto integral e para a página das editoras. O Medical Subject Headings⁵ (MeSH), vocabulário controlado criado pela NLM para a indexação dos artigos da MEDLINE – característica que

³ A MEDLINE é a principal base de dados bibliográficos de artigos científicos da área das Ciências da Saúde, com cerca de 19 milhões de referências bibliográficas, publicadas desde 1948. Disponível na WWW: <http://www.ncbi.nlm.nih.gov/pubmed>.

⁴ A PubMed é uma plataforma desenvolvida e mantida pelo Centro Nacional de Informações sobre Biotecnologia (NCBI), na Biblioteca Nacional de Medicina (NLM). Disponível na WWW: <http://pubmed.gov>.

⁵ O Medical Subject Headings compreende um conjunto de mais de 213.000 entradas que ajudam a encontrar o termo mais apropriado, tendo mais de 26.000 descritores. Além destes encabeçamentos, existem mais de 214.000 entradas, que se inscrevem num outro tesouro, relacionadas com conceitos complementares. Disponível na WWW: <http://www.nlm.nih.gov/mesh>.

distingue esta base de dados de outras afins – consiste numa lista de encabeçamentos de matérias, que traduzem conceitos com vários níveis de especificidade desta área do conhecimento e se estrutura de forma hierárquica.

No decurso deste levantamento, houve a perceção de que o número de palavras-chave recolhidas nos metadados dos artigos depositados no Estudo Geral eram em número reduzido e decidiu-se, então, fazer uma primeira comparação com as palavras-chave presentes na versão publicada pelo editor da revista, identificando-se, primeiro, os nomes das publicações periódicas a que pertenciam e confirmando-se, de seguida, o número de palavras-chave existentes em ambos os recursos. Num momento posterior e, reunidas todas as palavras-chave existentes para cada um dos artigos, compararam-se, então, com os com os descritores MeSH atribuídos aos títulos indexados na base de dados MEDLINE.

As palavras-chave recolhidas a partir dos metadados dos registos dos artigos depositados no repositório e da versão final da editora, foram organizadas, para cada artigo, numa grelha, juntamente com os termos eleitos pela MEDLINE. Para além destes elementos, foram adicionados campos de descrição, como, título, autor e ano de publicação. Na determinação quantitativa das relações identificadas entre as palavras-chave e os descritores MeSH, foram consideradas as relações de equivalência (idênticas), correspondendo à classificação de 1 ponto, e as relações associativas (relacionadas), correspondendo a meio ponto (0,5), conforme demonstra a Tabela 1.

Para a identificação das palavras-chave equivalentes aos descritores foram definidos os seguintes critérios: termos simples com a mesma grafia; termos com variantes em género ou número, termos compostos com a mesma grafia, ainda que com ordem invertida, e termos com o mesmo significado do ponto de vista semântico. De modo idêntico foram considerados os termos representados por acrónimos ou símbolos químicos, desdobrados, e os elementos textuais do alfabeto grego, transliterados para o alfabeto português.

Palavras-chave (PCH)	Descritores MeSH (DE)	Equivalentes (E)	Associadas (A)
Asthma	Asthma	1	
Prion	Prions	1	
Rheumatoid arthritis	Arthritis, rheumatoid	1	
ATP	Adenosine Triphosphate	1	
Amyloid- β peptide	Amyloide beta-peptides	1	
Kainate	Kainic acid	1	
Liver targeting	Liver		0,5
Aggregation	Platelet Aggregation		0,5
Biodistribution	Tissue Distribution		0,5
Glucocorticoid injection	Glucocorticoid		0,5
Elderly	Aged		0,5
Fatal forensic medicine	Forense medicine		0,5

Tabela 1 – Exemplo dos critérios definidos nas relações

Foi então criada uma segunda tabela (Tabela 2) com os resultados obtidos para todos os 207 artigos, registrando o número de relações (equivalências e associações) entre as palavras-chave usadas pelos autores e os descritores da MEDLINE e com a classificação atribuída a cada tipo de relação.

Artigo	PCH	DE	#PCH	#DE	#Usadas	#E	#A	Total (E+A)
Todo-Bom, 2012	Asthma	Adult	7	14	6	1	5	3,50
	Elderly	Aged						
	Allergy	Aged, 80 and over						
	T cell	Aging						
	CD45RA	Antigens, CD29						
	CD29	Antigens, CD45						
	CD95	Antigens, CD95						
		Apoptosis						
		Asthma						
		Female						
		Humans						
		Male						
		Middle Aged						
		T-Lymphocyte Subsets						

Ribeiro, 2012	Adipose Tissue	Adipose Tissue	7	10	4	3	1	3,50
	Cell Line	Aged						
	Cell Proliferation	Cell Proliferation						
	Cell tracking	Culture Media, Conditioned						
	Obesity	Humans						
	Periprostatic	Male						
	Prostate cancer	Matrix Metalloproteinase 2						
		Matrix Metalloproteinase 9						
		Obesity						
		Prostatic Neoplasms						
Moura-Ramos, 2012	Infertility	Adult	6	19	3	1	2	2
	Infertility stress	Cross-Sectional Studies						
	Emotional distress	Emotions						
	Indirect effects	Family Characteristics						
	Path analysis	Female						
	Couple analysis	Fertilization in Vitro						
		Humans						
		Infertility						
		Male						
		Parents						
		Portugal						
		Predictive Value of Tests						
		Questionnaires						
		Rural Health Services						
		Self Report						
		Social Class						
		Spouses						
		Stress, Psychological						
		Urban Health Services						

Tabela 2 – Dados dos artigos publicados em 2012

3. RESULTADOS E DISCUSSÃO

Dos 207 artigos selecionados, 76 não possuíam qualquer informação nos metadados referentes ao “assunto” no registo dos artigos depositados no Estudo Geral, embora essa informação constasse no ficheiro anexado, o que representa apenas 56% das palavras-chave dos autores no registo dos metadados. Tal significa que em termos quantitativos, e de acordo com os resultados obtidos, se identificaram 655 palavras-chave nos metadados do registo dos

artigos no Estudo Geral e 1174 na versão da editora. Aos mesmos artigos, indexados na MEDLINE, atribuíram-se 3061 descritores MeSH, conforme referido na Figura 1.

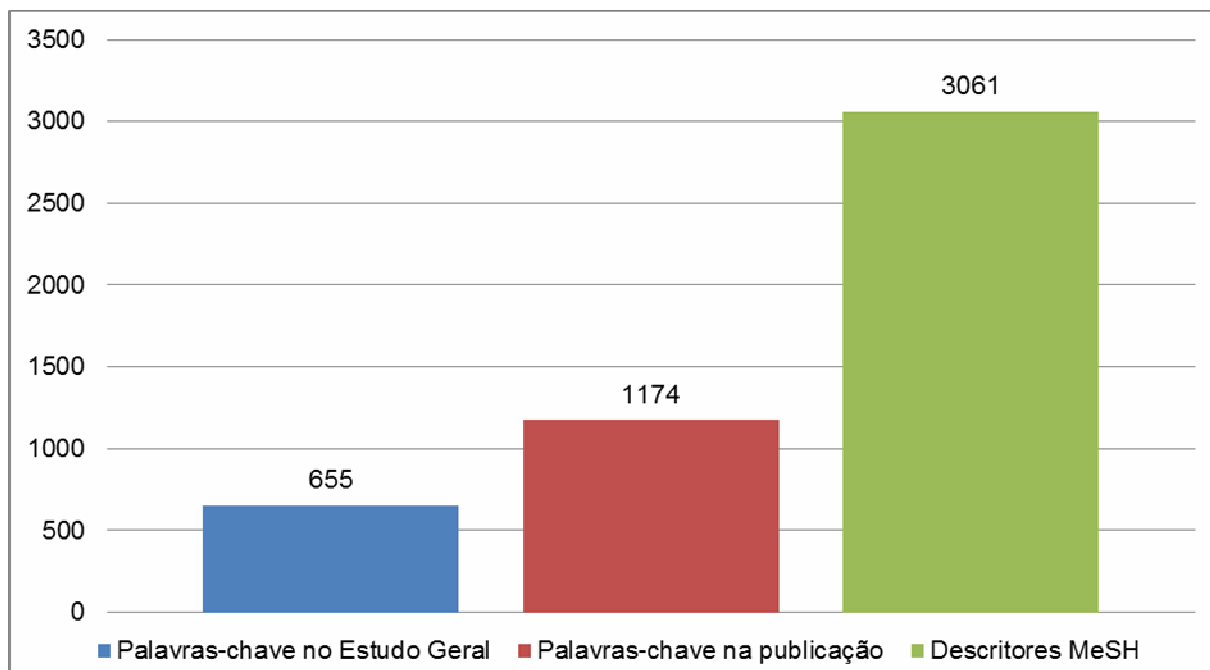


Figura 1 - Número de palavras-chave dos artigos e descritores da MEDLINE

Estes dados confirmam a percepção de que as palavras-chave recolhidas nos metadados dos artigos depositados no Estudo Geral, ainda que sendo as mesmas, são em número mais reduzido do que as presentes na versão publicada pela editora da revista. Este facto determina que a análise dos resultados seguintes incida apenas nas palavras-chave existentes na versão da editora, comparando-as com os descritores MeSH atribuídos aos títulos indexados na base de dados MEDLINE.

Assim, no que se refere à frequência do número de palavras-chave por artigo, a maioria dos artigos apresenta entre 4 a 6 palavras-chave, valores registados em 70% dos artigos, com uma média de 5,67 palavras-chave. Na atribuição dos descritores da MEDLINE, a frequência é mais dilatada, variando entre 11 e 20 descritores, o que corresponde a 66% dos artigos, com uma média de 14,79 descritores, e um pico de 39 descritores atribuídos a um único artigo, conforme a Tabela 3.

	PCH	Freq. Absoluta	Freq. Relativa (%)	DE	Freq. Absoluta	Freq. Relativa (%)
	2	2	1%	3	1	0%
	3	16	8%	4	0	0%
	4	37	18%	5	1	0%
	5	57	28%	6	3	1%
	6	50	24%	7	6	3%
	7	20	10%	8	6	3%
	8	9	4%	9	7	3%
	9	4	2%	10	11	5%
	10	7	3%	11	15	7%
	12	2	1%	12	27	13%
	14	2	1%	13	12	6%
	19	1	0%	14	17	8%
				15	14	7%
Total	1174	207	100%	16	17	8%
Média	5.67			17	13	6%
Moda	5			18	13	6%
				19	12	6%
				20	15	7%
				21	5	2%
				22	4	2%
				23	2	1%
				25	1	0%
				26	2	1%
				33	1	0%
				36	1	0%
				39	1	0%
			Total	3061	207	100%
			Média	14.79		
			Moda	12		

Tabela 3 – Frequência do número de palavras-chave na publicação e de descritores na MEDLINE

Do ponto de vista semântico, no que respeita às relações de equivalência e de associação entre as palavras-chave e os descritores da MEDLINE, foram identificadas 698 palavras-chave com algum tipo de relação: 345 equivalentes (29%) e 353 associadas (30%), como se indica na Figura 2. Regista-se, ainda, um conjunto de 476 palavras-chave (41%) que não apresentam qualquer tipo de relação.

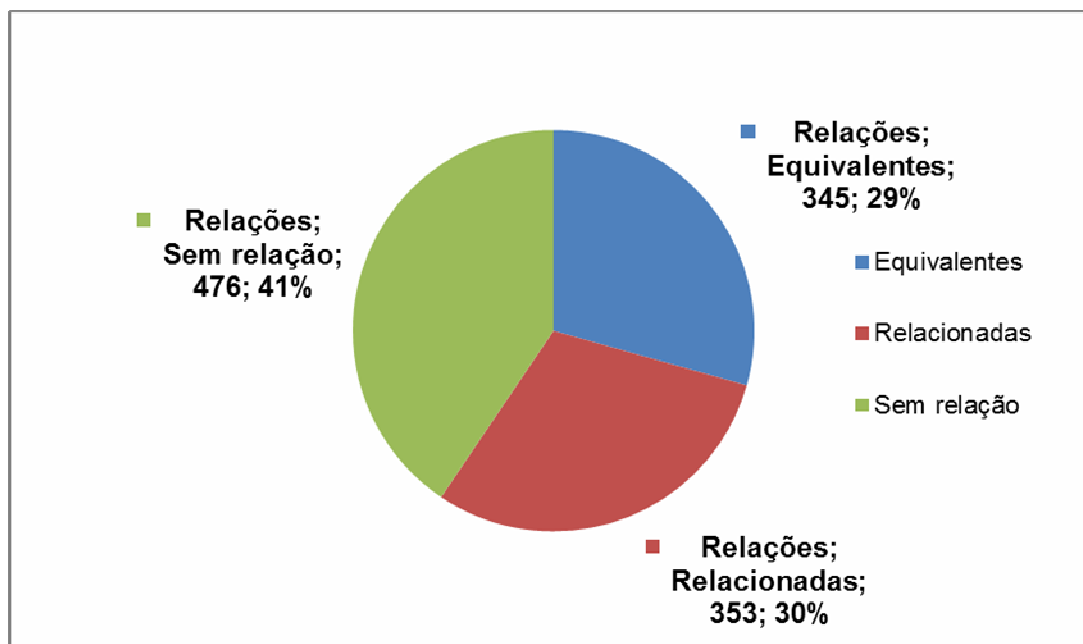


Figura 2 – Relações identificadas na comparação das palavras-chave com os descritores MeSH

Na tabela seguinte (Tabela 4) indica-se a distribuição dos 207 artigos selecionados para este estudo, por anos civis. O período temporal abrangido é de 20 anos, entre 1992 e 2012, mas contínuo apenas a partir de 1997, concentrando-se 145 dos artigos (70%) nos anos 2003 a 2008. Na classificação atribuída às relações estabelecidas entre as palavras-chave e os descritores da MEDLINE, tendo em conta a pontuação definida (1 ponto para as de equivalência e 0,5 para as associativas), as percentagens obtidas foram de 27% de equivalências e 17% de associações, num total de 44,49%.

Ano	#Artigos	#PCH	#Usadas	#DE	E	A	Total (E+A)	%E	%A	Total % (E+A)
1992	1	3	1	14	0	0,5	0,5	0%	17%	17%
1997	2	10	6	28	4	1	5	40%	10%	50%
1998	5	27	13	61	3	5	8	11%	19%	30%
1999	7	51	23	109	14	4,5	18,5	27%	9%	36%
2000	12	58	41	196	23	9	32	40%	16%	55%
2001	8	43	28	92	17	5,5	22,5	40%	13%	52%
2002	4	17	13	46	6	3,5	9,5	35%	21%	56%
2003	22	111	77	302	41	18	59	37%	16%	53%
2004	27	150	88	390	48	20	68	32%	13%	45%
2005	22	124	69	376	26	21,5	47,5	21%	17%	38%
2006	16	106	51	254	26	12,5	38,5	25%	12%	36%
2007	26	151	97	396	51	23	74	34%	15%	49%
2008	32	205	109	478	57	26	83	28%	13%	40%
2009	10	52	33	154	12	10,5	22,5	23%	20%	43%
2010	2	6	6	17	1	2,5	3,5	17%	42%	58%
2011	8	40	30	105	11	9,5	20,5	28%	24%	51%
2012	3	20	13	43	5	4	9	25%	20%	45%
Total	207	1174	698	3061	345	176,5	521,5			
Média								27,13	17,35	44,49

Tabela 4 – Dados gerais obtidos por ano de publicação dos artigos

Relativamente à frequência das relações estabelecidas pelas palavras-chave, que se apresenta na Tabela 5, esta situa-se, maioritariamente, entre 1 a 2 conceitos equivalentes (57%) e 1 a 2 conceitos associados (64%) por artigo, o que confirma a importância das palavras-chave escolhidas pelos autores para a representação da informação.

Equivalentes	Freq. Absoluta	Freq. Relativa (%)	Associadas	Freq. Absoluta	Freq. Relativa (%)
0	44	21%	0	31	15%
1	54	26%	1	60	29%
2	65	31%	2	73	35%
3	24	12%	3	30	14%
4	14	7%	4	8	4%
5	4	2%	5	5	2%
6	1	0%			
7	1	0%			

Total	345	207	100%	353	207	100%
Média	1.66			1.7		
Moda	2			2		

Tabela 5 – Frequência do número de palavras-chave equivalentes e associadas

4. CONCLUSÕES

Face aos resultados apurados, é possível afirmar que as palavras-chave representam uma importante fonte de acesso aos artigos científicos, o que foi comprovado pelo estudo realizado dos artigos indexados pela base de dados MEDLINE. Em média, cada artigo regista cinco palavras-chave, em que uma a duas apresentam relações com os descritores empregues pela MEDLINE. Assim, nos 207 artigos analisados, 29% das palavras-chave empregues são equivalentes e 30% são associadas aos descritores da MEDLINE, o que representa um conjunto de 59% das palavras-chave existentes nestes artigos. O número de relações equivalentes (353) e associativas (345) é aproximado, mas o critério aplicado determina que as relações de equivalência têm um valor superior (27%) ao valor atribuído às relações associativas (17%), com um total de 44,49%. Ainda que as percentagens não sejam muito elevadas, não deixam de apresentar um dado significativo e que vai ao encontro dos resultados obtidos por Gil-Leiva e Alonso-Arroyo (2005 e 2007) que permitem concluir por uma influência direta ou indireta na escolha dos descritores, que promove a utilização das palavras-chave na aprendizagem da indexação por assuntos e nos esforços a desenvolver para automatizar o processo de indexação automática.

Ao utilizar o Estudo Geral como suporte a este trabalho, defende-se o propósito de chamar a atenção para o repositório institucional da Universidade de Coimbra e para a importância dos autores, enquanto intervenientes ativos no processo de representação e recuperação da informação, através das palavras-chave registadas nos seus artigos. A pesquisa dos artigos científicos, armazenados em repositórios institucionais de acesso aberto, faz-se a partir de metadados que incluem as palavras-chave, e do conteúdo de texto integral, aplicando

técnicas que representam e expõem as relações semânticas entre documentos, em que os dados são relacionados. A combinação e a associação da informação permitem, hoje, a emergência de um conjunto alargado de aplicações que podem explorar os dados fornecidos, com destaque para a informação fornecida pelas palavras-chave.

Reconhecer a importância das palavras-chave para a representação e recuperação de informação, é uma forma de promover o autodepósito e de estimular o completo preenchimento dos metadados dos documentos depositados, como forma de contribuir para uma maior visibilidade, probabilidade de aumento de citações e impacto da produção científica dos autores depositantes, e da própria instituição.

REFERÊNCIAS BIBLIOGRÁFICAS

BORGES, Maria Manuel - Dos sistemas isolados aos partilhados : o repositório institucional da Universidade de Coimbra. In BORGES, M. M.; SANZ CASADO, E. (coord.) - **A Ciência da informação criadora do conhecimento : actas do IV Encontro Ibérico EDIBCIC**. Coimbra : Imprensa da Universidade, 2009. p. 157-166.

GIL-LEIVA, I.; ALONSO-ARROYO, A. - Keywords given by authors of scientific articles in database descriptors. **Journal of the American Society for Information Science** [em linha]. Vol. 58, nº 8 (2007) p. 1175–1187. [Consult. 27 Fev. 2013]. Disponível em WWW:< <http://dx.doi.org/10.1002/asi.20595>>.

GIL-LEIVA, I.; ALONSO-ARROYO, A. - La relación entre las palabras clave aportadas por autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. **Revista Española de Documentación Científica**. Vol. 28, nº 1 (2005), p. 62-79. [Consult. 27 Fev. 2013] Disponível em WWW:<<http://redc.revistas.csic.es/index.php/redc/article/view/165/219>>.

GONÇALVES, Aline Lima - Uso de resumos e palavras-chave em Ciências Sociais: uma avaliação. **Encontros Bibli** [em linha]. Vol. 13, nº 26 (2008). [Consult. 1 Mar. 2013]. Disponível em WWW:<<http://www.redalyc.org/articulo.oa?id=14712794006>>. ISSN 1518-2924

GRANT, M. J. - Key words and their role in information retrieval. **Health Information and Library Journals** [em linha]. Vol. 27, nº 3, (2010), p. 173-5. [Consult. 11 Mar. 2013]. Disponível na WWW:

<<http://onlinelibrary.wiley.com/doi/10.1111/j.1471-1842.2010.00904.x/pdf>>.

MENDES, Maria Teresa Pinto; SIMÕES, Maria da Graça - **Indexação por assuntos : princípios gerais e normas**. Lisboa : Gabinete de Estudos a&b, 2002. ISBN 9729882703

MIGUÉIS, Ana [et al.] - **Desenvolvimento e gestão do Estudo Geral : repositório da produção científica da Universidade de Coimbra** [em linha]. Braga : [s.n], 2010. [Consult. 20 Mar. 2013]. Disponível em WWW:<URL:<http://hdl.handle.net/10316/14216>>.

U.S. NATIONAL LIBRARY OF MEDICINE - **Fact Sheet Medical Subject Headings (MeSH®)** [em linha]. Bethesda, Maryland : NLM. [Consult. 20 Mar. 2013]. Disponível na WWW:<URL:<http://www.nlm.nih.gov/pubs/factsheets/mesh.html>>.

___ **MEDLINE** [em linha]. Bethesda, Maryland : NLM. [Consult. 2 Mar. 2013]. Disponível na WWW:<URL: <http://www.ncbi.nlm.nih.gov/pubmed>>.

___ **MeSH** [em linha]. Bethesda, Maryland : NLM. [Consult. 2 Mar. 2013]. Disponível na WWW:<URL:<http://www.nlm.nih.gov/mesh.html>>.

NP 3715. 1989, Documentação. **Método para análise de documentos, determinação do seu conteúdo e seleção de termos de indexação**. Lisboa: IPQ.

NP 4036. 1992, Documentação. **Tesaurus monolíngues: directivas para a sua construção e desenvolvimento**. Lisboa: IPQ.

SIMÕES, Maria da Graça - **Da abstração à complexidade formal : relações conceptuais num tesouro**. Coimbra : Ed. Almedina, 2008. ISBN 9789724033747.

SHARMA, K. K.; MEDIRATTA, P. K. - Importance of keywords for retrieval of relevant articles in MEDLINE search. **Indian Journal of Pharmacology** [em linha]. Vol. 34, nº 5 (2002), p. 369-371. [Consult. 5 Mar. 2013]. Disponível em WWW:<URL:<http://medind.nic.in/ibi/t02/i5/ibit02i5p369.pdf>>.

TONELLO, I.; LUNARDELLI, R.; ALMEIDA JUNIOR, O. - Palavras-chave: possibilidades de mediação da informação. **PontodeAcesso** [em linha]. Vol. 6 (2012), p. 21-34. [Consult. 1 Mar 2013].

Disponível em

WWW:<URL:<http://www.portalseer.ufba.br/index.php/revistaici/article/view/4524>>.

UNIVERSIDADE DE COIMBRA – **Estudo Geral** [em linha]. Coimbra : UC. [Consult. 2 Mar. 2013].

Disponível na WWW:<URL: <https://estudogeral.sib.uc.pt/>>.