

Tendências e Desafios na Organização do Conhecimento: Homenagem a Maria da Graça Simões

Coordenação
Maria Manuel Borges

**Tendências e Desafios na
Organização do Conhecimento:
Homenagem a Maria da Graça Simões**

Ficha Técnica

Título:

Tendências e Desafios na Organização do Conhecimento:
Homenagem a Maria da Graça Simões

Coordenação:

Maria Manuel Borges

Fotografia da capa:

Leão Borges

@2024

**Tendências e Desafios na
Organização do Conhecimento:
Homenagem a Maria da Graça Simões**

24 e 25 de outubro de 2024

Faculdade de Letras da Universidade de Coimbra

Coord.
Maria Manuel Borges

Sumário

<i>Nota Introdutória</i>	9
<i>O fazer da ciência: entre os saberes e os afetos. Uma reflexão breve sobre a obra de Maria da Graça Simões, Maria Manuel Borges</i>	11
<i>Organização e representação da informação por assuntos na formação em Ciência da Informação em Portugal, Fernanda Ribeiro</i>	23
<i>Revisitando Da abstração à complexidade formal: reflexões sobre convergência e divergência na normalização da organização do conhecimento, Olívia Pestana</i>	37
<i>Metodología e investigación sobre la Clasificación Decimal Universal, Daniel Martínez Ávila</i>	51
<i>Vocabulários controlados, descontrolados e degradados: uma breve análise dialógica do impacto da inteligência artificial no fenômeno da linguagem, pesquisa, ensino e aprendizagem, Renato Rocha Souza</i>	65
<i>O Resumo Documental como Estrutura Lógica na Organização do Conhecimento, José Augusto Chaves Guimarães</i>	83
<i>Sistemas de Organización del Conocimiento incluyentes. Revertir la invisibilización de la mujer, Blanca Rodríguez-Bravo</i>	95
<i>Uso de chatbots na produção de resumos científicos: o que os resumos nos dizem da investigação? – Uma revisão sistemática, Helena Leitão</i>	105
<i>Classificar, um ato imprescindível à organização do universo, Luis Miguel Oliveira Machado</i>	115
<i>Um silêncio e uma voz: a Literatura como desafio à Organização do Conhecimento, Patrícia de Almeida</i>	123
<i>“Taxio... o quê?": Um estudo sobre a taxonomia através dos Sistemas de Organização do Conhecimento, Fernanda Daniel da Silva</i>	129
<i>Uma breve análise bibliométrica ao legado científico de Maria da Graça Simões (1962-2019), Silvana Roque de Oliveira</i>	149

Nota Introdutória

Na conclusão da Conferência Internacional de homenagem a Maria da Graça Simões

Quando a Graça nos deixou no final de agosto de 2019, tinha em mãos vários projetos a que se dedicava com as limitações que os seus tratamentos impunham, mas que não lhe diminuía interesse ou entusiasmo. Um desses projetos resultou na publicação de “Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências”, livro editado já depois do seu falecimento, em plena emergência pandémica. Não será de estranhar, portanto, que esta Conferência Internacional que a homenageia retome de muito perto o tema e o interesse dessa obra, como se a Graça nos traçasse um caminho que, uma vez mais, acompanhamos.

Na impossibilidade de a lembrarmos e homenagearmos em tempo mais próximo ao do seu falecimento, nesta forma que nestes dois dias aqui nos congrega, gostaríamos que esta conferência nos ajudasse a visitar os diferentes aspetos que a lembrança da sua vida e seu trabalho despertam. Essas dimensões, que apenas artificialmente se conseguem separar, evidenciam a sua colaboração com os pares, com colegas da Universidade de Coimbra, tanto do corpo docente como de profissionais da área, e com estudantes dos três ciclos de estudos oferecidos pela Universidade de Coimbra.

A sua primeira dimensão, a do envolvimento com os pares, tem expressão ao nível nacional, mas também, e sobretudo, ao nível internacional, com foco muito particular em Espanha e no Brasil, refletida na produção

científica conjunta, na posição de professora visitante em várias universidades brasileiras e na participação em Grupos de investigação.

A segunda dimensão, que tem em conta a produção científica com colegas da Universidade de Coimbra, encontra idêntica expressão, o que é particularmente relevante para fomentar um espírito de grupo e de rede de colaboração fundamentais para a afirmação de uma área de conhecimento.

Quanto à terceira dimensão, a sua preocupação com as estudantes e os estudantes é visível em várias facetas: antes de mais, pelo seu cuidado em produzir instrumentos pedagógicos que permitissem a quem frequentasse o 1.º ciclo de estudos poder estudar as matérias do seu interesse no seu idioma materno, cruciais tanto para o desempenho de funções ligadas à profissão, como para o aprofundamento curricular em estudos de 2.º e 3.º ciclo. Mas não apenas: estimulava as estudantes e os estudantes destes ciclos de estudos que com ela desenvolviam os seus trabalhos de mestrado e de doutoramento a produzir e a seguir novos caminhos, como se facilmente se pode confirmar em atas de conferências ou em artigos científicos publicados.

Não é possível, nem sequer desejável, separar a investigadora da profissional da área. Também aí teve o cuidado de constituir parcerias de colaboração na Universidade de Coimbra para robustecer a prática profissional, em obras publicadas pela Imprensa da Universidade.

Esta conferência internacional ilustra bem o modo como estas dimensões se concretizaram, particularmente no capítulo que reflete sobre o seu legado científico, que torna evidentes as preocupações e as redes de colaboração de vária natureza, a que não são alheios, nem nunca poderiam tê-lo sido, os afetos.

A Comissão Organizadora

Maria Manuel Borges, Maria Cristina Freitas,

António Tavares Lopes, Ana Miguéis,

Maria João Padez e Bruno Neves

**O fazer da ciência: entre os saberes e os afetos.
Uma reflexão breve sobre a obra de
Maria da Graça Simões¹**

Maria Manuel Borges

ORCID:0000-0002-7755-6168

Univ Coimbra, CEIS20, FLUC, Portugal

mmb@fl.uc.pt

RESUMO: Este texto destina-se a dar uma nota breve sobre o percurso profissional de Maria da Graça Simões, Professora Auxiliar da Faculdade de Letras da Universidade de Coimbra e da sua obra na área da sua especialidade, a representação do conhecimento.

PALAVRAS-CHAVE: Maria da Graça Simões; Percurso profissional; Percurso académico.

ABSTRACT: The purpose of this text is to give a brief overview of the professional and academic career of Maria da Graça Simões, Assistant Professor at the Faculty of Letters of the University of Coimbra, and her work in the field of her speciality, the representation of knowledge.

KEYWORDS: Maria da Graça Simões; Professional career; Academic career.

¹ Texto apresentado por convite no V Congresso ISKO Espanha-Portugal, 25-26 de novembro de 2021, Universidade de Lisboa.

Percurso profissional e académico

Quero expressar, antes de mais, o meu profundo agradecimento ao capítulo Espanha-Portugal da ISKO na pessoa da sua Presidente, Doutora Ana Lúcia Terra, e à Comissão Organizadora deste encontro na Universidade de Lisboa, na pessoa do seu coordenador geral, Doutor Carlos Guardado, por incluírem esta homenagem à Doutora Graça Simões, Professora Auxiliar da FLUC, minha querida colega e amiga. Compreenderão, por isso, que ao longo desta brevíssima memória, a tratarei pelo seu nome. Se há fórum onde a Graça procurou garantir a sua presença foi efetivamente a ISKO, particularmente o seu capítulo ibérico, por ser a sociedade científica da sua área de investigação e também, por essa mesma razão, onde mais amizades reunia. Nada disto era fruto do acaso, naturalmente, antes a expressão de que também a malha da investigação é formada pelas redes que se estabelecem entre as pessoas, tecida a partir de um conjunto de interesses comuns que progride para parcerias e cumplicidades, dando origem, tantas vezes, a afetos sólidos, daqueles que duram para a vida. Por isso, a Graça, que nunca gostou de destaques em terrenos que não sentia como seus, que sempre privilegiou a essência sobre a aparência, haveria de ficar emocionada e sensibilizada por este momento. Partilhamos dessa emoção, como havemos de partilhar convosco o momento em que nos reuniremos de novo para a recordar na Universidade de Coimbra, na sua Escola, na Casa que fez sua, e à volta dessa memória mais uma vez nos debruçarmos sobre o futuro, como exigiria a Graça e como aqui fazemos.

Coincidi com a Graça em Lisboa, num encontro internacional, *Subject indexing: principles and practices in the 90's*, um *IFLA satellite meeting* realizado em 1993. Não nos conhecemos nessa altura e só viríamos a cruzar-nos dois anos mais tarde quando concorri a um lugar de docente de carreira na FLUC para lecionar num curso de pós-graduação, o Curso de Especialização em Ciências Documentais. À época este curso era a formação obrigatória para o ingresso na carreira em Portugal como Bibliotecário, Arquivista ou Documentalista. Tínhamos similitude de interesses científicos que deu origem a longas conversas e que foi também geradora de empatias, as quais facilmente transitam da razão para o

coração porque a Graça era, essencialmente, uma pessoa generosa e de afetos. Iniciámos a docência no Curso de Especialização em Ciências Documentais no mesmo ano, 1995, e daí prosseguimos para a criação de formações conferentes de grau na FLUC, do 1.º ao 3.º ciclo de estudos. A este trajeto de formação corresponde um espaço de tempo relativamente lato, duas décadas para ser mais precisa, ao longo do qual também a Graça se foi desligando do mundo estritamente profissional para se comprometer com o académico. O passo mais relevante neste sentido foi o que a levou ao doutoramento (esta fotografia corresponde ao dia das suas provas de doutoramento na Universidade de Salamanca) e à vinda definitiva, em 2014, para o campo da docência e da investigação a que se dedicou de corpo e alma, até ao limite das suas forças, julho de 2019.



Ao longo da sua carreira profissional, e mais tarde puramente académica, especializou-se na área de Organização do Conhecimento, particularmente a Representação da Informação por assunto, da qual foi o rosto na Universidade de Coimbra. Foi também nessa área que mais investiu nos domínios da docência e da orientação de dissertações de mestrado e de teses de doutoramento e na transferência de conhecimento, para além da sua própria produção científica, tanto em Portugal como no estrangeiro.

Pautou-se sempre por um grande rigor e uma grande preocupação em deixar Escola. Os vários livros que publicou tiveram por alvo os seus estudantes para que mais facilmente pudessem entrar nos meandros, nunca fáceis, da representação da informação por assunto e das múltiplas formas de que se pode revestir. A sua paixão, o seu entusiasmo, mas também a sua impaciência para os mais distraídos, a que logo procurava cativar e chamar a si, é recordada por todos nós. Naturalmente que ela própria vinha de uma Escola, a de Coimbra, pela mão dessa senhora inesquecível e extraordinária que foi a Dra. Maria Teresa Pinto Mendes que tive o gosto de conhecer no Curso de Especialização em Ciências Documentais. Foi pela sua insistência e empenho que se começou a embrenhar, diaria-

mente, nas questões da representação da informação por assunto, fazendo do estudo, ao contrário do que é habitual, uma rotina diária da sua vida profissional. É este espaço que falta na maioria das vezes aos profissionais desta área, mas que a Dra. Maria Teresa Pinto Mendes, com a sua visão e inteligência, sabia que era absolutamente necessário. E não hesitou um segundo, na hora em que se despediu da docência, em confiar nas mãos da Graça o legado da sua vida. Este trabalho conjunto é visível numa publicação conjunta em 2001 nas *Páginas a&b*, republicado posteriormente em 2002 pelo Gabinete de Estudos a&b (Mendes & Simões, 2002), expressa bem, sobretudo no capítulo dedicado aos *Princípios*, a preocupação com a qualidade da indexação por assuntos, tendo sempre em mente a questão da qualidade da recuperação da informação.

A vida profissional e a docência entrelaçavam-se na Graça com a sua vida pessoal, não sendo esta sem as outras. As suas funções na Biblioteca Geral da Universidade de Coimbra, num primeiro momento, mas sobretudo aquelas que desempenhou no Departamento de Arquitetura da FCTUC assim o demonstram, pelo seu envolvimento profissional e pessoal com colegas, docentes, investigadores e estudantes. De 2005 a 2014, o Departamento de Arquitetura da FCTUC foi, de facto, a sua casa e a sua família de coração.

Percurso académico

O doutoramento na Universidade de Salamanca, no qual ingressou em 2007, com muitas dúvidas e reservas sobre o que lhe podia trazer, foi uma decisão da qual nunca se arrependeu, tendo, também aqui, sido decisivo, para além do apoio de amigos e colegas, o apoio pessoal e institucional do Departamento de Arquitetura. No cumprimento desse trajeto preparou e defendeu provas de mestrado, tendo publicado em 2008 o livro *Da Abstração à complexidade formal: relações conceptuais num tesouro* (M. da G. Simões, 2008). Nesta obra, que tem como pano de fundo o que designa como ‘modelos ortodoxos da representação e recuperação da informação’, como as listas de termos, as classificações e os tesouros, e ainda os ‘novos modelos’ de que são exemplos os mapas

conceptuais, as taxonomias e as ontologias, propôs-se abordar a correlação entre a fundamentação teórica sobre a qual assentam estes modelos e as exigências e complexidades que se colocam à sua aplicação à realidade. A obra está dividida em 3 capítulos: o 1.º capítulo centra-se na revisão da literatura que contextualiza e ancora os capítulos seguintes. É o capítulo de maior dimensão no qual ganham destaque os autores considerados clássicos da Indexação, além dos pioneiros como Cutter ou Ranganathan. Clarificam-se as razões que concorrem para o aparecimento do Tesouro que tem uma importância fundamental, justificada tanto pelo volume da informação no período subsequente à II Guerra Mundial como pelo valor que a informação assume, o que torna premente repensar a sua representação de modo a aumentar a eficácia da sua recuperação. O 2.º capítulo, o mais pequeno, centra-se na descrição dos objetivos e da metodologia. O 3.º e último capítulo constitui a parte de análise e discussão de resultados a partir da norma ISO 2788:1986 sobre a construção de tesouros monolíngues que confronta com outras normas nacionais correspondentes, nomeadamente as NP 4036:1992, UNE 50-106-90, AFNOR NFZ47-100-1981, ANSI/NISO Z309.19-2005 e BS 8723-2:2005, para sinalizar a sua convergência relativamente à estrutura e conteúdo que termina com propostas de melhoria para a construção dos tesouros para aumentar “a eficácia na representação e recuperação da informação” (p. 230). É um dos principais textos para os estudantes da licenciatura, em funcionamento desde 2005.

Neste seu novo percurso profissional na academia, mais formal e exigente, ficou bem clara a elevada preparação e maturidade intelectual da Graça na área a que se dedicou e que no doutoramento se traduziu pela atribuição do Prémio Extraordinário de Doutoramento do curso de 2010/2011 da Universidade de Salamanca, com a tese intitulada *A representação de etnia e a sua evolução na Classificação Decimal Universal* (M. da G. Simões, 2010). Com esta tese inaugura-se um novo capítulo onde predomina claramente o interesse científico sobre o profissional, um efeito muito visível na abrangência das suas preocupações. A primeira parte da tese dá origem a um novo livro publicado pela Almedina, “Classificações bibliográficas: percurso de uma teoria” (M. da G. Simões, 2011) no qual aborda as classificações bibliográficas como modelos dinâmicos de organi-

zação do conhecimento, exemplificando com alguns dos sistemas de classificação mais utilizados em bibliotecas nacionais e estrangeiras. Pretende avaliar as vantagens e as fragilidades deste tipo de linguagem, no que respeita à representação e à recuperação do conhecimento. Esta obra, composta por 7 capítulos, está dividida em duas partes: a primeira, intitulada *Classificações bibliográficas: fragmentos teóricos*, visa construir o contexto teórico de inteligibilidade da segunda parte, a que chamou *Classificações bibliográficas: fundamentos e dinâmica estrutural*. Nesta procedeu a um estudo analítico-sintético, como o designa, sobre as classificações bibliográficas de tipo enciclopédico, com destaque para as de tipo enumerativo, misto e facetado pela sua importância na organização do conhecimento ao longo do século XX. A primeira parte está dividida em três capítulos nos quais discorre nos dois primeiros sobre os conceitos de Classificar e Indexar procurando, sempre quando apropriado, apontar em que medida se podem complementar, sendo o 3.º capítulo dedicado à Organização do conhecimento. Com este capítulo trouxe para a discussão o modo como se organizou o conhecimento partindo de planos de classificação e em que medida esse estes influenciam as classificações em geral e as classificações bibliográficas em particular. Já na segunda parte, o 4.º capítulo incide sobre as classificações precursoras dos sistemas de classificação de tipo enciclopédico para terminar o capítulo com os objetivos e função da classificação bibliográfica e discussão da sua estrutura que permite distinguir entre as classificações enumerativas, mistas e facetadas. Esta distinção será abordada nos 3 capítulos seguintes.

Em 2015 publica um novo livro, que terá uma nova impressão em 2018, desta vez dedicado aos resumos, *Resumo documental: uma incursão à desconstrução concetual na literatura científica* (M. da G. Simões, 2015). Na apresentação desta obra destaca os principais aspetos que a conduzem a este tema. A obra está dividida em 4 capítulos, constituindo o primeiro o pano de fundo que contextualiza os capítulos seguintes de ordem instrumental. No primeiro capítulo é realizada uma breve sinopse histórica que nos dá conta não apenas da antiguidade desta prática, mas também dos seus usos que extravasam aqueles que dele fazemos maioritariamente hoje. Evidencia, ainda, a sua aplicação nas diferentes áreas científicas e

o modo como se vai ampliando à medida que a ciência se massifica, notando já um crescimento significativo em meados do século XIX. É esta a origem dos serviços de resumos no final do mesmo século que se inicia com as áreas de Química e de Engenharia. A sua importância, a que não é alheio o ‘boom’ de literatura após a II Guerra Mundial, é expressiva pelo número de serviços de resumos contabilizados em 1969: 1500, dos quais apenas uma pequena parte era relativa às Ciências Sociais e Humanidades. E termina o 1.º capítulo com uma reflexão sobre o papel dos resumos no processo de filtragem da informação, tendo em mente que a maioria da informação científica transacionada hoje é de natureza digital. O 2.º capítulo tem um propósito claramente pedagógico, distinguindo a função, tipologia, formas de apresentação e localização dos resumos. Trata-se de um capítulo que apresenta muitos exemplos, os quais contribuem para a clarificação da mensagem e permitem perceber claramente distinções que podem ser mais subtis para o leitor menos familiarizado com o tema. O 3.º capítulo é fundamental para a compreensão do processo. A sua profundidade e o detalhe colocado na sua apresentação permitem perceber o papel que esta obra desempenha na formação dos/das estudantes em Ciência da Informação, a quem primordialmente se destina. Nesse sentido, é dada uma ênfase muito particular a aspetos que interessam particularmente a esta área científica e que extravasam o que poderia ser o objeto do capítulo. O 4.º e último capítulo apresenta uma visão geral e comparativa entre o resumo documental e outros recursos, nomeadamente o catálogo alfabético de assuntos e o catálogo sistemático, apresentando as suas similitudes e diferenças. Este livro, que é fundamentalmente dedicado a uma unidade curricular da Licenciatura em Ciência da Informação, revela os seus interesses e contributo para ciclos de formação avançada, nomeadamente doutoramento, ao mesmo tempo que servirá como mote para o desenvolvimento de teses de doutoramento. Será, por isso, objeto de várias publicações, nacionais e internacionais, tanto em autoria única como em coautoria.

Em 2017 tem oportunidade de refletir quer sobre as transformações da formação oferecida na Universidade de Coimbra, quer sobre os conteúdos programáticos da área que leciona num Encontro de Representação Documental em São Paulo a que deu o título *Representação da Informação*

por Assunto: do curso de Bibliotecário Arquivista à Licenciatura, Mestrado e Doutoramento em Ciência da Informação na Universidade de Coimbra. Centrando-nos no 1.º ciclo, a Licenciatura em Ciência da Informação Arquivística e Biblioteconómica criada em 2004 com a duração de 4 anos, com início em 2004/2005, sofre, no ano letivo de 2007/2008, uma reforma para responder à adequação ao processo de Bolonha. Este processo origina um duplo constrangimento: por um lado, a redução temporal do curso para 3 anos, e, por outro lado, o enquadramento no *Euroreferencial ID*. Daqui resulta a redução do número de unidades curriculares, sem que a área de representação da informação perca representatividade, mantendo-se como uma das áreas essenciais da licenciatura. Em 2015/2016, sob proposta da FLUC no âmbito da chamada *Reforma da Oferta Formativa* e com a aprovação da Agência de Avaliação e Acreditação do Ensino Superior, inicia-se um novo plano de estudos, com uma designação mais consentânea com os restantes ciclos de estudos em funcionamento, Licenciatura em Ciência da Informação, no qual desaparece a uc de *Análise Documental* e é introduzida a uc de *Laboratórios de Representação de Conteúdos* que tem uma vertente fortemente aplicacional. Nesta área é ainda introduzida uma nova uc obrigatória, *Organização da Informação em Bibliotecas*, que ficará também a seu cargo. A lecionação de *Organização da Informação e do Conhecimento Digital*, um seminário oferecido no âmbito do *Mestrado em Informação, Comunicação e Novos Media*, criado em 2008/2009, reforçou a sua imersão no meio digital. A criação do *Mestrado em Ciência da Informação em 2013/2014*, conduziu à extinção deste mestrado, mas manteve alguns dos seminários desta antiga oferta, de que esta é um exemplo, ao mesmo tempo que trouxe novos desafios incluídos em *Seminários em Ciência da Informação* para o qual a Graça fez convergir colaborações nacionais e internacionais para abrir perspetivas de investigação aos e às estudantes que frutificaram em novos projetos. No 3.º ciclo em Ciência da Informação, que tem a sua primeira edição em 2015/2016, participa no seminário de *Organização e Gestão do Conhecimento*.

A sua dedicação plena à docência desde 2014, como Professora Auxiliar da área de Ciência da Informação, permitiu, assim, ampliar e reforçar o conjunto de unidades curriculares oferecidas do 1.º ao 3.º ciclo de estudos,

sempre com ênfase na área da Representação da Informação por Assunto. A partir desta altura, também a sua atividade internacional ganhou corpo e impacto, traduzida por intensa atividade nas universidades às quais se deslocou.

Apesar de as tarefas de gestão nunca terem interessado à Graça, aceitou a direção da Licenciatura e a do Mestrado em Ciência da Informação, funções que desempenhou, como tudo o que fazia, com empenhamento e interesse. Foi, ainda, Coordenadora Sócrates/Erasmus com missões na Europa, mas também em Angola e sobretudo no Brasil, país que sentia como seu e que conhecia muito bem, tendo procurado ativamente implementar protocolos com universidades de relevo na área de Ciência da Informação. Esteve envolvida em vários grupos de pesquisa, tendo sido igualmente professora colaboradora no Curso de Ciência da Informação da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista-UNESP (Marília, São Paulo, Brasil) e no programa de Pós-Graduação em Gestão e Organização do Conhecimento, do Curso de Ciência da Informação da Universidade Federal de Minas Gerais (Brasil).

Também em 2015 publica um capítulo num livro de atas, por altura da comemoração dos 500 anos da biblioteca da Universidade de Coimbra, num tom confessional, que testemunha o seu percurso profissional, mas também os seus afetos na Biblioteca Geral da Universidade de Coimbra, a que dá o título, precisamente *Quinze anos na Biblioteca Geral: um percurso de aprendizagem e de afetos* (Simões, 2015, pp. 383-392).

A sua preocupação com os estudantes de Ciência da Informação manifestava-se também pela disponibilidade de textos no seu idioma, bem como os profissionais da área, constituindo uma parte muito importante do seu esforço para a publicação destas obras. A familiarização com a Classificação Decimal Universal levou-a a publicar dois livros, um em 2008, reimpresso em 2017, e outro em 2018. O primeiro, *Classificação Decimal Universal: fundamento e procedimentos* (M. da G. Simões & Ferreira, 2008), dá uma visão geral, funcional e sistemática dos fundamentos e procedimentos da CDU. A abordagem teórica é complementada com uma vertente aplicada visando facilitar a compreensão das regras e das técnicas do ato de classificar. Já o segundo livro, *Dos princípios da classi-*

ficção decimal universal a uma prática harmonizada (M. da G. Simões et al., 2018), dividido em 3 partes – princípios, estrutura e classes principais –, pretende igualmente constituir um guia para a aplicação e utilização da CDU e auxiliar na compreensão dos seus fundamentos teóricos, incluindo, para esse fim, exercícios de aplicação. Tem, por essa razão, como principais destinatários os/as estudantes de Ciência da Informação, mas também os profissionais da área. O envolvimento de alguns antigos/as estudantes do Curso de Especialização em Ciências Documentais, agora responsáveis pelo tratamento técnico das distintas bibliotecas da Universidade de Coimbra integradas no catálogo coletivo da UC, recorda a preocupação que a Dra. Maria Teresa Pinto Mendes tinha relativamente à harmonização de princípios, não sendo, assim, por acaso, que este livro lhe é dedicado.

Ainda em 2017, regista mais duas publicações no Brasil, um capítulo de livro, “Do conceito e dos princípios metodológicos de indexação: subsídios para uma análise crítica” (Borges, 2017a) e uma comunicação num livro de atas onde faz uma análise da área de representação da informação por assunto desde o curso de bibliotecário arquivista à licenciatura, mestrado e doutoramento em Ciência da informação da Universidade de Coimbra, já referido.

As suas publicações em revistas nacionais e internacionais com arbitragem científica e em atas de conferências internacionais de relevo evidenciam as múltiplas parcerias que consegue congrega, tanto em termos internos – colaborando com colegas da Universidade de Coimbra, orientandos de mestrado e doutoramento –, mas também colegas de outras universidades portuguesas e estrangeiras. A análise das publicações em revistas ou atas de conferências ao longo deste período (setembro de 2014 a julho de 2019) em que se encontra a trabalhar a tempo integral evidencia uma produção intensa e uma evolução nas coautorias escolhidas, as quais partem desta proximidade natural que tem como base tanto os afetos quanto a partilha de interesses científicos. No que se refere aos colegas estrangeiros, foi com Espanha e o Brasil que esta proximidade foi mais frutífera. Na verdade, no período referido, o volume de publicações tende a envolver muitos colegas do Brasil, sendo particularmente evidente entre 2016 e 2018, e em 2018 as únicas publicações registadas. Isto é o reflexo da intensa atividade

com universidades com as quais existiam já relações de grande proximidade, concretizadas em acordos bilaterais de cooperação, e com novas parcerias, concretizadas também em novos acordos. Os temas escolhidos são diversificados, desde o uso ético da informação até às questões da indexação automática, além de temas mais clássicos, teóricos, numa perspectiva de lacunas reveladas pela revisão da literatura e pela necessidade de as debater. Esta variedade ilustra de igual modo como a Graça soube identificar as parcerias que ampliavam claramente as perspectivas de investigação e potenciavam uma colaboração internacional fecunda.

É esse também o sentido de organizar um evento do capítulo ibérico da ISKO. Assumimos, assim, o compromisso de trazer a Coimbra o *III Congresso ISKO Espanha-Portugal XIII Congresso ISKO Espanha* (Simões & Borges, 2017). A realização deste encontro representou um esforço muito significativo para todos nós, mas trouxe à Graça o ânimo de que precisava para enfrentar os tempos sombrios que vivia e que, com ela, estávamos a viver. A alegria de rever os colegas e amigos, de tempos mais antigos ou mais recentes, que nos deram o gosto de estar connosco em Coimbra também nos animou a festejar com ela essa realização. Foi também nesse encontro que a Graça lançou o seu último projeto. O derradeiro esforço foi concentrado na publicação de um texto em parceria com colegas do Brasil. Esse livro, *Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências* (Simões et al., 2020), ganhou corpo já tardiamente no tempo, outubro de 2020, mas expressa bem, pela sua dimensão e qualidade dos capítulos, os laços intelectuais, e de afetos, que a Graça tanto prezava nas suas relações.

Nota final

A vida é sempre demasiado curta para o que sonhamos, mas a Graça partiu precisamente quando os seus projetos, que também eram nossos porque sempre partilhamos todas as conquistas, começavam a frutificar amplamente. Não é demais recordar que os livros que publicou na área da Representação por assunto – e abstraímos propositadamente de outras publicações – são raros no país. Dela recordamos o rigor, a exigência,

a vontade férrea, mas também a generosidade e a proximidade afetiva que eram marcas distintivas do seu perfil científico e humano que ficam para sempre na nossa memória.

Referências bibliográficas

- Mendes, M. T. P., & Simões, M. D. G. (2002). *Indexação por assuntos: Princípios gerais e normas*. <https://estudogeral.sib.uc.pt/bitstream/10316/20805/1/Indexacao%20por%20assuntos.pdf>
- Simões, M. da G. (2008). *Da abstração à complexidade formal: Relações conceituais num tesouro*. Almedina.
- Simões, M. da G., & Ferreira, C. (2008). *Classificação decimal universal: Fundamentos e procedimentos*. Almedina.
- Simões, M. da G. (2010). *A representação de etnia e a sua evolução na classificação decimal universal*.
- Simões, M. da G. (2011). *Classificações Bibliográficas: Percorso de uma Teoria* (1ª edição). Almedina.
- Simões, M. da G. (2015). *Resumo documental: Uma incursão à (des)construção concetual na literatura científica*. Imprensa da Universidade de Coimbra.
- Simões, M. da G. (2015). Quinze anos na Biblioteca Geral: um percurso de aprendizagem e de afetos. Em J. A. Bernardes, A. M. E. Miguéis, & C. A. S. Ferreira (Eds.), *A biblioteca da Universidade: Permanência e metamorfoses* (pp. 197-206). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-1045-0_11
- Simões, M. da G. (2017). Do conceito e dos Princípios Metodológicos de Indexação. Subsídios para uma Análise Crítica. (pp. 151-178). ISBN 978 857983904 (digital). ISBN 978 8579839030. Disponível em: <https://www.marilia.unesp.br/Home/Publicacoes/una-mirada-a-la-ciencia-de-la-informacion---completa.pdf>
- Simões, M. da G., & Borges, M. M. (coord.) (2017). *Tendências atuais e perspectivas futuras em organização do conhecimento: atas do III Congresso ISKO Espanha-Portugal XIII Congresso ISKO Espanha*. Universidade de Coimbra: Centro de Estudos Interdisciplinares do Século XX, CEIS20. ISBN 978-972-8627-75-1. Disponível em: <https://purl.org/sci/atas/isko2017.pdf>
- Simões, M. da G., Lima, G. Â. de, Maculan, B. C. M. dos S., & Dias, C. da C. (Eds.). (2020). *Do tratamento à organização da informação: Reflexões sobre concepções, perspectivas e tendências*. Imprensa da Universidade de Coimbra.
- Simões, M. da G., Rodríguez Bravo, B., Ferreira, C., Moura, M. de F., Carvalho, M. J., & Pais, S. de S. (Eds.). (2018). *Dos princípios da Classificação Decimal Universal a uma prática harmonizada*. Imprensa da Universidade de Coimbra.

Organização e representação da informação por assuntos na formação em Ciência da Informação em Portugal

Fernanda Ribeiro

ORCID iD 0000-0002-5641-9199

Universidade do Porto – Faculdade de Letras

CITCEM (doi.org/10.54499/UIDB/04059/2020), Portugal

fribeiro@letras.up.pt

RESUMO: Este texto visa dar uma panorâmica do que tem sido o ensino das técnicas de organização e representação da informação (ORI) por assuntos na formação dos profissionais da área da Ciência da Informação em Portugal. Após uma resenha histórica do tema no contexto da formação ministrada até ao momento em que começou a haver formação graduada no ensino superior português (2001), analisam-se os *currícula* dos diversos cursos de licenciatura e mestrado existentes atualmente no nosso País e procura-se perceber em que medida o ensino de matérias, que deveriam estar sempre presentes naquilo que é o *core* da Ciência da Informação, permanece (ou não) como indispensável na formação dos licenciados e mestres que estão a ser formados. A análise efetuada baseia-se nos planos de estudo dos cursos criados desde 2001, embora se circunscreva àqueles que se encontram atualmente a funcionar e que estão acreditados pela Agência de Avaliação e Acreditação do Ensino Superior (A3ES), pois é a partir da informação que consta dos relatórios de avaliação/acreditação que podemos ter acesso às estruturas curriculares, planos de estudo e programas das unidades curriculares, de forma sistemática, informação esta que é complementada com o que é disponibilizado nos *sites* das instituições que ministram os ciclos de estudo.

PALAVRAS-CHAVE: Ciência da Informação, ensino superior; Indexação por assuntos/ensino; Representação da informação por assuntos.

ABSTRACT: This text aims to give an overview of the teaching of information organisation and representation (IOR) techniques by subject in the training of information science professionals in Portugal. After a historical review of the subject in the context of the training provided up until the moment when graduate training began in Portuguese higher education (2001), the curricula of the various bachelor's and master's degree courses currently existing in our country are analysed and an attempt is made to understand to what extent the teaching of subjects that should always be present in what is the core of Information Science remains (or not) indispensable in the training of the bachelor's and master's graduates who are being trained. The analysis is based on the syllabuses of the programmes created since 2001, although it is limited to those that are currently running and that are accredited by the Higher Education Assessment and Accreditation Agency (A3ES), since it is from the information contained in the assessment/accreditation reports that we can systematically access the curricular structures, study plans and contents of the curricular units, information that is complemented with what is available on the websites of the institutions that are responsible for the study cycles.

KEYWORDS: Information Science; Higher education; Teaching subject indexing; Subject information representation.

1. De 1887 a 2001 – traços históricos da formação em ORI para bibliotecários e arquivistas

O ensino formal na área da Biblioteconomia e da Arquivística em Portugal iniciou-se com a criação, pelo Decreto de 29 de dezembro de 1887, artigo 13.º, do “curso de instrução superior, denominado de bibliotecario-archivista” (Portugal, 1888, p. 19), destinado a formar o pessoal habilitado para desempenhar funções nas bibliotecas e arquivos. O curso era tutelado pela Inspeção Geral das Bibliotecas e Arquivos e era constituído por sete disciplinas, umas de carácter geral (História Pátria e Universal, Filologia, Literatura latina e grega, Literatura moderna da Europa e especialmente a portuguesa), lecionadas no Curso Superior de Letras, e outras específicas (Diplomática, Numismática e Bibliologia), ministradas no Real Arquivo da Torre do Tombo e na Biblioteca Nacional de Lisboa, por funcionários destas instituições. O Regulamento do curso de bibliotecário-arquivista, no que dizia respeito a estas cadeiras “técnicas” foi promulgado em 3 de outubro de 1902 (Regulamento, 1902), mas apenas

contém disposições sobre aspetos funcionais e não faz menção aos conteúdos das referidas cadeiras.

Pouco depois de implantada a República, o Curso Superior de Letras foi integrado na Faculdade de Letras da Universidade de Lisboa, criada em 22 de março de 1911, e, por essa via, o Curso de Bibliotecário-Arquivista (CBA) passou para a esfera universitária. A primeira reforma republicana do setor das bibliotecas e arquivos, promulgada em 18 de março de 1911, não tem quaisquer disposições relevantes sobre o curso nem sobre as matérias lecionadas no mesmo. Mas, curiosamente, o Decreto de 11 de setembro de 1913, que instituiu o “Estágio de Arquivistas”, também sob a tutela da Inspeção das Bibliotecas e Arquivos, explicita as “matérias professadas” e elenca os respetivos conteúdos programáticos. É, pelo que conhecemos, o primeiro documento publicado que contém informação sobre as matérias ensinadas na formação dos profissionais desta área.

Para o tema que aqui nos ocupa, apenas selecionámos conteúdos da disciplina de Biblioteconomia, verificando-se que entre os vários tópicos do programa, são referidos expressamente os que constam do Quadro 1:

Quadro 1 – Conteúdos programáticos do Estágio de Arquivistas, instituído em 1913

“Catálogos metódicos. Classificação bibliográfica. Classificação decimal. Sistemas vários e seu desenvolvimento. Sistema de Brunet. Sistema de classificação actualmente usado na Biblioteca Nacional de Lisboa. Verbetes para o catálogo metódico. Indicações necessárias. Classificação dos verbetes. Catálogos ideográficos ou de informações, suas vantagens” (Organização, 1913, p. 14-15).

É interessante verificar que os conteúdos lecionados, apesar de pouco desenvolvidos, estavam em sintonia com aquilo que eram as tendências internacionais da época, incorporando matérias de grande atualidade ao tempo, como seja o ensino dos vários sistemas de classificação, de criação e desenvolvimento ainda muito recente.

A primeira remodelação do CBA, no contexto do novo regime republicano, ocorre com o Decreto n.º 4.885, de 5 de outubro de 1918, que promulga o *Regulamento do curso superior de bibliotecário-arquivista*

(Portugal, 1918). Esta reforma, justificada pela pouca frequência do curso em virtude de se verificar um excesso de disciplinas teóricas, veio reforçar a componente tecnicista do mesmo, através de um maior número de disciplinas lecionadas na Biblioteca Nacional e na Torre do Tombo, designadamente: Paleografia (1.^a parte) e Bibliologia, no 1.^o ano; Diplomática, Numismática Geral e Medalhística, Biblioteconomia, e um curso trimestral sobre higiene das espécies bibliácas, no 2.^o ano; Paleografia (2.^a parte), Arquivologia, Numismática Portuguesa e Esfragística, no 3.^o ano (Portugal, 1918, p. 603). Embora o regulamento estipulasse que os professores deveriam apresentar, no fim de cada ano letivo, os programas para o ano letivo seguinte, a verdade é que não temos conhecimento desses programas. Só em 1927, no âmbito do Decreto n.º 13.724, de 27 de maio, que promoveu uma reforma do setor das bibliotecas e arquivos, já imbuída da ideologia que inspirou a revolução do 28 de maio de 1926, tais programas foram publicados. Nesse diploma, constam os “programas das cadeiras especiais” do CBA, sendo que apenas encontramos uma brevíssima menção à organização e representação da informação por assuntos na cadeira de Biblioteconomia, num ponto intitulado “Elaboração de catálogos”, onde, entre os vários tipos de catálogos, é referido o “sistemático ou metódico”. De forma igualmente lacónica, surge o tópico “Sistemas vários de classificação” (Portugal, 1927, p. 900).

O Decreto n.º 19.952, de 27 de junho de 1931, promulgou uma profunda reorganização dos “serviços bibliotecários e arquivísticos”, sendo considerado uma verdadeira “lei de bases” do Estado Novo para este setor. Naturalmente, neste contexto, o CBA sofreu também uma remodelação e, em 16 de novembro do mesmo ano, a Portaria n.º 7.261, aprovou os programas do curso, elaborados pela Junta Consultiva das Bibliotecas e Arquivos, com uma enumeração exaustiva das matérias a lecionar em cada cadeira. É na cadeira de “Bibliologia e Bibliografia” que é dada alguma atenção à representação da informação por assuntos (Quadro 2).

Na cadeira de “Biblioteconomia”, apenas há referência a “Catálogos ideográfico e metódico”, dentro de um tópico sobre catálogos gerais.

Passados cerca de vinte anos sobre os programas que serviam de base à leção no Estágio de Arquivistas (1913), verifica-se que, nos

Quadro 2 – Conteúdos da cadeira de “Bibliologia e Bibliografia”, do CBA (1932)

- Classificação das espécies quanto às matérias para a sua distribuição pelas secções das bibliotecas – Sistemas bibliográficos.
- As classificações das ciências e os sistemas bibliográficos. O sistema de Brunet. As divisões e secções da Biblioteca Nacional.
- Bibliografias especiais. Sua classificação: [...] c) Sob o ponto de vista sistemático ou por assuntos. Bibliografias das matemáticas, das ciências físicas. Bibliografias históricas, etc.; d) Sob o ponto de vista ideográfico ou por assuntos. [...] (Portugal., 1932, p. 11).

pormenorizados programas das diversas cadeiras do CBA, que ocupam dezoito páginas do decreto, só há exíguas menções (uma dezena de linhas) ao tratamento da informação por assuntos, numa época em que na Europa e nos Estados Unidos estavam em pleno desenvolvimento os sistemas de classificação bibliográfica, cujos exemplos mais emblemáticos eram a *Dewey Decimal Classification* (DDC), a *Library of Congress Classification* (LCC), a *Classification Decimale Universelle* (CDU) ou a *Bibliographic Classification* (BC) de Henry Evelyn Bliss, todos eles já bem conhecidos à época¹. Será que esta estagnação e desfasamento face à evolução internacional é já um sinal do fechamento do novo regime ao exterior, em contraste com a abertura que caracterizou os governos da Primeira República, em que é notória a participação de técnicos, como Raul Proença, em encontros internacionais promovidos pelo Instituto Internacional de Bibliografia, ou em contactos com personalidades como Paul Otlet (Ribeiro, 2008, p. 93-94), ou mesmo as visitas técnicas a instituições estrangeiras de funcionários superiores do Ministério da Instrução Pública, como António Ferrão? A avaliar pela evolução que caracterizou as décadas seguintes até meados dos anos sessenta (Ribeiro, 2008, p. 111-152), não temos dúvida em afirmar que sim, embora em termos epistemológicos a formação ministrada seja, obviamente, reflexo natural da visão paradigmática custodial e historicista que se afirmou em meados do século XIX e consolidou ao longo do século XX (Silva & Ribeiro, 2012).

¹ Sobre estas classificações bibliográficas pode ver-se, a título meramente ilustrativo, uma obra clássica (Mills, 1973) ou o ensaio de Maria da Graça Simões, que contém igualmente uma resenha histórica (Simões, 2011).

Depois de ter sido transferido da tutela da Faculdade de Letras da Universidade de Lisboa para a da Inspeção das Bibliotecas e Arquivos, o CBA voltou novamente a ser integrado na academia em 1935, mas agora na Faculdade de Letras da Universidade de Coimbra, onde funcionou durante quase meio século², até ser extinto em 1982 (Portugal, 1982).

O diploma legislativo que aprovou a transferência do curso para Coimbra não faz qualquer referência aos programas das cadeiras, que na nova estrutura curricular se limitavam a cinco: Paleografia e Diplomática, Numismática e Esfragística, Bibliografia e Biblioteconomia, no 1.º ano; Curso de Aperfeiçoamento de Paleografia, Arquivologia e Arquivoeconomia, no 2.º ano. Naturalmente, as matérias do tratamento da informação por assuntos eram lecionadas na cadeira anual de Bibliografia e Biblioteconomia. Não existem publicados os programas do curso e só uma pesquisa no arquivo do mesmo poderá eventualmente trazer alguma luz sobre o assunto e ajudar a perceber a evolução que o ensino destas temáticas foi sofrendo ao longo dos anos. Temos, contudo, conhecimento dos programas lecionados no ano letivo de 1979-1980 (ano em que frequentámos o curso como aluna), os quais permitem perceber a importância que era dada à “Indexação” no âmbito da cadeira de Bibliografia e Biblioteconomia, numa época em que os organismos internacionais produziam já abundante normalização, as universidades – sobretudo as americanas, mas também as europeias, principalmente as do Reino Unido e dos países nórdicos – tinham formação em *Library Science* e *Information Science* e desenvolviam investigação científica com a correspondente produção de literatura (livros e artigos), e em Portugal, por ação de muitos profissionais e da BAD (Associação Portuguesa de Bibliotecários, Arquivistas e Documentalistas), se começava a ter um conhecimento mais efetivo dos desenvolvimentos da área (Ribeiro, 2003, vol. 1, p. 505-509).

² Além do Curso de Bibliotecário-Arquivista, ministrado na Faculdade de Letras da Universidade de Coimbra, funcionou em Lisboa, entre 1969 e 1974, um “Estágio de Preparação Técnica de Bibliotecários, Arquivistas e Documentalistas”, que conferia também habilitação própria para as funções de bibliotecário e arquivista, criado pelo Decreto n.º 49.009, de 7 de maio de 1969, mas não foi possível obter informação sobre os respetivos conteúdos programáticos.

Quadro 3 – Conteúdos programáticos do tópico “Indexação” no âmbito da cadeira “Bibliografia e Biblioteconomia” do Curso de Bibliotecário-Arquivista (1979-1980)

Indexação	
<p>1. Introdução: características da representação do conteúdo e da recuperação pelo assunto; definição de indexação.</p> <p>2. Análise documental</p> <p>2.1. Operações</p> <p>2.1.1. Classificação (a desenvolver)</p> <p>2.1.2. Indexação em sentido estrito (a desenvolver)</p> <p>2.1.3. Condensação</p> <p>2.2. Fases</p> <p>2.2.1. Apreensão do conteúdo total</p> <p>2.2.2. Identificação dos conceitos que o representam</p> <p>2.2.3. Seleção dos conceitos necessários à investigação ulterior</p> <p>3. Linguagens</p> <p>3.1. Linguagem natural</p> <p>3.2. Linguagem documental</p> <p>3.2.1. Sobre palavras</p> <p>3.2.2. Sobre conceitos</p> <p>3.2.3. Sobre assuntos</p> <p>3.2.4. Indexação coordenada</p> <p>3.3. O utilizador e a linguagem documental</p> <p>3.4. Tipos de linguagens</p> <p>3.4.1. Categoriais; Classificação</p> <p>3.4.2. Combinatórias; Indexação no sentido estrito</p> <p>3.4.3. Instrumentos: tabela classificatória; thesaurus</p>	<p>4. Classificação</p> <p>4.1. Definição, origens e estrutura de base; tabelas; notação e índices</p> <p>4.2. Críticas</p> <p>4.3. A Classificação Decimal Universal (CDU)</p> <p>4.4. Outras classificações; evolução (encarada particularmente sob o ponto de vista da multidimensionalidade): Brunet, Congresso, Dewey, Bliss, Cordonnier, Colon; a análise por facetas</p> <p>5. Indexação no sentido estrito: definição, origens e thesaurus</p> <p>5.1. Multidimensionalidade</p> <p>5.2. Estrutura dum thesaurus; relações semânticas</p> <p>5.2.1. Relações de substituição</p> <p>a) Equivalência</p> <p>b) Qualificação</p> <p>5.2.2. Relações de enriquecimento</p> <p>a) Hierarquia</p> <p>b) Associação</p> <p>5.2.3. Formas sob que se expressam, em várias línguas</p> <p>5.3. Outros princípios</p> <p>5.4. Tipos de thesaurus</p> <p>6. Qualidades da indexação</p>

Os conteúdos programáticos do tópico “Indexação. Linguagens documentais”, inserido na cadeira de Bibliografia e Biblioteconomia (Quadro 3) e lecionado pela bibliotecária da Biblioteca Geral da Universidade de Coimbra, Maria Teresa Pinto Mendes, são ilustrativos da preocupação em alinhar o ensino com as mais atualizadas tendências internacionais. Estes conteúdos programáticos eram acompanhados de 32 referências bibliográficas, sendo que 12 eram datadas da década de 1970, ou seja, muito atuais.

O CBA foi substituído pelo Curso de Especialização em Ciências Documentais (CECD), cujo funcionamento se iniciou em 1983, nas Faculdades de Letras de Coimbra e de Lisboa e, em 1985, na do Porto³. A atenção dada neste curso ao tratamento da informação por assuntos foi consideravelmente reforçada, desde logo pela inclusão no plano de estudo de duas disciplinas (Indexação por Assuntos I e II), no tronco comum do 1.º ano, e mais duas disciplinas (Indexação por Assuntos III e IV), na opção “Documentação e Biblioteca” do 2.º ano⁴.

A legislação regulamentadora do CECD teve ampla aplicação, uma vez que este curso de especialização foi instituído em variadas universidades públicas e privadas, contabilizando-se 17 cursos em funcionamento em 2007 (Pinto, 2008), ano em que a implementação do Processo de Bolonha veio produzir mudanças estruturais no ensino superior em Portugal.

Os conteúdos programáticos das disciplinas de ‘Indexação’ lecionadas no CECD não são de fácil acesso por não estarem publicados, embora em diversas instituições constassem de folhetos ou volumes policopiados para divulgação interna aos estudantes⁵. Seguiam *grossa modo*, as temáticas dos CECD das Universidades de Coimbra, Lisboa e Porto.

Apesar do qualificativo de “especialização”, o CECD era, na verdade, um curso de formação inicial, que conferia a habilitação profissional exigida por lei para o provimento dos lugares de bibliotecário e arquivista na administração pública, embora os detentores desta pós-graduação tivessem também possibilidade de trabalhar em centros de documentação, bibliotecas ou arquivos de organizações privadas.

Por alturas da viragem do século, a exigência que a profissão requeria já não era passível de ser obtida num curso com a duração de apenas dois anos e não conferente de grau. Daí que, a par do CECD, a partir de 2001,

³ O CECD foi criado na Faculdade de Letras da Universidade de Coimbra, pelo Decreto n.º 87/82, de 13 de julho, o qual extinguiu, em simultâneo, o CBA instituído na mesma Faculdade em 1935. A regulamentação do CECD nas Faculdades de Letras das Universidades de Coimbra, de Lisboa e do Porto efetivou-se pelas Portarias n.º 448/83, de 19 de abril, 449/83, de 19 de abril, e 851/85, de 9 de novembro, respetivamente.

⁴ No CECD da FLUP, as disciplinas eram anuais e, por isso havia apenas duas e não quatro.

⁵ A título de exemplo, pode referir-se o *Guia do estudante: 1999-2000* do CECD da FLUP, que compila os programas e a bibliografia básica das disciplinas lecionadas no curso (Universidade, 1999).

Quadro 4 – Conteúdos programáticos das disciplinas de Indexação por Assuntos I e II do CECD da FLUP (1999-2001)

Indexação por Assuntos I	Indexação por Assuntos II
<p style="text-align: center;">I</p> <ol style="list-style-type: none"> 1. Índices – função, elementos constituintes e papel na recuperação da informação 2. Indexação (em geral) – definição, objetivo e aplicação à Ciência da Informação 3. Indexação por assuntos – definição e posicionamento na cadeia de tratamento da informação <ol style="list-style-type: none"> 3.1. A Norma Portuguesa 3715 <ol style="list-style-type: none"> 3.1.1. Fases da indexação por assuntos <ol style="list-style-type: none"> a) Identificação/Reconhecimento dos conceitos: <ul style="list-style-type: none"> – Análise de conteúdo – metodologia (exercícios práticos) – Descrição do conteúdo – Seleção dos conceitos (condicionantes) b) Representação dos conceitos: <ul style="list-style-type: none"> – Termo de acesso – Noção de coordenação – Linguagens de indexação – tipos; princípios e métodos de construção 	<p style="text-align: center;">1.º semestre</p> <ol style="list-style-type: none"> I. Introdução (revisão de conceitos do ano anterior) <ol style="list-style-type: none"> 1. Tipos de linguagens de indexação: combinatórias (já estudadas) e categoriais 2. Linguagens categoriais: classificações enumerativas e facetadas <ol style="list-style-type: none"> 2.1. Princípios básicos 2.2. Estrutura 2.3. Método de construção II. Estudo das linguagens categoriais <ol style="list-style-type: none"> 1. Estudo de um sistema de classificação: a Classificação Decimal Universal 2. Comparação entre as linguagens categoriais e as combinatórias: vantagens e inconvenientes 3. Evolução das classificações. Caracterização de vários sistemas de classificação <p style="text-align: center;">2.º semestre</p> <p>Trabalhos elaborados pelos alunos: apresentação oral nas aulas e por escrito no final do ano – aulas com discussão generalizada (em seminário)</p>
<p style="text-align: center;">II</p> <ol style="list-style-type: none"> 1. As linguagens combinatórias – seu posicionamento no campo das linguagens de indexação <ol style="list-style-type: none"> 1.1. Filosofia geral; princípios e regras – a Norma Portuguesa 4036 1.2. Instrumentos de apoio – listas e tesauros 2. Exercícios práticos – uso da linguagem combinatória para representação do conteúdo informativo de diversas espécies documentais 3. A avaliação da qualidade da indexação 	

se tenha iniciado um novo modelo formativo em Portugal (Silva & Ribeiro, 2002), com a criação de licenciaturas para formação de base, passando os mestrados, que começavam a surgir, a configurar formação especializada.

2. O ensino da ORI por assuntos na atualidade

As primeiras licenciaturas na área da Ciência da Informação foram instituídas no ano letivo de 2001-2002, uma na Universidade do Porto, numa parceria das Faculdades de Letras e de Engenharia, e a outra na Escola Superior de Estudos Empresariais e de Gestão, do Instituto Politécnico do Porto. A estas duas licenciaturas seguiram-se outras, maioritariamente em universidades públicas, coexistindo com os CECD e com diversos mestrados. No ano letivo de 2007-2008, momento em que ocorreu a reestruturação do ensino superior no âmbito do Processo de Bolonha, existiam em Portugal, aprovados ou em funcionamento, 12 licenciaturas, 15 mestrados e 17 pós-graduações (modelo CECD) (Pinto, 2008, p. 57), uma oferta bastante excessiva para um país como Portugal. Porém, uma boa parte desses cursos não se submeteu à acreditação, ou não foi acreditada, pela Agência de Acreditação e Avaliação do Ensino Superior (A3ES), criada em finais de 2007.

No primeiro ciclo de avaliação, que ocorreu em 2013-2014, apenas se submeteram à avaliação, na área de Biblioteconomia, Arquivo e Documentação (BAD), 11 (onze) ciclos de estudo – 4 licenciaturas, 6 mestrados e 1 doutoramento – tendo os restantes deixado de funcionar por decisão das instituições que os ministravam. No segundo ciclo de avaliação, que teve lugar em 2019-2020, foram avaliados 8 (oito) ciclos de estudo – 3 licenciaturas, 4 mestrados e 1 doutoramento – os quais ainda se encontram em funcionamento.

Os dados que a seguir apresentamos, relativos às unidades curriculares em que a representação da informação por assuntos é lecionada, dizem respeito às licenciaturas e mestrados atualmente a funcionar e que constam do Quadro 5. A recolha de informação foi feita a partir dos processos de avaliação submetidos e aprovados pela A3ES e também do que está disponível nos *sites* das instituições que administram os ciclos de estudo.

Quadro 5 – Instituições com formação de 1.º e 2.º ciclos na área de Ciência da Informação (2024)

Instituição	Ciclo de estudo
Instituto Politécnico do Porto – Instituto Superior de Contabilidade e Administração do Porto	Licenciatura em Ciências e Tecnologias da Documentação e Informação
Instituto Politécnico do Porto – Instituto Superior de Contabilidade e Administração do Porto	Mestrado em Informação Empresarial
Universidade de Coimbra – Faculdade de Letras	Licenciatura em Ciência da Informação
Universidade de Coimbra – Faculdade de Letras	Mestrado em Ciência da Informação
Universidade de Lisboa – Faculdade de Letras	Mestrado em Ciências da Documentação e Informação
Universidade do Porto – Faculdade de Letras e Faculdade de Engenharia	Licenciatura em Ciência da Informação
Universidade do Porto – Faculdade de Engenharia e Faculdade de Letras	Mestrado em Ciência da Informação

Relativamente às três licenciaturas atualmente existentes, identificamos em todas elas unidades curriculares, cujos conteúdos incidem exclusivamente nas temáticas da organização e representação da informação (ORI) por assuntos, especificamente nos seguintes tópicos:

- Linguagens categoriais / Linguagens codificadas / Classificações bibliográficas / Classificações facetadas
- Linguagens vocabulares / Linguagens combinatórias
- Elaboração de resumos

Estes conteúdos são lecionados em UC com as seguintes designações:

Quadro 6 – UC com conteúdos exclusivos de ORI (nas licenciaturas)

Unidade Curricular	N.º de ECTS	Obrigatória / Opcional
Análise e Representação da Informação I	6	Obrigatória
Análise e Representação da Informação II	6	Obrigatória
Controlo de Autoridade e Indexação	9	Obrigatória
Laboratório de Representação de conteúdo	6	Opcional
Representação da Informação: Linguagens vocabulares	6	Opcional
Representação da Informação: Resumos e Linguagens codificadas	6	Opcional

Além disso, existem outras UC em cujos conteúdos programáticos se identificam tópicos relacionados com ORI, embora não constituam o *core* dos respetivos programas.

No que respeita aos quatro cursos de mestrado, também se identificam conteúdos programáticos relativos ao tratamento da informação por assuntos, em UC específicas (e alguns tópicos em outras UC) nomeadamente sobre:

- Análise de Conteúdo / Análise de Domínio
- Classificações bibliográficas / Classificações facetadas
- Folksonomias / Etiquetagem colaborativa
- Linguagens vocabulares
- Ontologias
- Resumos
- Sistemas de Organização do Conhecimento
- Taxonomias / Tesouros
- Vocabulários controlados
- Web semântica

Estes conteúdos são lecionados em UC com as seguintes designações:

Quadro 7 – UC com conteúdos exclusivos de ORI (nos mestrados)

Unidade Curricular	N.º de ECTS	Obrigatória / Opcional
Análise de Conteúdo e Indexação	6	Obrigatória
Organização do Conhecimento em Ambientes Digitais	6	Obrigatória
Organização da Informação I	6	Obrigatória
Organização da Informação II	6	Obrigatória
Organização da Informação e do Conhecimento Digital	10	Obrigatória
Seminários em Ciência da Informação	10	Obrigatória
Web Semântica	6	Obrigatória

3. Considerações finais:

Esta panorâmica da formação em organização e representação da informação por assuntos, no âmbito dos vários modelos formativos existentes em Portugal desde 1887 até hoje, é um pequeno recorte de um trabalho em curso, de maior fôlego, sobre aspetos múltiplos da formação em termos globais na área da Ciência da Informação no nosso país, que selecionamos

para esta homenagem à Doutora Graça Simões, que dedicou tanto do seu labor a ensinar, a investigar e a escrever precisamente sobre a temática do tratamento da informação por assuntos.

Desta singela abordagem, podemos concluir que:

- 1 – A temática da ORI esteve, desde sempre, presente nos *curricula* dos vários cursos, com maior ou menor desenvolvimento, mas em todos com carácter obrigatório;
- 2 – Essa presença continua a verificar-se na formação atualmente ministrada em licenciaturas e mestrados, em unidades curriculares obrigatórias, sendo os respetivos conteúdos curriculares dedicados totalmente ao tema do tratamento da informação por assuntos;
- 3 – O peso que esta temática tem no conjunto das unidades curriculares dos planos de estudo é objetivamente pequeno, pois em 180 ECTS das licenciaturas ou em 120 ECTS dos mestrados ORI ocupa apenas 6, 9 ou 12 ECTS, havendo apenas um caso em que ultrapassa os 12 ECTS.

Numa época em que as avalanches de informação que nos sufocam, cada vez mais precisam de ser triadas, avaliadas, pesquisadas de forma rápida e eficiente, será que a formação que estamos a ministrar não deveria dedicar mais tempo à análise de conteúdos, à indexação dos mesmos, à organização de esquemas de classificação, às técnicas de condensação dos conteúdos? Estamos já a tirar partido das potencialidades da Inteligência Artificial (IA) para acelerar a organização e representação dos conteúdos? Na aplicação de linguagens controladas, que exigem uma prévia análise intelectual dos conteúdos, compensa realmente recorrer a IA, deixando ao profissional da informação a tarefa final de validação da qualidade dos produtos elaborados automaticamente? Ou continuaremos a viver com os problemas da chamada “indexação automática”, que resulta eficaz para áreas do conhecimento que têm uma terminologia consensualmente fixada pela respetiva comunidade científica e infrutífera para as ciências sociais e humanas, onde a metáfora, o simbólico, a prolixidade são marcas incontornáveis do discurso científico?

Não estamos ainda a incorporar estas questões na formação académica que ministramos, mas urge que o façamos porque a realidade evolui e não está nas nossas mãos travar essa evolução...

Referências bibliográficas

- Mills, J. (1973). *A Modern outline of library classification*. Chapman and Hall.
- Organização do Estágio de Arquivistas (1913). In *Inspeção das Bibliotecas Eruditas e Arquivos (Compil.)*, *Legislação de 1913* (pp. 12-15). Imprensa da Universidade de Coimbra.
- Pinto, M. M. G. A. (2008). A Formação em Informação e Documentação: Portugal na contemporaneidade. *Páginas a&b: arquivos & bibliotecas*. Série 2, 1, p. 7-62.
- Portugal. Leis, decretos, etc. (1982). Ministério da Educação e das Universidades: Decreto n.º 87/82, de 13 de julho. *Diário do Governo. I série*. 159, p. 2.089-2.090.
- Portugal. Leis, decretos, etc. (1932). Portaria n.º 7:261 [de 16 de novembro de 1931]. *Diário do Governo. 1.ª série*. 1 (2 jan.) p. 9-18.
- Portugal. Leis, decretos, etc. (1927). Ministério da Instrução Publica: Direcção Geral do Ensino Superior: Decreto n.º 13:724 [de 27 de maio de 1927]. *Diário do Govêrno. 1.ª série*. 114 (3 jun.) p. 881-901.
- Portugal. Leis, decretos, etc. (1918). Decreto n.º 4:885 [de 5 de outubro de 1918]. *Colecção Oficial de Legislação Portuguesa*. (2.º sem.) p. 603-606.
- Portugal. Leis, decretos, etc. (1888). Direcção Geral de Instrucção Publica: 1.ª Repartição: [Decreto de 29 de dezembro de 1887]. *Diario do Governo*. Lisboa. 3 (4 jan.) p. 18-20.
- Regulamento do curso de bibliothecario archivista nas cadeiras professadas no Real Archivo da Torre do Tombo e na Bibliotheca Nacional de Lisboa [de 3 de outubro de 1902]. *Collecção Official de Legislação Portuguesa*. Lisboa. p. 970-971.
- Ribeiro, F. (2008). *Para o estudo do paradigma patrimonialista e custodial: a Inspeção das Bibliotecas e Arquivos e o contributo de António Ferrão (1887-1965)*. Edições Afrontamento.
- Ribeiro, F. (2003). *O Acesso à informação nos arquivos*. Vol. 1. Fundação Calouste Gulbenkian; Fundação para a Ciência e a Tecnologia, Ministério da Ciência e do Ensino Superior.
- Silva, A. M., & Ribeiro, F. (2002). *Das "Ciências" Documentais à Ciência da Informação: ensaio epistemológico para um novo modelo curricular*. Edições Afrontamento.
- Silva, A. M., & Ribeiro, F. (2012). Documentation/Information and their paradigms: characterization and importance in research, education, and professional practice. *Knowledge Organization*, 39(2), p. 111-124.
- Simões, M. G. (2011). *Classificações bibliográficas: percurso de uma teoria*. Almedina.
- Universidade do Porto. Faculdade de Letras. Departamento de Ciências e Técnicas do Património (1999). *Curso de Especialização em Ciências Documentais: guia do estudante, 1999-2001*. FLUP. Texto policopiado.

Websites:

- <https://www.a3es.pt/pt/acreditacao-e-auditoria/resultados-dos-processos-de-acreditacao>
- <https://www.iscap.ipp.pt/cursos/licenciatura/808>
- <https://www.iscap.ipp.pt/cursos/mestrado/833>
- <https://apps.uc.pt/courses/PT/course/5543>
- <https://apps.uc.pt/courses/PT/course/3921>
- <https://www.letras.ulisboa.pt/pt/ensino/mestrados/cursos#ci%C3%AAncias-da-documenta%C3%A7%C3%A3o-e-informa%C3%A7%C3%A3o>
- https://sigarra.up.pt/flup/pt/cur_geral.cur_planos_estudos_view?pv_plano_id=34044&pv_ano_lectivo=2023&pv_tipo_cur_sigla=L&pv_origem=CUR
- https://sigarra.up.pt/feup/pt/cur_geral.cur_planos_estudos_view?pv_plano_id=2503&pv_ano_lectivo=2023&pv_tipo_cur_sigla=M&pv_origem=CUR

Revisitando *Da abstracção à complexidade formal*: reflexões sobre convergência e divergência na normalização da organização do conhecimento

Olívia Pestana

ORCID 0000-0002-5485-3143

CITCEM, Faculdade de Letras, Universidade do Porto,

Via Panorâmica, s/n, 4150-564 Porto, Portugal

opestana@letras.up.pt

RESUMO: Maria da Graça Melo Simões estudou, analisou e perspectivou trajetos do passado, da contemporaneidade e do futuro de uma das principais áreas da Ciência da Informação, a organização do conhecimento. Deixou-nos as suas marcas indeléveis de uma dedicação à investigação rigorosa e de grande abertura científica nas publicações que deu à estampa e que continuam a orientar-nos e a inspirar-nos no desafio de estudar uma área do conhecimento tão vasta e em constante evolução.

Neste capítulo de homenagem, revisitamos a obra *Da abstracção à complexidade formal: relações conceptuais num tesouro*, pelo pertinente, e ainda atual, contributo para o estudo dos métodos de construção de tesouros, particularmente quanto à expressão das relações conceptuais, enquadrando no âmbito da normalização internacional. Apresentamos, ainda, a replicação do estudo empírico disponível na obra em apreço relativamente à convergência e divergência no tesouro EuroVoc face aos princípios postulados pela, agora em vigor, norma ISO 25964-1:2011, especificamente no âmbito das relações semânticas.

Desta forma, homenageamos a Prof. Doutora Maria da Graça Simões e partilhámos o seu entendimento da necessidade de estudos críticos no domínio da normalização aplicada à organização do conhecimento.

PALAVRAS-CHAVE: Organização do conhecimento; Normalização; Tesouro.

ABSTRACT: Maria da Graça Melo Simões studied, analysed and envisioned the past, present and future of one of the most important areas of Information Science, the knowledge organisation. She left us her indelible mark of dedication to rigorous research and great scientific openness in her publications, which continue to guide and inspire us in the challenge of studying such a vast and constantly evolving field of knowledge.

In this tribute book chapter, we revisit *Da abstracção à complexidade formal: relações conceptuais num tesouro* (From abstraction to formal complexity: conceptual relationships in a thesaurus) for its pertinent and still current contribution to the study of thesaurus construction methods, particularly in terms of expressing conceptual relationships, in the context of international standardisation. We also present the replication of the empirical study available in the book on the convergence and divergence of the EuroVoc Thesaurus in relation to the principles postulated by the current ISO 25964-1:2011 standard, specifically the semantic relations.

In this way, we pay tribute to Professor Maria da Graça Simões and share her understanding of the need for critical studies in the field of standardisation applied to the knowledge organisation.

KEYWORDS: Knowledge organisation; Standardisation; Thesaurus.

Introdução

Maria da Graça Melo Simões estudou, analisou e perspetivou trajetos do passado, da contemporaneidade e do futuro de uma das principais áreas da Ciência da Informação, a organização do conhecimento. Deixou-nos as suas marcas indeléveis de uma dedicação à investigação rigorosa e de grande abertura científica nas publicações que deu à estampa e que continuam a orientar-nos e a inspirar-nos no desafio de estudar uma área do conhecimento tão vasta e em constante evolução.

Relembrando os princípios que subjazem aos estudos sobre organização do conhecimento (OC), citemos Mazzocchi (2018) por considerar que o termo “organização do conhecimento”, tal como hoje o utilizamos, foi retomado por Dahlberg (1974), dando visibilidade ao termo alemão *Wissensordnung* (ordem do conhecimento) ao ser utilizado para designar a organização conceptual e sistemática do conhecimento humano, termo esse adotado internacionalmente na sua adaptação para inglês, *knowledge*

*organisation*¹. É de salientar que, conforme identificado por Mazzocchi (2018), o termo remonta a trabalhos anteriores aos anos 70, tendo as suas raízes em diversos trabalhos com destaque para os de Bliss (1929, 1933), nos quais o autor utilizou a expressão *The Organization of Knowledge*. Naturalmente que os trabalhos de Dahlberg (1993, 2006, 2009) trouxeram uma outra projeção a esta área que, principalmente a partir dos estudos desta autora, ficou definitivamente estabelecida como uma área científica.

Ainda no contexto desta área de estudos, importa referir que, propostos por muitos investigadores de OC, como Broughton et al. (2005) e Hjørland (2003, 2008,), há dois elementos essenciais que caracterizam a OC: os processos de organização do conhecimento, como a indexação, a catalogação, a análise por assunto ou a classificação; e os sistemas de organização do conhecimento, ou seja, os instrumentos direcionados para as atividades incluídas nos processos, como os tesouros, a que, neste trabalho, nos reportamos.

Importa, ainda, sublinhar que, tal como apontado por Hjørland (2016), a consolidação do termo organização do conhecimento, bem como o desenvolvimento de estudos alicerçados nas diversas correntes teóricas que lhe dão corpo, partiram, inequivocamente, do aparecimento, em 1974, da publicação científica intitulada *International Classification*, a qual foi designada de *Knowledge Organization*, a partir de 1993, e têm tido significativa expressão desde a constituição da International Society for Knowledge Organization (ISKO), em 1989, tendo a referida publicação assumido o papel de órgão oficial da ISKO.

Esta brevíssima clarificação sobre as origens da organização do conhecimento, enquanto campo de estudo da Ciência da Informação, serve, apenas, de enquadramento ao trabalho que, aqui, apresentamos. Na verdade, o trajeto percorrido pela área em apreço reveste-se de muitas interações disciplinares, bem como de um frequente serpentear das abordagens

¹ Importa referir que coexistem duas versões do termo, sendo uma na versão do inglês britânico (*knowledge organisation*) e a outra na versão do inglês americano (*knowledge organization*).

terminológicas, sendo esse caminho aprofundado e detalhadamente descrito por Hider (2018).

No âmbito do seu percurso de investigação desenvolvido na Universidade de Salamanca, Maria da Graça Simões realizou trabalho de análise e reflexão sobre aspetos determinantes para a construção e funcionamento de instrumentos de organização do conhecimento, como as classificações bibliográficas e os tesouros, por onde iniciou a sua produção científica nesse âmbito. A obra *Da abstracção à complexidade formal: relações conceptuais num tesouro* constitui um imprescindível estudo para a compreensão dos métodos de construção de tesouros, particularmente quanto à expressão das relações conceptuais, vistas pela lente da normalização internacional. Neste trabalho, após uma exposição teórica sobre os tesouros em geral, baseada numa revisão bibliográfica, a autora desenrola um estudo comparativo entre as normas ISO 2788-1986, ANSI/NISO Z39.19-2005, NF Z47-100, NP 4036:1992, UNE 50-106-90 e BS 8723-2:2005², tendo selecionado a norma ISO 2788-1986, intitulada *Documentation – Guidelines for the establishment and development of monolingual thesauri*, para desenvolver um estudo sobre a sua aplicação em dois tesouros, concretamente no EuroVoc e no Macrothesaurus da OCDE/ONU. A sua análise concentrou-se, essencialmente, no estudo do ponto da norma intitulado *8 Basic relations in a thesaurus* e na verificação do seu reflexo, visível ou não, nos referidos vocabulários controlados.

A norma sobre a qual o trabalho se debruçou foi, entretanto, substituída pela norma ISO 25964-1:2011, intitulada *Information and documentation – Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*, mas tal não veio colocar em segundo plano, por eventual desatualização, o estudo de Maria da Graça Simões, bem pelo contrário. Passada mais de uma década sobre a publicação da obra da

² Para facilitar a leitura, mencionamos, apenas, os códigos das normas, devendo ser procuradas as referências bibliográficas pelas entidades emissoras, respetivamente International Organization for Standardization, National Information Standards Organization, AFNOR, Instituto Português da Qualidade, Asociación Española de Normalización e, por último, British Standards Institution.

autora que, aqui, evocamos, bem como sobre a publicação da norma em vigor, tendo sido confirmada a sua atualização pela International Organization for Standardization, em 2022, entendemos como pertinente replicar o estudo desenvolvido a respeito das relações conceptuais. No presente capítulo, apresentamos a replicação do estudo empírico descrito na obra em análise relativamente à convergência e divergência no tesauro Eurovoc face aos princípios postulados pela, agora em vigor, norma ISO 25964-1:2011, especificamente no âmbito das relações semânticas que são estabelecidas entre os termos.

Da norma ISO 2788:1986 à norma ISO 25964-1:2011

A evolução dos vocabulários controlados, surgida a partir de meados do século XIX, motivou a inevitável necessidade de estabelecimento de princípios padronizados, os quais se consolidaram no século XX. Como refere Maria da Graça Simões (2008, p. 56),

“Face ao impacto que esta nova linguagem, o tesauro, obteve no âmbito das novas tecnologias, nomeadamente no desenvolvimento de catálogos colectivos virtuais ao tempo ainda embrionários, houve a necessidade de normalizar critérios para a sua elaboração, com vista à intercomunicabilidade e compatibilidade das linguagens controladas.

Entre os vários factores que concorreram directamente para a necessidade da criação de textos normativos para a construção de um tesauro, teve particular relevo a criação das bases de dados”.

E a autora prossegue, afirmando:

“Esta preocupação normativa prendia-se com a vontade de controlar o vocabulário nas suas duas vertentes:

- controlo formal [normalização morfológica e sintáctica dos termos];
- controlo semântico [criação de uma tipologia de relações entre os termos e normalização das mesmas]” (Simões, 2008, p. 61).

A criação da norma ISO 2788:1986, surgida na sua primeira versão em 1974 e no seguimento do estabelecimento de princípios orientadores desenvolvidos por entidades como a UNESCO (1971), ilustrou essa necessidade, vindo a constituir o modelo padrão por ser emanada pela International Standard Organization, tendo, conseqüentemente, reflexo internacional ao ser independente dos aspetos linguísticos de um dado país em particular, mas respeitando as eventuais necessidades de harmonização nesse contexto.

A norma ISO 2788:1986 esteve em vigor durante 25 anos, até ao aparecimento da norma ISO 25964-1, em 2011. Cada organismo nacional de normalização publicou a sua própria versão, traduzida e adaptada a cada país.

Como refere Dextre-Clarke (2018), permaneceram, ao longo dos tempos, alguns princípios básicos na normalização relativa à construção de tesauros. Para além das relações de equivalência, um tesouro deve reconhecer explicitamente as relações hierárquicas e associativas entre conceitos. A autora revela, ainda, que todas as normas apresentam alguma dificuldade na definição da relação associativa. Por vezes, apresenta-se como uma relação de exclusão ao não se qualificar como uma relação hierárquica ou de equivalência. No entanto, está subjacente o princípio de que a sua função é a de orientar os utilizadores para termos que podem complementar ou substituir aquele em que pensaram inicialmente.

No que se refere às relações semânticas entre os termos de um tesouro, a norma ISO 2788:1986 apresentava a descrição de três tipos de relações no ponto 8 da norma, como já referimos, sendo explicitadas as relações semânticas acima identificadas. Maria da Graça Simões (2008) apresentou a seguinte análise crítica a respeito da descrição das relações semânticas que a norma ilustrava:

- relações de equivalência – a norma apresenta uma lista extensa de categorias de termos sinónimos e quasi-sinónimos passíveis de relações de equivalência, sem apresentar uma sistematização de critérios;
- relações hierárquicas – a norma apresenta as situações lógicas em que pode ocorrer este tipo de relação semântica, ou seja, relação genérica, relação partitiva ou todo-parte, relação de instância e, ainda, as relações poli-hierárquicas, numa estrutura explícita, faltando,

no entanto, algum destaque gráfico do texto no caso em que é justificado o recurso à aplicação de relações associativas;

- relações associativas – a norma descreve e exemplifica as relações entre termos da mesma categoria, bem como de categorias diferentes, todavia os exemplos não são suficientemente elucidativos.

A necessidade de atualização da norma ISO 2788:1986 para que conseguisse fazer face à evolução tecnológica, bem como à diversidade de meios para além do textual, a par da evidente necessidade de correspondência linguística, levou a International Organization for Standardization a desenvolver uma nova norma, a ISO 25964 que veio substituir um conjunto de normas existentes e transformar a abordagem normativa relativamente à construção de vocabulários controlados. Segundo Sánchez-Cuadrado, Colmenero-Ruiz e Moreiro (2012, p. 231-232),

“É provavelmente a renovação mais profunda da norma ISO relativa aos tesouros, há muito solicitada pelos profissionais. Funde as normas anteriores para os tesouros monolíngues e multolíngues. Propõe, igualmente, um formato de intercâmbio baseado num esquema XML. O seu objetivo é fornecer recomendações para o desenvolvimento e manutenção dos tesouros. Verifica-se um aumento da utilização de facetas para a organização do esquema de conhecimento. Destina-se a diferentes tipos de recursos de informação, independentemente do meio utilizado (texto, multimédia, bases de conhecimento, etc.), que permitam a sua utilização em aplicações de recuperação de informação”³.

A nova norma ISO 25964 está dividida em duas partes, sendo a primeira, a ISO 25964-1:2011, dedicada à construção de tesouros para a recuperação da informação e a segunda parte, a ISO 25964-2:2013, dedicada à interoperabilidade com outros vocabulários. A primeira parte da norma é a que apresenta as relações semânticas que podem ser estabelecidas entre os termos. Considerando que esta norma veio substituir as distintas normas que anteriormente se reportavam em separado aos tesouros monolíngues

³ Tradução da autora.

e aos tesouros multilingues, encontramos uma estrutura diferente no que respeita à apresentação das relações de equivalência. O ponto intitulado *8 The equivalence relationship, in a monolingual context* é dedicado exclusivamente à descrição das relações de equivalência num contexto monolíngue. O ponto seguinte é inteiramente dedicado à equivalência entre línguas, muito oportunamente criado e desenvolvido, face ao necessário estabelecimento de princípios que assegurem as práticas linguísticas de cada país, bem como a eventual influência de fatores culturais nas opções de seleção dos termos. O ponto intitulado *10 Relationships between concepts* apresenta as relações hierárquicas e as associativas. Apesar de ser uma nova norma, incorpora muitos dos exemplos que constavam na norma ISO 2788:1986, fazendo com que as dúvidas na seleção dos termos, bem como na criação das relações semânticas, persistam em muitos dos casos. Ou seja, apesar da visível evolução da norma, no que respeita à definição, clarificação e exemplificação das relações semânticas entre os termos, a norma fica muito aquém do expectável e a análise crítica efetuada por Maria da Graça Simões (2008) permanece atual.

Construção das relações semânticas no tesouro EuroVoc e a norma ISO 25964-1:2011

Com o objetivo de conhecer as eventuais diferenças na atualidade relativamente ao tesouro EuroVoc, bem como à aplicação dos princípios da norma ISO 25964-1:2011, replicámos o estudo desenvolvido por Maria da Graça Simões (2008) sobre esse tesouro. A descrição disponível na página do Serviço das Publicações da União Europeia⁴ refere o seguinte:

“O EuroVoc é um tesouro multilingue e multidisciplinar que abrange as atividades da UE. Contém termos em 24 línguas da UE (...) e em três línguas de países candidatos à adesão à UE (albanês, macedónio e sérvio).

⁴ Informação disponível em: <https://op.europa.eu/pt/web/eu-vocabularies/thesauri>. Tradução da autora.

O EuroVoc é gerido pelo Serviço das Publicações da União Europeia, que avançou para uma gestão de tesouro baseada em ontologias e para tecnologias da Web semântica em conformidade com as recomendações do W3C, bem como com as últimas tendências em matéria de normas de tesouros. (...)

Os utilizadores do EuroVoc incluem as instituições da União Europeia, o Serviço das Publicações da UE, os parlamentos nacionais e regionais da Europa, bem como governos nacionais e utilizadores privados em todo o mundo”.

O tesouro EuroVoc é composto por vinte e um campos temáticos, tendo sido utilizada no nosso trabalho a última versão⁵.

Seguimos o mesmo critério de levantamento dos termos utilizado no estudo, ou seja, tendo como base a apresentação alfabética do tesouro, dado que a apresentação sistemática apenas permite visualizar as relações hierárquicas. A seleção dos termos a consultar corresponde à lista de termos que constam no primeiro estudo, considerando, portanto, os primeiros oitenta termos que, então, o tesouro apresentava alfabeticamente. A análise das relações estabelecidas entre os descritores seguiu, também, o mesmo procedimento do primeiro estudo no que se refere à aplicação da norma, neste caso a ISO 25964-1:2011, pelo que verificámos se o tipo de relações semânticas entre os termos estavam de acordo com os princípios da norma. Para efeitos do presente estudo, apenas registámos as alterações ocorridas nos termos do tesouro relativamente ao volume de relações semânticas estabelecidas e não desenvolvemos uma análise percentual de valores de convergência e divergência. Tal como no primeiro estudo (Simões, 2008), não analisámos o nível de reciprocidade entre os termos que apresentavam relações semânticas entre si, pelo que não é apresentado o valor de consistência das relações consideradas.

O estudo desenvolvido por Maria da Graça Simões (2008) levou à conclusão de que, em termos gerais, se destacava no EuroVoc o estabelecimento de relações de equivalência, especialmente as de reenvio para o

⁵ Versão revista em 11 de julho de 2024.

genérico e as de termos quasi-sinónimos, sendo escassos os casos em que não estavam determinadas de acordo com a norma ISO 2788:1986. Relativamente às relações hierárquicas, também ocorriam em elevado volume, mas, neste caso, uma parte não seguia o postulado na norma, devendo os termos em questão estar sob o estabelecimento de relações associativas, nuns casos, e noutros casos no princípio de quasi-sinonímia. No caso das relações hierárquicas entre termos que correspondiam ao preceituado na norma, a percentagem das relações partitivas era preponderante relativamente à das relações genéricas. Por fim, o estudo da investigadora concluiu, ainda, que, na amostragem analisada, havia relações associativas apenas entre termos de diferentes categorias, estabelecidas de acordo com o postulado na norma, mas o seu volume poderia ser aumentado se fossem corretamente estabelecidas as relações semânticas incorretamente colocadas em relações hierárquicas.

Os resultados do atual estudo revelam que 17 dos oitenta termos analisados foram alvo de atualização, sendo que um dos termos deixou de ser utilizado em 2016 ao ser substituído por outro⁶. Também foram identificadas a substituição do termo de uma e a eliminação de duas relações hierárquicas. A atualização dos termos originou quarenta novas relações semânticas, distribuindo-se da seguinte forma:

- relações de equivalência – assinalam-se nove novas relações de equivalência, das quais quatro de tipo reenvio para o genérico, uma de variante ortográfica, um sinónimo, um quasi-sinónimo e duas relações que, de acordo com a norma ISO 25964-1:2011 deveriam ser relações hierárquicas de tipo genérico (o total de relações de equivalência identificadas no primeiro estudo era o de 97);
- relações hierárquicas – registam-se oito novas relações, sendo uma de tipo todo-parte, duas que não se enquadram em qualquer dos tipos identificados na norma e cinco que deveriam ser relações

⁶ Os termos que foram alvo de atualização no estabelecimento de relações semânticas são os seguintes: abandono de exploração agrícola; abate de animais; aborto; abuso de confiança; abuso de poder; ação comum (substituído por "decisão UE"); ação financeira; ação judicial; acesso à informação; acesso ao emprego; acidente de trabalho; acidente de transporte; acidente nuclear; ácido; ácido inorgânico; acordo ADN; acordo bilateral.

associativas (o total de relações hierárquicas identificadas no primeiro estudo era o de 89);

- relações associativas – assinalam-se 23 novas relações, sendo de destacar o facto de 12 não serem enquadráveis nos tipos de relação associativa que a norma ISO 25964-1:2011 exemplifica, mas o significado dos termos, para os quais somos remetidos, justifica a criação de uma relação; há, ainda, sete relações por dependência causal, outras três que ilustram uma ação e o produto da ação e, por último, uma que corresponde a disciplina ou área de estudo e os objetos ou fenómenos estudados (o total de relações associativas identificadas no primeiro estudo era o de 39).

Conclui-se, portanto, que a entrada em vigor de uma nova norma não trouxe alterações significativas à atualização do tesauro EuroVoc, concentrando-se as alterações no estabelecimento mais alargado de relações semânticas de associação, sendo de sublinhar que esta necessidade havia sido identificada por Maria da Graça Simões no seu estudo. Aliás, mantém-se, também, a identificação de termos em que as relações deveriam ser associativas e não hierárquicas. Em geral, verifica-se um elevado cumprimento do postulado na norma ISO 25964-1:2011, havendo, portanto, um sentido de convergência, mas devemos reconhecer que, apesar de o Serviço das Publicações da União Europeia indicar que o tesauro segue as últimas tendências em matéria de normas, ainda se identifica alguma imprecisão na sua aplicação, especificamente da norma em causa. Tal poderá dever-se à, ainda, difícil interpretação das categorias das relações semânticas, bem como à presença de exemplos pouco elucidativos na norma.

Conclusão

Encerramos estas linhas com a convicção de que a produção científica desenvolvida por Maria da Graça Simões deve ser revisitada e estudada, pois a análise crítica que nos entregou nos seus textos dá-nos pistas para o desenvolvimento de investigação essencial para a evolução da organização do conhecimento.

Este estudo sobre o qual nos debruçámos revela que quer a normalização relativa à construção dos tesouros, quer as tomadas de decisão subjacentes à construção dos tesouros existentes, nos suscitam alguma reflexão, sendo perceptível que muito caminho existe, ainda, para ser possível atingir na totalidade o cumprimento dos princípios estabelecidos neste âmbito.

Esperamos, com este capítulo, não só ter honrado a memória da homenageada, mas também, e na sequência dos seus estudos, tracejar caminhos para a investigação teórica e, também, baseada em estudos empíricos no campo da organização do conhecimento.

Referências bibliográficas

- AFNOR. (1981). *Règles d'établissement des thésaurus monolingues* (AFNOR Standard No. NF Z47-100). <https://www.boutique.afnor.org/fr-fr/norme/nf-z47100/regles-detablissement-des-thesaurus-monolingues/fa015919/14806> (Norma revogada e substituída pela norma ISO 25964-1:2011)
- Asociación Española de Normalización. (1990). *Documentación. Directrices para el establecimiento y desarrollo de tesouros monolingües* (UNE Standard No. UNE 50-106-90). <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0005074> (Norma revogada e substituída pela norma UNE-ISO 25964-1:2014)
- Bliss, H. E. (1929). *The Organization of Knowledge and the System of the Sciences*. Henry Holt and Company.
- Bliss, H. E. (1933). *The Organization of Knowledge in Libraries and the Subject-Approach to Books*. The H. W. Wilson Company.
- British Standards Institution. (2005). *Structured vocabularies for information retrieval. Guide – Thesauri* (BSI Standard No. BS 8723-2:2005). (Norma revogada e substituída pelas normas BS ISO 25964-1:2011 e BS ISO 25964-2:2013)
- Broughton, V., Hansson, J., Hjørland, B., & Huertas, M. J. L. (2005). Knowledge Organization. In L. Kajberg & L. Lørring (Eds.), *European Curriculum Reflections on Library and Information Science Education* (pp. 133-48). Royal School of Information Science. <https://arkiv.iva.ku.dk/hyperion/423.pdf>
- Dahlberg, I. (1974). *Grundlagen universaler Wissensordnung: Probleme und Möglichkeiten eines universalen Klassifikationssystems des Wissens*. (DGD-Schriftenreihe 3). Verlag Dokumentation.
- Dahlberg, I. (1993). Knowledge Organization: Its Scope and Possibilities. *Knowledge Organization*, 20(4), 211-222. <https://doi.org/10.5771/0943-7444-1993-4-211>
- Dahlberg, I. (2006). Knowledge Organization: A New Science? *Knowledge Organization*, 33(1), 11-19. <https://doi.org/10.5771/0943-7444-2006-1-11>
- Dahlberg, I. (2009). Concepts and Terms: ISKO's Major Challenge. *Knowledge Organization*, 36(2-3), 169-177. <https://doi.org/10.5771/0943-7444-2009-2-3-169>
- Dextre-Clarke, S. G. (2018). Knowledge organization system standards. In J. D. McDonald & M. Levine-Clark (Eds.), *Encyclopedia of library and information sciences* (4th ed., pp. 2665-2676). CRC Press.

- Hider, P. (2018). The Terminological and Disciplinary Origins of Information and Knowledge Organization. *Education for Information*, 34(2), 135-161. <https://doi.org/10.3233/EFI-180165>
- Hjørland, B. (2003). Fundamentals of knowledge organization. *Knowledge Organization*, 30(2), 87-111. <https://doi.org/10.5771/0943-7444-2003-2-87>
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2-3), 86-101. <https://doi.org/10.5771/0943-7444-2008-2-3-86>
- Hjørland, B. (2016). Knowledge organization. *Knowledge Organization*, 43(6), 475-484. <https://doi.org/10.5771/0943-7444-2016-6-475>
- Instituto Português da Qualidade. (1992). *Documentação: tesouros monolíngues, directivas para a sua construção e desenvolvimento*. (Norma 4036:1992) (Norma anulada pelo IPQ).
- International Organization for Standardization. (1986). *Documentation – Guidelines for the establishment and development of monolingual thesauri* (ISO Standard No. 2788:1986). <https://www.iso.org/standard/7776.html> (Norma revogada e substituída pela norma ISO 25964-1:2011)
- International Organization for Standardization. (2011). *Information and documentation – Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval* (ISO Standard No. 25964-1:2011). <https://www.iso.org/standard/53657.html>
- International Organization for Standardization. (2013). *Information and documentation – Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies*. (ISO Standard No. 25964-2:2011). <https://www.iso.org/standard/53658.html>
- Mazzocchi, F. (2018). Knowledge Organization System (KOS): An Introductory Critical Account. *Knowledge Organization*, 45(1): 54-78. <https://doi.org/10.5771/0943-7444-2018-1-54>
- National Information Standards Organization. (2005). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (ANSI/NISO Standard No. Z39.19-2005 R2010). <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>
- OECD, ONU (1998). *Macrothesaurus for Information Processing in the Field of Economic and Social Development* (5th Ed.). OECD Development Centre. <https://doi.org/10.1787/9789264162990-en>
- Sánchez-Cuadrado, S., Colmenero-Ruiz, M. J., & Moreira, J. A. (2012). Tesouros: estándares y recomendaciones. *El profesional de la información*, 21(3), 229-235. <http://dx.doi.org/10.3145/epi.2012.may.02>
- Simões, M. G. (2008). *Da abstracção à complexidade formal: relações conceptuais num tesouro*. Almedina.
- UNESCO. (1971). *Guidelines for the Establishment and Development of Monolingual Thesauri for Information Retrieval*. SC/WS/500. <https://unesdoc.unesco.org/ark:/48223/pf0000005951.locale=en>

Metodología e investigación sobre la Clasificación Decimal Universal

Daniel Martínez Ávila

<https://orcid.org/0000-0003-2236-553X>

Universidad de León, España

dmarta@unileon.es

RESUMEN: La Clasificación Decimal Universal es uno de los principales sistemas de clasificación bibliográfica del mundo, siendo Maria da Graça de Melo Simões una especialista sobre el sistema. Desde 1969, los organismos encargados de la gestión de la Clasificación Decimal Universal han organizado encuentros especializados denominados *International UDC Seminars*. En este trabajo se analizan las posiciones epistemológicas y los métodos de investigación de los trabajos publicados en la emblemática edición de 2015. Se presenta una metodología desarrollada ad-hoc para este tipo de análisis y se concluye que la mayoría de los trabajos analizados siguen una epistemología racionalista y utilizan como método de investigación el estudio de caso.

PALABRAS CLAVES: Clasificación Decimal Universal; Métodos de investigación; Posiciones epistemológicas.

ABSTRACT: The Universal Decimal Classification is one of the most important bibliographic classifications in the world, with Maria da Graça de Melo Simões being a specialist on the system. Since 1969, the organizations in charge of the Universal Decimal Classification have organized specialized meetings called International UDC Seminars. This chapter analyzes the epistemological positions and research methods of the works published in the emblematic 2015 edition of the event. A methodology developed ad-hoc for this type of analysis is presented and it is concluded that the majority of the works analyzed follow a rationalist epistemology and use a case study as a research method.

KEYWORDS: Universal Decimal Classification; Research methods; Epistemological stances.

Introducción

La Clasificación Decimal Universal (CDU) es uno de los sistemas de clasificación bibliográfica más importantes del mundo, actualmente traducida a 57 idiomas y utilizada en 135 países (UDC Consortium 2024). La CDU es estudiada en universidades y cursos de Biblioteconomía, Información y Documentación de todo el mundo como uno de los aspectos básicos de la historia y realidad bibliotecaria. En la Universidad de León, la CDU se imparte en la asignatura *Sistemas de representación normalizada del contenido*, donde el libro *Dos princípios da Classificação Decimal Universal a uma prática harmonizada* (Simões et al. 2018) es utilizado como obra de cabecera para su aprendizaje. Maria da Graça de Melo Simões ha sido sin duda una de las mayores especialistas y referentes internacionales sobre la CDU, habiendo publicado además del citado otros trabajos esenciales como el libro *Classificação Decimal Universal: fundamentos e procedimentos* (Simões, 2008) o su tesis doctoral titulada *A representação e evolução de etnia na Classificação Decimal Universal* (Simões, 2010).

Conocí a Graça personalmente en el año 2015, cuando trabajaba en la Universidade Estadual Paulista en Brasil. Ella había sido invitada a nuestra universidad como especialista en clasificación e inmediatamente conectamos personal y profesionalmente. Desde el principio Graça demostró ser una persona integradora y generosa, siempre pensando en nuevas ideas para publicaciones y compartiendo contactos. De esta forma, ella me puso en contacto no sólo con colegas de la Universidade de Coimbra, como las profesoras Cristina Freitas y Maria Manuel Borges, sino también con investigadoras de otras instituciones, como mi actual colega Blanca Rodríguez Bravo de la Universidad de León, en España, y Luciana Gracioso de la Universidade Federal de São Carlos, en Brasil, además de estupendos alumnos de postgrado de los cuales de muchos terminé siendo codirector, como Luís Miguel Oliveira Machado, Patrícia de Almeida, Edina Rodrigues Lima y Rita Isabel Figueira Costa. Desde que conocí a Graça puedo decir que gran parte de mis publicaciones y contactos estuvieron coordinados por ella.

En 2015 también se celebró en Lisboa el *International UDC Seminar* (Martínez-Ávila, 2016), el encuentro bianual especializado en clasificación y la CDU que reúne a los principales especialistas del mundo y que es organizado por el *UDC Consortium* desde 1969. Pese a que no nos dio tiempo a preparar un trabajo para esta edición de 2015, sí fue aquí donde plantamos la semilla de nuestra contribución para el siguiente *International UDC Seminar* de 2017 junto a la actual doctora Patrícia de Almeida (Almeida et al., 2017). La edición de 2017, celebrado en Londres, es el último *International UDC Seminar* que ha sido organizado hasta la fecha.

En el presente capítulo analizo los métodos de investigación y las posturas epistemológicas de los trabajos publicados en el *International UDC Seminar* de 2015, el encuentro que impulsó nuestra colaboración y amistad. La importancia personal del estudio es doble: por un lado, el interés de Graça en investigar no sólo la clasificación y la CDU sino también las perspectivas y los principios metodológicos en la Ciencia de la Información y la Organización del Conocimiento (véase por ejemplo, Freitas y Simões, 2011; Simões, 2017); y, por otro lado, como tributo doble ya que sigue un método similar al que utilicé en el *festschrift* de mi codirectora de tesis Hope Olson (Martínez-Ávila y Beak 2016), otra de las personas que más me han marcado académica y profesionalmente. Además, el análisis de las posturas epistemológicas tiene una gran relación con la propia formación de Graça. El análisis se realiza sobre el citado encuentro de 2015 (cuyo tema era *Classification & Authority Control: Expanding Resource Discovery*) y no sobre el de 2017 en el que participamos ya que el tema del segundo fue las clasificaciones facetadas, (*Faceted Classification Today: Theory, Technology and End Users*) y este es un aspecto que iba a limitar el análisis epistemológico de los trabajos tal como se discute en Hjørland (2013a; 2014).

Metodología

Se analizaron los 17 trabajos completos publicados en el libro de actas del *International UDC Seminar* de 2015 (Slavic y Cordeiro, 2015). Aquellas presentaciones cuyo texto completo no fue incluido en el libro de actas

fueron descartadas para el análisis (ver Tabla 1), ya que a fecha de hoy tampoco están disponibles para consulta en la página web del congreso (<https://seminar.udcc.org/2015/programme.php>).

Tabla 1. Trabajos presentados en el *International UDC Seminar* de 2015

Autor(es)	Título	Texto completo
Buckland, Michael K.	Classification, links and contexts	Sí
Tillett, Barbara B.	Complementarity of perspectives for resource descriptions	Sí
Cordeiro, Maria Inês	Libraries, classifications and the network: bridging past and future	No
Freire, Nuno	Linking library data: contributions and role of subject data	No
Žumer, Maja; Zeng, Marcia Lei	Application of FRBR and FRSAD to classification systems	No
Green, Rebecca	Relational aspects of subject authority control: the contributions of classificatory structure	Sí
Ilik, Violeta	Distributed person data: using Semantic Web compliant data in subject name headings	Sí
Soergel, Dagobert; Popescu, Denisa	Organization authority database design with classification principles	Sí
Schöneberg, Ulf; Sperber, Wolfram	Machine-learning methods for classification and content authority control in mathematics	Sí
Balíková, Marie	Subject authority control supported by classification: the case of National Library of the Czech Republic	No
Pika, Jiri; Pika-Biolzi, Milena	Multilingual subject access and classification-based browsing through authority control: the experience of the ETH-Bibliothek	Sí
Vukadin, Ana	Development of a classification-oriented authority control: the experience of the National and University Library in Zagreb	Sí
Frâncu, Victoria; Dediu, Liviu-Iulian	TinREAD – an integrative solution for subject authority control	Sí
Pestana, Olívia	Alignment in medical sciences: towards improvement of UDC	Sí
Gnoli, Claudio; De Santis, Rodrigo; Pusterla, Laura	Commerce, see also Rhetoric: cross-discipline relationships as authority data for enhanced retrieval	Sí

Autor(es)	Título	Texto completo
Golub, Koraljka; Hansson, Joacim; Soergel, Dagobert; Tudhope, Douglas	Managing classification in libraries: a methodological outline for evaluating automatic subject indexing and classification in Swedish library catalogues	Sí
Piros, Attila	Automatic Interpretation of Complex UDC Numbers: Towards Support for Library Systems	Sí
Scharnhorst, Andrea; Smiraglia, Richard P.; Guéret, Christophe; Salah, Alkim Almila Akdag	Knowledge maps for libraries and archives – uses and use cases	No
Wang, Shenghui; Koopman, Rob	A second life for authority records	No
Freire, Nuno; Charles, Valentine; Isaac, Antoine	Subject information and multilingualism in European bibliographic datasets: experiences with Universal Decimal Classification	No
Barbalet, Suzanne	Enhancing subject authority control at the UK Data Archive: a pilot study using UDC	Sí
Cardillo, Elena; Solodovnik, Iryna; Taverniti, Maria	Towards the creation of integrated authority files in the domain of science and technology: an Italian use case	Sí
Ledl, Andreas	The Basel Register of Thesauri, Ontologies & Classifications (BARTOC)	Sí
Rozman, Darija	Experience with UDC updates: the Slovenian perspective	Sí
Kowalczyk, Agnieszka Maria; Skonieczny, Łukasz; Wornbard, Małgorzata	Visualization of a library collection based on UDC: research in the Warsaw University of Technology Main Library	No
Šauperl, Alenka	UDC as a standardisation method for providing titles of documents	No

Para el análisis de los textos se utilizó una metodología desarrollada en trabajos previos sobre análisis de posiciones epistemológicas y métodos de investigación en organización del conocimiento (Beak et al., 2015; Mattos et al., 2015; Martínez-Ávila y Beak, 2016). Para cada trabajo se determinó la posición epistemológica, los métodos o técnicas utilizadas y la existencia o no de sección de metodología. Mientras que la importancia del vínculo entre la epistemología y la metodología ha sido señalada por

diversos trabajos (por ej. Bradley, 1993; Martínez-Ávila et al., 2016), para el presente análisis de la posición epistemológica se trabajó con las cuatro escuelas epistemológicas utilizadas por Hjørland (por ej. 2009; 2013b) en la organización del conocimiento: empiricismo, racionalismo, historicismo y pragmatismo. Según este autor en el empiricismo se da importancia a aspectos como las observaciones, los datos obtenidos a través de los sentidos y las inducciones. En el racionalismo se da importancia a aspectos como la lógica, los modelos matemáticos, el raciocinio y la deducción, el modelado informático, los axiomas, teoremas, etc. En el historicismo se da importancia al contexto y desarrollo histórico, las perspectivas evolutivas y todo lo que rodea a las condiciones de producción. En el pragmatismo se da importancia a las metas y valores de la investigación y de los sujetos (dentro de esta corriente se incluirían muchas de las perspectivas críticas que adoptan puntos de vista de grupos específicos como el feminismo, etc.). Respecto a la clasificación de los métodos o técnicas utilizadas se utilizó una lista no cerrada ad-hoc desarrollada con base en la literatura (Järvelin and Vakkari, 1990; Creswell, 2007; Hider y Pymm, 2008; Cibangu, 2010; Chu, 2015). La lista de métodos contemplados a priori (aunque no todos identificados en los textos) incluye: análisis conceptual, análisis de contenido, escrito crítico, deconstrucción, análisis del discurso, análisis de documentos, enfoque teórico, bibliometría, análisis de redes sociales, análisis estadístico, método Delphi, experimento, entrevista, observación, cuestionario, análisis del modelo de bloques, estudio de caso, análisis de conversación, análisis de facetas, teoría fundamentada y otros. En aquellos casos en los que se identificó otro método no contemplado se incluyó en los resultados. En los estudios en los que más de un método fue utilizado se registraron todos de forma no excluyente.

Resultados

La mayoría de los trabajos analizados se encuadran dentro de una posición epistemológica racionalista (11 trabajos, 65%), seguido de una posición empiricista (4 trabajos, 23%) y una posición pragmatista (2 trabajos,

12%); ninguno de los trabajos fue identificado dentro de la corriente historicista. Estos datos pueden ser visualizados proporcionalmente en la Figura 1.

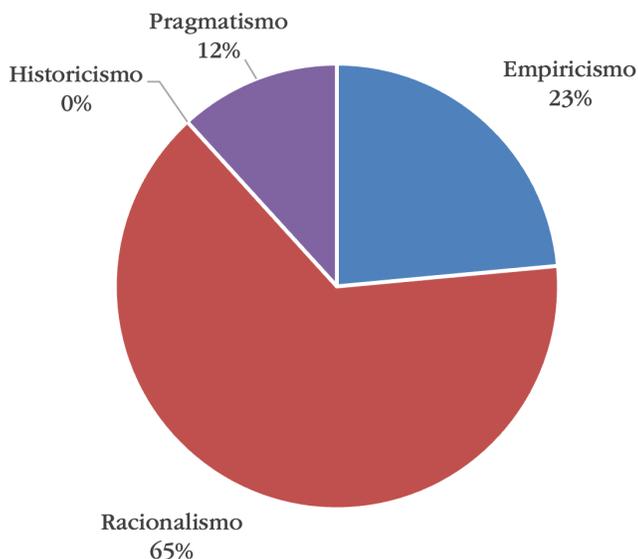


Fig. 1 – Posiciones epistemológicas

La predominancia de la corriente racionalista se considera un resultado esperado teniendo en cuenta la naturaleza facetada de la CDU y la epistemología subyacente a este aspecto (Hjørland, 2013a; 2014). Específicamente, los trabajos identificados con epistemología racionalista son los de Green, Ilik, Sorgel y Popescu, Schöneberg y Sperberg, Pika y Pika-Biolzi, Vukadin, Frâncu y Dediu, Gnoli et al., Piros, Cardillo et al. y Ledl, que abarcan temas como las estructuras y relaciones de clasificaciones, modelos conceptuales de datos de autoridad como FRASAD, ligación de datos de autoridad por bibliotecas y la web semántica, diseño de bases de datos, métodos de aprendizaje automático y matemáticas, navegación y búsqueda en OPACs, clasificación automática, construcción de notaciones con tecnologías semánticas y otros temas relacionados con aspectos computacionales y formales.

Los trabajos empiricistas aparecen en menor medida. Aquí se incluyen los trabajos de Tillet, Pestana, Barbalet y Rozman, que tratan temas relacionados con experiencias con datos bibliográficos y de autoridad,

revisión, actualización y comparación de clases de la CDU e incorporación de la CDU para datos de autoridad de archivos. Pese a que algunos de estos trabajos pudiesen presentar cierto eclecticismo epistemológico, todos ellos tienen como característica predominante la presentación y adquisición de conocimiento a través de las experiencias relatadas.

La corriente pragmatista se ve representada por los trabajos de Buckland y Golub et al.: el primero discute la tensión entre las relaciones estandarizadas de clasificaciones y la web semántica y los contextos culturales locales de los sujetos que las utilizan; y el segundo realiza una evaluación de la clasificación e indexación automática basándose, entre otros aspectos, en una perspectiva de análisis de dominio. Este segundo trabajo, realizado en coautoría por cuatro autores de diversas instituciones, se adscribe globalmente a una epistemología pragmatista debido a la importancia que se da en el texto al análisis de dominio como marco de referencia (concepto desarrollado por Birger Hjørland, considerado pragmatista), a pesar de que ciertos aspectos podrían ser interpretados como empiricistas o racionalistas. Este grado de eclecticismo del trabajo de Golub et al. podría ser explicado por la variedad de autores involucrados en la construcción del mismo y sus diversas posiciones epistemológicas.

La ausencia de trabajos desde una perspectiva historicista puede justificarse por el relativo bajo número de trabajos que componen la muestra de este evento. No obstante, en términos comparativos, este menor grado puede servir también para ilustrar la distribución epistemológica de la investigación sobre la CDU.

En lo que respecta a los métodos utilizados (Figura 2), la amplia mayoría de los trabajos (14, 82%) hace uso del estudio de caso. En el trabajo de Ilik se presenta el caso de VIVO, una herramienta de descubrimiento basada en tecnologías de la web semántica para enriquecer registros de autoridad desarrollada por la Cornell University; Soergel y Popescu presentan el diseño de la base de datos de autoridad del World Bank Group; Schöneberg y Sperberg revisan el caso de la base de datos bibliográfica especializada en matemáticas zbMATH, incluyendo las iniciativas de herramientas y prácticas automáticas de control de autoridades y clasificación en este contexto, así como la aplicación de aprendizaje automático para el desarrollo

de vocabularios controlados; Pika y Pika-Biolzi relatan el caso de NEBIS, el sistema de navegación de control de autoridades de materias y acceso de materias multilingüe de la ETH-Bibliothek, en Suiza; Vukadin revisa los problemas y desafíos en el desarrollo de la base de datos de autoridades de la CDU de la Biblioteca Nacional y Universitaria de Zagreb, en Croacia; Frâncu y Dediu presentan el caso de TinREAD, una solución integrada para control de autoridades de materias; Pestana estudia la revisión del número 61 de medicina de la CDU, comparándolo también con la National Library of Medicine Classification; Gnoli et al. presentan, entre otras cuestiones, el caso de SciGator, una interfaz para explorar relaciones interdisciplinarias en la Clasificación Decimal de Dewey desarrollado en la Universidad de Pavia, en Italia; Golub et al. discuten un proyecto para evaluar la indexación automática de recursos textuales suecos utilizando la Clasificación Decimal de Dewey y los Swedish Subject Headings; Piros presenta un XML schema específico de la CDU para construir números complejos; Barbalet investiga un estudio piloto de aplicación de la CDU para mejorar el control de autoridades de materias en el UK Data Archive; el estudio de Cardillo et al. relata la creación de un fichero de autoridades de nombres locales que integra listas de autoridades en el ámbito de ciencia y tecnología para su aplicación en el proyecto CNR S&TDL; Ledl revisa el caso de BARTOC, una base de datos bibliográfica de sistemas de organización del conocimiento desarrollada por la biblioteca de la Universidad de Basel, en Suiza; y Rozman presenta los casos de prácticas relacionadas con la CDU en Eslovenia. Dentro de los 14 trabajos que utilizan estudios de caso, las posiciones epistemológicas se distribuyen en porcentajes semejantes a los de la distribución total: 10 trabajos de epistemología racionalista (71%), 3 trabajos de epistemología empiricista (21%) y 1 trabajo de epistemología pragmatista (7%).

El resto de métodos utilizados aparecen en mucha menor medida: 2 trabajos (12%) hacen uso de análisis de documentos (Golub et al. y Cardillo et al., siguiendo posiciones pragmatista y racionalista respectivamente); 2 trabajos (12%) aplican análisis de facetas (Vukadin y Piros, ambos de epistemología racionalista); 2 trabajos (12%) usan un enfoque teórico (Buckland y Tillett, de epistemología pragmatista y empiricista

respectivamente); 1 trabajo (5%) utiliza análisis de dominio (Golub et al., pragmatista); y 1 trabajo (5%) aplica un análisis conceptual (Green, de epistemología racionalista).

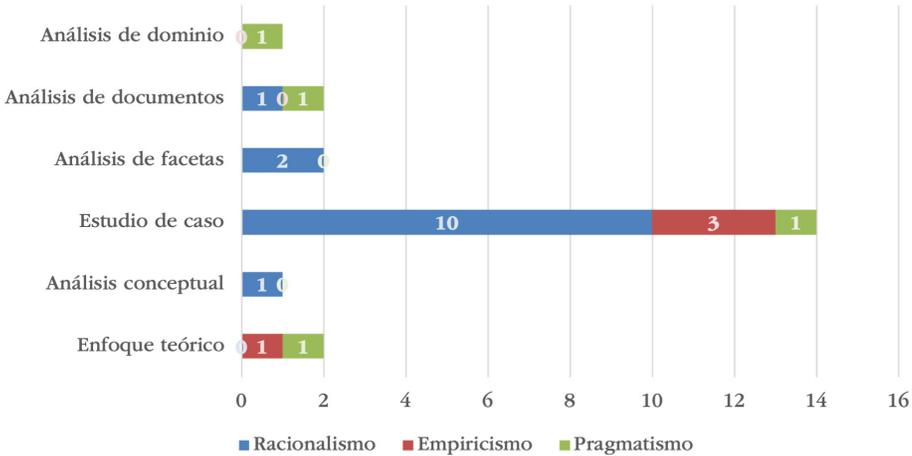


Fig. 2 – Frecuencia de métodos y posiciones epistemológicas

Teniendo en cuenta que se trata de un encuentro especializado sobre la CDU, no es sorprendente encontrar que la mayoría de los trabajos utilizan estudios de caso en los que se colocan, mayoritariamente, propuestas formales y experiencias relacionadas con el sistema. Pese a la tradicional lógica inductiva de los estudios de caso, en los que a partir de un caso particular se infieren las conclusiones generales, y que a priori sugeriría una mayoría de trabajos de epistemología empiricista, la predominancia de trabajos formales y computacionales, en los que primero se colocan los principios y a partir de ellos se deducen los casos propuestos, explica el resultado de una presencia mayoritaria de posiciones racionalistas en los trabajos analizados.

En lo que respecta a la sección de metodología, ésta aparece explícita en apenas dos estudios de caso (Golub et al. y Cardillo et al., de posiciones pragmatista y racionalista respectivamente), que además coinciden con los dos únicos trabajos que utilizan el método de análisis de documentos (y en el caso de Golub et al. también análisis de dominio) (ver Figura 3).

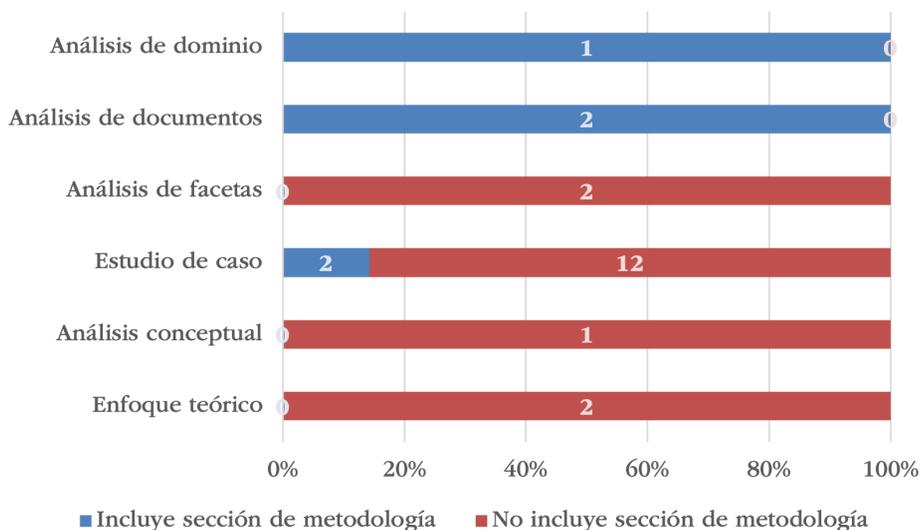


Fig. 3 – Porcentaje y frecuencia de trabajos que incluyen sección de metodología por método

Los resultados sobre la sección de metodología muestran que, pese a que en términos generales los trabajos no suelen incluir una sección dedicada a explicar cómo se ha realizado el estudio, en todos los casos en los que se identificó como método el análisis de documentos o el análisis de dominio sí se incluyó dicha sección, ya sea por la necesidad de explicar un método complejo o por la importancia que se dio al mismo. En los trabajos en los que se aplicó un análisis de facetas, análisis conceptual o enfoque teórico, y en la mayoría de los estudios de caso, no se incluyó dicha sección ya sea porque no se le dedicó suficiente importancia o porque no se necesitó tanto espacio para explicar el método.

Conclusión

La Clasificación Decimal Universal es un sistema bibliográfico internacional cuyos avances son desarrollados también a partir de las contribuciones científicas de una comunidad de autores de todo el mundo. Maria da Graça de Melo Simões es una de las autoras que han contribuido al cuerpo científico de este sistema no sólo con la elaboración de su

tesis doctoral y diversas publicaciones, sino también con la participación en uno de sus principales foros científicos sobre el sistema: los *International UDC Seminars*. La epistemología y la metodología de la investigación sobre el sistema afecta el desarrollo teórico y práctico del mismo. Por lo tanto, el presente análisis de una de las ediciones de este evento contribuye a entender las bases teóricas del mismo, revelando que la mayoría de las publicaciones analizadas sigue una epistemología racionalista y usa estudios de caso como principal método de investigación.

Referencias bibliográficas

- Almeida, Patrícia de, Simões, Maria da Graça, & Martínez-Ávila, Daniel. (2017). The Contribution of Ranganathan's Facets to the Determination of Aboutness in Novels. In Aida Slavic, & Claudio Gnoli (Eds.), *Faceted Classification Today: Theory, Technology and End Users: Proceedings of the International UDC Seminar 2017, London (UK), 14-15 September* (pp. 93-107). Ergon Verlag.
- Beak, Jihee, Choi, Inkyung, & Lee, Tae Hee. (2015). Research Methods in KO Domain: Pilot Study. Poster presented at *The Big Deal: 3rd Milwaukee Conference on Ethics in Knowledge Organization, Milwaukee, United States, May 28-29*.
- Bradley, Jana. (1993). Methodological Issues and Practices in Qualitative Research. *Library Quarterly* 63(4), 431-449.
- Chu, Heting. (2015). Research Methods in Library and Information Science: A Content Analysis. *Library & Information Science Research*, 37(1), 36-41. <https://doi.org/10.1016/j.lisr.2014.09.003>
- Cibangu, Sylvain K. (2010). Paradigms, Methodologies, and Methods. *Library & Information Science Research*, 32(3), 177-8. <https://doi.org/10.1016/j.lisr.2010.03.006>
- Creswell, John W. (2007). *Qualitative Inquiry & Research Design: Choosing Among Five Approaches*. Second edition. Sage.
- Freitas, Maria Cristina Vieira de, & Simões, Maria da Graça de Melo. (2011). Perspectivas metodológicas inovadoras para a ciência da informação: Dois estudos desenvolvidos no âmbito do programa de doutoramento da universidade de Salamanca. In Antonio Pulgarín Guerrero, & Agustín Vivas Moreno (Eds.), *Límites, fronteras y espacios comunes encuentros y desencuentros en las Ciencias de la Información* (pp. 438-448). Universidad de Extremadura.
- Hider, Philip, & Pymm, Bob. (2008). Empirical Research Methods Reported in High-Profile LIS Journal Literature. *Library & Information Science Research*, 30(2), 108-114. <https://doi.org/10.1016/j.lisr.2007.11.007>
- Hjørland, Birger. (2009). Concept Theory. *Journal of the American Society for Information Science and Technology*, 60(8), 1519-1536. <https://doi.org/10.1002/asi.21082>
- Hjørland, Birger. (2013a). Facet Analysis: The Logical Approach to Knowledge Organization. *Information Processing and Management*, 49(2), 545-557. <https://doi.org/10.1016/j.ipm.2012.10.001>
- Hjørland, Birger. (2013b). Theories of Knowledge Organization—Theories of Knowledge. *Knowledge Organization*, 40(3), 169-181. <https://doi.org/10.5771/0943-7444-2013-3-169>

- Hjørland, Birger. (2014). Is Facet Analysis Based on Rationalism? A Discussion of Satija (1992), Tennis (2008), Herre (2013), Mazzocchi (2013b), and Dousa & Ibekwe-SanJuan (2014). *Knowledge Organization*, 41(5), 369-376. <https://doi.org/10.5771/0943-7444-2014-5-369>
- Järvelin, Kalervo, & Vakkari, Pertti. (1990). Content Analysis of Research Articles in Library and Information Science. *Library & Information Science Research*, 12(4), 395-422.
- Martínez-Ávila, Daniel. (2016). Classification & Authority Control: Expanding Resource Discovery. *Knowledge Organization*, 43(1), 64-68. <https://doi.org/10.5771/0943-7444-2016-1-64>
- Martínez-Ávila, Daniel, & Beak, Jihee. (2016). Methods, Theoretical Frameworks and Hope for Knowledge Organization. *Knowledge Organization*, 43(5), 358-366. <https://doi.org/10.5771/0943-7444-2016-5-358>
- Martínez-Ávila, Daniel, Semidão, Rafael, & Ferreira, Marcio. (2016). Methodological Aspects of Critical Theories in Knowledge Organization. *Knowledge Organization* 43(2). 118-125, <https://doi.org/10.5771/0943-7444-2016-2-118>
- Mattos, Lorena Rocha de, Favaretto, Danton Wesley, Martínez-Ávila, Daniel, & Beak, Jihee. (2015). Fundamentos teóricos y aspectos metodológicos en el dominio de la International Society for Knowledge Organization (ISKO). Presented at *IX Encontro Internacional de Informação, Conhecimento e Ação (EIICA), Marília, Brasil, December, 2-4*.
- Simões, Maria da Graça. (2008). *Classificação Decimal Universal – Fundamentos e Procedimentos*. Almedina.
- Simões, Maria da Graça de Melo. (2010). A representação de Etnia e a sua evolução na Classificação Decimal Universal. [Tesis doctoral, Universidad de Salamanca]. <https://doi.org/10.14201/gredos.83331>
- Simões, Maria da Graça de Melo. (2017). Do Conceito e dos Princípios Metodológicos de Indexação. Subsídios para uma Análise Crítica. In Maria José Vicentini Jorente, & Dunia Llanes Padrón (Eds.), *Una Mirada a la ciencia de la información desde los nuevos contextos paradigmáticos de la posmodernidad* (pp. 151-176). Oficina Universitária; Cultura Acadêmica.
- Simões, Maria da Graça, Rodríguez-Bravo, Blanca, & Ferreira, Carla. (2018). *Dos princípios da Classificação Decimal Universal a uma prática harmonizada*. Imprensa da Universidade de Coimbra. <https://doi.org/10.14195/978-989-26-1378-9>
- Slavic, Aida, & Cordeiro, Maria Inês. (2015). *Classification & authority control: expanding resource discovery: proceedings of the International UDC Seminar, 29-30 October 2015, Lisbon, Portugal*. Ergon.
- UDC Consortium. (2024). *UDC Fact Sheet*. <https://udcc.org/index.php/site/page?view=factsheet>

Vocabulários controlados, descontrolados e degradados: uma breve análise dialógica do impacto da inteligência artificial no fenômeno da linguagem, pesquisa, ensino e aprendizagem

Renato Rocha Souza

ORCID 0000-0002-1895-3905

Fundação Getulio Vargas/CPDOC, Universidade de Viena, Brasil / Áustria

renato.souza@fgv.br

RESUMO: O artigo discute as implicações da Inteligência Artificial na organização do conhecimento, com foco especial na evolução dos vocabulários controlados e no surgimento de novas ferramentas como os Large Language Models. O texto contém uma homenagem à pesquisadora Maria da Graça de Melo Simões, que dedicou sua carreira à organização do conhecimento e aos vocabulários controlados. Em seguida, compara a precisão e estrutura dos vocabulários controlados com a flexibilidade das folksonomias, destacando os benefícios e desafios de cada abordagem. A principal tese do artigo é que a Inteligência Artificial, especialmente os modelos de linguagem, está revolucionando a forma como organizamos e acessamos a informação. Estes são capazes de gerar textos coerentes e relevantes, realizar traduções, responder a perguntas complexas e até mesmo criar novos conteúdos. No entanto, o uso indiscriminado dessas ferramentas levanta preocupações sobre a originalidade, a confiabilidade e a diversidade da produção acadêmica.

PALAVRAS-CHAVE: Inteligência Artificial (IA); Organização do Conhecimento; Vocabulários Controlados; Large Language Models (LLMs); Produção Acadêmica.

ABSTRACT: The article discusses the implications of Artificial Intelligence on knowledge organization, with a special focus on the evolution of controlled vocabularies and the emergence of new tools such as Large Language Models. The text includes a tribute to the researcher Maria da Graça de Melo Simões, who dedicated her career to knowledge organization and controlled vocabularies. It then compares the precision and structure of controlled vocabularies with the flexibility of folksonomies, highlighting the benefits and challenges of each approach. The main thesis of the article is that Artificial Intelligence, especially language models, is revolutionizing the way we organize and access information. These models are capable of generating coherent and relevant texts, translating languages, answering complex questions, and even creating new content. However, the indiscriminate use of these tools raises concerns about originality, reliability, and diversity in academic production.

KEYWORDS: Artificial Intelligence (AI); Knowledge Organization; Controlled Vocabularies; Large Language Models (LLMs); Academic Production.

Este artigo começa com uma nota pessoal. Será apresentado em um evento em homenagem a uma grande amiga, a pesquisadora Maria da Graça de Melo Simões da Universidade de Coimbra, por ocasião dos 5 anos de sua partida prematura de nossa convivência. Graça, como a chamávamos, tinha em sua agenda de pesquisa o amplo campo da organização de conhecimento, e especial predileção pelo tópico dos vocabulários controlados. Gostava também de aludir às linguagens “descontroladas”, tais como folksonomias e outros tipos de *social tagging*, que se tornaram tão comuns com o advento das *hashtags* em redes sociais eletrônicas. Para uma melhor contextualização, **os vocabulários controlados desempenham um papel crucial na ciência da informação, fornecendo uma estrutura organizada e padronizada para organizar e recuperar informações. Eles aumentam a recuperabilidade da informação, promovendo consistência, precisão, interoperabilidade, gerenciamento de conhecimento e mineração de dados. Esses benefícios são essenciais para a comunicação eficaz, a colaboração, a integração de sistemas, a pesquisa precisa, a preservação do conhecimento e a análise de dados.**¹ Folksonomias, por sua vez, são sistemas

¹ Describe Controlled vocabularies in one paragraph, highlighting the importance for Information Science. Translate to Portuguese.

de classificação de informações criados colaborativamente por um grupo de pessoas. Ao invés de seguirem uma estrutura rígida e pré-definida, como em uma taxonomia tradicional, permitem que cada usuário atribua etiquetas (tags) livremente aos conteúdos, como textos, imagens ou vídeos. Essas tags, escolhidas de forma natural e espontânea, servem como marcadores para organizar e recuperar informações de forma personalizada e intuitiva. A grande vantagem das folksonomias é a sua flexibilidade e adaptabilidade, pois elas evoluem de acordo com as necessidades e interesses da comunidade que as utiliza.²

Como um olhar treinado pode ter notado, o trecho anterior de texto grafado em verde, com sua estrutura asséptica e ausência de “alma”, foi originalmente gerado pelo Gemini³, um *Large Language Model* (LLM) da Google, com *prompts* em Português e Inglês, e posteriormente modificados pelo autor. Nesse artigo, em consonância com a homenagem aos temas de pesquisa de nossa amiga e colega, propõe-se uma abordagem dialógica em que o autor apresentará o fenômeno da transformação dos vocabulários decorrente do uso de Inteligência Artificial (IA) entremeando o discurso e a apresentação de ideias com *prompts* que geraram respostas de LLMs. Todo o texto original que for gerado por IA será marcado em laranja, e quando apresentarmos texto gerado por IA que foi posteriormente modificado pelo autor, este será marcado em verde. Também serão apresentados, nas notas de rodapé, os *prompts* originais que foram utilizados para gerar os textos. Desta forma, tornar-se-ão claras a distinção e a alegoria almejadas no artigo, permitindo tanto a reflexão e a reprodutibilidade. Este ensaio também aborda, de maneira marginal, as transformações da produção acadêmica por consequência destas ferramentas, como atitudes de docentes e educadores frente ao uso indiscriminado de tecnologias por alunos e o decaimento da diversidade linguística devido ao processo de treinamento de modelos cada vez maiores.

Pouco sabíamos que, uma década após as preocupações sobre o efeito do social tagging para recuperação da informação – em contraste aos

² Explique o que são folksonomias em um parágrafo conciso.

³ <https://gemini.google.com/app>

vocabulários controlados e ontologias formais, criados por especialistas em assuntos – que estes modelos de linguagem viessem a mudar novamente a forma com que nos relacionamos com informações, fatos e documentos. Esta nova tecnologia, ao invés de interferir somente na forma como recuperamos documentos, os transforma em abstratas representações binárias, armazenadas em colossais matrizes de números, com conceitos embutidos em seus parâmetros internos. Em última instância, tornam por vezes desnecessários os documentos originalmente utilizados para treiná-los, a ponto de gerar uma perspectiva de degradação e empobrecimento às linguagens humanas, decorrentes da adoção desenfreada de LLMs para a produção de conhecimento de forma ampliada.

Estes modelos *tornaram-se cada vez mais comuns na produção acadêmica, oferecendo aos pesquisadores uma ferramenta poderosa para tarefas como revisões de literatura, redação de propostas e até mesmo a elaboração de artigos. Embora os LLMs possam agilizar esses processos e aumentar a produtividade, seu uso levanta preocupações significativas. Os críticos argumentam que os LLMs podem introduzir vieses na pesquisa, pois são treinados em vastos conjuntos de dados que podem conter vieses inerentes. Além disso, a dependência dos LLMs pode diminuir a originalidade e as habilidades de pensamento crítico dos pesquisadores, levando a uma homogeneização da produção acadêmica. Além disso, o potencial de uso indevido, como plágio e a geração de conteúdo enganoso, apresenta desafios éticos.*⁴

Educadores (UNESCO, s.d., Holmes, 2023) têm debatido a necessidade de adaptarmos os modos de ensino e aprendizagem tradicionais às ferramentas de IA gerativa. Kobak et al. (2024) apresentam uma análise contemporânea bastante abrangente, *analisando a frequência de palavras de estilo em resumos acadêmicos para estimar o uso de modelos avançados de linguagem (LLMs) na literatura científica. Os resultados sugerem que pelo menos 10% dos resumos publicados em 2024 foram processados com LLMs, com variações entre disciplinas, países e periódicos. Esse impacto*

⁴ Discuss in one paragraph the use of Large Language Models in academic production with a critical point of view. Then translate to Portuguese.

é considerado sem precedentes, superando o de grandes eventos mundiais.⁵

Os autores também identificam palavras que, por conta de sua inesperada frequência, indicam a autoria “artificial”, embora esta informação nunca seja explicitada. Muitos outros pesquisadores (Baskara, 2023; Rifaina & Sulisworo, 2023; Zirar, 2023, Kalla et al., 2023) exploram os efeitos do uso de modelos de linguagem na produção acadêmica, ensino e pesquisa, explicitando as vantagens e potenciais problemas.

Bhaskara (2023) defende que o uso de ChatGPT no aprendizado de idiomas pode levantar preocupações sobre a possível substituição de professores de idiomas humanos e as implicações éticas do uso de um sistema de aprendizado de máquina para geração de texto. Rifaina & Sulisworo (2023) defendem que o ChatGPT pode ser uma ferramenta útil para aprendizado, ajudando com tarefas como pesquisa, tradução e geração de perguntas. No entanto, ressaltam a importância de verificar as informações fornecidas pela ferramenta com fontes confiáveis e manter uma abordagem crítica. Concluem que embora tenha limitações, os usuários geralmente consideram que o ChatGPT pode melhorar a produtividade e a eficiência do aprendizado. Zirar (2023) conclui que que modelos de linguagem podem influenciar a forma como os alunos aprendem e são avaliados. Embora possam ser úteis, a dependência excessiva pode prejudicar o aprendizado. Também afirma que modelos podem gerar material de ensino e avaliação, mas devem ser usados com cautela, e que alunos e educadores precisam verificar a qualidade do material gerado para que

⁵ Traduza e resuma o texto abaixo:

===

Recent large language models (LLMs) can generate and revise text with human-level performance, and have been widely commercialized in systems like ChatGPT. These models come with clear limitations: they can produce inaccurate information, reinforce existing biases, and be easily misused. Yet, many scientists have been using them to assist their scholarly writing. How wide-spread is LLM usage in the academic literature currently? To answer this question, we use an unbiased, large-scale approach, free from any assumptions on academic LLM usage. We study vocabulary changes in 14 million PubMed abstracts from 2010-2024, and show how the appearance of LLMs led to an abrupt increase in the frequency of certain style words. Our analysis based on excess words usage suggests that at least 10% of 2024 abstracts were processed with LLMs. This lower bound differed across disciplines, countries, and journals, and was as high as 30% for some PubMed sub-corpora. We show that the appearance of LLM-based writing assistants has had an unprecedented impact in the scientific literature, surpassing the effect of major world events such as the Covid pandemic.

a integração seja benéfica. Kalla et al. (2023) apresentam um profundo estudo de impactos de uso de LLMs para a academia, cibersegurança, suporte ao cliente, desenvolvimento de software, empregos e tecnologia da informação, bem como suas possíveis aplicações para pesquisadores e estudantes. Dentre as visões positivas e negativas, fica claro a paralisia e falta de consenso em como agir, o que se reflete nos frequentes debates nas cenas universitárias e de pesquisa. Muitos acadêmicos tradicionais têm defendido o banimento do uso de ferramentas de IA, e a punição aos estudantes, utilizando-se de ferramentas para detecção de textos gerados por inteligência artificial – uma abordagem radical que ignora completamente a mudança paradigmática em curso. Para este público, muitas empresas têm apresentado ferramentas para detecção de IA, mas os resultados nem sempre são confiáveis. Segundo Harwell (2023) a ferramenta Turnitin⁶ foi usada para avaliar 38 milhões de ensaios estudantis para detectar uso de IA em apenas 3 meses (entre Abril e Junho de 2023), mas a empresa admite que seu algoritmo sinaliza incorretamente 4% das frases analisadas. Isso representa potencialmente 1,8 milhão de ensaios falsamente rotulados como “conteúdo de IA”. Em um outro experimento, reportado na mesma fonte, descobriu-se que a Turnitin errou em mais de 50% das vezes. Um documento da OpenAI (2024) admite que os detectores de IA não funcionam de forma confiável, e embora a própria OpenAI tenha lançado ferramentas que afirmam detectar conteúdo gerado por IA, nenhuma delas provou ser capaz de distinguir de forma inequívoca o conteúdo gerado por IA e conteúdo gerado por seres humanos. Fleckenstein et al. (2024) reforçam o fato que professores têm pouca capacidade de distinguir textos gerados por humanos e por IA, gerando conflitos, falsas acusações e insegurança acadêmica. Este novo horizonte é, por si só, um campo bastante profícuo de pesquisas. Mas existe um outro efeito subjacente, com consequências de longo prazo, fundamentado na própria maneira como se treinam os modelos de linguagem.

LLMs, como se sabe, são treinados em vastas quantidades de texto extraído da web, como livros, artigos, código e páginas da web. Esse processo

⁶ <https://www.turnitin.com/>

envolve a coleta de corpora com bilhões de palavras, permitindo que aprendam padrões linguísticos, gramática, fatos e até mesmo nuances de estilo. Ao serem expostos a essa diversidade de informações, os LLMs desenvolvem a capacidade de gerar textos coerentes e contextualmente relevantes, traduzir idiomas, responder a perguntas e realizar diversas outras tarefas de processamento de linguagem natural. É como ensinar uma criança a ler e escrever, mas em uma escala muito maior e com uma velocidade exponencial.⁷ Uma vez treinados, tornam-se uma espécie de “caixa-preta” que utiliza estatística de frequências de palavras para geração de sentenças, além de utilizar técnicas de RLHF (Reinforcement Learning from Human Feedback) para calibragem. RLHF é uma técnica de treinamento de modelos de linguagem que permite que eles aprendam a gerar texto de forma mais alinhada com as preferências humanas. Em vez de apenas serem treinados em grandes volumes de dados, esses modelos são refinados através de interações com humanos. Os humanos fornecem feedback sobre as respostas geradas pelo modelo, indicando quais são mais relevantes, precisas e úteis. Com base nesse feedback, o modelo ajusta seus parâmetros internos para produzir resultados que sejam mais satisfatórios para os usuários. Esse processo iterativo permite que os modelos de linguagem evoluam e se tornem cada vez mais capazes de gerar texto que seja coerente, relevante e seguro. O RLHF é um componente crucial na criação de chatbots e assistentes virtuais que podem interagir de forma natural e útil com os seres humanos.⁸

A característica gerativa, notadamente estatística, e emulada a partir de trilhões de parâmetros, faz com que seja impossível entender e prever completamente o funcionamento de LLMs, o que pode levar a alucinações. Alucinações em modelos de linguagem são um fenômeno em que o modelo gera texto que parece coerente e relevante, mas que na verdade é falso ou sem fundamento. Isso acontece porque, embora esses modelos sejam treinados em vastas quantidades de dados, eles não compreendem

⁷ Descreva em um parágrafo como grandes modelos de linguagem usam dados disponíveis na web para treinamento.

⁸ Explique RLHF em um parágrafo.

o mundo da mesma forma que um humano. Em vez disso, aprendem a identificar padrões nos dados e a gerar texto que se encaixe nesses padrões. Por exemplo, se um modelo é perguntado sobre um evento histórico que nunca ocorreu, ele pode gerar uma resposta detalhada e convincente, mas completamente falsa. Essa capacidade de gerar texto falso pode ser problemática, especialmente em aplicações onde a precisão da informação é crucial, como na geração de notícias ou na resposta a perguntas complexas. A compreensão das alucinações é fundamental para o desenvolvimento de modelos de linguagem mais robustos e confiáveis.⁹ A impossibilidade de realizar assessoria na capacidade gerativa dos modelos faz com que seja impossível garantir que alucinações sejam eliminadas em qualquer futuro próximo, usando as tecnologias atuais. Um interessante estudo da Anthropic¹⁰, criadora do LLM Claude¹¹ (Templeton, 2024) mostra como os padrões de ativação das redes neurais subjacentes podem ser mapeados em termos de conceitos e *features*, mas este tipo de acesso às “entranhas” de um modelo demanda recursos computacionais consideráveis e não está disponível para o pesquisador comum. A tendência é de que modelos sejam cada vez maiores, o que os torna cada vez mais opacos. Esta necessidade de treinar modelos cada vez maiores – atualmente envolvendo algumas centenas de trilhões de parâmetros – demanda que mais e mais documentos multimídia sejam utilizados para a parametrização adequada, o que desencadeia a necessidade de geração de textos sintéticos para treinar estes novos e gigantes modelos. Enquanto dados reais são coletados a partir de bases de dados, redes sociais, bases de conhecimento, dentre outras milhares de fontes, dados sintéticos são gerados artificialmente a partir de dados reais. Um conjunto de dados sintéticos de qualidade pode apresentar estruturas e distribuições estatísticas semelhantes aos dados originais, tornando-os indistinguíveis. Quando os dados reais são insuficientes em qualidade, volume ou variedade, dados sintéticos podem ser usados para treinar

⁹ Explique alucinações em modelos de linguagem em um parágrafo conciso.

¹⁰ <https://www.anthropic.com/>

¹¹ <https://claude.ai/>

modelos de IA ou Machine Learning (ML). Além disso, dados sintéticos podem ser uma ferramenta crucial para iniciativas de IA empresarial, especialmente quando os dados existentes não atendem aos requisitos comerciais ou comprometem a privacidade.¹² Pelo menos em teoria.

Na prática, estas distribuições estatísticas apreendidas pelos modelos para geração de novos textos são diferentes, e tendem a reduzir a riqueza, criatividade e plasticidade do vocabulário original. Hao et al. (2024) explicam que o uso de dados sintéticos em modelos de IA apresenta desafios e riscos significativos como a limitação da capacidade dos modelos de generalizar para o mundo real, levando a resultados imprecisos e viesados. Além disso, levanta questões éticas sobre a geração de conteúdo falso e a responsabilidade da IA. Em termos de segurança, a introdução de dados sintéticos torna os modelos mais vulneráveis a ataques, comprometendo a confiabilidade dos sistemas. Por fim, o uso de dados sintéticos em setores regulamentados, como o financeiro, pode gerar conflitos com as exigências legais de transparência e interpretabilidade.¹³ Numa outra

¹² Crie uma justificativa pelo qual modelos de linguagem têm usado dados sintéticos para treinamento. Use o texto abaixo como contexto

===

“In contrast to “real” data gathered through real-world surveys or events, synthetic data, as its name implies, is essentially “fake” data that is artificially generated. Real data is used to create this synthetic data. An excellent synthetic dataset can capture the underlying structure and show identical statistical distributions as the original data, making them indistinguishable. When real historical data sets are insufficient in quality, volume, or variety, artificially created data might be used to train AI/ML models in their place. When existing data doesn’t satisfy business requirements or could compromise privacy if utilized to develop ML models, test software, or the like, synthetic data can be a crucial tool for enterprise AI initiatives.”

¹³ Crie um resumo em português em apenas um parágrafo para o texto abaixo:

===

“Widespread use of synthetic data may constrain the generalization performance of AI models. For instance, datasets like PersonX, generated within a gaming data engine, may deviate significantly from real-world data. In natural language processing, relying on fine-tuning with large language models may restrict downstream tasks to the performance and biases of the selected model. In healthcare, an abundance of non-real cases during model training may undermine confidence in diagnostic results among healthcare professionals and patients. The use of synthetic data may give rise to ethical and social concerns. Creating fictional characters or scenarios through synthetic data raises questions about AI responsibility in generating fictional content, potentially leading to misinformation, misunderstandings, or the dissemination of false information with detrimental societal impacts. The introduction of synthetic data brings forth security and adversarial attack risks. Malicious use of synthetic data may render AI models unstable during adversarial attacks, as models may not adequately learn the complexity and diversity of the real world during training. This susceptibility increases the likelihood of deception or manipulation, posing threats to the credibility and security of the

perspectiva, Guo et al. (2023) revelam uma diminuição consistente na diversidade das respostas dos modelos de linguagem após iterações sucessivas, especialmente notável para tarefas que exigem altos níveis de criatividade. Essa tendência destaca os riscos potenciais do treinamento de modelos de linguagem em texto sintético, particularmente no que diz respeito à preservação da riqueza linguística. Também destacam a necessidade de consideração cuidadosa dos efeitos de longo prazo de tais abordagens de treinamento nas capacidades linguísticas dos modelos de linguagem. Segundo estimativa (Villalobos et al., 2022), o suprimento total de texto de alta qualidade no mundo chega a cerca de 17 trilhões de tokens, com uma taxa de crescimento anual de 4 a 5%. Isso inclui todos os livros do mundo, artigos científicos, notícias, páginas da Wikipédia, código disponível e o restante do conteúdo filtrado da web. Mas nem todo este repertório pode ser suficiente, sendo que podemos estar nos aproximando do esgotamento de todo o estoque mundial de dados de treinamento de linguagem utilizáveis, potencialmente em uma ordem de magnitude. Consequentemente, as gerações subsequentes de modelos estão sendo pré-treinadas com dados heterogêneos, dentre textos criados por seres humanos, textos sintéticos criados por máquinas e textos sintéticos filtrados para excluir exemplos ruidosos e de má qualidade.

Como mostrado na Fig. 1, *com as iterações sucessivas de treinamento, os modelos de linguagem podem apresentar uma diminuição na diversidade sintática dos textos gerados. Isso pode ser devido ao esquecimento catastrófico e à falta de exposição a novos elementos de aprendizado, como ocorre com os resumos escritos por humanos.*¹⁴ O esquecimento catastró-

system. Legal Compliance Challenges: In certain domains, the use of synthetic data may present challenges regarding legal compliance. For instance, employing synthetic data for risk assessment in the financial sector may encounter regulatory hurdles. Regulatory authorities often require transparent and interpretable models, and the synthetic data generation process may face difficulties in meeting these standards.”

¹⁴ Traduza e resuma o texto abaixo:

===

“However, as the iterations advance, the syntactic diversity scores of the texts produced by the models eventually decline, dropping below those of human-written abstracts. This trend might be partly attributed to catastrophic forgetting (Mc-Closkey and Cohen, 1989). Additionally, while human-written abstracts may have limited syntactic diversity, their structure is markedly different from the pretraining data, thus introducing new learning elements for the model.

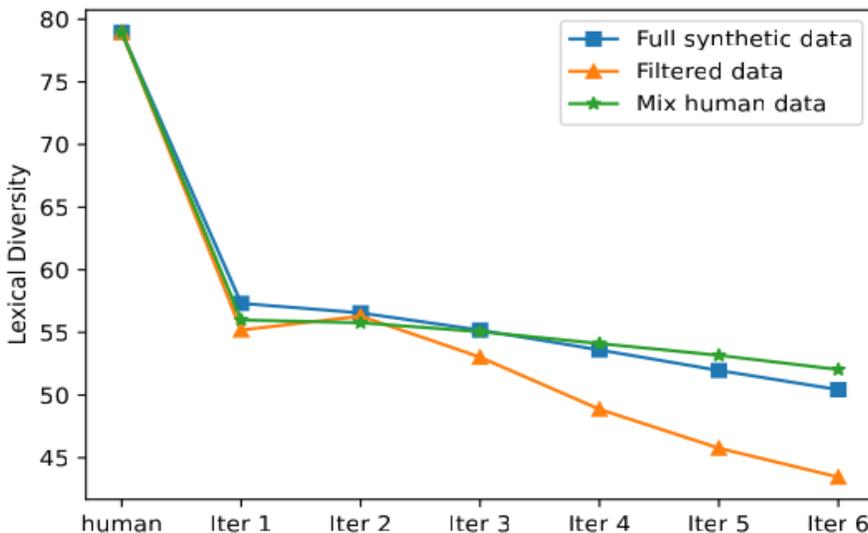


Fig. 1 – Diminuição da diversidade lexical na medida em que se treinam modelos com diferentes tipos de dados. Fonte: Guo et al. (2023)

fico é um fenômeno no aprendizado de máquina em que um modelo treinado em uma tarefa esquece como executar tarefas nas quais foi previamente treinado. No contexto de grandes modelos de linguagem (LLMs), isso significa que, à medida que o modelo é treinado em novos dados, ele pode começar a gerar texto que é inconsistente ou até mesmo contraditório com as informações que aprendeu anteriormente.¹⁵

Guo et al., 2023 concluem fazendo ressalvas a uma tendência preocupante: à medida que os modelos de linguagem dependem cada vez mais do texto gerado por predecessores para treinamento, há um risco tangível de diminuição da riqueza e variedade linguística em suas saídas, e enfatizam a necessidade de uma abordagem mais cautelosa e teleológica no desenvolvimento de modelos de linguagem, enfatizando a importância de preservar a diversidade linguística ao lado da melhoria do desempenho técnico.

In contrast, the synthetic data produced by Model (2), despite its marginally higher internal syntactic diversity, closely mirrors the model's own training distribution. This lack of novel information leads to a subsequent reduction in variation."

¹⁵ Explain what is "catastrophic forgetting" in the context of Large Language Models and translate the response to Portuguese.

Em um outro ponto de vista, mais positivo e propositivo, podemos reconhecer que estas ferramentas condensam uma amostra tão rica do vocabulário humano, que é extremamente fácil desempenhar tarefas comuns e historicamente custosas, no campo da Ciência da Informação. Para além das bastante decantadas sumarização, tradução e correção ortográfica, para as quais há inúmeros exemplos, até mesmo neste pequeno artigo, há exemplos não tão triviais como a) realizar o levantamento de termos pertinentes em um campo de conhecimento; ou b) extrair entidades nomeadas de um determinado documento; ou mesmo c) extrair relações semânticas e construção de triplas de (sujeito, predicado, objeto).

À guisa de exemplo, vamos mostrar os resultados obtidos para algumas destas tarefas utilizando o modelo ChatGPT 4o-mini, da OpenAI.

a) Levantamento de termos pertinentes em um campo de conhecimento.

Utilizamos o seguinte prompt:

“Assume you are a specialist in knowledge organization Please provide a list of terms to compose a controlled vocabulary about archival sciences Present the result as a table with the original term and the translations to Portuguese and Spanish. Limit your results to the 30 most pertinent concepts.”

Obtivemos o resultado:

As a specialist in knowledge organization, here is a list of 30 pertinent terms to compose a controlled vocabulary in archival sciences, along with their translations to Portuguese and Spanish:”

English Term	Portuguese Translation	Spanish Translation
Archival Description	Descrição Arquivística	Descripción Archivística
Archival Processing	Tratamento Arquivístico	Tratamiento Archivístico
Archive	Arquivo	Archivo
Archivist	Arquivista	Archivero/Archivera
Collection	Coleção	Colección

English Term	Portuguese Translation	Spanish Translation
Conservation	Conservação	Conservación
Digital Archive	Arquivo Digital	Archivo Digital
Digital Preservation	Preservação Digital	Preservación Digital
Document	Documento	Documento
Finding Aid	Instrumento de Pesquisa	Guía de Archivos
Fonds	Fundo	Fondo
Metadata	Metadados	Metadatos
Original Order	Ordem Original	Orden Original
Permanent Retention	Guarda Permanente	Retención Permanente
Preservation	Preservação	Preservación
Provenance	Proveniência	Procedencia
Record	Registro	Registro
Record Group	Grupo de Arquivos	Grupo de Archivos
Records Management	Gestão de Documentos	Gestión Documental
Reference Code	Código de Referência	Código de Referencia
Retention Schedule	Tabela de Temporalidade	Tabla de Retención
Series	Série	Serie
Storage	Armazenamento	Almacenamiento
User Access	Acesso do Usuário	Acceso del Usuario
Appraisal	Avaliação	Valoración
Confidentiality	Confidencialidade	Confidencialidad
Data Migration	Migração de Dados	Migración de Datos
Disposition	Destinação	Disposición
Ephemeral Records	Documentos Efêmeros	Documentos Efímeros
Transfer Protocol	Protocolo de Transferência	Protocolo de Transferencia

This controlled vocabulary aims to cover essential concepts in archival sciences, facilitating multilingual access and understanding in the field.

Obviamente, esta análise poderia se estender à exaustão, permitindo a coleta de termos de virtualmente qualquer área do conhecimento, com traduções automáticas e qualquer tipo de formato a ser exportado.

Vamos experimentar com uma segunda tarefa:

b) extrair entidades nomeadas de um determinado documento

Utilizamos o seguinte prompt:

“Assume you are an expert analyst, capable of identifying named entities in news documents

Extract the named entities in the excerpt of text below.

===

“A Universidade de Coimbra GCSE (UC) é uma universidade pública localizada na cidade de Coimbra, em Portugal. É uma das universidades mais antigas do mundo ainda em operação, sendo a mais antiga e uma das maiores do país.[3] Composta por 3 polos, 8 faculdades e 18 museus, a instituição conta ainda com o Jardim Botânico e o Estádio Universitário de Coimbra em um espaço com 25 188 alunos em 2020.[carece de fontes] A sua história remonta ao século seguinte ao da fundação da nação portuguesa, dado que foi criada a 1 de Março de 1290, quando o rei D. Dinis I assinou na cidade de Leiria o documento Scientiae thesaurus mirabilis, criando a universidade, o qual foi intermediado e confirmado pelo Papa. Fixada definitivamente na cidade de Coimbra em 1537, sete anos depois, todas as suas faculdades se instalam no antigo Paço Real da Alcáçova (denominado Paço das Escolas, após a sua aquisição pela Universidade de Coimbra em 1597).[4] ”

===

Use the common named entities tags for identifying each of the entities found Present the results in a table format.”

Obtivemos o resultado:

Here is a table of the named entities extracted from the provided text excerpt, with common named entity tags:

Named Entity	Tag
Universidade de Coimbra GCSE (UC)	ORG (Organization)
Coimbra	LOC (Location)
Portugal	LOC (Location)
Jardim Botânico	LOC (Location)
Estádio Universitário de Coimbra	LOC (Location)
D. Dinis I	PER (Person)
Leiria	LOC (Location)
Scientiae thesaurus mirabilis	MISC (Miscellaneous)
Papa	PER (Person)
Paço Real da Alcáçova	LOC (Location)
Paço das Escolas	LOC (Location)
Universidade de Coimbra	ORG (Organization)
1 de Março de 1290	DATE (Date)
1537	DATE (Date)
1597	DATE (Date)

These named entities include organisations, locations, persons, miscellaneous entities, and dates, which are all important for understanding the context of the news document.

Outra transformação célere de informações, impensável há cerca de 3 anos. O potencial de auxílio na indexação de documentos e criação de pontos de acesso é tremendo, aumentando em demasia a encontrabilidade de documentos.

c) extrair relações semânticas e construção de triplas de (sujeito, predicado, objeto).

Utilizamos o seguinte prompt:

Assume you are a Semantic Web Specialist

Create triples of subject-predicate-object for feeding a triplestore from the text below

===

“A Torre da Universidade de Coimbra, tem 33,5 m de altura, constitui o emblema tradicional de Coimbra. Começou a construir-se em 1728 e foi terminada em 1733. No topo sobre o relógio, abre-se um miradouro do qual se desfruta uma panorâmica esplendorosa da cidade e do vale do Mondego. Nesta Torre está, entre outros sinos, a célebre «Cabra», que marcava as horas do despertar e do recolher dos estudantes. No reinado de D. José I, a Universidade sofreu uma profunda alteração. Em 28 de Junho de 1772 o rei ratifica os novos estatutos (Estatutos Pombalinos), que marcam o início da Reforma. Esta manifestava, sobretudo, um grande interesse pelas ciências da natureza e pelas ciências do rigor, que tão afastadas se encontravam do ensino universitário.”

Obtivemos o resultado:

Here are the triples in the subject-predicate-object format extracted from the provided text for feeding a triplestore:

1. Torre da Universidade de Coimbra – hasHeight – 33.5 meters
2. Torre da Universidade de Coimbra – isEmblemOf – Coimbra
3. Torre da Universidade de Coimbra – constructionStarted – 1728
4. Torre da Universidade de Coimbra – constructionCompleted – 1733
5. Miradouro – isLocatedAt – Top of Torre da Universidade de Coimbra
6. Miradouro – providesViewOf – Cidade de Coimbra
7. Miradouro – providesViewOf – Vale do Mondego
8. Torre da Universidade de Coimbra – contains – Sinos
9. Sino – isNamed – Cabra
10. Cabra – markedTimeFor – Estudantes
11. Universidade de Coimbra – underwentReformDuring – Reign of D. José I
12. Reform – ratifiedBy – D. José I
13. D. José I – ratifiedOnDate – 28 de Junho de 1772
14. Reform – introduced – Estatutos Pombalinos
15. Estatutos Pombalinos – markedStartOf – Reforma da Universidade de Coimbra

16. Reforma – emphasizedInterestIn – Ciências da Natureza
17. Reforma – emphasizedInterestIn – Ciências do Rigor
18. Ciências da Natureza – wasDistancedFrom – Ensino Universitário
19. Ciências do Rigor – wasDistancedFrom – Ensino Universitário

These triples capture the key relationships and entities mentioned in the text, suitable for use in a Semantic Web or Linked Data context.

Sob qualquer ponto de vista, estes resultados impressionam pela acuidade, praticidade e rapidez, sendo passíveis de automatização com o uso de ambientes de linguagens de programação. Para a Ciência da Informação: é momento de refletirmos sobre a necessidade de evoluir uma área de conhecimento que se baseia, e tem sua razão de existência, nos métodos e ferramentas para coletar, organizar, armazenar, recuperar e disseminar informação, de forma eficiente e eficaz. Há que se repensar os currículos e formação, assim como os processos e os perfis que queremos atrair para nossos quadros docentes.

De maneira mais geral, assistimos a uma mudança paradigmática que trará impactos mais gerais nas formas de ensino, aprendizagem e cognição, com consequências ainda não totalmente claras. Cabe a nós construir este futuro.

Bibliografia

- Baskara, Risang. “Exploring the implications of ChatGPT for language learning in higher education.” *Indonesian Journal of English Language Teaching and Applied Linguistics* 7.2 (2023): 343-358.
- Firaina, Radha, and Dwi Sulisworo. “Exploring the usage of ChatGPT in higher education: Frequency and impact on productivity.” *Buletin Edukasi Indonesia* 2.01 (2023): 39-46. Acessível em: <https://journal.iistr.org/index.php/BEI/article/view/310>
- Fleckenstein, Johanna, et al. “Do teachers spot AI? Evaluating the detectability of AI-generated texts among student essays.” *Computers and Education: Artificial Intelligence* 6 (2024): 100209.
- Guo, Y., Shang, G., Vazirgiannis, M., & Clavel, C. (2023). The curious decline of linguistic diversity: Training language models on synthetic text. *arXiv preprint arXiv:2311.09807*.
- Hao, S., Han, W., Jiang, T., Li, Y., Wu, H., Zhong, C., ... & Tang, H. (2024). Synthetic data in AI: Challenges, applications, and ethical implications. *arXiv preprint arXiv:2401.01629*.
- Harwell, Drew. “Turnitin’s AI Cheating Detector, Launched in a Panic, Is Now Ubiquitous.” *The Washington Post*, 2 junho 2023, <https://www.washingtonpost.com/technology/2023/06/02/turnitin-ai-cheating-detector-accuracy/>.

- Holmes, Wayne, and Fengchun Miao. *Guidance for generative AI in education and research*. UNESCO Publishing, 2023.
- Kalla, Dinesh, et al. "Study and analysis of chat GPT and its impact on different fields of study." *International journal of innovative science and research technology* 8.3 (2023).
- Kobak, D., Márquez, R. G., Horvát, E. Á., & Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv:2406.07016*. Acessível em: <https://arxiv.org/abs/2406.07016>
- OPENAI. Educator FAQ. OpenAI Help Center, n.d. Disponível em: <https://help.openai.com/en/collections/5929286-educator-faq>. Acesso em: 31 ago. 2024.
- Templeton, Adly. *Scaling monosemanticity: Extracting interpretable features from claude 3 sonnet*. Anthropic, 2024.
- UNESCO. (s.d.). AI and the future of learning. UNESCO. <https://www.unesco.org/en/digital-education/ai-future-learning>
- Villalobos, Pablo, et al. "Will we run out of data? an analysis of the limits of scaling datasets in machine learning." *arXiv preprint arXiv:2211.04325* (2022).
- Zirar, Araz. "Exploring the impact of language models, such as ChatGPT, on student learning and assessment." *Review of Education* 11.3 (2023): e3433.

O Resumo Documental como Estrutura Lógica na Organização do Conhecimento

José Augusto Chaves Guimarães

ORCID: 0000-0002-0310-2331

Universidade Estadual Paulista (UNESP), Brasil

chaves.guimaraes@unesp.br

RESUMO: O resumo científico, vem sendo, ao longo de décadas, objeto de estudos na Ciência da Informação, desempenhando importante papel nos processos de recuperação da informação e de comunicação científica. No entanto, observa-se o **problema** de sua elaboração ser antes encarada como mera aplicação de uma norma do que como reflexo de uma lógica textual. Desse modo, o presente trabalho tem por **objetivo** resgatar conceitual e metodologicamente o processo de condensação documental – e seu produto, o resumo – no âmbito da organização e representação do conhecimento, com ênfase no meio científico, no intuito de evidenciar como esse processo possui uma natureza hermenêutica e como seu produto traz consigo uma lógica intrínseca. Para tanto, em termos **metodológicos**, realiza-se uma revisão crítica de literatura sobre a temática, com ênfase na produção científica em língua portuguesa. Os **resultados** evidenciam que: a) o resumo científico, enquanto produto do processo de condensação documental, integra a fase analítica da organização e representação do conhecimento; b) constitui um texto sintético representativo do texto original, com o qual guarda relação de semelhança e contiguidade, c) sua estruturação pressupõe textualidade, evidenciando coesão e coerência, e d) tem por características a concisão, a seletividade, a clareza, a precisão e a completeza. Ademais, sua lógica de estruturação demanda a representação do problema (por que?), do tema (o que?), dos objetivos (para que?), da metodologia (como?), dos resultados (o novo o que) e das conclusões (para onde?). Essa lógica de estruturação, em termos procedimentais, decorre de um processo hermenêutico desenvolvido pelo resumidor, no qual são desenvolvidas

atividades interpretativas que vão além do tema em si para atingir os contextos de produção e de recepção do documento. **Conclui-se** que o reconhecimento do processo de condensação documentária como um exercício hermenêutico balizado pelo conteúdo e pelo contexto do documento científico permeia-se por uma lógica intrínseca, que decorre do próprio contexto de pesquisa de tal forma que o respeito a tal lógica, refletida na estrutura do resumo, constitui um imperativo para que o processo de comunicação científica ocorra em bases cada vez mais amigáveis e profícuas.

PALAVRAS-CHAVE: Condensação documental; Organização do conhecimento; Hermenêutica.

ABSTRACT: The scientific summary has been, for decades, the object of studies in Information Science, playing an important role in the processes of information retrieval and scientific communication. However, there is a problem with its elaboration being seen as the mere application of a norm rather than as a reflection of textual logic. Thus, the present work aims to conceptually and methodologically rescue the process of document condensation – and its product, the summary – within the scope of the organization and representation of knowledge, with an emphasis on the scientific environment, in order to highlight how this process has a hermeneutic nature and how its product brings with it an intrinsic logic. To this end, in methodological terms, a critical review of the literature on the topic is carried out, with an emphasis on scientific production in Portuguese. The results show that the scientific summary is a product of the documentary condensation process that integrates the analytical phase of the organization and representation of knowledge and constitutes a condensed text representative of the original text, with which it has a relationship of similarity and contiguity, its structuring presupposes textuality, evidencing cohesion and coherence and its characteristics are conciseness, selectivity, clarity, precision and completeness. Its structuring logic demands the representation of the problem (why?), the theme (what?), the objectives (for what?), the methodology (how?), the results (the new what) and the conclusions (where?). This structuring logic, in procedural terms, results from a hermeneutic process developed by the summarizer, in which interpretative activities are developed that go beyond the theme of the document itself to reach the contexts of production and reception of the document. It is concluded that the recognition of the documentary condensation process as a hermeneutic exercise guided by the scientific document and its context is permeated by an intrinsic logic, which arises from the research context itself in such a way that respect for such logic, reflected in the Summary structure is imperative for the scientific communication process to occur on an increasingly friendly and fruitful basis.

KEYWORDS: Abstracts; Abstracting; Knowledge organization; Hermeneutics.

Introdução

O resumo, especialmente no âmbito da produção científica, vem sendo, ao longo de décadas, objeto de estudos na Ciência da Informação, desempenhando importante papel nos processos de recuperação da informação e de comunicação científica.

Para tanto, teóricos da área como Borko & Bernier (1975), Lancaster (1991, 1999)¹, Pinto Molina (1991, 1992), Iglesias Maturana (1992), Moreiro González (1993) e, em língua portuguesa, Calado (1972), Macedo & Moreira (1978); Kobashi (1994, 1997); Luz (1996); Guimarães (2004); Azevedo (2012); Costa & Moura (2013); Araújo & Santos (2021); González & Cervantes (2021) e outros dedicaram especial atenção aos aspectos históricos e metodológicos dessa temática.

Nesse contexto, especial destaque merece a atuação científica de Maria da Graça Melo Simões, que contribuiu significativamente para a Ciência da Informação por meio de uma profícua e verticalizada produção científica sobre resumos (Simões et al., 2013; Simões & Ferreira, 2014; Simões, 2015 a,b).

Em um momento que se reverencia a memória dessa importante pesquisadora da Ciência da Informação, cuja contribuição científica ocupa lugar de destaque na área, o presente capítulo busca resgatar conceitual e metodologicamente o processo de condensação documental – e seu produto, o resumo – no âmbito da organização e representação do conhecimento, com ênfase no meio científico, no intuito de evidenciar como esse processo possui uma natureza hermenêutica e como seu produto traz consigo uma lógica intrínseca.

¹ A referida obra foi objeto de várias edições e atualizações em inglês, e, inclusive, teve duas edições em português, cuja última foi: Lancaster, F.W. *Indexação e resumos: teoria e prática*. 2.ed. Brasília: Briquet de Lemos/Livros, 2004.

A condensação documental como processo na organização do conhecimento

A organização do conhecimento, constitui área de estudos que, em um sentido lato, refere-se às formas como a humanidade, ao longo da história buscou categorizar o conhecimento produzido pelo homem, como é o caso da Classificação dos Seres Vivos, de Lineu, da Tabela Periódica dos elementos químicos, da *Encyclopédie*, de Diderot e D´Alembert e outras tantas iniciativas. Por outro lado, essa área de estudos, em um sentido estrito – ao qual nos referimos neste trabalho – constitui campo de pesquisa e de formação e atuação profissional que se encontra prioritariamente afiliado à Biblioteconomia e Ciência da Informação, e cuja ênfase reside em descrever, analisar, representar, armazenar e organizar documentos, assuntos e conceitos, valendo-se da atuação humano ou de processos automatizados (Hjørland, 2008, 2016).

Como destaca Guimarães (2015, p. 15), essa área de estudos se constrói e se desenvolve a partir de uma trílice estrutura, em que se tem uma dimensão epistemológica, “voltada para a consolidação de teorias, metodologias, paradigmas, escolas de pensamento”, uma dimensão aplicada, “fortemente ligada ao avanço tecnológico e ao desenvolvimento de novos instrumentos”, e uma dimensão cultural, “voltada para os contextos, as comunidades e os sujeitos que interagem na OC”. Tais dimensões, por sua vez, produzem efeitos nos processos (e.g., classificação, indexação, etc.) que são desenvolvidos, utilizando instrumentos (e.g., esquemas de classificação, tesauros, etc.) para gerar produtos (notações, descritores, índices, etc.) (Guimarães & Dodebei, 2012).

Especificamente no âmbito dos processos inerentes à organização e representação do conhecimento, em especial no que se refere ao tratamento de conteúdo dos documentos, tem-se dois tipos de operações consecutivas e interdependentes: a análise e a síntese documentária. A etapa analítica consiste em apreender (identificar e selecionar) o conteúdo de um documento, ao passo que a etapa sintética consiste em representar tal conteúdo em uma linguagem padronizada para que o mesmo possa ser recuperado. Na etapa analítica, procede-se à leitura

técnica do documento (a partir de sua), a identificação de conceitos e a seleção dos conceitos que efetivamente serão colocados à disposição para pesquisa a partir da importância que ocupam no documento (Chaumier, 1980).

Uma vez identificados e selecionados os conceitos de um documento, chega-se à fase da síntese, na qual são gerados os produtos documentários (ou informações documentárias, como denomina Kobashi, 1994), que podem ser simbólicos ou textuais. Os produtos documentários simbólicos, ou índices, são obtidos por meio da atribuição de termos descritores baseados em esquemas de vocabulários controlados, tesouros e outros tipos de sistemas de organização do conhecimento. Por outro lado, o documento pode ser representado de forma textual, por um novo documento, fruto de um processo de condensação. Esse processo de condensação, por sua vez, dá origem a um produto específico – o resumo – que constitui uma representação sintética do documento valendo-se de um número limitado palavras, sentenças ou frases expressivas de sua substância (Luz, 1999, p. 1).

O resumo com produto da condensação documental²

Em que pese o fato de as atividades de condensação documentária ocorrerem em distintas áreas, tais como nas resenhas literárias ou nas sinopses de filmes, entre outros, é na esfera científica que o resumo assume maior protagonismo pois constitui chave para a pesquisa e para a comunicação científica, sendo imprescindível em artigos de revistas e em bases de dados. Nesse âmbito, o resumo deve apresentar um conjunto de características que lhe garantirão identidade e utilidade, a saber: concisão, seletividade, clareza, precisão e completeza.

A concisão pressupõe que a essência temática do documento seja devidamente representada e que tudo o que for secundário ou supérfluo seja dispensado (em outras palavras, deve dizer o máximo valendo-se do menor

² Destaque-se, especialmente, a minuciosa *Evolução histórica do resumo documental* feita por Simões (2015).

número possível de palavras). Isso pressupõe, por definição, a seletividade, pois devem ser descartadas do resumo as informações que possam desempenhar papel secundário no âmbito do documento e que se constituíam em elementos de somenos importância para o leitor, quais sejam: repetição dos elementos já contidos no título, explicações, expressões introdutórias que nada dizem e termos excessivamente genéricos para serem significativos; optando-se por frases curtas que se tornem facilmente apreensíveis (Lakatos, 1986; Lancaster, 1993).

A clareza é igualmente imprescindível ao resumo, resultando na facilidade de interpretação e rapidez na apreensão de seu conteúdo (garantir o máximo de compreensão do texto com o menor tempo de leitura). Desse modo, o resumo não pode conter pontos obscuros, contradições (que porventura gerem interpretações dúbias) ou ainda vocabulário não condizente com o público a que se destina, devendo constituir-se em um documento cuja compreensão independe da leitura do documento original.

A precisão de um resumo, por sua vez, decorre de um criterioso processo de seleção de palavras que se configurem adequadas à representação dos conceitos do documento enquanto a completeza revela a capacidade de o resumo fazer-se entender por si mesmo, sem leitura complementar ao documento original (Luz, 1999).

Moreiro González (1993), refere-se, ainda, à necessidade de o resumo evidenciar: entropia (assegurando sua função comunicativa), coerência (por meio de uma linearidade em suas orações, em uma relação de continuidade); correção (em observância às regras linguísticas para garantir univocidade dos enunciados); e um distanciamento tolerável, (na medida em que o resumo deve ser produzido quase que imediatamente após o original para evitar obsolescência).

A lógica de estruturação do resumo científico como resultado de um processo hermenêutico

O resumo científico, como já mencionado, constitui um dos principais pontos de acesso para artigos e demais textos, sendo amplamente utilizado em bases de dados como elemento para busca de informação para

subsidiar pesquisas. Para tanto, é fundamental que esse resumo represente adequadamente o *iter* investigativo que se materializa em um texto científico, pois aí reside sua lógica intrínseca. Nesse sentido, Kobashi (1994) propõe uma macroestrutura textual para o resumo científico, composta de Problema (Indagação); Hipótese (Conjetura); Metodologia (Observação); Resultados (Interpretação) e Conclusão (Comentário final).

Para tanto, é fundamental que o resumo se inicie com a questão que deu origem à pesquisa e, como decorrência, com o tema que daí emerge (em outras palavras, o *por que* e o *que* da pesquisa). Em seguida, o resumo deve deixar claro onde se pretende chegar a partir dessa questão inicial (o *para que*), explicitando os objetivos. Para que se atinjam os objetivos almejados, é necessário que se estabeleça e se cumpra um percurso investigativo que balize – e avalize – a pesquisa em termos de universo, corpus, procedimentos, materiais etc., ou, em outras palavras, a metodologia empregada (o *como*). O emprego da metodologia visando a atingir os objetivos almejados conduz a um conjunto de resultados (o *novo o que*), os quais, por sua vez, evidenciam o surgimento de fatos novos, de descobertas significativas, de contradições e teorias anteriores, e de relações e efeitos novos verificados, entre outros aspectos. Tais resultados são objeto de interpretação e, com um olhar para o futuro, levam a um conjunto de recomendações, aplicações, sugestões, avaliações e à propositura de novas relações e de investigações em continuidade, objeto da conclusão.

Cumpra destacar que muitas revistas, em especial nas ciências exatas e nas ciências biológicas, exigem a apresentação dos denominados resumos estruturados em que essas etapas da investigação: Problema, Objetivos, Metodologia, Resultados e Conclusão devem ser apresentadas em parágrafos distintos.

Vale destacar que essa lógica de estruturação, em termos procedimentais, decorre de um processo hermenêutico desenvolvido pelo resumidor, tendo-se em conta que a hermenêutica, enquanto arte ou ciência da interpretação oriunda da Filosofia, transcende à simples compreensão textual para atingir significados em contextos mais amplos, razão pela qual o resumidor deve sempre permear sua atuação pela garantia do

estabelecimento do processo comunicativo entre o autor e leitor, devendo estar suficientemente inteirado do contexto de produção do texto original (o autor, o contexto institucional, o tema, a área de especialidade etc) e do contexto de recepção (as comunidade usuárias que podem ser atingidas).

Autores como Friedrich Schleiermacher (1768-1834), Wilhelm Dilthey (1833-1911), Martin Heidegger (1889-1976) e Hans-Georg Gadamer (1900-2002) foram seminais para a consolidação da hermenêutica como campo de estudo científico, indo além dos até então usos da hermenêutica para fins teológicos ou jurídicos.

Para Schleiermacher (2003), a hermenêutica envolve um processo tanto técnico quanto intuitivo, na medida em que uma construção original tem seu significado reconstruído para compreender também o contexto histórico e cultural em que a obra foi concebida. Essa compreensão contextual é reafirmada por Dilthey (1984) para quem o contexto de produção se incorpora ao objeto a ser interpretado. Heidegger (2006), por sua vez, recorda que interpretar é uma atividade inerente ao ser humano e essa interpretação, como destaca seu discípulo Gadamer (2002; 2005) pressupõe um diálogo contínuo entre o passado e o presente, em que a hermenêutica, como uma verdadeira fusão de horizontes, estabelece uma interlocução entre o intérprete e a tradição.

A hermenêutica, ao ser permeada pelo contexto histórico-cultural (do texto, da intenção do autor e também do intérprete) vale-se da linguagem para promover um processo de interação com o texto. A linguagem, como destacam Oliveira & Guimarães (2023, p. 9), “não é apenas um meio de expressão, mas é fundamental para a maneira como entendemos o mundo”.

A caracterização da condensação documental – visando à elaboração do resumo científico – enquanto um processo hermenêutico tem início já no primeiro momento da análise, quando se desenvolve a leitura documental para fins de seleção dos conceitos que integrarão o resumo. Tal seleção, por sua vez, como destaca Guimarães (2004) se rege por dois princípios: o grau de importância de um dado tema no conteúdo e, ainda, pela ocorrência de algum tema que, ainda que secundário, seja de especial interesse para a comunidade usuária. Tem-se, aí, a evidência al daquilo

que Hope Olson (2002) denomina como “poder de nomear” (*power to name*), enquanto “um poder interpretativo que lhe é conferido pela sociedade para que crie substitutos de conhecimentos (*surrogates of knowledge*) que propiciem o acesso a um conhecimento registrado e socializado” (Oliveira & Guimarães, 2023, p. 11).

Esse poder interpretativo se torna ainda mais evidente na escolha dos termos que serão utilizados para a composição do resumo visto que estes não são, necessariamente, os mesmos que integram o documento original, pois decorrem de um processo de síntese e de normalização terminológica com vistas à maior clareza possível e a propiciar em mais larga medida a efetivação de um processo comunicativo.

Tem-se, assim, no resumo científico, a incorporação da dupla dimensão de conteúdo prevista por Beghtol (2006): a tematicidade intrínseca de um documento (*aboutness*) e os, muitas vezes representada como assunto principal, com os distintos significados (*meanings*) que um dado conteúdo informacional pode ter para diferentes comunidades e contextos de usuários.

Um resumo, enquanto representação de um documento (este considerado como materialização e socialização de um dado conhecimento) “traz consigo a compreensão de significados subjacentes e de conexões culturais em uma abordagem mais flexível e adaptável, reconhecendo que as informações não são estáticas, mas fluidas e contextualmente dependentes” (Oliveira & Guimarães, 2023, p. 18).

Desse modo, a representação do conteúdo de um texto científico em um resumo será tanto mais elucidativa e promotora de um processo comunicativo mais amplo entre o autor e a(s) comunidade(s) a que o texto puder atingir se for assumido, no momento da elaboração, que esse resumo pressupõe uma lógica própria, lógica essa que, por sua vez, decorre de um processo eminentemente interpretativo.

Conclusão

Reportando-nos a importantes considerações de Maria da Graça Simões (2015), devemos considerar o imprescindível papel do resumo na investi-

gação e na pesquisa científica, ao “alertar o consumidor de informação, antecipar a sua publicação e, eventualmente substituir o documento original”, ao que se alia o seu papel educativo como elemento de rápida atualização, “pelo facto de a informação ser representada de uma forma breve e precisa” (Simões, 2015, p. 14). Tal aspecto se torna ainda mais significativo quando se consideram todos os esforços que vêm sendo empreendidos pelas revistas científicas – especialmente nas Ciências Biológicas e Exatas – para o aperfeiçoamento dos resumos, de modo a garantir-lhe cada vez mais clareza, coerência e completeza.

Hoje, os resumos científicos deparam-se com o fenômeno da inteligência artificial, como é o caso do Chat GPT (*Generative Pre-trained Transformer*), o que pode, por um lado, contribuir para a agilização do processo de elaboração de resumos estruturados (desde que com parâmetros prévios muito bem estabelecidos) mas, por outro, pode levar a implicações éticas de grande monta.

Por fim, e resgatando a proposta do presente artigo, não se pode esquecer que a elaboração de resumos científicos, enquanto fruto de um processo interpretativo, permeia-se por uma lógica intrínseca, lógica essa que decorre do próprio contexto de pesquisa³. Reconhecer e respeitar tal lógica, refletindo-a na estrutura do resumo é algo imperativo para que o processo de comunicação científica ocorra em bases cada vez mais amigáveis e profícuas.

Referências

- Araujo, V. J., & Santos, C. A. C. M (2021). Elaboração do resumo documentário: qualidade, parâmetros e normas. *InCID: Revista de Ciência da Informação e Documentação*, 12(1) 2021.
- Azevedo, A. R. A função do resumo na comunicação científica: uma comparação entre resumos tradicionais e estruturados. *Perspectivas em Ciência da Informação*, 17(4), 2012.

³ Essa lógica do resumo a partir do contexto de geração do documento original é ressaltada por Guimarães (2004), em outro âmbito – o dos resumos (ementas, no Brasil) de sentenças e acórdãos judiciais – em que o referido autor propõe quatro categorias temáticas – Fato, Instituto Jurídico, Entendimento e Argumento – para a composição do enunciado, categorias essas que emergem das partes fundamentais da decisão: Relatório, Voto e Dispositivo.

- Beghtol, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42(2), 84-113.
- Borko, H., & Bernier, C.L. (1975). *Abstracting concepts and methods*. New York: Academic Press.
- Calado, A. A. (1972). *O problema das análises bibliográficas nos centros de documentação*. Luanda: Instituto de Investigação Científica de Angola.
- Chaumier, J. (1980). *Travail et méthodes du/de la documentaliste*. Paris, Editions ESF.
- Costa, M. U. P., & Moura, M. A. (2013) A representação da informação em contextos de comunicação científica: a elaboração de resumos e palavras-chave pelo pesquisador – autor. *Informação & Informação*, 18(3), 45-67.
- Dilthey, W. (1984). Origens da hermenêutica. In: Magalhães, R. *Textos de hermenêutica*. Porto: Rés Editora.
- Gadamer, H. G. (1989). *Truth and Method*. 2. Ed. rev. New York: Crossroad.
- Gadamer, H. G. (2003). *O problema da consciência histórica*. 2.ed. Rio de Janeiro: FGV, 2003.
- Gonçalez, P. R. V. A. & Cervantes, B. M. N. (2021). Perfil de aplicação para resumos estruturados (PARE). *Informação & Informação*, 26(1), 483-511.
- Guimarães, J. A. C. (2015). Apresentação. In: Guimarães, J. A. C. & Dodebei, V. (org.). *Organização do conhecimento e diversidade cultural / Marília: ISKO-Brasil; Fundepe*.
- Guimarães, J. A. C. (2004) *Elaboração de ementas jurisprudenciais: elementos teórico-metodológicos*. Brasília: Conselho da Justiça Federal.
- Guimarães, J. A. C. & Dodebei, V. (org.). (2012). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade*. Marília: ISKO-Brasil: Fundepe, 2012.
- Heidegger, M. (2006). *Ser e tempo*. São Paulo: Vozes.
- Hjørland, B. (2008). "What is Knowledge Organization (KO)?" *Knowledge Organization* 35(2/3), 86-101.
- Hjørland, Birger. 2016. Knowledge organization. *Knowledge Organization* 43, no. 6: 475-84. (Also available in Hjørland, Birger, ed. *ISKO Encyclopedia of Knowledge Organization*, https://www.isko.org/cyclo/knowledge_organization)
- Iglesias Maturana, M. T. (1992). *El analisis y recuperación de información documental*. Santiago (Chile): Instituto Profesional de Santiago.
- Kobashi, N. Y. (1997). Resumos documentários: uma proposta metodológica. *Revista de Biblioteconomia de Brasília* 21(2), 201-210.
- Kobashi, N.Y. (1994) A elaboração de informações documentárias: em busca de uma metodologia. São Paulo, 1994. Tese (Doutorado em Ciências) Escola de Comunicações e Artes, Universidade de São Paulo.
- Lakatos, E., Marconi, M.A. (1986) *Fundamentos de metodologia científica*. São Paulo: Atlas.
- Lancaster, F. W. (1991). *Indexing and Abstracting in Theory and Practice*. London: Library Association.
- Lancaster, F. W., & Pinto, M. (1999). Abstracts and abstracting in knowledge discovery. *Library Trends*, 48 (1) 234-248.
- Lancaster, F. W., (2004). *Indexação e resumos: teoria e prática*. 2.ed. Tradução por Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos/Livros.
- Luz, A. C. O. (1996). Critérios para a elaboração de resumos. *Revista de Biblioteconomia de Brasília*, 20(10), 27-40.

- Macedo, N. D. de, Moreira, M. De F. G. (1978). Resumos: subsídios para sua elaboração. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, 11(1/2, jan./jun.), 65-72.
- Moreiro Gonzalez, J. A. (1993). *La aplicación de las ciencias del texto al resumen documental*. Madrid: Universidad Carlos III; B.O.E.
- Oliveira, W. C., & Guimarães, J. A. (2023). C. Hermenêutica e Organização e Representação da Informação: transversalidade e verticalidade na Ciência da Informação. *Brazilian Journal of Information Science: Research Trends*, 17, e023062.
- Pinto Molina, M. (1991). Hacia un modelo de representación documental: la técnica de resumir. *Investigaciones bibliotecológicas*, México, D.F., 5(10), 17-28.
- Pinto Molina, M. (1992). *El resumen documental: principios y métodos*. Madrid: Fundación Germán Sánchez Ruiperez.
- Schleiermacher, F. D. E. (2003). *Hermenêutica: arte e técnica da interpretação*. Bragança Paulista: Editora Universitária São Francisco.
- Simões, M. da G. et al. (2013) O Resumo como recurso privilegiado na divulgação da produção científica: origem e evolução do resumo estruturado. In: *I Congresso ISKO Espanha e Portugal*. Porto: FLUP, CETAC.MEDIA, 879-900.
- Simões, M. da G. (2015). Resumo documental e literatura científica: origem, desenvolvimento e consolidação. *Páginas a&b: Arquivos E Bibliotecas*, 3(3), 15-36.
- Simões, M. da G. (2015b.). *Resumo documental: uma incursão à (des)construção concetual na literatura científica*. Coimbra: Imprensa da Universidade de Coimbra.
- Simões, M. da G., & Ferreira, C. (2014). Resumo e abstract: evolução morfológica e semântica. *Biblos*. 1(2), 531-550.

**Sistemas de Organización del
Conocimiento incluyentes.
Revertir la invisibilización de la mujer**

Blanca Rodríguez-Bravo
ORCID 0000-0002-9476-7602
Universidad de León, España
blanca.rodriguez@unileon.es

RESUMO: Se presentan los principales sesgos y exclusiones relativos a la mujer y lo femenino detectados en análisis de diversos sistemas de organización del conocimiento y se sugieren posibles mejoras y/o soluciones.

PALAVRAS-CHAVE: Mujeres; Sesgos; Sistemas de Organización del Conocimiento.

ABSTRACT: The main biases and exclusions related to women and femininity detected in analyzes of various knowledge organization systems are presented and possible improvements and/or solutions are suggested.

KEYWORDS. Biases; Knowledge Organization Systems; Women.

Objetivo

Con este trabajo pretendo contribuir al homenaje de una muy querida amiga, la Profesora Doctora Graça Simões, y servirme para ello de un texto relacionado con una disciplina que la apasionaba, la Organización del Conocimiento, y sobre la que compartimos investigaciones tras la defensa de su tesis (Salamanca, 2010): “La representación de Etnia y su evolución en la Clasificación Decimal Universal” que co-dirigí junto al doctor Frías.

Antecedentes

En otros trabajos anteriores realizados algunos en solitario y otros en colaboración con colegas españolas y portuguesas, como es el caso de la Dra. Graça Simões, hemos puesto de relieve diversas inconsistencias y exclusiones de los sistemas clasificatorios centradas en cuestiones de género fundamentalmente, y tanto de carácter estructural como lingüístico, y tanto sobre vocabularios jerárquicos, en particular la Clasificación Decimal Universal (CDU) y la Clasificación Decimal Dewey (CDD), como sobre vocabularios alfabéticos, en este caso Listas de Encabezamientos y Tesoros.

En el primer estudio que realizamos (Morán Suárez y Rodríguez Bravo, 2001) comparábamos la edición de la CDU abreviada en lengua española de 1995 con la de 2000 y poníamos de relieve que en la edición de 2000 la presencia de la mujer había disminuido con respecto a la de 1995 por la desaparición del número 396 (Feminismo. Situación de la mujer), y que se realizaba un esfuerzo en la corrección de sesgos estructurales y terminológicos. No obstante, quedaba camino por recorrer para evitar la discriminación y minusvaloración de las mujeres.

En un trabajo posterior (Simões, Rodríguez-Bravo y Pestana, 2018) analizábamos comparativamente la representación de la mujer en la Clasificación Decimal Dewey y en la Clasificación Decimal Universal, en esta ocasión la 22.^a edición integral en lengua inglesa para la CDD y la edición abreviada de la CDU en español de 2016.

En este estudio se observaba que, aunque ambos sistemas habían realizado un esfuerzo por evolucionar en la línea de evitar estereotipos, continuaban mostrando desigualdad y anacronismo en el caso del concepto de mujer. Como ya se detectó en la edición de 2000 de la CDU se apreciaba una tendencia a que la presencia de la mujer disminuyera, prácticamente no aparecía en las tablas principales estando relegada a la tabla de auxiliares de personas. Por el contrario, la mujer estaba más presente en las clases principales de la CDD. El plano en el que se refleja a la mujer en ambos sistemas es de inferioridad respecto al del hombre.

Asimismo, hemos realizado en solitario estudios sobre la inclusión de la mujer en vocabularios asociativos, el primero (Rodríguez-Bravo, 2006)

centrado en tres vocabularios: Library of Congress Subject Headings (LCSH), Tesauro de la UNESCO y Women's European Thesaurus. El segundo, algo posterior, circunscrito a Listas de encabezamientos de materia y tesauros en lengua española y de uso principalmente en España (Rodríguez-Bravo, 2007).

En los Subject Headings de la Library of Congress y en el Tesauro de la UNESCO se apreciaba la hegemonía de lo masculino que obligaba a reservar un espacio especial para la mujer. Este espacio era siempre escaso, si bien hay que precisar que la atención dedicada a las mujeres era superior en la LCSH que en el Tesauro de la Unesco. Asimismo, se observaban sesgos en la representación del género. En ambos vocabularios puede notarse, todavía hoy, que a las mujeres se les otorga un estatus de inferioridad y de dependencia respecto a los hombres y se les limita a roles tradicionales.

Tanto en las clasificaciones analizadas como en los vocabularios alfabéticos se constataba el énfasis en el estatus marital de la mujer, en su sexualidad, en los procesos vinculados con la maternidad, la belleza y otros aspectos de trascendencia discutible.

Principales sesgos y exclusiones detectados

Para dar algún ejemplo de sesgos y exclusiones podemos sistematizar estos siguiendo la tipología utilizada por Olson y Schlegl (2021) que los divide en cinco modalidades: el tratamiento del tema como excepción; la guetoización del tema; la omisión del tema; la estructura inapropiada de la norma; y la terminología sesgada.

1 – Tratamiento del tema como excepción: el ejemplo paradigmático es el de “mujer como matemática, física, etc.” dándose a entender que una mujer que se dedique a las matemáticas es una excepción. Un ejemplo de este sesgo está presente en el Tesauro de la Unesco: “Women in politics”

2 – Guetoización del tema: consiste en agrupar y posteriormente aislar un tema en vez de integrarlo. Hemos hallado ejemplos frecuentes de guetoización, en concreto la acumulación de entradas bajo el encabezamiento mujeres en la LCSH, y la concentración de los estudios de la mujer en el número 396 en versiones antiguas de la CDU.

3 – Omisión del tema: la omisión podría deberse a una falta de vigencia de este, pero también puede ocurrir por presuposiciones o desinterés por la materia. La misma desaparición del 396 en la CDU implicó que haya disminuido la visibilidad de la mujer en esta clasificación que ahora se encuentra muy limitada en su presencia: el Feminismo en la clase 1, las comadronas y la ginecología en la medicina (61) y como heroínas y mujeres fatales en el cine (79).

4 – Estructura inapropiada de la norma: este desequilibrio se materializa en poco espacio para una materia, agrupación o cercanía de temas sin similitud, mecanismos inadecuados de división, etc. Esta situación afecta directamente a la percepción de diversas materias y es la que la literatura sobre sesgos en las clasificaciones ha tratado con mayor frecuencia. También hemos encontrado diversos ejemplos de estructura inconveniente que ayuda a realizar una lectura de determinadas materias equivocada. El mismo número 396 que hemos comentado, se ubicaba en las ediciones antiguas de la CDU entre el Ceremonial y la etiqueta (395) y los Pueblos primitivos (397).

5 – Terminología sesgada: modalidad que también ha sido muy estudiada en los vocabularios asociativos, sobre todo en las listas de encabezamientos de materia, pero que está ganando protagonismo también en las clasificaciones, debido al uso de las tecnologías y la abundancia de recursos en línea. En este contexto podemos indicar que se utiliza el femenino para señalar la presencia de la mujer en oficios menos considerados que los de los hombres o podemos hallar encabezamientos claramente displicentes para la mujer del tipo “overweight women” frente a “strong men” que encontramos en los encabezamientos de la LCSH.

Ante la exclusión femenina de los vocabularios de pretensión universal de mayor uso es comprensible que las bibliotecas y los centros de documentación especializados en mujeres hayan procurado la construcción de sistemas propios de organización del conocimiento, como son: European Women's thesaurus (EWT) (IIAV); On equal terms: a thesaurus for nonsexist indexing and cataloguing (Marshall); Tesouro “Mujer” del Instituto español de la mujer, el Listado de descriptores en el tema de la mujer chileno (ISIS Internacional), etc.

Estas iniciativas suponen el paso desde la representación androcéntrica de la realidad a una representación femenina de la misma, las materias se nombran por lo general en femenino. Sin embargo, su empleo queda restringido a ámbitos muy concretos: centros de documentación sobre la mujer y similares.

El caso de las clasificaciones ¿universales?

Son varios los autores que han puesto de relieve las limitaciones de las clasificaciones como sistemas de organización del conocimiento, entre ellos destacaríamos los trabajos pioneros de Hope Olson, de la Universidad de Alberta.

Los sistemas de organización del conocimiento documental tienen generalmente una presunción de universalidad. Esta presunción queda de relieve en el nombre de diversas clasificaciones y es una finalidad de otros sistemas de organización del conocimiento como tesauros y listas de encabezamientos ya mencionadas.

Muchas de estas presunciones se justifican desde la literatura y sus postulados con el argumento de que la diversidad en el lenguaje causa problemas de comunicación, y en el proceso de normalización –intrínseco en los vocabularios controlados– a veces es necesario sacrificar algunos usos y conveniencias minoritarias a favor del beneficio de la mayoría. Así, la solución tradicional en organización del conocimiento ha sido traducir el lenguaje natural a uno sindético universal. Sin embargo, como señala Olson (2001a), la representación de la información de este modo se muestra como un ideal inalcanzable y además el resultado es frecuentemente excluyente.

La visión de universalidad seguida en los sistemas de organización del conocimiento es lo que Olson llama “la tiranía de la mayoría”. Con todo, la suma de las voces de las minorías a veces sí supera en volumen a la de la mayoría. Así, se han venido sucediendo de forma continuada, numerosas críticas desde distintos ámbitos que han puesto de manifiesto el concepto de invalidez universal en las clasificaciones bibliotecarias (Martínez-Ávila, et al., 2022).

Los sistemas de organización del conocimiento universales no son válidos para todas las comunidades a las que pretenden dar servicio. Ello obedece a que, en el caso de las clasificaciones bibliográficas, están construidas sobre fundamentos culturales establecidos que primarán unas facetas sobre otras en el desarrollo lineal de sus clases, y esto se plasmará en una serie de sesgos lingüísticos y estructurales que afectarán de forma negativa a la población que los utiliza (Martínez-Ávila y Guimarães, 2013).

Asimismo, se ha esgrimido comúnmente el argumento de que la invalidez de una organización del conocimiento de carácter universal deriva de que la sistematización que realizan de la realidad y del conocimiento es provisional y caduca en consonancia con el hecho de que las disciplinas científicas están sujetas a numerosas concepciones y cambios. Y además a que la estructuración que presentan las clasificaciones está constreñida a la concepción del mundo de aquellos que han ordenado el conocimiento.

Como ponía de relieve Kwasnik (1999) las clasificaciones nunca son creadas en el vacío social o político y en ellas es inherente la jerarquía. Desde el momento en que una clasificación organiza los conceptos en unas determinadas clases y con una determinada dependencia, indirectamente está describiendo y presentando una determinada visión del mundo, contribuyendo así a conformar en la mente de las personas que la consultan esa misma percepción.

Ante esta situación, muchos estudios que critican la neutralidad de las clasificaciones se han focalizado en la revisión de la estructura y terminología de estas en persecución de la eliminación de sesgos y asimetrías.

Sin embargo, últimamente se cuestiona que sea posible eliminar todos los sesgos existentes en los sistemas de organización del conocimiento en todos los posibles contextos, ya sea mediante revisiones y correcciones, como se ha hecho tradicionalmente, o ya sea a través de la construcción de nuevos esquemas (Martínez-Ávila, et al., 2022). Esto se debe a que, durante la construcción de un esquema de clasificación, al tratarse de un sistema lineal, el desarrollo de las facetas según los intereses de un grupo o contexto concreto inevitablemente siempre va a perjudicar los intereses del resto de grupos o contextos cuyas facetas de interés no hayan sido primadas.

Es por ello por lo que resulta necesario escuchar las voces de todos los grupos minoritarios y conocer las necesidades y críticas realizadas desde cada contexto de una forma independiente e individual.

Olson y Schlegl (2001) ya indicaban que los estándares alternativos, la adaptación de estándares y los usos creativos (o incluso subversivos) de estos podrían contribuir a una aminoración del sesgo negativo en el acceso por materias. Olson (2001b) habla de rechazar soluciones universales, pero sin abandonar las clasificaciones existentes en la esperanza de alcanzar una mayor objetividad (Olson, 1998).

Por tanto, una medida a implementar sería la creación de contextos y adaptaciones locales en vez de la revisión autorizada y centralizada en un único organismo como puede ser el Consorcio de la CDU, OCLC o la Library of Congress. Esta revisión debería ser el fruto de las diversas comunidades interesadas en el uso de los estándares (Martínez-Ávila, et al., 2022).

Olson opina que todos los sistemas tienen un límite que excluye o margina algo o alguien. Reemplazar el viejo sistema con uno nuevo puede cambiar los límites, pero no los elimina. Un enfoque mucho más útil sería hacer el límite permeable, y de esta forma permitir que las voces de aquellos grupos previamente excluidos o marginados hablen con el sistema (Olson, 1997).

Una propuesta real fue el proyecto titulado “Mapping and Expanding Library Classification for Marginalized Domains: Feminism as a Prototype” a cargo de la profesora Olson entre 1998 y 2001. Los resultados de dicho proyecto se publicaron en el artículo “Adapting dominant classifications to particular contexts” (Kublik et al., 2003), aunque la materialización práctica de dicha expansión nunca llegó a hacerse efectiva debido a las restricciones legales y la política privativa de gestión de la Clasificación Decimal de Dewey por parte de OCLC.

En un trabajo todavía reciente (Martínez-Ávila, et al., 2022) se ha planteado que los problemas de universalidad en los sistemas de organización del conocimiento universales se pueden paliar con la aplicación de adaptaciones locales que se acomoden a los diferentes grupos y contextos, es decir, realizando adaptaciones locales de los sistemas de

organización del conocimiento y estableciendo un vínculo entre las tablas de estas adaptaciones teóricas y las tablas oficiales de la CDU (el Master Reference File – MRF). Esta solución iría de la mano de una aplicación informática que permita la interoperabilidad semántica de unos contextos y variantes con otros utilizando ese mapeo con las tablas oficiales del Master Reference File ya mencionado.

Reflexiones finales y posibles soluciones para los lenguajes asociativos

En los estudios realizados se han encontrado todos los sesgos apuntados anteriormente pese a que se ha ido avanzando en pro de la búsqueda de la equidad algo que se aprecia en el análisis de diferentes versiones de la CDU; Igualmente, se observan diferencias en el tratamiento de género que realizan los distintos lenguajes asociativos, no todos tienen el mismo sesgo sexista pero la mayoría no están libres de él.

En la mayor parte de los instrumentos terminológicos la presencia femenina es escasa porque el lenguaje ha erigido lo masculino como universal, como genérico, y, asimismo, es inapropiada porque su discurso de representación del conocimiento mantiene los estereotipos femeninos, ofreciendo una imagen anacrónica de la mujer y dando muestras de sexismo.

Los vocabularios asociativos no reflejan los contenidos emergentes de la temática feminista y ocultan todo lo referido a la mujer. Es frecuente así la guetoización de las materias bajo una única entrada, “Mujeres”, tan general que dificulta la recuperación de información sobre cualquier temática específica. Además, se aprecia claramente que no existe paralelismo con la entrada “Hombres”.

Cabe señalar que si se compara el tratamiento de género de vocabularios en lengua española y en lengua inglesa se constata que en español existen más problemas terminológicos, fundamentalmente por dos razones. De un lado, porque el español utiliza con frecuencia el masculino como genérico donde el inglés distingue –parents, father and mother–. De otro, porque la distinción de género, tanto en sustantivos como en adjetivos,

ha conducido a la frecuente utilización del masculino para representar a la humanidad completa, en muchos casos de forma escasamente acertada.

Las posibles soluciones son de alcance parcial. Existen recursos o medidas no difíciles de arbitrar: la autorización –por medio de notas– a utilizar encabezamientos/descriptores en femenino si es pertinente, la utilización de aclaradores para distinguir el masculino del genérico, o la utilización de fórmulas similares a la propuesta por el European women's thesaurus que cuando precisa referirse al varón añade el calificador (-W) (no women).

Como ya propusimos en trabajos previos sería factible establecer dos aclaradores de género a usar junto con encabezamientos y descriptores cuando se requiera especificar que el documento de la temática X no se refiere a la humanidad al completo, sino solo a la mujer o solo al varón (M) y (V), por ejemplo.

En resumen, tenemos que seguir trabajando en pro de la equidad en el terreno del género y de otros muchos ámbitos de exclusión y para ello parece que el camino puede venir de la mano de la creación de contextos y variantes como las que se plantean con las clasificaciones, sin olvidar la más que necesaria revisión constante no solo de la estructura de los Sistemas de Organización del Conocimiento, sino también del lenguaje empleado, eliminando sesgos y asimetrías. Tratar de evitar la exclusión es una obligación habida cuenta del papel de los sistemas de organización del conocimiento en la configuración del pensamiento.

Referencias bibliográficas

- Kublik, A., Clevette, V., Ward, D. B. & Olson, H. A. (2003). Adapting dominant classifications to particular context. *Cataloging & Classification Quarterly*, 37(1/2), 13-31.
- Kwasnik, B. H. (1999). The role of classification in knowledge: representation and Discovery. *Library Trends*, 48(1), 22-47.
- Martínez-Ávila, D., San Segundo, R., Frías Montoya, J. A., & Rodríguez-Bravo, B. (2022). Propuesta para la adaptación de la CDU en entornos multicontextuales. *Revista EDICIC*, 2(4), 1-18.
- Martínez-Ávila, D., & Guimarães, J. A. C. (2013). Library classifications criticisms: universality, poststructuralism and ethics. *Scire: Representación y Organización del Conocimiento*, 19(2), 21-26.
- Morán Suárez, M. A., & Rodríguez Bravo, B. (2001). La imagen de la mujer en la Clasificación Decimal Universal (CDU). En A. I. Extremeño Placer (ed.). *La representación y organización*

del conocimiento: metodologías, modelos y aplicaciones. ISKO-España. Universidad de Alcalá de Henares.

Olson, H. A. (1997). Thinking professionals: teaching critical cataloguing. *Technical Services Quarterly*, 15(1/2), 51-66.

Olson, H. A. (1998). Mapping beyond Dewey's boundaries: constructing classificatory space for marginalized knowledge domains. *Library Trends*, 47(2), 233-254.

Olson, H. A. (2001a). Patriarchal structures of subject Access and subversive techniques for change. *Canadian Journal for Information and Library Science*, 26(2/3), 1-29.

Olson, H. A. (2001b). Sameness and difference. A cultural foundation of classification. *Library Resources & Technical Services*, 45(3), 115-122.

Olson, H. A., & Schlegl, R. (2021). Standardization, objectivity and user focus: a meta-analysis of subject Access critiques. *Cataloging & Classification Quarterly*, 32(2), 61-80.

Rodríguez-Bravo, B. (2006). *The visibility of women in indexing languages*. En *Knowledge organization for a global learning society*. Ergon Verlag, 413-422.

Rodríguez-Bravo, B. (2007). *La integración de la mujer en los lenguajes documentales: una utopía necesaria en la sociedad del conocimiento*. BiD: textos universitaris de biblioteconomia i documentació, 18, <https://raco.cat/index.php/BiD/article/view/72761>

Simões, M. da G., Rodríguez-Bravo, B., & Pestana, O. (2018). Representação do conceito de mulher na Classificação Decimal Dewey (CDD) e na Classificação Decimal Universal (CDU): duas perspectivas sobre o mesmo conceito? *Liinc Em Revista*, 14(2). <https://doi.org/10.18617/liinc.v14i2.4340>

Uso de *chatbots* na produção de resumos científicos: o que os resumos nos dizem da investigação? – Uma revisão sistemática

Helena Leitão

0000-0002-8417-6657

Universidade Católica Portuguesa – Centro Regional do Porto, Portugal

hleitao@ucp.pt

RESUMO: Os resumos científicos são versões abreviadas, concisas e fiéis do conteúdo intelectual de um documento original. Eles podem ter uma estrutura explícita ou implícita e desempenham funções de alerta, antecipação, substituição e seleção de informação. Tradicionalmente, são produzidos por operadores humanos, mas, com o avanço da tecnologia, sistemas automáticos também passaram a ser utilizados, especialmente a partir da segunda metade do século XX. Os sistemas automáticos, inicialmente baseados na frequência e posição das palavras, evoluíram com a inteligência artificial. *Chatbots*, como o *Chat GPT*, lançado em 2022, revolucionaram a produção de texto, inclusive de resumos científicos, ao gerar conteúdo realista com base em grandes volumes de texto pré-existente. No entanto, há preocupações éticas sobre a autoria e a precisão científica desses resumos, uma vez que IA não pode ser responsável pelo conteúdo gerado.

Este estudo objetiva, através de resumos de artigos, analisar a utilização de ferramentas de IA na construção de resumos científicos, avaliando temas, contextos, métodos e conclusões de estudos recentes.

Metodologicamente, procedeu-se a uma revisão sistemática da literatura, utilizando-se as bases *Web of Science*, *Scopus* e *Ebsco Discovery Service*, focando em artigos dos últimos três anos.

A quantidade de estudos que existem sobre o tema, e que cujos resumos foram objeto de análise, demonstram que ainda há muita necessidade de

estudo e reflexão acerca deste tema, Destaca-se a dificuldade em distinguir resumos gerados por IA dos humanos e a controvérsia sobre a qualidade e ética do uso de IA. Apesar das melhorias na tecnologia, a validação humana ainda é crucial para garantir a precisão e a qualidade dos resumos, reforçando a importância do papel humano na supervisão e adaptação dos conteúdos gerados por IA.

PALAVRAS-CHAVE: Resumos científicos; Inteligência artificial; Resumos automáticos; Resumos manuais.

ABSTRACT: Scientific abstracts are abbreviated, concise, and faithful versions of the intellectual content of an original document. They can have an explicit or implicit structure and have functions such as alerting, anticipating, substituting, and selecting information. Traditionally, they are produced by human operators, but with technological advancements, automated systems have also been used, especially since the second half of the 20th century. Initially, these automated systems were based on word frequency and position, but they have evolved with artificial intelligence. *Chatbots*, like *Chat GPT*, launched in 2022, have revolutionized text production, including scientific abstracts, by generating realistic content based on large volumes of pre-existing text. However, there are ethical concerns about authorship and the scientific accuracy of these abstracts since AI cannot be held responsible for the generated content.

This study aims to analyse the use of AI tools in constructing scientific abstracts through article summaries, evaluating themes, contexts, methods, and conclusions of recent studies.

Methodologically, a systematic literature review was conducted using the Web of Science, Scopus, and Ebsco Discovery Service databases, focusing on articles from the last three years.

The existing studies, whose summaries were analyzed, demonstrate a significant need for further study and reflection on this topic. There is a noted difficulty in distinguishing AI-generated abstracts from human ones and controversy regarding the quality and ethics of using AI. Despite technological improvements, human validation remains crucial to ensure the accuracy and quality of abstracts, reinforcing the importance of human oversight and adaptation of AI-generated content.

KEYWORDS: Scientific abstracts; Artificial intelligence; Automatic abstracts; Human abstracts.

Introdução

Um resumo científico consiste numa versão abreviada, concisa e fiel do conteúdo intelectual do documento que lhe deu origem (Bernier, 1968; Lancaster & Neway, 1982; Leitão et al., 2015; Maizell et al., 1979; Moore, 2009; Pinto Molina, 2001; Simões, 2015).

Tendo como funções alertar, antecipar, substituir e selecionar informação (Simões, 2015), no que se refere à sua construção, apesar de tradicionalmente estar associada a operadores humanos, particularmente ao autor do documento primário, fruto de um trabalho intelectual (Moreiro Gonzalez, 2004; Pinto Molina, 2001), pode ser produzida por outros operadores humanos (especialista na matéria ou resumidor profissional), ou por sistemas automáticos.

Os sistemas automáticos de produção de resumos surgiram na segunda metade do século XX (Cunha et al., 2009; Gillick, 2011) com Hans Peter Luhn (Kobashi, 1996). A técnica de construção de resumos automáticos até ao surgimento dos softwares atuais de inteligência artificial (IA), estava relacionada com uma lógica de significância das frases (Luhn, 1958).

Os *chatbots*, criados com o objetivo de compreenderem, interpretar e gerarem texto idêntico ao produzido pelo ser humano (Rossetini et al., 2023), têm atraído os investigadores e revolucionado a forma como a informação científica é produzida (Osama & Afridi, 2023), e vêm colmatar as lacunas que existiam nos processos automáticos de produção de resumos baseados em métodos extrativos, em que o produto final era uma junção de extratos do texto original.

O lançamento do *chatbot Chat GPT*, em 2022, revolucionou a forma como a informação é produzida (Hua et al., 2023), na medida em que gera texto realista em resposta a solicitações específicas do utilizador (Else, 2023). Constituindo-se como um *large language model* (LLM), baseia-se “on neural networks that learn to perform a task by digesting huge amounts of existing human-generated text” (p. 423).

Uma questão importante, e para a qual o *Committee on Publication Ethics* (COPE) salienta é a autoria da informação. Para o COPE (2023),

as ferramentas de IA, não podem ser indicadas como autores, já que “AI tools cannot meet the requirements for authorship as they cannot take responsibility for the submitted work”.

A verdade é que frequentemente são apresentadas vantagens e desvantagens estimadas do uso destes sistemas, havendo uma constante referência às alucinações dos sistemas de IA, que acabam por produzir texto que não é cientificamente válido (Else, 2023). Para Rossettini et al. (2023), a forma como direcionamos o *chatbot*, ou seja, as questões e as indicações que lhe damos, são determinantes para o produto final, sendo uma das suas *fraquezas* a dependência que tem das orientações humanas.

Vários estudos e discussões têm sido promovidas acerca da aplicabilidade da IA e dos *chatbots* na análise e produção de informação científica. Assim, questiona-se que investigação está a ser feita e que conclusões já se alcançaram acerca da sua utilização na produção de resumos científicos.

Objetivo e Metodologia

O trabalho tem como objetivo concluir acerca dos temas e contextos dos estudos mais atuais no que se refere ao uso de ferramentas de IA – particularmente *chatbots* – na construção de resumos científicos. Esta aferição é relevante na medida em que nos permite compreender, por um lado, que pontos-chave estão a ser estudados, com que objetivos, que métodos estão a ser utilizados, que resultados têm sido alcançados e, principalmente, que conclusões podem ser retiradas; e por outro contribuir para a reflexão sobre o tema no sentido de desenvolvimento de estudos mais aprofundados, ou de reflexão acerca de tópicos que estão a ser marginalizados nas análises.

Do ponto de vista metodológico, optou-se por uma revisão sistemática da literatura, incidindo a análise nos resumos de artigos. Utilizaram-se as bases de dados referenciais Web Of Science e Scopus e a plataforma Ebsco Discovery Service. Com base nas palavras-chave em inglês *scientific abstract*, *artificial intelligence*, *chatbots* criou-se a expressão “*scientific abstract**” AND (“*artificial intelligence*” OR IA OR *Chatbots*). A pesquisa incidiu nos campos de *título*, *resumo* e *palavras-chave* e optou-se por

inserir um limitador cronológico, restringindo-se a documentos publicados nos últimos 3 anos (2022 atualidade-junho de 2024).

Foram recuperados 46 artigos, sendo identificados 20 duplicados. Foi ainda removido um documento que sugeria ser um duplicado e dois documentos que não apresentavam resumos. Após triagem dos 12 resultados restantes através dos resumos, foram excluídos 5 resultados que não estavam relacionados efetivamente com o tema em análise. Assim, concluiu-se um total de 7 resumos de artigos para revisão.

Quadro 1 – Estudos incluídos na revisão

Autor	Título
Kim H.J. et al.	<i>Assessing the Reproducibility of the Structured Abstracts Generated by ChatGPT and Bard Compared to Human-Written Abstracts in the Field of Spine Surgery: Comparative Analysis</i>
Eppler M.B. et al.	<i>Bridging the Gap Between Urological Research and Patient Understanding: The Role of Large Language Models in Automated Generation of Layperson's Summaries</i>
Howard F.M. et al.	<i>Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023</i>
Cheng S.-L. et al.	<i>Comparisons of Quality, Correctness, and Similarity Between ChatGPT-Generated and Human-Written Abstracts for Basic Research: Cross-Sectional Study</i>
Theocharopoulos P.C. et al.	<i>Detection of Fake Generated Scientific Abstracts</i>
Hong-Uyen Hua, et al.	<i>Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots</i>
Muhammad Osama; Sabah Afridi	<i>ChatGPT: A New Era in Research Writing Assistance</i>

No que se refere à análise dos resumos, do ponto de vista formal, analisou-se a sua autoria – se identificada como humana ou produzida por IA. Do ponto de vista do conteúdo, analisou-se a área do conhecimento geral do artigo, com vista a concluir se há uma tendência para a publicação

sobre este tema nalgum contexto científico em particular; os elementos estruturantes – contexto, procurando perceber o que desencadeou o estudo; os objetivos que nortearam os estudos; a metodologia seguida as conclusões retiradas dos estudos; e as palavras-chave, aferindo que termos são mais utilizados na representação destes assuntos.

Para a organização da informação recolhida, utilizou-se um quadro.

Resultados

Após análise dos resumos dos artigos, verificou-se que, no que se refere à autoria, nenhum dos artigos identificava a responsabilidade pela construção do texto. Pelo facto de não haver uma referência ao uso de ferramentas de IA na sua construção, e pelas suas características textuais e estruturais, infere-se que os 7 resumos foram construídos por um operador humano, no caso o/os autores do artigo original.

Relativamente às áreas do conhecimento, 6 artigos enquadram-se na área da medicina tanto tematicamente quanto ao nível da revista onde foram publicados. O artigo restante, não identifica nenhuma área em concreto.

No que se refere ao enquadramento, especificamente ao pressupostos que potenciaram estes estudos: a) dificuldade na distinção entre resumos produzidos por IA e por humanos; b) controvérsia no uso da IA para a produção de conteúdo científico particularmente resumos; c) falta de estudos acerca das diferenças na capacidade dos vários softwares de IA; d) impacto significativo dos *chatbots*, tanto para a literacia em saúde como para a investigação; e) falta de estudos acerca das desvantagens dos *softwares* de IA na produção de resumos, particularmente da criação de conteúdo impreciso proveniente de alucinações; b) inevitabilidade da integração da IA no quotidiano da população.

No que se refere aos objetivos, podem ser resumidos em 5: a) comparação da similaridade entre o produto da IA e o operador humano; b) avaliação da probabilidade dos resumos terem sido produzidos por AI; c) comparação da qualidade do produto humano e de AI; d) Capacidade de o software de AI gerar resumos adequados ao público leigo; e) avaliação da aplicabilidade de um modelo de AI na construção de resumos;

Do ponto de vista da metodologia, verificou-se quatro métodos distintos relacionados com produção de resumos e com a avaliação do produto final: a) comparação entre resumos produzidos pela IA e pelo homem com base em instruções aos autores; b) utilização de validadores humanos, especialistas na matéria, para a deteção cega de resumos produzidos por IA e por operadores humanos; c) aplicação de métodos automáticos para a deteção de conteúdo produzido por IA; d) Utilização algoritmos para a verificação da similaridade entre os dois tipos de resumos.

A utilização de validadores humanos e de métodos automáticos para deteção de conteúdo produzido por IA, surgem em na maioria dos artigos.

Importa salientar que, do ponto de vista de *chatbots*, 6 artigos (em 7), analisaram o Chat GPT. Apenas dois artigos tiveram como base outros *chatbots* e um dos artigos não refere nenhum em concreto.

Relativamente às conclusões dos estudos, podemos sistematizá-las em 9 pontos-chave: a) na comparação entre o *Chat GPT* e o *Bard*, ambos podem ser usados para ajudar a produzir resumos, mas a maioria dos resumos são considerados antiéticos devido às altas taxas de plágio e deteção de IA; b) tendo em consideração as diretrizes de formatação, os resumos gerados pelo *Chat GPT*, em comparação com outros *softwares*, demonstram superioridade; c) o *ChatGPT* pode criar resumos precisos para o público leigo a partir de resumos científicos, com instruções bem elaboradas que melhoram a facilidade de utilização da informação; d) o uso do *ChatGPT* para criação de resumos científico pode não apresentar problemas de similaridade quando a sua fonte são textos completos reais escritos por humanos; e) tendo em consideração que os revisores humanos não distinguem com precisão os resumos produzidos por humanos e os produzidos por IA, é fundamental avaliar os limites éticos do uso de programas de IA; f) verifica-se uma presença crescente de conteúdo gerado por IA; g) conclui-se que o tamanho do resumo ou o tipo de artigo não afetam a probabilidade de o conteúdo ser gerado por IA; h) recomenda-se prudência no uso destes recursos de IA para fins de educação em saúde ou académicos; i) embora os resumos produzidos por IA sejam satisfatórios, é necessária uma verificação especializada para melhorar a precisão.

No que se refere às palavras-chave, 2 dos 7 resumos não apresentam nenhuma. Os termos presentes nos restantes 5 artigos, foram sistematizados num quadro. Salientamos a quantidade de formas de representação dos conceitos que foram a base da nossa pesquisa: 3 termos representativos do conceito de IA, 4 representativos do conceito de modelos de linguagem, um representativo dos *Chatbots* e 3 representativos dos resumos científicos. As variações estão relacionadas com abreviaturas, plurais e singulares e relações entre conceitos mais gerais e mais específicos.

Conclusão – O que os resumos nos dizem?

O uso de ferramentas de IA está claramente disseminado na produção de informação científica. Ao nível dos estudos sobre a produção de resumos, conclui-se, em primeiro lugar, que o número de estudos publicados ainda são muito limitados.

Um aspeto importante, tem a ver com o facto de nenhum dos resumos se identificar como produzido por IA, inferindo-se que são produto do trabalho humano. Esta inferência justifica-se, tendo em consideração que seria essencial essa informação estar descrita na metodologia caso um sistema de IA tivesse sido utilizado, indo ao encontro do entendimento do COPE (2023) de que “Authors who use AI tools in the writing of a manuscript, production of images or graphical elements of the paper, or in the collection and analysis of data, must be transparent in disclosing in the Materials and Methods (or similar section) of the paper how the AI tool was used and which tool was used”.

Estes *resumos* também reforçam o facto de continuar a ser a área da saúde a pioneira na investigação acerca dos resumos científicos, produzindo quase todos os estudos inovadores. O artigo acerca da produção de resumos para o público leigo relembra-nos um aspeto fundamental – a literacia da população. Apesar de, historicamente, a literatura científica ter como foco a comunicação entre pares (Borges, 2006), a informação produzida deve estar à mercê do uso e discussão pública. No caso da saúde, por exemplo, “Medical journals, intended primarily for specialists, have widened their audience with the advent of the World Wide Web” (Splendiani et al., 2014, p. 419).

Constata-se o facto de que os sistemas de IA estão implantados e demonstram melhores resultados e mais alternativas a cada dia, mas o *Chat GPT* continua a ser o *chatbot* mais reconhecido.

Do ponto de vista metodológico, conclui-se que os estudos têm uma vertente tendencialmente comparadora dos dois tipos de resumos. Da mesma forma, utilizam validadores, sendo que os validadores humanos têm um papel muito relevante, particularmente numa fase em que a aplicabilidade de *chatbots* na construção de resumos apresenta taxas significativas de erro e/ou apresenta claramente características que fazem com que o conteúdo seja imediatamente identificado como produzido por IA.

Salienta-se a dependência que os *chatbots* têm da interação com o homem, e o quanto as orientações humanas são importantes para a precisão. No caso dos resumos, este aspeto é muito significativo, podendo ser determinante na escolha do sistema a utilizar, na medida em que permite que possamos indicar algumas especificidades necessárias, como as orientações das revistas, a estrutura pretendida, dimensões, entre outros aspetos. Assim, verifica-se que o operador humano continua a ter um papel central na garantia da qualidade dos processos automáticos – pela intervenção que tem através das *prompts* – e do produto final, através da sua adaptação, correção e validação. Com isto, o princípio da prudência deve ser sempre considerado, principalmente quando a aplicação destes sistemas afeta aspetos essenciais da vida humana e a responsabilidade final pelo produto de informação é sempre do autor humano que recorreu à ferramenta.

Referências bibliográficas

- Bernier, C. (1968). Abstracts and abstracting. In *Encyclopedia of library and information science* (Vol. 1, pp. 16-38). Marcel Dekker Inc.
- Borges, M. M. (2006). *A Esfera: Comunicação Académica e Novos Media* [University of Coimbra]. <http://hdl.handle.net/10316/8557>
- COPE. (2023, February 13). *COPE position statement*. <https://publicationethics.org/cope-position-statements/ai-author>
- Cunha, I., Torres-Moreno, J. M., Velazquez-Morales, P., & Vivaldi, J. (2009). Un algoritmo lingüístico-estadístico para resumen automático de textos especializados. *Linguamática*, 2, 67-80. <http://www.linguamatica.com/index.php/linguamatica/article/view/33/0>

- Else, H. (2023). Abstracts written by ChatGPT fool scientists. *Nature*, 613. <https://doi.org/10.1101/2022.12.23.521610>
- Gillick, D. J. (2011). The Elements of Automatic Summarization. *EECS UC Berkeley Thesis*, 53(9), 1689-1699. <https://doi.org/10.1017/CBO9781107415324.004>
- Hua, H. U., Kaakour, A. H., Rachitskaya, A., Srivastava, S., Sharma, S., & Mammo, D. A. (2023). Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots. *JAMA Ophthalmology*, 141(9), 819-824. <https://doi.org/10.1001/jamaophthalmol.2023.3119>
- Kobashi, N. Y. (1996). Análise documentária e representação da informação. *Informare: Cadernos Do Programa de Pós-Graduação Em Ciência Da Informação*, 2(2), 5-27. https://edisciplinas.usp.br/pluginfile.php/359680/mod_resource/content/1/Kobashi AD.pdf
- Lancaster, F. W., & Neway, J. M. (1982). The future of indexing and abstracting services. *Journal of the American Society for Information Science*, 33(3), 183-189. <https://doi.org/10.1002/asi.4630330315>
- Leitão, H., Borges, M. M., & Simões, M. da G. (2015). Resumos Científicos em Saúde – Estudo exploratório de resumos de artigos das revistas Nature e PLOS One. *Ciência e Sociedad*. <https://doi.org/10.13140/RG.2.1.3056.2641>
- Luhn, H. P. (1958). The Automatic Creation of Literature Abstracts. *IBM Journal of Research and Development*, 2(2), 159-165. <https://doi.org/10.1147/rd.22.0159>
- Maizell, R., Smith, J., & Singer, T. (1979). *Abstracting scientific and technical literature : an introductory guide and text for scientists, abstractors, and management*. R. E. Krieger Pub. Co.
- Moore, A. (2009). What's in an abstract? *BioEssays*, 31(3), 261. <https://doi.org/10.1002/bies.200900023>
- Moreiro Gonzalez, J. (2004). *El contenido de los documentos textuales: su analisis y representación mediante el lenguaje natural*. Ediciones Trea.
- Osama, M., & Afridi, S. (2023). ChatGPT: A New Era in Research Writing Assistance. In *Journal of the Pakistan Medical Association* (Vol. 73, Issue 9, pp. 1929-1930). Pakistan Medical Association. <https://doi.org/10.47391/JPMA.9183>
- Pinto Molina, M. (2001). *El resumen documental: Paradigmas, modelos y métodos*. Fundación Germán Sánchez Ruipérez.
- Rossettini, G., Cook, C., Palese, A., Pillastrini, P., & Turolla, A. (2023). Pros and Cons of Using Artificial Intelligence Chatbots for Musculoskeletal Rehabilitation Management. In *Journal of Orthopaedic and Sports Physical Therapy* (Vol. 53, Issue 12, pp. 728-734). Movement Science Media. <https://doi.org/10.2519/jospt.2023.12000>
- Simões, M. da G. (2015). *Resumo Documental: um incursão à (des)construção concetual na literatura científica*. Imprensa da Universidade de Coimbra.

Classificar, um ato imprescindível à organização do universo

Luis Miguel Oliveira Machado

ORCID: 0000-0003-3403-5618

Universidade de Coimbra, Centro de Estudos Interdisciplinares, Portugal
lmmachado@yahoo.com

RESUMO: O processo de classificar é a base dos sistemas de organização do conhecimento. A forma como esse processo é executado resulta em diferentes artefactos de representação. Dentro da abordagem ontológica, a aplicação de rigorosos princípios ontológicos e formais do processo classificatório resulta num modelo ontológico, que se distingue de sistemas mais permissivos. Uma das características destes modelos rigorosos é o seu potencial de interoperabilidade semântica, mesmo em contextos interdisciplinares. Esta introdução pretende chamar à atenção para a necessidade da inclusão da abordagem ontológica na organização do conhecimento

PALAVRAS-CHAVE: Classificar; Modelo ontológico; Abordagem ontológica.

ABSTRACT: The process of classifying is the basis of knowledge organization systems. The way this process is executed results in different representation artifacts. Within the ontological approach, applying rigorous ontological and formal principles of the classificatory process results in an ontological model, distinguished from more permissive systems. One of the characteristics of these rigorous models is their potential for semantic interoperability, even in interdisciplinary contexts. This introduction draws attention to the need to include the ontological approach in the organization of knowledge.

KEYWORDS: Classifying; Ontological model; Ontological approach.

Síntese introdutória à abordagem ontológica na organização do conhecimento

“Classificar assume-se como um acto imprescindível à organização do universo.” (Simões, 2011, p. 28), estas palavras servem de mote para uma breve incursão à abordagem ontológica na organização do conhecimento. Uma abordagem não tradicional numa matéria de estudo de eleição de Maria da Graça Simões. Desta forma, pretende-se prestar uma singela homenagem ao espírito pioneiro daquela que foi uma grande estudiosa da teoria da classificação.

O ato de classificar é imprescindível tanto como uma necessidade cognitiva de apreensão do mundo, como uma necessidade organizativa do conhecimento. De facto, a miríade de artefactos desenvolvidos intencionalmente para a organização do conhecimento humano, apesar de poderem resultar de distintos processos organizacionais, têm em comum o ato de classificar. Nesse sentido, será possível considerar os sistemas de organização do conhecimento (SOC) como sistemas de classificação. Importa enfatizar que esta designação genérica é distinta da designação clássica aplicada apenas a um tipo específico de sistemas, comumente chamados de classificações bibliográficas.

Tradicionalmente as classificações bibliográficas seguem o que se pode chamar de uma abordagem epistemológica ou disciplinar. Em classificações como a Classificação Universal Decimal, por exemplo, classifica-se as áreas de estudo e não os objetos de estudo. Uma abordagem ontológica, por seu lado, procura classificar o que existe e como essas entidades se relacionam. A Integrative Leves Classification (ILC) é um raro exemplo dessa abordagem no contexto das classificações bibliográficas (Gnoli et al., 2022). No âmbito dos SOC são as chamadas ontologias computacionais que monopolizam a abordagem ontológica, embora nem todas o façam no sentido rigoroso do termo. Neste ponto é necessário introduzir algumas definições por forma a clarificar os diferentes artefactos em questão:

Análise ontológica = def. uma categorização dos tipos de entidades existentes em algum domínio que podem ser objetivamente subsumidos sob categorias distinguíveis.

Sistema ontológico = def. um artefacto de representação cujo processo classificatório segue uma análise ontológica.

Modelo ontológico = def. um sistema ontológico que segue rigorosos princípios ontológicos e formais na definição dos tipos de entidade que incorporam, assim como as relações entre estes, refletindo uma adequação representativa à realidade exterior ao próprio modelo.

Partindo destas definições e do vasto e diferenciado grupo de artefactos a que termo *ontologia computacional* é aplicado, pode-se afirmar que o mesmo integra alguns modelos ontológicos, mas também vários que apenas são sistemas ontológicos e, até, alguns artefactos de representação que nem nesse grupo podem ser incluídos (vide Fig.1).

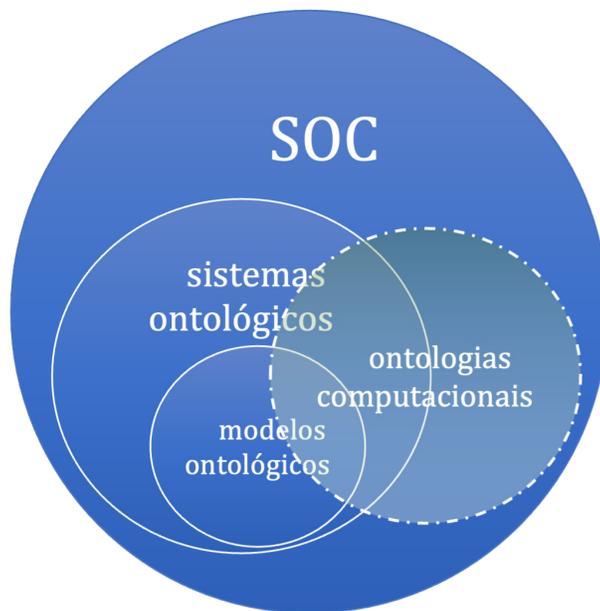


Figura 1 – Relação entre as ontologias computacionais e os sistemas e modelos ontológicos

Pelas especificações da ILC esta situa-se dentro dos sistemas ontológicos, mas fora dos modelos ontológicos (Machado, 2021a). Estes modelos apresentam como características distintas os rigorosos princípios ontológicos

e formais que aplicam no processo classificatório. Importa, portanto, discorrer um pouco mais sobre os mesmos. Ir-se-á, de forma breve, abordar os seguintes princípios de um modelo ontológico:

- i. Apresenta definições precisas e ontologicamente consistentes.
- ii. Diferencia claramente as instâncias das classes, nomeadamente no que respeita às relações entre ambas.
- iii. Classifica com base nos traços pertencentes às respetivas entidades, i.e., às suas características intrínsecas.
- iv. Não representa os dados de uma base em particular, mas as classes de entidade existentes no domínio ao qual os dados reportam.
- v. Na modelação ontológica, os conceitos são auxiliares de representação não os representados.

A importância do princípio i. – a precisão e consistência ontológica das definições incorporadas nos modelos ontológicos, decorre da necessidade de se alcançar interoperabilidade semântica entre SOC (Arp et al., 2015). O caso real, ocorrido entre 2000 e 2007, da construção da aeronave de passageiros A380 é demonstrativo desta importância. Construída por partes em diferentes países, no momento da junção das mesmas tal não foi possível devido a uma diferença milimétrica nos orifícios da cablagem. Sendo a razão para tal o facto dos sistemas usados tratarem de forma ontologicamente diferente esses orifícios. Para um sistema o orifício era uma entidade com três dimensões e para o outro a mesma entidade tinha apenas duas dimensões. Um desacordo ontológico que resultou em avultados prejuízos para os responsáveis do projeto.

No princípio ii. outro aspeto da precisão necessária num modelo ontológico é colocado em destaque. Trata-se de diferenciar a relação *é_um(a)*, de subordinação entre classes, da relação *instância_de*, entre um particular e uma classe. A primeira é uma relação efetuada a um nível de abstração de particularidades individuais. A segunda faz a ponte entre esse nível de abstração e o nível concreto das instâncias particulares. Tratar as duas relações como iguais, como as comumente denominadas de ontologias ligeiras o fazem, é um erro ontológico com implicações na compatibilidade semântica e erros de inferência quando se procura interoperar com estes artefactos (Machado, 2021b). Este é apenas um dos critérios não cumpridos

por estas ontologias computacionais para que possam ser consideradas modelos ontológicos.

O princípio iii. tem uma aplicação mais alargada que não apenas aos modelos ontológicos. Ele pode ser usado como auxiliar na decisão de critérios de classificação. Tarefa considerada por alguns autores como subjetiva, afirmando não existir forma objetiva de estabelecer prioridades sobre as propriedades dos objetos a classificar (Hjørland, 2017). O autor citado apresenta, como exemplo, três figuras geométricas (cf. Fig. 2) afirmando que podem ser classificadas pela *forma* ou pela *cor* sem que uma dessas propriedades possa ser objetivamente considerada mais importante. Contudo, ao contrário da cor, a propriedade *forma* permite prever propriedades adicionais das figuras (por exemplo, o número de lados ou a amplitude dos ângulos). Esta característica pode ser encarada como um indicador da sua ligação intrínseca ao objeto. A relevância duma propriedade partilhada por um grupo de objetos aumenta quando providencia informação adicional sobre o respetivo grupo (Parsons & Wand, 2013).

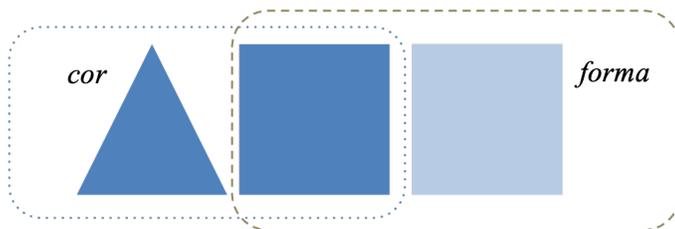


Figura 2 – Diferentes critérios de classificação, igualdade de relevância?

No princípio iv. volta-se novamente à questão recorrente da interoperabilidade semântica. Um SOC que na sua representação apenas considera um particular conjunto de dados, sem atender ao contexto mais alargado, dificilmente será compatível com outras classificações (Jansen, 2008). Um exemplo retirado do thesaurus do Instituto Nacional de Cancro dos Estados Unidos, do início dos anos dois mil, é demonstrativo da questão. Nele, a entrada ‘paciente’ é subdividida em ‘paciente com cancro’ e ‘paciente ambulatorio’. No contexto específico da classificação todos os pacientes são pacientes com cancro, mas a não especificação das classes em ‘paciente

internado com cancro' e 'paciente ambulatorio com cancro', provoca erros na interoperabilidade com outros SOC. Num contexto mais alargado, um paciente ambulatorio não é forçosamente um paciente com cancro e um paciente com cancro pode ser internado ou ambulatorio.

Por fim, o princípio v. foca um aspeto bastante controverso no que respeita às classificações – as suas unidades elementares. Enquanto artefactos de representação, a natureza conceptual destas unidades classificatórias será incontornável, contudo não se deverá confundir o representante pelo representado. Uma classificação ontológica necessita de distinguir existência real, o que possui energia e é capaz de mudança, de existência ideal, o que, por falta de energia própria, só muda por ação de terceiros (Machado et al., 2023). Uma distinção que uma entidade como o *conceito*, com tamanha ambiguidade semântica e ontológica (Machado, Veronez Júnior, & Martínez-Ávila, 2022), não se mostra adequada para a fazer. Para uma classificação ontológica, a referida distinção torna necessária a revisão da tendência dominante de apontar o *conceito* como unidade básica dos SOC.

Conclusão

Nesta breve introdução à abordagem ontológica enfatizou-se dois importantes aspetos. A demarcação entre sistemas e modelos ontológicos e sua relação com as ontologias computacionais, devido à tendência para um uso demasiado abrangente e vago do termo ontologia no contexto dos SOC (Machado, Almeida, et al., 2020). Em segundo lugar, a importância da abordagem ontológica para a interoperabilidade semântica entre SOC, principalmente em contextos interdisciplinares (Machado, Simões, et al., 2020; Machado & Borges, 2022).

Ao dar ênfase à abordagem ontológica para a organização do conhecimento não se está a advogar a sua total dominância, antes a sua equivalência em termos de importância com a dominante abordagem epistemológica. As duas têm o seu papel no âmbito dos SOC que, de outra forma, não se ocuparão de todo o espectro possível do conhecimento.

Referências bibliográficas

- Arp, R., Smith, B., & Spear, A. D. (2015). *Building ontologies with Basic Formal Ontology*. MIT Press. <https://doi.org/10.7551/mitpress/9780262527811.001.0001>
- Gnoli, C., Almeida, P. de, Machado, L. M. O., & Ridi, R. (2022). Taiga Penguins: Expressing Existence and Fictionality in a Phenomenon-Based Classification. In M. Lykke, T. Svarre, D. Haynes, M. Skov, M. Thellefsen, & D. Martínez-Ávila (Eds.), *Knowledge Organization across Disciplines, Domains, Services and Technologies* (pp. 101-110). Ergon – Nomos. <https://doi.org/10.5771/9783956509568-101>
- Hjørland, B. (2017). Classification. *Knowledge Organization*, 44(2), 97-128. <https://doi.org/10.5771/0943-7444-2017-2-97>
- Jansen, L. (2008). Classifications. In K. Munn & B. Smith (Eds.), *Applied Ontology: An Introduction* (pp. 159-172). Ontos Verlag. <http://doi.org/10.1515/9783110324860>
- Machado, L. M. O. (2021a). Ontologias, dos sistemas aos modelos: Uma abordagem introdutória no contexto dos sistemas de organização do conhecimento. *Fronteiras da Representação do Conhecimento*, 1(1), 1-18. <https://periodicos.ufmg.br/index.php/fronteiras-rc/article/view/33869>
- Machado, L. M. O. (2021b). Ontologies in Knowledge Organization. *Encyclopedia*, 1(1), 144-151. <https://doi.org/10/gk2dds>
- Machado, L. M. O., Almeida, M. B., & Souza, R. R. (2020). What researchers are currently saying about ontologies: A review on recent Web of Science articles. *Knowledge Organization*, 47(3), 199-219. <https://doi.org/10.5771/0943-7444-2020-3-199>
- Machado, L. M. O., & Borges, M. M. (2022). A partilha interdisciplinar de conhecimento: Algumas questões teóricas e operacionais em torno dos sistemas de organização do conhecimento. *Revista Estudos do Século XX*, 22(digital humanities), Article 22. https://doi.org/10.14195/1647-8622_22_2
- Machado, L. M. O., Martínez Ávila, D., Almeida, M. B., & Borges, M. M. (2023). Towards a Moderate Realist Foundation for Ontological Knowledge Organization Systems: The Question of the Naturalness of Classifications. *Journal of Information Science*, 49(e-print), 1-18. <https://doi.org/10.1177/01655515231160031>
- Machado, L. M. O., Simões, M. da G., Gnoli, C., & Souza, R. R. (2020). Can an Ontologically-Oriented KO Do Without Concepts? In M. Lykke, T. Svarre, M. Skov, & D. Martínez-Ávila (Eds.), *Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, 2020 Aalborg, Denmark* (pp. 502-506). Ergon–Nomos. <https://doi.org/10.5771/9783956507762-502>
- Machado, L. M. O., Verenez Júnior, W. R., & Martínez-Ávila, D. (2022). A indeterminação ontológica dos conceitos: Interpretações linguísticas e psicológicas. In T. H. B. Barros & R. do C. F. Laipelt (Eds.), *Organização e representação do conhecimento em múltiplas abordagens* (pp. 94-123). Pimenta Cultural. <https://doi.org/10.31560/pimentacultural/2022.95613.3>
- Parsons, J., & Wand, Y. (2013). Extending Classification Principles from Information Modeling to Other Disciplines. *Journal of the Association for Information Systems*, 14(5), 245-273. <https://doi.org/10.17705/1jais.00332>
- Simões, M. da G. (2011). *Classificações bibliográficas: Percorso de uma teoria*. Almedina.

**Um silêncio e uma voz:
a Literatura como desafio à
Organização do Conhecimento**

Patrícia de Almeida

ORCID 0000-0002-8061-8622

Centro de Estudos Interdisciplinares – Universidade de Coimbra, Portugal

mebpatricia@gmail.com

RESUMO: As operações semânticas de análise e representação do assunto são reconhecidas como áreas de grande relevo, fundamentais à organização e recuperação da informação. Todavia, verifica-se que os principais sistemas de classificação bibliográfica silenciaram o conteúdo intelectual da Literatura. Embora enviesada, esta prática impôs-se na tradição da Organização do Conhecimento, face aos constrangimentos técnicos e conceptuais da área ficcional, razão pela qual a consideram um desafio. Apesar de silenciada, a representação por assunto dos documentos literários e ficcionais mostra-se necessária e algumas vozes o reafirmam, principalmente nos últimos 50 anos. Historicamente, vêm-se fortalecendo novas abordagens e, na atualidade, é já possível reconhecer caminhos consolidados e estabelecer os horizontes que dão a merecida voz à Literatura. Estará, pois, criada a oportunidade para o surgimento de um novo paradigma na Organização do Conhecimento.

PALAVRAS-CHAVE: Assunto literário; Literatura; Ficção; Organização do Conhecimento.

ABSTRACT: Semantic operations of subject analysis and representation are highly important, and fundamental to the organisation and retrieval of information. However, major bibliographic classification systems have silenced the intellectual content of Literature. Such practices have persisted within the tradition of Knowledge Organisation, facing both technical and conceptual challenges

in the fictional domain, hence being deemed a challenge. While silenced, the subject representation of literary and fictional documents remains necessary, with increasing reaffirmation voices over the past 50 years. Historically, new approaches have strengthened, and today, it is possible to recognise established paths and set horizons that give Literature a due voice. Therefore, an opportunity for a new paradigm in Knowledge Organisation emerges.

KEYWORDS: Literary subject; Literature; Fiction; Knowledge Organisation.

A Organização do Conhecimento (OC) trata os modelos organizativos e representativos de realidades materializadas, tendo em vista uma futura recuperação da informação e a consequente produção de conhecimento. Esta organização de ideias sempre acompanhou o ser humano, destacando-se a sua natureza predominantemente cognitiva (Gnoli, 2014; Hjørland, 2001). Enquanto área de atuação da Ciência da Informação, sobressaem as operações de cunho semântico relacionadas com a análise e a representação documental, em concreto os múltiplos sistemas de classificação bibliográfica e de indexação por assunto.

Verifica-se, todavia, que a pesquisa por assunto é tida como a de maior complexidade e os seus resultados mostram-se frequentemente dececionantes (Hudon & El Hadi, 2010), com particular destaque para o domínio das Humanidades, designadamente para a Literatura de cariz artístico ou ficcional. A distinção no tratamento documental entre domínios do conhecimento está já há muito reconhecida, destacando-se Langridge (1976), que explora as tradicionais diferenças entre as áreas das ciências humanas e sociais e as das ciências exatas.

Esta forma convencional de divisão assenta numa perspetiva maniqueísta e redutora para todos os campos do conhecimento, pois coloca os documentos ora no domínio das faculdades imaginativas ora no domínio das faculdades racionais (Beghtol, 1994). Na verdade, os sistemas tradicionais de organização do conhecimento colocam a tónica neste último, observando-o como a voz, isto é, o paradigma da informação, remetendo ao quase silêncio o que se assume literário e ficcional. Assim, a tradição da não representação por assunto da Literatura poderá assentar na ideia repetida de que os documentos de ficção servem o entretenimento e as

emoções, enquanto os documentos de não ficção servem a informação e a razão (Lancaster, 1993).

Apesar de esta perspectiva se revelar altamente redutora para ambas as partes, poderá explicar o motivo pelo qual, de uma forma generalizada, a representação documental da Literatura (ainda) assenta no género ou forma literária, em detrimento do seu conteúdo intelectual – o assunto. Por outro lado, constata-se que o cenário da representação documental se afigura mais complexo na área ficcional e compreende-se que a OC a entenda como menos objetiva, mais interpretativa e quase evasiva, comparando-a com áreas mais técnicas e propícias a modelos de representação com menor grau de ambiguidade.

A representação documental poderá abarcar tanto uma dimensão formal como uma dimensão temática, onde os documentos de Literatura representam um desafio aos investigadores e aos profissionais da informação, no que respeita aos modelos de análise e de representação de assunto (Saarti, 1999). Portanto, imaginar um possível encontro – *match* – entre o utilizador e a organização do conhecimento ficcional permanece um exercício incrivelmente atrativo (Kelly, 2012) para a investigação em OC.

Numa perspetiva histórica, constata-se que os cinco principais sistemas de classificação bibliográfica – *Dewey Decimal Classification*, *Library of Congress Classification*, *Universal Decimal Classification*, *Bliss Bibliographic Classification* e *Colon Classification* – consolidam um paradigma que dá voz à forma do documento literário e silêncio ao seu assunto, estabelecendo uma tradição na OC. Sob o olhar da contemporaneidade, julga-se esta abordagem à Literatura como uma atitude enviesada, pouco adequada não só às necessidades dos utilizadores, mas também às atuais bibliotecas públicas e escolares onde predomina o documento ficcional. O reconhecimento ativo do estatuto “subprivilegiado” da ficção enquanto fonte de conhecimento terá a sua primeira voz forte com Annelise Mark Pejtersen, nos anos 80 do século passado, ecoando em Clare Beghtol (1994), Jarmo Saarti (1999), Rune Eriksson (2010), João Baptista Ernesto de Moraes (2011), Patrícia de Almeida (2021), entre outros.

No que se refere às capacidades expressivas do assunto, os sistemas e os processos de organização do conhecimento poderão ser arbitrários,

mas não aleatórios, existindo necessariamente uma relação conceptual entre si e a realidade que representam, seja ela ficcional ou não. Por conseguinte, defende-se que tanto os sistemas de classificação como os de indexação documental deverão atentar e refletir as especificidades das obras literárias, inclusivamente as respeitantes ao seu conteúdo intelectual. Todavia, reconhecem-se constrangimentos de ordem conceptual e técnica nas teorias e nas práticas de atribuição de assunto em Literatura.

Verifica-se uma permanente instabilidade semântica dos conceitos de “análise de assunto” e “assunto”, este último multiplicado em termos correlatos (Hjørland, 2001), entre eles “aboutness”, cujo tradução em língua portuguesa se mostra impraticável ou, no mínimo, de difícil consenso. Paralelamente, e na literatura da OC, registam-se questões problemáticas bastante pertinentes associadas à leitura técnica e escassez de elementos paratextuais (Mendes & Simões, 2002), aos modelos não consensuais de representação de assunto e controlo vocabular (Morales, 2011), bem como ao perfil dos profissionais de informação (Almeida, 2024). Por conseguinte, não será surpreendente que os estudos nesta área busquem outros trilhos, tais como as práticas de participação cidadã e etiquetagem colaborativa (Almeida, 2018) ou, mais recentemente, a abertura à indexação automática (Falk, 2023), explorando as atuais ferramentas sustentadas pelos *large language models*.

Não recusando estes percursos, contudo reconhecendo a impossibilidade de se estabelecer um conceito único de assunto em Literatura, propõe-se alcançar a sua substância através de uma perfiguração semanticamente dinâmica, com seis variáveis operacionais: assunto forma/conteúdo; assunto uni/multidimensional; assunto objetivo/subjetivo; assunto humano/auto-mático; assunto vulgar/especializado; assunto profissional/social. Cada uma corresponderá a um ponto de vista distinto sobre a representação do assunto, todavia passível de diálogo e, por isso, complementar de um todo agregador das propostas mais significativas para a representação da Literatura (Almeida, 2021). Considera-se que a teoria facetada do conhecimento proporciona o devido enquadramento teórico. Encontram-se algumas aplicações exploratórias desta proposta de representação de assunto à ficção, em Almeida e Gnoli (2021) e Almeida (2023), com explanação em

Romance, Novela, Literatura Infantil (Conto) e Literatura Oral (Lenda); no primeiro trabalho, opta-se por uma abordagem interartes, pelo que se engloba o Cinema e a Banda Desenha (novela gráfica).

Esta proposta, ainda em desenvolvimento, constitui mais uma voz representativa daqueles que não querem seguir a tradição, se isso significar o silêncio do assunto em Literatura. A verdade é que esta voz sempre existiu, estando documentados diversos esquemas de classificação de ficção desde 1899 (Baker & Shepherd, 1987), muito embora os primeiros resultados não se tenham mostrado perduráveis. Os resultados posteriores, com reflexo no momento atual, poderão constituir pelo menos três horizontes na representação do assunto literário ficcional (Almeida, 2021): horizonte anglo-saxónico – Estados Unidos da América, Inglaterra e Canadá; horizonte norte-europeu – Dinamarca, Suécia, Finlândia e Polónia; horizonte luso-brasileiro (quantitativamente menos significativo) – Brasil e Portugal. O primeiro e último unem-se pelo fator língua e pela influência ou parceria científico-intelectual, concretizada ou em potencial; o segundo une-se pelo fator geográfico, pela profundidade na análise e na representação do assunto ou pelo alcance dos serviços que prestam às suas comunidades.

Considera-se que a carência de uma resposta para as necessidades de informação dos utilizadores tem obrigado a um interesse constante pelo domínio literário e ficcional, particularmente nas últimas décadas. Por outro lado, o potencial das ferramentas digitais do século XXI trouxe um novo impulso para fazer face ao desafio que a Literatura representa na OC. No atual enquadramento, e como sistema de organização do conhecimento em destaque neste domínio, agrega-se aqui a *Integrative Levels Classification* (<http://www.iskoi.org/ilc/>), onde a tradição da OC é repensada e há lugar para a voz (assunto) da Literatura. Afigura-se, pois, que estará criada a oportunidade para um novo paradigma na representação documental.

Com efeito, algumas práticas e estudos sobre os processos de análise e de representação do assunto literário e ficcional mostram desenvolvimento contínuo e consolidação, o que, acredita-se, já não será compatível com o generalizado silêncio. Hoje, cale-se o tradicional conforto intelectual das ideias feitas. Hoje, e adulterando o verso ao poeta português renascentista António Ferreira, fale, cante e oiça-se a Literatura!

Referências bibliográficas

- Almeida, P. (2018). Etiquetagem colaborativa nas bibliotecas: o caso da Literatura. *Perspectivas em Ciência da Informação*, 23(2), 50-70. Acesso em <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2741/2068>
- Almeida, P. (2021). *Indexação de Literatura: quadro teórico e princípios gerais*. Tese de Doutorado. Faculdade de Letras da Universidade de Coimbra. Acesso em <https://hdl.handle.net/10316/95425>
- Almeida, P. (2023). Assim reza a Lenda: uma proposta para a representação de assunto. In *Informação e Conhecimento para o desenvolvimento da Amazônia*. Brasil: Universidade Federal do Pará.
- Almeida, P. (2024). A organização do conhecimento como profissão: um olhar sobre o perfil de competências do indexador. *The Frontiers of Knowledge Representation*, 1(1), 58-78. Acesso em <https://periodicos.ufmg.br/index.php/fronteras-rc/article/view/48580>
- Almeida, P., & Gnoli, C. (2021). Fiction in a phenomenon-based classification. *Cataloging & Classification Quarterly*, 59(5). Acesso em <https://doi.org/10.1080/01639374.2021.1946232>
- Baker, S. L., & Shepherd, G. W. (1987). Fiction Classification Schemes: The principles behind them and their success. *RQ*, 27(2), 245-251.
- Beghtol, C. (1994). *The Classification of Fiction: The Development of a System Based on Theoretical Principles*. Metuchen: Scarecrow Press.
- Eriksson, R. (2010). *Klassifikation og indeksering af skønlitteratur: Et teoretisk og historisk perspektiv*. Tese de doutorado. Danmarks Biblioteksskole. Acesso em https://komm.ku.dk/ansatte/?pure=files%2F47028127%2FEriksson_phd_2010.pdf
- Falk, O. (2023). Representing the aboutness of fiction: A comparison of manual and computational perspectives. Presented at the *Cultivating Information, Organizing Culture: Exploring Classification Practices in Modern Culture*, Copenhagen, December 7-8, 2023. Acesso em <https://urn.kb.se/resolve?urn=urn:nbn:se:hb:diva-31051>
- Gnoli, C. (2014). Knowledge Organization Systems. In Golub, K. (Ed.). *Subject access to information: an interdisciplinary approach* (43-65). California: Libraries Unlimited.
- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content... and relevance. *Journal of the American Society for Information Science and Technology*, 52(9), 774-778. Acesso em <https://doi.org/10.1002/asi.1131>
- Hudon, M., & El Hadi, W. M. (2010). Organisation des connaissances et web 2.0 de l'organisation hiérarchique centralisée à l'organisation sociale distribuée. *Les Cahiers du Numérique*, 6, 9-38. Acesso em <https://www.cairn.info/revue--2010-3-page-9.htm>
- Kelly, M. A. (2012). *Library Encounters: textuality and the institution*. Tese de Doutorado. University of Sydney. Acesso em <https://ses.library.usyd.edu.au/handle/2123/14380>
- Lancaster, F. W. (1993). *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos.
- Langridge, D.W. (1976). *Classification and indexing in the Humanities*. London: Butterworth.
- Mendes, M. T. P.; Simões, M. G. (2002). *Indexação por assuntos: princípios gerais e normas*. Lisboa: Gabinete de Estudos a&b.
- Moraes, J. B. E. (2011). *A questão do aboutness no texto narrativo de ficção: perspectivas metodológicas para a ciência da informação*. Tese de Livre Docência. Faculdade de Filosofia e Ciências da Universidade Estadual Paulista.
- Saarti, J. (1999). *Kaunokirjallisuuden Sisällönkuailun Aspektit. Kirjastoammattilaisten ja kirjastonkäyttäjien tekemien romaanien tiivistelmien ja asiasanoitusten yhdenmukaisuus*. Tese de Doutorado. Oulun Yliopisto. Acesso em <http://jultika.oulu.fi/Record/isbn951-42-5476-7>

“Taxio... o quê”?: Um estudo sobre a taxonomia através dos Sistemas de Organização do Conhecimento

Fernanda Daniel da Silva

0000-0002-4027-8264

Universidade de Coimbra, Brasil

fernanda.silva@student.fl.uc.pt

RESUMO: A taxonomia é um sistema de organização do conhecimento que sistematicamente organiza itens de informação em ambientes digitais. Isso permite aos usuários dos sistemas de informação acessar e recuperar informações com rapidez e eficiência. Organizar e classificar são atos intrínsecos ao ser humano, possuindo um profundo significado como atividades sociais. Essas ações são naturais e essenciais para estruturar nosso entendimento do mundo. O objetivo deste trabalho é elucidar o conceito, relevância, aplicação e tipos de taxonomia, além de apresentar sucintamente o conceito de Sistema de Organização do Conhecimento. No contexto da organização da informação e do conhecimento, as taxonomias são reconhecidas como sistemas fundamentais. Em ambientes digitais, elas desempenham um papel crucial ao satisfazerem necessidades recuperação da informação, contribuindo significativamente para representação eficaz e recuperação otimizada de dados.

PALAVRAS-CHAVE: Taxonomia; Sistemas de organização do conhecimento; Organização do conhecimento; Organização da informação.

ABSTRACT: Taxonomy is a knowledge organization system that systematically organizes information items in digital environments. This allows users of information systems to access and retrieve information quickly and efficiently. Organizing and classifying are intrinsic acts of human nature, holding profound significance as social activities. These actions are natural and essential for

structuring our understanding of the world. The aim of this work is to elucidate the concept, relevance, application, and types of taxonomy, as well as to briefly introduce the concept of Knowledge Organization System. In the context of information and knowledge organization, taxonomies are recognized as fundamental systems. In digital environments, they play a crucial role in meeting information needs, significantly contributing to effective representation and optimized retrieval of data.

KEYWORDS: Taxonomy; Knowledge organization systems; Knowledge organization; Information organization.

Introdução

O objetivo deste trabalho é esclarecer o conceito de taxonomia, sua relevância, aplicação e diferentes tipologias. Para isso, será necessário também apresentar sucintamente o conceito de Sistema de Organização do Conhecimento (SOC), já que uma taxonomia é um tipo de SOC e esse termo também pode causar estranheza.

Desta forma, quando se ouve o termo taxionomia pela primeira vez, pode parecer estranho. Isso ocorre porque o termo não é amplamente utilizado no cotidiano das pessoas. Desde o seu primeiro uso, oriundo da biologia até o seu uso na organização do conhecimento, as taxonomias têm assumido um papel de destaque em ambientes digitais. No contexto da organização da informação e do conhecimento, as taxonomias são reconhecidas como um SOC mais relevantes. Em ambientes digitais, elas desempenham um papel crucial ao atenderem às necessidades de informação, contribuindo significativamente para uma representação eficaz e uma recuperação otimizada de dados.

1.1. SOC?! O que é isso?

Organizar e, especialmente, classificar, são atos inerentes ao ser humano, com um significado profundo como atividades sociais. Não apenas isso, o ato de classificar é necessário e natural, como um modo de organizar, na verdade, poucas coisas nos parecem tão ‘naturais’, óbvias e indiscutíveis quanto as classificações dos seres, fatos e eventos que moldam nossos quadros mentais (Pombo, 1998).

O ato de classificar ocorre também no campo do conhecimento, sendo este o motivo da existência dos sistemas de organização do conhecimento. Contudo, por que chamá-los de sistemas de organização do conhecimento e não de sistemas de organização da informação? Para compreender isso, é necessário realizar um breve estudo e diferenciar os conceitos de informação e conhecimento.

1.1.1. A Informação e o Conhecimento: não é tudo a mesma coisa?

Na busca pela definição do termo 'informação' logo encontra dificuldades porque 'informação' se refere a se informar, reduzir a ignorância e a incerteza. Ironicamente, o termo 'informação' é ambíguo e utilizado de diversas maneiras. Desta forma, o termo 'informação' possui várias definições, portanto não há uma única definição universalmente aceita para ele (Buckland, 1991).

Para Cunha & Cavalcanti (2008) a informação poder ser definida como um registro de um conhecimento essencial para uma tomada de decisão. A palavra 'registro' abrange não apenas documentos impressos, mas também cópias reproduzidas e qualquer outro formato adequado para armazenamento e posterior uso. Informação, em seu sentido mais abrangente, é uma evidência que sustenta ou apoia um fato; registro de conhecimento para uso futuro; dados numéricos, alfabéticos ou alfanuméricos processados por computador.

McGarry (1999) considera informação um quase-sinônimo do termo 'fato'; um reforço do conhecimento existente; a liberdade de escolha ao selecionar uma mensagem; a matéria-prima da qual se extrai o conhecimento; algo que é intercambiado com o mundo exterior, não apenas recebido passivamente; definido pelos seus efeitos no receptor; algo que reduz a incerteza em uma situação específica. Costa (1995), considera que a informação vem da necessidade de transmitir coisas abstratas geradas e baseadas pelo intelecto.

Alguns autores usam o termo conhecimento para definir o que é informação. Como Iglesias-Rebollo (2012) que define informação com o conhecimento adquirido por um ser humano a partir dos dados que percebemos. Ela constitui a base fundamental da nossa relação com o mundo

exterior e com os outros. Contudo, para McGarry (1999) a informação e conhecimento não são sinônimos, entre informação e conhecimento existe uma diferença, que o conhecimento deve, de alguma forma, depender da informação; frequentemente, esses termos são intercambiáveis.

O conhecimento é uma característica essencialmente humana, podendo ser tanto individual quanto coletivo. O conhecimento individual é possuído ou acessível por um indivíduo, enquanto o conhecimento coletivo é aquele registrado e disponível publicamente. Em muitos casos, o conhecimento humano pertence simultaneamente aos níveis individual e coletivo, sendo o conhecimento coletivo tanto originador quanto resultado do conhecimento individual (Lopes, 1998).

O conhecimento vai além de documentos: está embutido em ferramentas e artefatos (cadeiras nos ensinam a sentar, martelos a martelar preciso), em ambientes projetados (auditórios ensinam a projetar a voz, cafés incentivam conversas), e em práticas cotidianas, rituais e hábitos. Não é uma entidade abstrata, isolada em mentes, mas entrelaçada no mundo social e físico de objetos, situações, costumes, expressões e comunicações (Lambe, 2007).

A informação constitui um elemento essencial do conhecimento. Em poucas palavras, é o conhecimento que pode ser desmaterializado e registrado em algum tipo de suporte, seja escrito, impresso, digital, oral ou audiovisual. O significado da informação é transmitido através de mensagens registradas em um suporte espacial-temporal, podendo ser impresso ou um sinal elétrico (Le Coadic, 1996; Lima & Alvares, 2012).

Observa-se que o conceito de conhecimento frequentemente é associado ao conceito de informação, sendo até considerados quase sinônimos. É difícil dissociar completamente o conhecimento da informação. O conhecimento pode ser visto como o processo mental de assimilar informações úteis, que resulta em diversas formas de pensamento. Dessa forma, pode-se entender o conhecimento como algo abstrato e pessoal, essencialmente subjetivo.

Para além, pode-se concluir que eles não são sinônimos e não possuem definições únicas. 'Conhecimento' é a informação que foi 'processada' pela mente humana, o que torna seu gerenciamento mais complexo.

No entanto, podemos entender 'informação' como um elemento constituinte do conhecimento. Na área da CI, o conhecimento é adquirido quando o sujeito se apropria da informação através de uma organização intencional que lhe confere significado. Nesse contexto, o conhecimento é um processo contínuo e interligado à informação, já que o conhecimento impulsiona a criação de nova informação, e esta, por sua vez, leva à formação de novos conhecimentos.

Após estudarmos os conceitos de informação e conhecimento, surge a questão crucial de como organizar tanto a informação quanto o conhecimento.

1.1.2. A Organização da Informação e a Organização do Conhecimento

Para Sales & Café (2010) a organização da informação (OI) é um processo de organização de coleções tradicionais ou eletrônicas, realizado através da descrição física e de conteúdo (temático) de seus objetos informacionais. No entanto, para alcançar esses objetivos, é crucial realizar a descrição física e de conteúdo dos objetos informacionais.

Brascher & Café (2008) entendem OI como um processo que abrange a descrição física e de conteúdo dos objetos informacionais. O resultado desse processo descritivo é a representação da informação, que consiste em um conjunto de elementos descritivos que capturam os atributos de um objeto informacional específico. Algumas formas de representação da informação são desenvolvidas por meio de linguagens elaboradas especialmente para os propósitos de organização da informação.

Quanto ao que se refere a organização do conhecimento (OC), Dahlberg (1993) o fundamento teórico da OC reside no fato de que OC deve ser baseada em unidades de conhecimento – que não são nada além de conceitos. Conceitos consistem em elementos de conceito, também chamados de características do conceito, e são exatamente esses fatores pelos quais sistemas de conceitos – e sistemas de classificação são tais sistemas de conceitos – podem ser construídos. O conhecimento, por si só, não pode ser compreendido ou representado a menos que seja apresentado por unidades de conhecimento e suas muitas combinações possíveis em palavras/termos ou declarações.

Enquanto para Hjørland (2008) a OC diz respeito a atividades como descrição de documentos, indexação e classificação realizadas em bibliotecas, bases de dados bibliográficas, arquivos e outros tipos de ‘instituições de memória’ por bibliotecários, arquivistas, especialistas em informação, especialistas em assuntos, além de algoritmos de computador e leigos. Assim, a organização do conhecimento é o processo que abrange a descrição minuciosa de documentos, incluindo seus conteúdos, características e objetivos, e a organização dessas descrições de modo que possam ser facilmente acessadas e recuperadas pelos usuários. Além disso, essa atividade é contínua e cíclica, pois o conhecimento transmitido constantemente se transforma em novos saberes.

A OC é a outra face da recuperação da informação. Assim, a OC tem como finalidade possibilitar uma melhor recuperação do conhecimento, um conhecimento bem organizado é facilmente recuperado. Sendo assim, objetivo principal da organização do conhecimento é a recuperação da informação (Rowley & Hartley, 2008).

Le Coadic (1996) contribui para esta discussão ao usar o termo ‘Organização do Conhecimento e da Informação’ de forma integrada, ao invés de separada – OI e OC. Segundo o autor, o foco deve estar na Organização do Conhecimento e da Informação, não na OC isoladamente nem na OI isoladamente. Na CI, este conceito abrange documentos em seu sentido mais amplo, incluindo o conhecimento registrado neles. Esse processo envolve tanto análises quantitativas quanto qualitativas dos documentos, através de:

- (a) Representação descritiva (catalogação), que detalha elementos paratextuais como autores, título, editores, local de publicação, data e idioma;
- (b) Representação temática, que inclui a análise do texto (análise documentária ou análise de assunto), envolvendo indexação, classificação e elaboração de resumos.

Contudo, não existe um consenso terminológico sobre o que seja a OC. Brascher & Café (2008) divergem de Hjørland (2008) sobre o que seria OC: enquanto para Hjørland (2008) a OC trata das atividades de descrição, indexação e classificação de documentos, isto é, trata do

registro da informação, para Brascher & Café (2008), a OC trata da construção de modelos. Deste modo, a OC está no domínio dos conceitos, e não do registro da informação, papel que é desenvolvido pela OI. Dessa forma, "organização do conhecimento [...] visa à construção de modelos de mundo que se constituem em abstrações da realidade" (Brascher & Café, 2008, p. 7).

O desenvolvimento da capacidade humana para armazenar e recuperar grandes volumes de informações tem gerado um interesse crescente em explorar esses dados para avançar o conhecimento humano. Este interesse tem sido objeto de discussão entre estudiosos de diversos campos, desde a filosofia até a cibernética, por um longo período. Eles têm se dedicado a entender o conceito de conhecimento e os desafios de representá-lo em sistemas de informação.

Esse debate é alimentado pelo crescimento exponencial da literatura técnica e científica, assim como pela necessidade crescente de rapidez e eficiência na recuperação de informações, o que tem transformado a gestão de acervos, tanto físicos quanto digitais.

Por sua vez, nas organizações, a crescente demanda por gerenciar conhecimento de forma estratégica e competitiva tem impulsionado a busca por sistemas mais avançados e eficazes de organização e gestão de informações (Aganette, Alvarenga, & Souza, 2010; Kwasnick, 1999). Nesse contexto, associado à OC estão os SOC.

1.1.3. Os Sistemas de Organização do Conhecimento – SOC

Barité et al. (2015) explicam que o SOC é um sistema de conceitos que têm como objetivo principal fornecer designações precisas para representar tematicamente o conteúdo de documentos, dados e outros recursos informacionais, independentemente do suporte ou estrutura em que se encontram. Utilizando símbolos codificados ou expressões linguísticas, esses sistemas facilitam a busca e recuperação temática de maneira eficaz, pertinente e relevante.

Os SOC são essenciais para estruturar o conhecimento dentro de um determinado domínio. Como ferramentas de representação do conheci-

mento, desempenham um papel fundamental na organização das informações contidas nos documentos, sendo indispensáveis nas atividades de tratamento da informação em bibliotecas, museus e arquivos (Hodge, 2000).

Para Carlan & Medeiros (2011) os SOC são ferramentas semânticas com vocabulários estruturados e formalizados, utilizados para o tratamento e a recuperação de informação, tanto em ambientes web quanto tradicionais, inclusive no contexto do desenvolvimento da web semântica. Dessa forma, os SOC possibilitam o mapeamento, a representação e a organização do conhecimento, facilitando a recuperação de informação.

Os SOC, como taxonomias, sistemas de classificação, tesouros e ontologias, desempenham um papel crucial na representação e estruturação do conhecimento. Na CI, essas ferramentas são extensivamente estudadas para facilitar a organização e recuperação de informações, entre outros propósitos essenciais (Lima & Maculan, 2017).

Nos últimos anos, observa-se um aumento significativo na produção de informação, acompanhado pela crescente demanda por métodos eficazes de organização do conhecimento e da informação. Com o avanço das novas tecnologias, a informação migra cada vez mais para o ambiente digital. Surgem assim novos espaços como sites, repositórios digitais, bancos de teses e dissertações, entre outros, criando a necessidade premente de desenvolver novos SOC estruturar e facilitar o acesso a essas ferramentas. Luz (2018, p. 25) corrobora e destaca que “A informação digital, como forma estruturante do conhecimento contemporâneo e para ser tratada por agentes computacionais, precisa ser organizada em suas principais características, tais como os temas (taxonomias) ou a semântica (ontologias)”.

Os tipos mais comuns de SOC incluem os sistemas de classificação e vocabulários controlados, como tesouros, listas e taxonomias. Contudo, dada essa variedade de tipos de SOC, alguns autores trabalharam na identificação de agrupamentos. Considera-se que Hodge (2000) foi o primeiro a oferecer uma concepção linear para a classificação dos diversos tipos de SOC, compostos por: (1) listas de termos: arquivo de autoridade, glossários, gazetteers e dicionários; (2) classificações e categorizações: lista de cabeçalhos de assunto, sistemas de classificação bibliográfica,

taxonomias e sistemas de classificação bibliográfica facetados; (3) listas de termos e relacionamentos: tesouros, redes semânticas e ontologias. Partindo dessa lista, Hjørland, (2007) incluiu os mapas bibliométricos, mapas conceituais, topic maps e folksonomias. Pieterse & Kourie (2014,) alegam que foi possível identificar cinco amplas classes de SOC: listas, taxonomias, lattices, tesouros e ontologias.

1.2. Taxio... o quê?: Taxonomia

A palavra taxonomia possui sua raiz etimológica do grego *táxis* = ordem, ordenação e *onoma* = lei, norma, regra, método e baseou-se em um dos ramos da Biologia que trata da classificação lógica e científica dos seres vivos.

Inicialmente, as taxonomias eram usadas na botânica e na zoologia para classificar espécies vivas através de uma terminologia binária. Com o passar dos séculos, o conceito de taxonomia evoluiu significativamente. Na atualidade, especialmente na CI, a taxonomia é empregada para organizar informações e conhecimentos de maneira estruturada. Em ambientes digitais, ela é fundamental para criar estruturas de domínio e rótulos de metadados, facilitando a organização sistemática de itens de informação. Assim, o uso das taxonomias está diretamente ligado às técnicas automatizadas de organização de informações, assegurando a precisão e estabelecendo relações hierárquicas de forma eficiente.

Originalmente, o termo 'taxonomia' estava associado à sua função de classificar as espécies em botânica e zoologia, utilizando uma terminologia binária. Augustin Pyrame de Candolle, biólogo suíço do início do século XIX, foi o primeiro a empregar o termo, enquanto Carolus Linnaeus, médico sueco, se destacou como o expoente mais relevante da taxonomia e da Biologia sistemática em geral. Linnaeus introduziu um esquema hierárquico de classificação, agrupando espécies muito semelhantes em um mesmo gênero, os gêneros em famílias, e assim sucessivamente em ordens, classes, filos e espécies. Esse sistema proporcionou uma representação estruturada das relações entre as espécies, comparável à árvore de Porfírio e à metafísica clássica (Barquín et al., 2006).

A taxonomia biológica permite a classificação em categorias como filo, classe, ordem, família, gênero, espécie e variedade. Os biólogos reconhecem que essa taxonomia é fundamental para assegurar a precisão na comunicação sobre sua ciência, além de proporcionar uma compreensão detalhada da organização e das interrelações entre as diferentes partes do mundo animal e vegetal (Bloom, 1956). Desta forma, a taxonomia “surgiu como Ciência das leis da classificação de formas vivas e, por extensão, ciência das leis da classificação. No ambiente dos sistemas de classificação, das ontologias, da inteligência artificial, é entendida como classificação de elementos de variada natureza.” A taxonomia “é, por definição, classificação, sistemática” (Campos & Gomes 2007, p. 3).

Simões (2008) explica que o surgimento das taxonomias atuais pode ser atribuído a vários motivos. Primeiramente, o aumento exponencial da informação em bases de dados revelou a inadequação dos instrumentos tradicionais de recuperação de informação. Diante dessa dificuldade, tornou-se essencial desenvolver novas ferramentas que funcionassem como filtros para facilitar a busca de informações relevantes. Além disso, houve uma baixa propensão dos usuários em pesquisar em bases de dados devido à complexidade na utilização desses sistemas, resultando na recuperação de um alto volume de informações irrelevantes. Outro desafio era a falta de compatibilidade entre as terminologias dos tesouros e as terminologias utilizadas nos documentos, causando um desalinhamento entre a indexação feita pelo tesouro e a linguagem usada na documentação institucional. Por fim, a proliferação de intranets conectando diversas bases de dados que utilizam múltiplos tesouros contribuiu para a construção de taxonomias que integram termos de diferentes origens, formando um ‘megatesouro’ onde os termos são identificados conforme o tesouro de origem correspondente.

Simões (2008, p. 71) explica que “De acordo com a própria definição de taxonomia, estas constituem um esquema ordenado do conhecimento em várias áreas, segundo determinadas regras ou normas.” A ISO 25964 (2013) alude que uma taxonomia é apresentada como um vocabulário hierárquico, utilizado para classificar, organizar, navegar, buscar e/ou filtrar qualquer tipo de conteúdo em ambientes de rede.

A norma NISO Z39.19 (2005), define uma taxonomia como uma coleção de termos de vocabulário controlado organizados em uma estrutura hierárquica. Cada termo em uma taxonomia está em uma relação de pais/filhos (mais amplo/mais específico) com outros termos na taxonomia.

O conceito de taxonomia se torna cada vez mais crucial à medida que o volume de informações aumenta exponencialmente. Os usuários desempenham um papel fundamental não apenas na produção, mas também na utilização e na categorização dessas informações. As taxonomias são amplamente empregadas em portais institucionais, repositórios, Web Semântica, ontologias, gestão da informação e gestão do conhecimento. Elas representam um recurso essencial de consulta, complementando as ferramentas de pesquisa tradicionais (Braga, 2018).

Para Sousa & Araújo Junior (2013), no âmbito dos arquivos, a taxonomia se apresenta como uma linguagem documentária viável e representativa pois possui um caráter hierárquico mais simples e flexível do que os tradicionais tesouros. Os autores ainda entendem que uma taxonomia pode ser definida como uma estrutura organizacional em camadas, formada por uma hierarquia simples que pode ser detalhada de acordo com as necessidades específicas de classificação e indexação de documentos de arquivo. Essa estrutura permite categorizar abrangentemente objetos, seres vivos e coleções de documentos em grupos ordenados hierarquicamente, facilitando sua identificação, localização e acesso. Quando associada à classificação de documentos, possui o objetivo de estruturar e sistematizar os assuntos correspondentes às atividades desempenhadas pelas áreas de trabalho de uma organização, com o intuito de permitir uma tomada rápida de decisão e um rápido acesso à informação e documentos.

Para Vogel (2004, p. 1) a taxonomia é “um sistema para classificar informações – seus assuntos, naturezas, tipos, etc. O que significa que ela é a lista controlada de palavras-chave que irá rotular a informação que a empresa produz.” Para além, a taxonomia corporativa funciona como ferramenta, produto para a organização e classificação da informação, e como um processo. Dessa forma, no âmbito corporativo, a taxonomia não é apenas um produto, um fim, mas é todo um processo desde sua construção, manutenção e uso, isto é, todo seu processo de construção

é importante. No processo de construção de uma taxonomia, em uma corporação, permite compreender mais sobre o contexto informacional da corporação.

1.2.1. Taxonomia e suas Tipologias

Segundo Campos & Gomes (2007, p. 4) as taxonomias podem ser compreendidas “como estruturas classificatórias representam os propósitos de organização intelectual de um dado contexto”, representadas na forma hierárquica, que podem ser: a) taxonomia canônica, classificação binária (dicotômica), de unidades sistemáticas (família, gênero, espécie); b) taxonomia de domínio; c) taxonomia de processos e tarefas gerenciais.

Ainda segundo Campos & Gomes (2007) as taxonomias de domínio e a taxonomia de processos e tarefas gerenciais caracterizam-se por serem policotômicas, isto é, baseado em um domínio ou tarefa várias divisões são possíveis. Devido a sua complexidade, necessitam primeiro delimitamento por categorias e dentro destas várias divisões e subdivisões são possíveis em cada passo de divisão.

Ademias, na literatura também são encontradas outras tipologias, além da taxonomia hierárquica: em árvore, poli-hierárquicas, em matrizes, sistemas de mapas e facetadas (Lambe, 2007; Maculan, 2014) e corporativas (Aganette; Alvarenga; Souza, 2010). Lambe (2007) considera que a taxonomia mais utilizada é a hierárquica. Contudo, podem ser criadas outros tipos de estruturas: como a poli-hierárquica ou a facetada. Sendo que a facetada usa os princípios usados na construção de sistemas de classificação facetados e de tesouros, baseados na Teoria da Classificação Facetada desenvolvida por Ranganathan (Lima & Maculan, 2017).

Novo (2010, p. 146) destaca que a taxonomia de domínio “se configura como uma classificação de conceitos ordenados hierarquicamente para a representação de um dado domínio de conhecimento.”, já Campos & Gomes (2007) ressaltam que, “as taxonomias de domínio [...] têm seu recorte determinado pelas características da organização a que servem”.

De acordo com Maculan (2014, p. 71), a taxonomia facetada é uma taxonomia hierárquica que agregou os princípios da teoria da classificação

facetada, que “foi estruturada em facetas, dimensões, aspectos ou atributos. É um conjunto de características que podem descrever um mesmo objeto, seja ele um produto, [...] uma imagem [...] ou documento.” Maculan, Lima, & Penido (2011, p. 243) explicam que a “estruturação provida por uma taxonomia facetada facilita o entendimento de um domínio, pois torna esse domínio inteligível ao usuário.”

Segundo Lambe (2007), uma taxonomia facetada representa uma classificação expressa do conteúdo de um domínio sob diferentes dimensões, que são representadas pelas facetas. Sobre isso, Morville & Rosenfeld (2006, pp. 221-222) estabelecem que “a diferença entre um sistema de classificação tradicional e uma classificação facetada é que esta última é baseada no princípio de múltiplas taxonomias”.

Kwasnick (1999) enumera algumas vantagens das taxonomias facetadas como: a) não necessitar de conhecimento completo e ao desenvolver um esquema facetado, não é preciso conhecer nem a extensão total das entidades a serem acomodadas pelo esquema, nem a extensão total das relações entre as facetas.

Outro tipo de taxonomia é navegacional. Para Aquino et al. (2009), a taxonomia navegacional busca organizar as informações de maneira flexível, permitindo sua visualização pelo usuário ao acessar o site, através da navegação. A estruturação das informações é claramente perceptível ao usuário final, garantindo que ele descubra as informações de forma intuitiva durante sua navegação. Aplicadas como mecanismo de navegação, as características das taxonomias possibilitam aos usuários filtrar os resultados da busca com base nos critérios escolhidos pelo usuário, uma vez que os elementos da taxonomia são associados entre si (Maculan, 2014).

Já quando é acoplada no portal de uma organização, a taxonomia é denominada de corporativa, que possibilita uma uniformidade no uso de termos por seus funcionários. Segundo Woods (2004) uma definição simples de taxonomia é que ela é uma hierarquia de categorias usada para classificar documentos e outras informações. Uma taxonomia corporativa é uma forma de representar as informações disponíveis dentro de uma empresa. Nas organizações comerciais modernas, o conhecimento está contido em uma variedade de elementos de natureza heterogênea em

termos de criação, armazenamento e reutilização. Sharma, Foo, & Morales-Arroyo (2008) explicam que capturar conhecimento é de pouca utilidade se não for armazenado de maneira que possa ser compreendido, indexado, acessado facilmente, cruzado com referências, pesquisado, vinculado e manipulado de forma geral para o benefício máximo de todos os membros de uma empresa.

A taxonomia corporativa é definida como um vocabulário controlado que reflete o contexto, a audiência e o conteúdo de uma organização específica. Ela permite a representação de todos os seus objetos informativos para facilitar diversas funções nos sites corporativos, como organização de conteúdos, busca, navegação, pesquisa competitiva, entre outras (Argudo & Centelles, 2005).

As taxonomias corporativas possuem várias finalidades, e Mahemei, Thulasi, & Rajashekar (2001) destacam que as taxonomias oferecem uma abordagem muito útil para atender às necessidades arquitetônicas dos portais de informações. Os autores relatam que as taxonomias corporativas são, normalmente, por facetas – por exemplo, comunidades, assuntos, locais, – e cada faceta é composta de um conjunto de termos, organizador por hierarquias. As facetas possibilitam, assim, que as taxonomias corporativas sejam multidimensionais. Os autores também afirmam que as taxonomias corporativas serão sempre personalizadas para refletir os processos de negócios, ativos de conteúdo, idioma, cultura e objetivos de uma determinada empresa. E acrescentam que as taxonomias e metadados trabalham juntos, já que os metadados descrevem um ativo, relacionado a um conjunto significativo de atributos, como, por exemplo, sujeitos e tipo de conteúdo. Assim, as taxonomias corporativas, associadas aos metadados, possibilitam aos funcionários de uma empresa a marcação autorizada de conteúdo, desde a sua criação até o envio ao portal. Assim, a navegação pela taxonomia corporativa permite que os funcionários entendam o escopo e o contexto das informações disponíveis.

Ressalta-se que a taxonomia corporativa possui algumas vantagens, como a uniformidade do uso de termos utilizados na empresa por seus funcionários; a economia no tempo de classificação das informações, tanto em formato físico quanto virtual; celeridade no momento da localiza-

ção, seja por navegação ou busca; a empresa pode se preparar para organizar seu futuro arquivo histórico e saber mais facilmente o que precisa ficar no arquivo morto (Vogel, 2004).

É importante destacar que não há uma taxonomia correta ou incorreta, apenas uma taxonomia organizada a partir de um determinado ponto de vista, que serve como uma forma de classificação para compreender uma realidade específica conforme diferentes propósitos. Portanto, as taxonomias não são neutras, mas sim criadas com base em características que atendem a objetivos específicos. Elas são desenvolvidas com um propósito claro e uma finalidade determinada (Aquino et al., 2009, p. 206).

1.2.1. Taxonomias e Relações com outros SOC

As semelhanças entre taxonomias e tesouros residem no fato de que ambas compartilham uma utilidade em comum: servem como métodos de classificação. Além disso, são fundamentais para sistematizar o conhecimento de maneira lógica e coerente, seguindo normas preestabelecidas. Tanto os tesouros quanto as taxonomias funcionam como sistemas que podem ser pré-coordenados e pós-coordenados nas linguagens documentárias. Essas ferramentas desempenham um papel crucial no avanço da ciência, pesquisa e inovação humana, facilitando seu desenvolvimento contínuo (Currás, 2010).

Para além, dado ao fato das taxonomias empregam apenas as propriedades *broader term* – BT / *narrower term* – NT para construir uma hierarquia e não utilizam mais nenhuma propriedade além desta, portanto, pode-se considerar que cada tesouro possui uma taxonomia, isto é, a relação hierárquica (gênero-espécie; todo-parte) forma uma taxonomia que a estrutura o tesouro.

Entre tesouros e taxonomias, destacam-se diferenças significativas. Os tesouros podem ser estruturados manual ou mecanicamente, ao passo que as taxonomias são exclusivamente estruturadas por meio de tecnologias de informação. Além disso, as metodologias de ordenação variam consideravelmente: as taxonomias são predominantemente utilizadas por profissionais de informática em empresas, enquanto os

tesauros são ferramentas essenciais para profissionais da informação (Currás, 2010).

Hedden (2017) destaca duas diferenças entre tesauros e taxonomias que afetam os termos órfãos. A primeira diz respeito às relações hierárquicas: as taxonomias não incluem relações associativas. A segunda refere-se à estrutura das taxonomias. Enquanto os tesauros utilizam relacionamentos hierárquicos entre termos, as taxonomias limitam ou singularizam as hierarquias dos termos.

Isto posto, uma taxionomia pode se diferenciar de um tesouro por ter, normalmente, menos níveis hierárquicos e uma estrutura menos complicada. Geralmente, a taxonomia não inclui termos equivalentes (sinônimos ou termos variantes) ou termos relacionados (relações associativas) (Harpring, 2016).

As ontologias, tesauros e taxonomias possuem diferenças e semelhanças. A hipótese de restringir a linguagem natural é ponto comum entre as três (Gilchrist, 2003). Assim como os tesauros, as ontologias são modelos de representação do conhecimento que se fundamentam em controles terminológicos de domínios específicos. No entanto, apresentam divergências distintas. Ao contrário de um tesouro, uma ontologia não se limita a três tipos de relacionamento, termo mais amplo (Termo geral), termo mais restrito (Termo específico) e termo relacionado. Em uma ontologia a relação entre dois conceitos pode ser de qualquer tipo, como "é uma parte de", "é uma instância de", "é um tipo de", "é um produto de" e assim por diante". São assim mais aperfeiçoas que um tesouro ou mesmo uma taxonomia de estrutura de árvore, pois podem especificar a natureza exata da relação entre os dois conceitos descritos. As ontologias são criadas mais para máquinas do que para humanos, sendo mais complexas que taxonomias ou tesauros (Lambe, 2007; Sales & Café, 2009).

Outra similaridade entre as taxonomias e as ontologias vem do fato de todas as ontologias serem centradas em uma taxonomia, sendo a taxonomia uma espinha dorsal da ontologia. Guarino, (1998) explica que todas as ontologias são centradas em uma taxonomia, baseada em uma relação de ordenação parcial denominada de várias maneiras, como ISA, subsunção, hiperonímia/hiponímia. Essa taxonomia é o esqueleto principal

da ontologia, que pode ser enriquecida com a adição de atributos e outras relações entre os nós (como meronímia ou antonímia).

Conclusões

Para entender a construção dos SOC, foi necessário explorar os conceitos de conhecimento, informação, OC e OI. Ao longo deste trabalho, observou-se que, apesar das mudanças temporais, a necessidade de organizar e facilitar o acesso à informação e conhecimento permanece constante. A produção de informação aumentou significativamente, acompanhada pela busca contínua por formas de estruturar o conhecimento e a informação

Foi estudada a relação entre conhecimento e informação, concluindo que são conceitos distintos: a informação é a matéria-prima contextualizada e com propósito, enquanto o conhecimento é especializado e derivado da organização intencional de informações, conferindo-lhes significado.

Assim, esse trabalho explorou a OI e OC, com foco nos SOC, especialmente nas taxonomias. Outro ponto considerado é que não existe um consenso sobre as definições de informação e de conhecimento assim como uma concordância do que é OI e OC.

Observou-se que o SOC é um sistema conceitual que visa fornecer designações precisas para representar o conteúdo temático de documentos e recursos informacionais, independentemente de seu formato ou estrutura. Utilizando símbolos codificados ou expressões linguísticas, esses sistemas facilitam a busca e recuperação eficaz de informações relevantes. São fundamentais para estruturar o conhecimento em um domínio específico e desempenham um papel crucial na organização da informação em bibliotecas, museus e arquivos.

Foi observado que o conceito de taxonomia está passando por transformações significativas e está atualmente intimamente ligado ao contexto digital. Trata-se de um mecanismo composto por rótulos e metadados que permitem a organização sistemática dos recursos de informação disponíveis em bibliotecas e repositórios, utilizando relações hierárquicas. Esse sistema também desempenha um papel crucial no controle de sinonímia e ambiguidade dentro da área estudada.

As taxonomias são organizadas segundo perspectivas específicas para classificar realidades conforme diferentes propósitos. Não são neutras, sendo criadas com características que servem a objetivos específicos e desenvolvidas com um propósito e finalidade definidos.

Apesar de seu nome peculiar, a taxonomia desempenha um papel fundamental na organização e recuperação do conhecimento como um vocabulário hierárquico. Ela é utilizada para classificar, organizar, navegar, buscar e/ou filtrar todos os tipos de conteúdo em ambientes de rede. Imagens e tabelas devem ser precedidas e seguidas de uma linha em branco. O espaçamento entrelinhas entre os seus diferentes elementos (título, corpo, legenda) é de um espaço simples.

Referências bibliográficas

- Aganette, E., Alvarenga, L., & Souza, R. R. (2010). Elementos constitutivos do conceito de taxonomia. *Informação Sociedade*, 20(3), 77-93. Retrieved from <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3994>.
- Aquino, I. J., Carlan, E., & Brascher, M. B. (2009). Princípios classificatórios para a construção de taxonomias. *PontodeAcesso*, 3(3), 196-215.
- Argudo, S., & Centelles, M. (2005). Metodología para el diseño de taxonomías corporativas. *Investigación Bibliotecológica: archivonomía, bibliotecología e información*, 19(39), 158. <https://doi.org/10.22201/iibi.0187358xp.2005.39.4082>
- Barité, M., Colombo, S., Blanco, A. D., Simón, L., Castromán, G. C., Odella, M. L., & Vergara, M. (2015). *Diccionario de Organización del Conocimiento : Clasificación, Indización, Terminología* (6.a ed. co). Montevideo: csic.
- Barquín, B. A. R., Gonzáles, J. A. M., & Pinto, A. L. (2006). Construção de uma ontologia para sistemas de informação empresarial para a área de Telecomunicações. *DataGramaZero – Revista de Ciência Da Informação*, 7(2), 1-24. Retrieved from http://www.dgz.org.br/abr06/Art_04.htm
- Braga, F. d. R. (2018). Extração semiautomática de taxonomia para domínios especializados usando técnicas de mineração de textos. *Ciência da Informação*, 45(3). <https://doi.org/10.18225/ci.inf.v45i3.4056>.
- Bloom, B. S. (1956). *Taxonomy of education objectives: the classification of educational goals*. New York: David McKay Co Inc. Retrieved from [https://www.uky.edu/~rsand1/china2018/texts/Bloom et al -Taxonomy of Educational Objectives.pdf](https://www.uky.edu/~rsand1/china2018/texts/Bloom%20et%20al%20-%20Taxonomy%20of%20Educational%20Objectives.pdf)
- Brascher, M., & Café, L. (2008). *Organização da Informação ou Organização do Conhecimento?* (pp. 1-14).
- Buckland, M. (1991). *Information As Thing*, 5(June), 351-360.
- Campos, M. L. A., & Gomes, H. E. (2007). Taxonomia e classificação: a categorização como princípio. In *Encontro Nacional de Pesquisa em Ciência da Informação*. Salvador.
- Carlan, E., & Medeiros, M. B. B. (2011). Sistemas de Organização do Conhecimento na visão da Ciência da Informação. *Revista Ibero-Americana de Ciência Da Informação*, 4(2), 53-73.

- Costa, C. I. (1995). La teoría de la información en bibliotecología. *Informatio*, 1, 13-17.
- Cunha, M. B. da, & Cavalcanti, C. R. de O. (2008). *Dicionário de biblioteconomia e arquivologia*. Brasília: Briquet de Lemos/Livros.
- Currás, E. (2010). *Onologia, taxonomia e tesouros em teoria de sistemas e sistemática*. (Thesaurus, Ed.). Brasília.
- Dahlberg, I. (1993). Knowledge organization: its scope and possibilities. *Knowledge Organization*, 20(4), 211-222.
- Guarino, N. (1998). Some Ontological Principles for Designing Upper Level Lexical Resources. In *First International Conference on Language Resources and Evaluation*. Granada. Retrieved from <http://arxiv.org/abs/cmp-lg/9809002>
- Gilchrist, A. (2003). Thesauri, taxonomies and ontologies – an etymological note. *Journal of Documentation*, 59(1), 7-18. <https://doi.org/10.1108/00220410310457984>
- Harpring, P. (2016). *Introdução aos Vocabulários Controlados: terminologia para arte, arquitetura e outras obras culturais*. (Martha Baca, Ed.). São Paulo: ACAM Portinari. Retrieved from https://issuu.com/sisem-sp/docs/vocabul_rios_controlados_-_digital
- Hedden, H. (2017). *Orphan Terms in a Taxonomy*. Retrieved September 11, 2019, from <http://accidental-taxonomist.blogspot.com/2017/01/orphan-terms-in-taxonomy.html>
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 353(352), 86-101. <https://doi.org/10.1002/asi.23250>
- Hodge, G. (2000). Systems of knowledge organization for digital libraries. *Systems Engineering* (Vol. 11). https://doi.org/10.1207/S15327000EM0204_10.
- Hjørland, B., & Capurro, R. (2007). O Conceito da Informação. *Perspectivas Em Ciência da Informação*, 12(1), 148-207. Retrieved from <http://bogliolo.eci.ufmg.br/downloads/CAPURRO.pdf>
- International Organization for Standardization. (2005). ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. *Management* (Vol. 2003).
- International Organization for Standardization. (2013). *Thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies ISO 25964-2*. Geneva: International Standard Organization.
- Iglesias-Rebollo, C. (2012). Derecho e información. *El Profesional de La Información*, 21(5), 449-452. <https://doi.org/10.3145/epi.2012.sep.01>.
- Kwasnick, B. H. (1999). The role of classification in knowledge representation and discovery. *Library Trends*, 48(1), 22-47.
- Lima, G. Â. d., & Maculan, B. C. M. d. S. (2017). Estudo comparativo das estruturas semânticas em diferentes sistemas de organização do conhecimento. *Ci.Inf.*, 46(1), 60-72. <https://revista.ibict.br/ciinf/article/view/4014/3714>
- Lambe, P. (2007). *Organising Knowledge: taxonomies, knowledge and organizational effectiveness*. Oxford: Chandos.
- Le Coadic, Y.-F. (1996). *A Ciência da Informação*. Brasília: Briquet de Lemos/Livros.
- Lima, J. L. O., & Alvares, L. (2012). Organização e representação da informação e do conhecimento. In *Organização da Informação e do Conhecimento: conceitos, subsídios interdisciplinares e aplicações* (pp. 21-34). São Paulo: B4Editores.
- Lopes, M. I. (1998). As bibliotecas e a organização do conhecimento: evolução e perspectivas. *Leituras: Rev. Bibl. Nac.*, 3(2), 141-157.
- Luz, C. (2018). *Ontologia Digital Arquivística*. São Paulo: Bookess Editora.
- Maculan, B. C. M. dos S. (2014). *Taxonomia facetada e navegacional: um mecanismo de recuperação*. Curitiba: Appris.

- Maculan, B. C. M. S., Lima, G. A. B. O., & Penido, P. (2011). Taxonomia facetada como interface para facilitar o acesso à informação em bibliotecas digitais faceted taxonomy as interface to information's access facility in digital library. *Revista acb: biblioteconomia em santa catarina*, 16(1).
- Mahemei, L. K., Thulasi, K., & Rajashekar, T. B. (2001). Approaches to Taxonomy Development: Some Experiences in the Context of an Academic Institute Information Portal. *Indian Institute of Science. 2001. Vickery*, 1-18.
- McGarry, K. (1999). *O contexto dinâmico da informação*. Brasília: Briquet de Lemos/Livros.
- Morville, P., & Rosenfeld, L. (2006). *Information Architecture for the World Wide Web: Designing Large-Scale Web Sites*. O'Reilly Media, Incorporated.
- Novo, H. F. (2010). A taxonomia enquanto estrutura classificatória: uma aplicação em domínio de conhecimento interdisciplinar. *PontodeAcesso*, 4(2), 131-156.
- Pieterse, V., & Kourie, D. G. (2014). Lists, Taxonomies, Lattices, Thesauri and Ontologies: Paving a Pathway Through a Terminological Jungle. *Knowledge Organization*, 41(3), 217-229. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=101115523&site=ehost-live>.
- Pombo, O. (1998). Da classificação dos seres à classificação dos saberes. *Leituras. Revista Da Biblioteca Nacional de Lisboa*, 19-33.
- Rowley, J., & Hartley, R. (2008). *Organizing knowledge: An Introduction to Managing Access to Information* (4th ed.). London; New York: Routledge.
- Sales, R. de, & Café, L. M. A. (2010). Organização da informação: Conceitos básicos e breve fundamentação teórica. In *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento* (pp. 116-129). Brasília.
- Sharma, R. S., Foo, S., & Morales-Arroyo, M. (2008). Developing Corporate Taxonomies for Knowledge Auditability: A Framework for Good Practices. *Knowledge organization*, 35(1), 30-46. <https://doi.org/10.5771/0943-7444-2008-1-30>.
- Simões, M. da G. (2008). *Da abstração à complexidade formal: relações conceptuais num tesouro*. Coimbra: Almedina.
- Sousa, R. T. B. d., & Araújo Júnior, R. H. d. (2015). A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística. *Ciência da Informação*, 42(1). <https://doi.org/10.18225/ci.inf.v42i1.1400>.
- Vogel, M. J. M. (2004). *Taxonomia: produto ou processo?* São Paulo.
- Woods, E. (2004). The corporate taxonomy: creating a new order – *KMWorld Magazine*. Retrieved February 10, 2018, from <http://www.kmworld.com/Articles/Editorial/Features/The-corporate-taxonomy-creating-a-new-order-9566.aspx>

Uma breve análise bibliométrica ao legado científico de Maria da Graça Simões (1962-2019)

Silvana Roque de Oliveira

ORCID 0000-0001-6100-1777

Escola Nacional de Saúde Pública, ENSP, Universidade NOVA de Lisboa, Lisboa, Portugal;
CHAM, Faculdade de Ciências Sociais e Humanas, FCSH, Universidade NOVA de Lisboa
e CEIS20, Universidade de Coimbra, Portugal
silvana.oliveira@ensp.unl.pt

RESUMO: Esta pesquisa ensaia uma primeira análise bibliométrica ao perfil produtivo de Maria da Graça Simões, professora e investigadora da Universidade de Coimbra na área da Ciência da Informação. Para reconstituir o conjunto das suas publicações científicas foram recolhidos dados a partir do seu currículo na plataforma Ciência Vitae, completados por pesquisas feitas no repositório institucional da Universidade de Coimbra, no Google Scholar, na Scopus, na Web of Science, na Scielo e na Dimensions, identificando-se 73 documentos entre comunicações, artigos, livros, capítulos de livros, edições científicas, entre outras tipologias.

Procurando-se seguir as recomendações para um uso responsável dos indicadores bibliométricos aplicados ao estudo de um perfil individual, optou-se por realizar uma análise da predominância das diferentes tipologias documentais ao longo da carreira, uma caracterização das fontes de informação escolhidas para a disseminação do conhecimento e uma análise da colaboração.

PALAVRAS-CHAVE: Maria da Graça Simões; Análise bibliométrica; Perfil individual; Ciência da Informação.

ABSTRACT: This research attempts a first bibliometric analysis of the productive profile of Maria da Graça Simões, a professor and researcher of Information Science at the University of Coimbra. To reconstitute a comprehensive list of her scientific publications, data was collected from her CV on the *Ciência*

Vitae platform, supplemented by research carried out in the institutional repository of the University of Coimbra, Google Scholar, Scopus, Web of Science, Scielo, and Dimensions, identifying 73 documents including communications, articles, books, book chapters, scientific editions, among other document types.

Seeking to follow the recommendations for a responsible use of bibliometric indicators applied to the study of an individual profile, it was decided to carry out an analysis of the predominance of different documentary typologies throughout the career, a characterization of the sources of information chosen for knowledge dissemination and an analysis of collaboration.

KEYWORDS: Maria da Graça Simões; Bibliometric analysis; Individual profile; Information Science.

Introdução

Maria da Graça Simões foi professora da Universidade de Coimbra na área da Ciência da Informação desde 1995, primeiro como docente convidada, a par de uma carreira de técnica superior BAD (Biblioteconomia, Arquivística e Documentação) iniciada em 1990, desenvolvida também na Universidade de Coimbra, no Departamento de Engenharia Civil, na Biblioteca Geral e no Departamento de Arquitetura. A partir de 2014, ingressou de forma definitiva na carreira docente como Professora Auxiliar da Secção de Informação do Departamento de Filosofia, Comunicação e Informação da Faculdade de Letras, com dedicação exclusiva, onde exerceu vários cargos de direção pedagógica e científica relativamente aos cursos aí ministrados nos três ciclos do Ensino Superior, tendo ainda sido professora convidada em universidades do Brasil (Universidade Federal de S. Carlos (UFSCar), Universidade Estadual Paulista (UNESP) e Universidade Federal de Minas Gerais (UFMG)) e Espanha (Universidad de León). De 2013 a 2014, foi investigadora do Centro de Estudos de Tecnologia e Ciência da Comunicação (CETAC.MEDIA) da Universidade do Porto e, a partir de 2015, passou a integrar o CEIS20 – Centro de Estudos Interdisciplinares da Universidade de Coimbra.

Com formação inicial em História (1988), terminou, entre outras formações, o Curso de Especialização em Ciências Documentais, na opção de

Arquivo, em 1990, e na opção de Biblioteca e Documentação, um ano mais tarde, tendo concluído o doutoramento em Dezembro de 2010, na Universidade de Salamanca, com uma tese intitulada *A representação e evolução de etnia na Classificação Decimal Universal*, que lhe valeu o Prémio Extraordinário de Doutoramento do Curso 2010/2011 atribuído pela *Comisión de Doctorado y Posgrado* da mesma universidade (Simões, 2019a, 2019b).

Passados cinco anos sobre o seu precoce falecimento, ainda nos custa muito compreender a sua ausência, as palavras continuam a ser insuficientes e temos como certo que não há números que a descrevam, mas acreditamos que o ensaio de levantamento metódico e quantitativo do seu legado científico levado a cabo para apresentar uma primeira análise bibliométrica do seu perfil produtivo, como aqui se propõe, pode ser mais uma forma, mesmo que forçosamente singela, para recordar ou dar a conhecer uma figura ímpar da Ciência da Informação portuguesa.

Se a abordagem quantitativa constitui uma boa opção metodológica para fazer prevalecer a objetividade sobre a emoção, a verdade é que o caminho não passa, mesmo assim, a ser mais fácil. No seio da comunidade bibliométrica, são assertivos os alertas para os problemas que se levantam ao tratamento de perfis individuais (Wouters et al., 2013), independentemente da maior ou menor proximidade com a figura a ser trabalhada. De alguma forma, conhecer-se o percurso da autora surge mesmo como uma vantagem, na medida em que não se devem nunca separar os resultados obtidos do seu contexto produtivo.

Na maior parte dos casos, os indicadores bibliométricos conseguem recuperar documentos científicos com qualidade, mas nem todos os documentos com qualidade são recuperados pelos indicadores bibliométricos (Karolinska Institute. Research Strategy Committee, 2013). Esta é a maior fragilidade a maior fragilidade deste tipo de análises, pelos vieses geopolíticos, linguísticos e disciplinares a que estão sujeitas as suas fontes de informação mais comuns – as bases de dados internacionais disponíveis – vieses esses sentidos com particular acuidade nas Ciências Sociais e Humanas (Hicks, 2012; Sivertsen, 2014), onde a publicação se faz muito para além dos artigos científicos publicados em revistas difundidas exclusivamente em língua inglesa, a tipologia privilegiada

naquelas bases, e na abordagem aos casos individuais, que demandam uma muito maior atenção às condições específicas de cada percurso, além de um trabalho manual acrescido na recolha dos dados, para se alcançar um retrato mais fidedigno.

Antes mesmo do *Manifesto de Leiden* (Hicks et al., 2015), e de todo o movimento em prol das métricas responsáveis, do qual o relatório inglês *The Metric Tide* (Wilsdon et al., 2015) constituiu um momento charneira, já Cassidy Sugimoto e Blaise Cronin (2012) tinham proposto um modelo de perfis bibliométricos assente na diversidade, simplicidade e complementaridade de indicadores. De entre as suas opções, prescindiram, por exemplo, do recurso ao Índice-H – o indicador composto de produção e impacto, tão famoso quanto criticado (Bornmann & Daniel, 2007; Waltman & Van Eck, 2012) – recusando-se a caracterizar com um só número cada um dos seis autores de referência da Ciência da Informação internacional que selecionaram para ilustrar o seu intento. Por maioria de razão, e na senda do acordo difundido pela *Coalition for Advancing Research Assessment* (CoARA) (2022), que tem vindo a ganhar uma notável aceitação no contexto científico europeu, impõem-se escolhas no mesmo sentido para este exercício bibliométrico em torno do legado de Maria da Graça Simões.

Fontes de informação e métodos

Para se reconstituir da forma mais completa possível o conjunto das publicações científicas de Maria da Graça Simões, começou por se tomar como ponto de partida o seu currículo na plataforma Ciência Vitae, disponível em acesso aberto (Simões, 2019a), tendo-se selecionado e registado numa folha Excel todas as publicações de âmbito científico, sem exceção para nenhuma tipologia, excluindo-se apenas uma publicação de carácter mais técnico dos primeiros anos da sua atividade profissional enquanto bibliotecária (um catálogo de um Fundo Antigo) e a tese de doutoramento, tendo-se registado 62 documentos.

De seguida, com o duplo objetivo de se completar a pesquisa com potenciais publicações póstumas ou outras que, por lapso, não tivessem sido identificadas pela própria no seu currículo, e recolher dados relativos

à indexação das suas publicações, foram realizadas consultas no repositório institucional da Universidade de Coimbra (UC), nas bases de dados bibliométricas Scopus, Web of Science, Scielo e Dimensions e no universo da Google Scholar, o que resultou num acréscimo de mais 11 publicações, num conjunto de 73 documentos (Tabela 1) composto por: comunicações, artigos, livros, capítulos de livros, bem como outras tipologias menos comuns em análises bibliométricas, mas muito reveladoras da proeminência de um perfil docente e de investigação, como sejam as edições científicas, além de trabalhos de menor dimensão, classificados em “Outros” (um *poster*, um prefácio e uma entrevista).

Tabela 1 – Distribuição dos documentos selecionados nas fontes de informação por tipologias documentais (N=73)

Documentos selecionados	Comunicações	Artigos	Livros	Capítulos	Edições científicas	Outros	Total
Ciência Vitae	29	19	6	4	1	3	62
Google Scholar	19	17	4	2	1	2	44
Repositório UC	13	18	1	0	0	1	33
Dimensions	4	11	0	1	1	1	18
Scopus	3	11	0	0	0	0	14
Web of Science	1	7	0	0	0	0	8
Scielo	0	1	0	0	0	0	1
Seleção final	33	24	6	4	3	3	73

Relativamente às demais fontes de informação, os valores finais apresentados na Tabela 1 são o resultado das seguintes estratégias de pesquisa, que revelaram ser as que garantiam a maior cobertura possível, depois de algumas consultas exploratórias, tendo sido todas fechadas a 15 de julho de 2024:

- a) Google Scholar – recolha de dados com recurso ao *software* livre *Publish or Perish* (Harzing, 2007), através da pesquisa pelo nome da autora nas suas duas variantes: “Maria da Graça Simões” e “Maria da Graça de Melo Simões”, sem patentes, por serem irrelevantes. Procedeu-se à eliminação de publicações de uma homónima e à

- fusão dos duplicados com soma das citações de cada versão, tendo-se alcançado 44 documentos;
- b) Dimensions – pesquisa em *Researcher* por “Maria da Graça Simões”, tendo-se recolhido sucessivamente cinco resultados distintos nos três perfis sugeridos – “Maria Da Graça Simões – University of Coimbra”, “Maria Da Graça De Melo Simões – University of Coimbra” e “Maria Da Graça De Melo Simões” – num total de 15 documentos.
- c) Scopus – pesquisou-se por *Author last name* “Simões”, *Author first name* “Maria”, *Affiliation* “Universidade de Coimbra”, selecionando-se três dos nove perfis: AU-ID (“da Graça Simões, Maria” 57192928929) OR AU-ID (“de Melo Simões, Maria Da Graça” 57199717551) OR AU-ID (“Simões, Maria Da Graça De Melo” 57194830538), de onde resultaram 14 documentos;
- d) Web of Science (WoS) – pesquisa realizada na *Web of Science Core Collection*, em *Researchers*, por *Author Name* “Simoes, Maria da Graca”, tendo-se optado por combinar os dois perfis sugeridos – o de “Maria da Graça Simoes” e o de “Maria da Conceição Ribeiro Santos”, por este segundo, aparentemente errado, assumir também como variantes do nome os nomes por que assinou a nossa autora, tendo resultado num total de nove documentos;
- e) Scielo Citation Index (SciELOCI) – no SciELOCI não foi possível recuperar resultados nas pesquisas exploratórias, pelo que se recorreu à pesquisa avançada do portal da Scielo (<https://scielo.org/>), pesquisando-se por Autor “Maria da Graça Simões”; de seguida, identificado um único documento, voltou-se ao ScieloCI para se fazer a recuperação do registo pelo seu DOI;

Análise e discussão do perfil produtivo

Estabelecidos os procedimentos para a identificação e delimitação do *corpus* documental, passa-se agora a apresentar uma breve análise bibliométrica a partir dos 73 documentos e dos respetivos dados recuperados, de forma complementar, nas sete fontes de informação utilizadas. Conforme já delineado, procurou-se ir ao encontro das recomendações

cada vez mais unânimes sobre o uso responsável das métricas na análise de perfis individuais, pelo que se traçou um perfil produtivo através da análise da predominância das diversas tipologias documentais ao longo da carreira, da caracterização das fontes de informação escolhidas para a disseminação do conhecimento e, finalmente, uma ponderação dos tipos de colaboração que Maria da Graça Simões foi tecendo na sua rede de coautorias, ao longo de 28 anos de publicações estritamente científicas, entre 1994 e 2021, incluindo-se naturalmente as publicações póstumas.

Para dar contexto à nossa análise, importa ter presente que estamos perante uma autora que começou a publicar ainda como profissional BAD e vários anos antes de se doutorar e ingressar oficialmente na carreira da investigação. Por essa razão, na Tabela 2, onde se pode observar a evolução no número de documentos na sua distribuição por tipologias documentais, optou-se por sublinhar a diferença entre duas etapas fundamentais da sua carreira, que se convencionou distinguir com a conclusão do doutoramento no final de 2010.

É notória e compreensível a diferença no volume de produção anterior e posterior ao doutoramento, passando de 10 para 63 documentos (86%), respetivamente. A predominância das comunicações a congressos antes do doutoramento também se enquadra no que costuma ser um perfil mais profissional que Maria da Graça Simões tinha à época (Glänzel et al., 2006). No entanto, como acontece em quase todas as análises individuais, há sempre desvios ao mais expectável, sendo a presença de três livros publicados antes de 2010 um sinal inequívoco de que se estava já a desenhar uma carreira distinta, muito marcada pelas preocupações próprias de uma docente-investigadora em difundir sínteses bem refletidas e alargadas da sua matéria de especialização (Engels et al., 2018; Sugimoto & Cronin, 2012).

Realizado o doutoramento, e mau grado estarmos perante uma etapa malogradamente menor em número de anos, é realmente muito expressivo o aumento da produção, com a publicação de mais um livro logo em 2011, a partir dos capítulos mais teóricos da sua tese, seguido de outros dois, em 2015 e 2018, sendo a opção pela publicação de monografias uma das características singulares de Maria da Graça Simões ao longo

Tabela 2 – Evolução das publicações de Maria da Graça Simões
por tipologias documentais (N=73)

	Anos	Comuni- cações	Artigos	Livros	Capítulos	Edições científicas	Outros	Subtotal
Publicações pré-doutoramento	1994	1	0	0	0	0	0	1
	1995	0	0	0	0	0	0	0
	1996	1	0	0	0	0	0	1
	1997	1	0	0	0	0	0	1
	1998	0	0	0	0	0	0	0
	1999	0	0	0	0	0	0	0
	2000	1	0	0	0	0	0	1
	2001	0	2	0	0	0	0	2
	2002	0	0	1	0	0	0	1
	2003	0	0	0	0	0	0	0
	2004	0	0	0	0	0	0	0
	2005	0	0	0	0	0	0	0
	2006	0	0	0	0	0	0	0
	2007	0	0	1	0	0	0	1
	2008	0	0	1	0	0	0	1
	2009	1	0	0	0	0	0	1
	2010	0	0	0	0	0	0	0
Subtotal	5	2	3	0	0	0	10	
Publicações pós-doutoramento	2011	1	0	1	0	0	0	2
	2012	0	0	0	0	0	0	0
	2013	4	2	0	0	1	0	7
	2014	2	2	0	0	0	2	6
	2015	5	2	1	1	0	0	9
	2016	1	3	0	0	0	1	5
	2017	9	5	0	1	1	0	16
	2018	4	2	1	1	0	0	8
	2019	0	4	0	0	0	0	4
	2020	2	1	0	1	1	0	5
	2021	0	1	0	0	0	0	1
Subtotal	28	22	3	4	3	3	63	
Total	33	24	6	4	3	3	73	

de toda a carreira. Ao mesmo tempo, observa-se o crescimento muito marcado da publicação de artigos científicos (22), a par do fortalecimento da aposta nas comunicações (28). A confirmar a sua maturidade científica no período pós-doutoramento, sublinhe-se a presença de três edições científicas, uma de um livro com várias colaborações, uma de umas atas e uma outra de um número especial de uma revista.

Tratando-se de uma destacada especialista na área da Representação da Informação e Organização do Conhecimento, não surpreendeu ter-se apurado que a grande maioria das suas publicações (74%) tenha versado sobre essas temáticas, desde a Classificação Decimal Universal à *Web Semântica*, passando pelas questões correlatas da interoperabilidade ou da eficiência dos sistemas de pesquisa e recuperação da informação. Não obstante, o seu olhar de investigadora e docente da Ciência da Informação fê-la também refletir a partir do seu *locus* e publicar sobre questões epistemológicas relativas à própria disciplina, os cursos ministrados, as motivações dos alunos, as questões éticas da investigação, assim como também escreveu sobre as bibliotecas por onde passou – a Biblioteca do Departamento de Arquitetura ou a Biblioteca Geral, ambas da Universidade de Coimbra.

As fontes de disseminação do conhecimento

Para conhecer com mais pormenor as estratégias de publicação da autora, passa-se a identificar as fontes dos seus artigos e comunicações, a fim de avaliar a variância de revistas e congressos que foram sendo selecionados para a difusão dos seus resultados científicos.

Na Tabela 3, constata-se a grande variação na escolha das revistas, sendo que os 25 documentos, onde se incluem, para além dos artigos, uma entrevista e uma edição de um número especial, estão relacionados com 19 revistas, havendo apenas três – a revista brasileira *Ciência da Informação*, e as revistas portuguesas *Páginas a & b* e *Cadernos BAD*, onde a autora tem mais do que uma participação.

No lugar de perpetuar a avaliação das revistas através do seu Fator de Impacto, seguiu-se o repto de alguns autores consagrados da Bibliome-

Tabela 3 – Distribuição das revistas por documentos, país, modelo de AA e bases de indexação (N=25)

Títulos das revistas	N.º Docs	País	Modelo de AA	Bases de indexação
<i>Ciência da informação</i>	4	Brasil	Diamante	Scopus; Dimensions
<i>Páginas A & B</i>	3	Portugal	Diamante (2) e Fechada (1)	
<i>Cadernos BAD</i>	2	Portugal	Diamante	
<i>AtoZ: novas práticas em informação e conhecimento</i>	1 ^a	Brasil	Diamante	Dimensions
<i>Biblios: Revista de bibliotecologia y ciencias de la información</i>	1	EUA	Diamante	SSCI, SciELOCI, Scopus, Dimensions
<i>Biblos: revista FLUC</i>	1	Portugal	Diamante	
<i>Brazilian Journal of Information Science</i>	1	Brasil	Diamante	Dimensions
<i>Documentación de las Ciencias de la Información</i>	1	Espanha	Diamante	Dimensions
<i>El profesional de la información</i>	1	Espanha	Híbrida	SSCI, Scopus, Dimensions
<i>Iris: Informação, Memória e Tecnologia</i>	1	Brasil	Diamante	
<i>Journal of Documentation</i>	1	Inglaterra	Híbrida	SSCI, Scopus, Dimensions
<i>Journal of the Association for Information Science and Technology</i>	1	EUA	Híbrida	SSCI, Scopus, Dimensions
<i>Knowledge Organization</i>	1	Alemanha	Híbrida	SSCI, Scopus, Dimensions
<i>Liinc em Revista</i>	1	Brasil	Diamante	Dimensions
<i>O SBAL: Serviço de Bibliotecas e Apoio à Leitura</i>	1	Portugal	Fechada	
<i>Ponto de Acesso</i>	1 ^b	Brasil	Diamante	Dimensions
<i>Revista Ibero-Americana de Ciência da Informação</i>	1	Brasil	Diamante	ESCI
<i>Scire representación y organización del conocimiento</i>	1	Espanha	Diamante	ESCI
<i>Tendências da Pesquisa Brasileira em Ciência da Informação</i>	1	Brasil	Diamante	

Legenda: ^a entrevista; ^b edição de número

tria (Wouters et al., 2019), no sentido de se estabelecerem indicadores mais diretamente ligados às suas funções primordiais. No caso presente, preferiu-se escolher um indicador diretamente ligado à disseminação do conhecimento tal como esta é garantida pelas práticas da Ciência Aberta, aqui tomada no seu aspeto mais simples, o modelo de Acesso Aberto (AA) de cada revista no momento em que Maria da Graça Simões nelas publicou. Em 72% das suas publicações em revistas, a escolha recaiu em revistas de modelo diamante, as que garantem gratuidade plena tanto aos autores como aos leitores, promovendo um maior e mais imediato grau de disseminação. Deve também notar-se que só se encontravam em acesso fechado duas revistas portuguesas cujas publicações datavam de 2001, altura em que a sua edição se limitava ao formato impresso. As quatro revistas de modelo híbrido, onde a autora optou por assegurar o seu acesso aberto, são das mais prestigiadas deste conjunto, na medida em que estão indexadas nas três bases de dados internacionais (SSCI, Scopus e Dimensions), mas oneram os autores com o pagamento de uma taxa para que a sua disseminação seja aberta para os leitores.

Na Tabela 4, pode observar-se a concentração de comunicações em quatro congressos, onde se repetem publicações, sendo 52% nos dois congressos especializados em Organização do Conhecimento (o *ISKO Internacional* (32%) e o seu capítulo ibérico (20%)), seguidos das seis comunicações no *EDICIC Ibérico*, o congresso dedicado à educação e investigação em Ciência da Informação, e de duas nos *Encontros das Bibliotecas de Arte de Espanha e Portugal*.

Daqueles 34 documentos, 62% (21) encontram-se em Acesso Aberto, um valor um pouco menos expressivo do que o que foi observado para as revistas, mas, ainda assim, maioritário. Quanto à indexação, é notória a dificuldade em se conseguir o reconhecimento das bases de dados internacionais em relação aos congressos, por precisarem de terem normas editoriais muito profissionalizadas, como as que são atingidas pela *ISKO Internacional*, ou pelas publicações da Imprensa da Universidade de Coimbra, responsável pela edição da conferência *A Biblioteca da Universidade: permanência e metamorfoses*.

Tabela 4 – Distribuição dos congressos por documentos, AA e bases de indexação (N=34)

Títulos dos congressos	N.º Docs	N.º Docs AA	Bases de indexação
<i>ISKO Espanha-Portugal</i>	11	11	
ISKO Internacional	7	0	Scopus (4), Dimensions (3)
<i>EDICIC Ibérico</i>	6	6	
Encontro de Bibliotecas de Arte de Espanha e Portugal	2	0	
CONFIBERCOM Congresso Mundial de Comunicação Ibero-Americana	1	1	
<i>ENANCIB Encontro Nacional de Pesquisa em Ciência da Informação [Brasil]</i>	1	1	
<i>EnReDo – Encontro de representação documental [Brasil]</i>	1	0	
European Conference of Architecture School Librarians	1	0	
International UDC Seminar	1	0	
<i>Jornadas de Bibliotecas de Arte [Portugal]</i>	1	0	
A Biblioteca da Universidade [de Coimbra]: permanência e metamorfoses	1	1	CPCI-SSH, Dimensions
<i>Conferência Luso-Brasileira sobre Acesso Aberto</i>	1 ^a	1	
Total	34	21	

Legenda: ^a *poster*

Colaboração

A colaboração, nas suas dimensões mais diversas, está no coração da atividade científica e representa um aspeto imprescindível no percurso de qualquer perfil de investigação, contribuindo para tornar efetiva a participação na dimensão essencialmente comunitária da ciência. A colaboração expressa formalmente na coautoria dos diferentes tipos de publicações é um dos meios mais expeditos de que dispomos para se identificarem e fixarem algumas das redes informais de partilha de interesses e trabalho de pesquisa comuns ao longo de uma vida. Como não podia deixar de ser, também o legado de Maria da Graça Simões foi construído com e a partir de uma teia de coautorias que nos ajudam a perceber melhor todo o seu potencial e alcance.

Apontada como um indicador de maturidade, autoridade e inclusão plena na carreira científica (Beaver, 2004; Beaver & Rosen, 1979), importa pois começar por estabelecer o peso relativo da colaboração no conjunto das suas publicações¹, tendo-se optado por observar a taxa de colaboração e o índice de coautoria nos dois momentos de referência já gizados.

Tabela 5 – Evolução da taxa de colaboração e do índice de coautoria ao longo da carreira (N=73)

Etapas	Taxa de colaboração	Índice de coautoria
Pré-doutoramento (1994-2010)	70,0%	1,8
Pós-doutoramento (2011-2021)	85,7%	3,0
Totalidade da carreira	83,6%	2,9

Na Tabela 5, é muito expressiva a importância da colaboração nas duas principais etapas, sendo já um valor elevado (70,0%) até ao doutoramento, mas com uma clara reafirmação do seu perfil marcadamente colaborativo na década que se lhe seguiu, com 83,6%, uma taxa muito próxima da apresentada num estudo dos documentos de autores espanhóis da Ciência da Informação indexados na Scopus (82%) entre 1997 e 2007 (Vargas-Quesada et al., 2010, p. 109). Para além da crescente aposta na colaboração, é ainda mais notório o alargamento do número de coautorias implicadas nas publicações realizadas na etapa do pós-doutoramento e da entrada oficial na carreira académica, quase duplicando o índice de coautoria de 1,8 para 3,0 autorias por documento.

Através da Tabela 6, pode perceber-se a clara internacionalização das coautorias, que, ao atingirem 66,6% dos documentos publicados na última

¹ Faz-se notar que nesta pesquisa não se diferenciou a ponderação das distintas tipologias documentais, pelo que a maior ou menor relação de trabalho exigida pelos diferentes tipos de documentos, como sejam um livro, um poster ou uma edição científica, não está espelhada nesta análise da colaboração, valendo cada documento dos 73 recolhidos ou cada participação das 133 coautorias uma unidade.

década, vêm substituir o perfil de produção mais doméstico (50,0%) e individual (30%) da primeira etapa, uma época de formação e preparação para uma transição da carreira profissional BAD e docência em regime supletivo, para uma adesão plena à carreira acadêmica. São de notar os valores muito baixos da colaboração a um nível nacional, que passa de uns escassos 10,0% para uns residuais 1,6%, o que vem fortalecer o que Blaise Cronin (2008) referiu sobre como a sua própria prática colaborativa se estendia, preferencialmente, aos mais chegados, aos alunos e colegas investigadores da sua instituição.

Tabela 6 – Distribuição relativa do tipo de colaboração ao longo da carreira (N=73)

Tipo de colaboração	Totalidade da carreira	Pré-doutoramento (1994-2010)	Pós-doutoramento (2011-2021)
Internacional	58,9%	10,0%	66,6%
Nacional	2,8%	10,0%	1,6%
Intrainstitucional	22,5%	50,0%	17,5%
Sem colaboração	16,9%	30,0%	14,3%

Para dar uma leitura mais concreta aos valores gerais da colaboração, vale a pena identificarem-se as afiliações insitucionais das 133 coautorias que 48 autores desenvolveram com Maria da Graça Simões. Na Tabela 7, podemos ver como a colaboração internacional se deu no eixo iberoamericano, com prevalência quase exclusiva das instituições brasileiras (36,1%) e das instituições espanholas (21,8%). Em termos nominais, destacam-se a Universidade Federal de S. Carlos, no estado de S. Paulo (16,5%), e a Universidad de León (15,0%), seguidas, ainda que com valores abaixo da metade, pela Universidade Federal de Minas Gerais (6,0%), a Universidade Estadual Paulista (5,3%) – as quatro universidades onde foi professora convidada –, a Fundação Getúlio Vargas (5,3%), no Rio de Janeiro, e a Universidade de Salamanca (4,5%), a *alma mater* de Maria da Graça Simões, onde defendeu a sua tese de doutoramento, em 2010.

Ao nível nacional, confirma-se a predominância da colaboração intra-institucional com colegas da Universidade de Coimbra, a sua instituição de base, desde sempre, com um expressivo peso relativo no cômputo

geral das instituições (39,1%). Finalmente, e a título pontual, surge ainda uma colaboração com uma autoria italiana da Università de Pavia.

Tabela 7 – Distribuição absoluta e relativa das coautorias pelas afiliações institucionais (N=133)

País	Instituições	N.º Coautorias	% Coautorias
Portugal	Univ Coimbra	52	39,1%
	Univ Porto	1	0,8%
	Univ Porto	1	0,8%
	Fundação Calouste Gulbenkian	1	0,8%
	Subtotal das instituições portuguesas	55	41,4%
Brasil	UFSCar	22	16,5%
	UFMG	8	6,0%
	UNESP	7	5,3%
	Fundação Getúlio Vargas	7	5,3%
	UFParaíba	2	1,5%
	Bib Pública Ibate SP	1	0,8%
	U S Paulo	1	0,8%
	Subtotal das instituições brasileiras	48	36,1%
Espanha	Univ Leon	20	15,0%
	Univ Salamanca	6	4,5%
	Univ Granada	2	1,5%
	Univ Carlos III Madrid	1	0,8%
	Subtotal das instituições espanholas	29	21,8%
Itália	Univ Pavia	1	0,8%
	Subtotal das instituições italianas	1	0,8%

A um nível pessoal, dos 48 autores com quem colaborou fê-lo pontualmente (apenas uma vez) com 23, havendo uma ligeira preponderância para as coautorias repetidas (52%). Tomando novamente o total das coautorias (133), 68% foram do sexo feminino e 38% do masculino, o que confirma a tendência para uma maior colaboração entre colegas do mesmo sexo (Sugimoto & Cronin, 2012). Quanto ao contexto de colaboração, das 133 coautorias, 89% foram na academia e investigação e 11% no universo profissional BAD.

Tabela 8 – Distribuição relativa das coautorias académicas pelo tipo de relação com Maria da Graça Simões (n=118)

Tipo de relação	N.º Coautorias	% Coautorias
Mestres	14	12%
Discípulos	23	19%
Pares	81	69%

À semelhança do que Sugimoto e Cronin ensaiaram no seu modelo de perfis biobibliométricos (2012), na Tabela 8 distinguem-se as coautorias do meio académico pela relação que Maria da Graça Simões tinha com cada uma, dentro de uma categorização da hierarquia científica, revelando-se como predominaram as relações de parceria (69%), mas também se mantiveram os laços de colaboração com os seus mestres (orientadores, 12%), ainda que tendo sido suplantadas pelas coautorias com os seus discípulos (orientandos, 19%). Em todo o caso, ao identificar a um nível nominal as coautorias mais destacadas de cada uma destas categorias – 13 coautorias com a sua orientadora Blanca Rodríguez Bravo, da Universidad de León, número só igualado pelas 13 colaborações com a sua colega da Universidade de Coimbra, Maria Cristina Vieira de Freitas, e as 9 coautorias com o seu orientando Luís Miguel Machado – presente-se que, pela sua expressão, estaremos perante relações que foram todas convergindo para uma verdadeira relação *inter pares*.

Conclusão

Do perfil aqui ensaiado, ressalta a grande versatilidade de Maria da Graça Simões para disseminar os seus resultados científicos, tendo recorrido a um conjunto muito diversificado de tipologias documentais, incluindo a publicação de livros especializados no domínio da Representação da Informação e Organização do Conhecimento desde bastante cedo na sua carreira, a que se juntou uma participação muito regular e intensa nos principais congressos internacionais e regionais da mesma especialidade, a par de uma publicação consistente em revistas internacionais, com grande prevalência da publicação em Acesso Aberto.

Um outro traço muito característico foi a constituição de uma ampla rede de coautorias, onde prevaleceram tanto colegas da Universidade de Coimbra, predominantemente do contexto académico, mas também do meio profissional BAD, como colegas das universidades brasileiras e espanholas com quem foi selando, através das publicações em colaboração, os laços tecidos desde o seu doutoramento até aos projetos de investigação lançados como professora convidada.

Se há, no entanto, uma última conclusão a retirar deste breve périplo quantitativo pelas publicações de Maria da Graça Simões é a de que, felizmente, nenhuma vida, nenhum percurso científico se reduzem a um único número ou à medição estritamente quantitativa do seu impacto, que aqui optámos deliberadamente por não desenvolver, mesmo que os números nos ajudem a objetivar e comprovar muito do que, de alguma forma, já fora intuitivo. Na verdade, do legado científico da Professora Maria da Graça importa, mais do que nada, continuar a lê-lo e com ele continuar a aprender e a ir mais ao fundo das questões que tanto a apaixonaram.

Referências bibliográficas

- Beaver, D. (2004). Does collaborative research have greater epistemic authority? *Scientometrics*, 60(3), 399-408. <https://doi.org/10.1023/B:SCIE.0000034382.85360.cd>
- Beaver, D., & Rosen, R. (1979). Studies in scientific collaboration Part III. Professionalization and the natural history of modern scientific co-authorship. *Scientometrics*, 1(3), 231-245. <https://doi.org/10.1007/BF02016308>
- Bornmann, L., & Daniel, H. (2007). What do we know about the *b* index? *Journal of the American Society for Information Science and Technology*, 58(9), 1381-1385. <https://doi.org/10.1002/asi.20609>
- CoARA. (2022). *The Agreement on Reforming Research Assessment*. <https://coara.eu/agreement/the-agreement-full-text/>
- Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology*, 59(6), 1002-1006. <https://doi.org/10.1002/asi.20774>
- Engels, T. C. E., Istenič Starčič, A., Kulczycki, E., Pölonen, J., & Sivertsen, G. (2018). Are book publications disappearing from scholarly communication in the social sciences and humanities? *Aslib Journal of Information Management*, 70(6), 592-607. <https://doi.org/10.1108/AJIM-05-2018-0127>
- Glänzel, W., Schlemmer, B., Schubert, A., & Thijs, B. (2006). Proceedings literature as additional data source for bibliometric analysis. *Scientometrics*, 68(3), 457-473. <https://doi.org/10.1007/s11192-006-0124-y>
- Harzing, A.-W. (2007). *Publish or Perish* (Versão 8) [Software]. <https://harzing.com/resources/publish-or-perish>

- Hicks, D. (2012). One size doesn't fit all: On the co-evolution of national evaluation systems and social science publishing. *Confero*, 1(1). https://works.bepress.com/diana_hicks/33/
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431. <https://doi.org/10.1038/520429a>
- Karolinska Institute. Research Strategy Committee. (2013). *Recommendations concerning the use of bibliometric indicators in the assessment of publications of individual researchers* (2-1242/2013; p. 9). Karolinska Institutet, University Library Web Team. <https://kib.ki.se/en/publish-analyse/bibliometrics-verification/recommendations-concerning-use-bibliometric-indicators>
- Simões, M. da G. (2019a). *Curriculum Vitae* [Online]. *CiênciaVitae*. <https://www.cienciavitae.pt/4F1F-41FA-64FD>
- Simões, M. da G. (2019b). *Relatório de Atividades Científicas e Pedagógicas com vista à Manutenção do Contrato de Trabalho por Tempo Indeterminado: 2014-2019* (p. 43). Faculdade de Letras da Universidade de Coimbra. https://www.uc.pt/fluc/pessoal/docentes/relatoriospdf/atividades_pedag/graca_simoes.pdf
- Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. Em P. Wouters & E. Noyons (Eds.), *Context Counts: Pathways to Master Big and Little Data* (pp. 598-604). University of Leiden, CWTS. https://www.researchgate.net/profile/Tindaro_Cicero/publication/265396216_Research_quality_characteristics_of_publications_and_socio-demographic_features_of_Universities_and_Researchers_evidence_from_the_Italian_VQR_2004-2010_evaluation_exercise/links/540d89180cf2df04e754b658.pdf#page=610
- Sugimoto, C. R., & Cronin, B. (2012). Biobibliometric profiling: An examination of multifaceted approaches to scholarship. *Journal of the American Society for Information Science and Technology*, 63(3), 450-468. <https://doi.org/10.1002/asi.21695>
- Vargas-Quesada, B., Minguillo, D., Chinchilla-Rodríguez, Z., & Moya-Anegón, F. (2010). Estructura de la colaboración científica española en Biblioteconomía y Documentación (Scopus 1999-2007). *Revista Interamericana de Bibliotecología*, 33(1), 105-123. <https://revistas.udea.edu.co/index.php/RIB/article/view/6283>
- Waltman, L., & Van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406-415. <https://doi.org/10.1002/asi.21678>
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J., & Johnson, B. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management* (p. 163). <https://doi.org/10.13140/RG.2.1.4929.1363>
- Wouters, P., Glänzel, W., Gläser, J., & Rafols, I. (2013). The dilemmas of performance indicators of individual researchers: An urgent debate in bibliometrics. *ISSI Newsletter*, 9(3), 48-53. https://www.researchgate.net/publication/301853186_The_dilemmas_of_performance_indicators_of_individual_researchers-An_urgent_debate_in_bibliometrics
- Wouters, P., Sugimoto, C. R., Larivière, V., McVeigh, M. E., Pulverer, B., de Rijcke, S., & Waltman, L. (2019). Rethinking impact factors: Better ways to judge a journal. *Nature*, 569(7758), 621-623. <https://doi.org/10.1038/d41586-019-01643-3>

