

OLÍVIA MANUELA MARQUES PESTANA

**OS VOCABULÁRIOS CONTROLADOS E O DOMÍNIO DA
BIOMEDICINA: PERCURSOS DE (INTER)OPERABILIDADE**

Sumário pormenorizado da lição

Faculdade de Letras da Universidade do Porto
Julho de 2025

Para cumprimento do disposto no artigo 8º - alínea c) do Decreto-Lei nº 239/2007, de 19 de Junho, que aprova o regime jurídico do título académico de agregado, bem como do disposto no artigo 19º do Decreto-Lei nº 64/2023, de 31 de julho, que atualiza o regime jurídico do título académico de agregado.

Sumário

<i>Introdução.....</i>	<i>4</i>
<i>Enquadramento temático.....</i>	<i>7</i>
<i>Sumário da lição a apresentar.....</i>	<i>12</i>
<i>Referências citadas neste texto.....</i>	<i>13</i>
<i>Referências consultadas para preparação da lição</i>	<i>16</i>

Introdução

A seleção de um tema para a lição, em cumprimento do determinado legalmente no contexto da prestação de provas para a obtenção do título de agregado, reveste-se de alguma dificuldade, especialmente se tivermos em consideração a amplitude temática que a Ciência da Informação nos oferece, bem como a individual experiência que a docência na área e a investigação nos proporcionam. O tema que, aqui, é apresentado decorre de diversas paixões de que as vidas profissional e académica da autora desta lição a têm permitido usufruir. A abordagem à combinação do tratamento da informação por assuntos com a especialidade terminológica do domínio da biomedicina resulta de um percurso de três décadas de estudo em torno destes temas. A lição decorre, por isso, da experiência de investigação sobre o tema da aplicação de vocabulários controlados no domínio da biomedicina, com especial atenção à aplicação dos vocabulários à literatura científica. O facto de o domínio da biomedicina oferecer em livre acesso a pesquisa numa das bases de dados mundialmente mais acedida, a MEDLINE, e com maior oferta de modalidades de pesquisa, graças à sua robusta estrutura tecnológica e à aplicação consolidada dos princípios de construção e utilização de vocabulários controlados, torna este tema de referência obrigatória em aulas de qualquer ciclo de estudos que abordem o controlo de autoridade e a construção e utilização dos vocabulários controlados, com efetivo reflexo na recuperação da informação.

De acordo com González-Márquez et al. (2024), são atualmente publicados mais de 1 milhão de artigos científicos por ano no domínio das ciências biomédicas e da vida. Em 2021, esperava-se que menos de 200 indexadores da National Library of Medicine (NLM), dos EUA, indexassem 1 milhão de artigos (Rae et al., 2020), o que não seria possível sem a ajuda do sistema Medical Text Indexer (MTI), um sistema automatizado que combina aprendizagem automática e métodos baseados em regras para analisar títulos e resumos de artigos e gerar pontos de acesso por assunto de medicina recomendados - os Medical Subject Headings (MeSH). Em 2002, a NLM introduziu a indexação semiautomatizada da base de dados MEDLINE utilizando a tecnologia MTI e anunciou, mais tarde, em 2021, que iria, até meados de 2022, automatizar totalmente a indexação da base de dados com um MTI melhorado (Chen, Bullard & Giustini, 2023). A MEDLINE é a principal base de dados bibliográfica da NLM, que contém, atualmente, mais de 31 milhões de referências a artigos de publicações científicas no domínio das

ciências da vida, com especial incidência na biomedicina, e é a principal componente da PubMed, uma base de dados de acesso gratuito desenvolvida e mantida pelo Centro Nacional de Informação Biotecnológica (NCBI) da NLM¹. A par da utilização de vocabulário controlado para a indexação dos artigos da referida base de dados, muitas bibliotecas, incluindo a NLM, utilizam este vocabulário para indexar todos os recursos que disponibilizam, seja em catálogos de bibliotecas, em repositórios de trabalhos académicos ou de investigação. E fazem-no, também, para aconselhar autores ou utilizadores dos serviços, no sentido de melhor indexarem a sua produção científica ou de otimizarem a pesquisa de literatura.

Nas últimas décadas, têm sido feitos esforços para aproximar os sistemas dos utilizadores finais, fornecendo-lhes os recursos necessários para realizarem as suas próprias pesquisas através de interfaces e menus de fácil utilização que ajudam a melhorar o desenvolvimento das estratégias de pesquisa (Lopes, 2002). Sendo a informação crítica para o sucesso dos cuidados prestados aos doentes e para o controlo dos custos dos cuidados de saúde (Pestana, 2014), a qualidade do resumo e da indexação tem um impacto direto na otimização da recuperação da informação. Mas, apesar dos avanços tecnológicos, persiste a necessidade de aplicação de vocabulários controlados à informação, contribuindo para a recuperação da informação mais adequada a uma dada necessidade de informação, mesmo que o utilizador não se aperceba da existência desses vocabulários.

É neste cenário que se desenrola a evolução dos vocabulários controlados do domínio da biomedicina, havendo, nos últimos anos, uma tendência no sentido de aproximar as linguagens do contexto bibliográfico das linguagens utilizadas em contexto da prestação de cuidados de saúde. Nesta lição serão abordados os vocabulários controlados no domínio de especialidade da biomedicina, a fundamentação teórica que subjaz à sua construção e à sua operacionalização, quer relativamente à evolução da construção das suas estruturas e delimitações terminológicas, quer ao exercício da indexação e serão, ainda, ilustradas experiências de interoperabilidade entre diversos vocabulários controlados neste domínio. A finalizar a lição será equacionada a evolução tecnológica

¹ Informação confirmada em 30 junho de 2025, através do endereço: https://www.nlm.nih.gov/medline/medline_overview.html. A última revisão do conteúdo desta página é de 5 de fevereiro de 2024.

recente no âmbito da inteligência artificial e o seu reflexo na construção e (inter)operabilidade dos vocabulários controlados em biomedicina.

Enquadramento temático

Segundo Doris H. Clack (1990), o controlo da autoridade de assunto visa assegurar “a utilização de um vocabulário coerente que corresponda exatamente tanto ao do utilizador como ao do documento”. Na tradição biblioteconómica norte-americana, a expressão “vocabulários controlados” é frequentemente associada apenas a tesouros e a listas de cabeçalhos de assuntos² (Wyman, 1999). No entanto, a noção de controlo de termos na indexação e recuperação de assuntos é relevante para todas as linguagens de indexação, uma vez que todas funcionam com vocabulário e regras de sintaxe previamente preparados. O facto de este “vocabulário” ser expresso em linguagem ou em símbolos de uma notação não altera o processo de análise ou indexação de assuntos. Hjørland (1997) salienta que as linguagens de recuperação da informação representam os resultados da análise do assunto e, por conseguinte, não existe uma diferença real entre a indexação e a classificação do assunto. Persiste algum debate sobre os termos “vocabulário controlado”, “linguagens de indexação” e “sistemas de organização do conhecimento”, sobretudo tendo em consideração que alguns autores os apresentam como sinónimos, enquanto outros autores estabelecem uma distinção e hierarquização entre os termos³.

Conforme identificado por Zeng e Chan (2004), e utilizando a sua opção terminológica, a evolução dos sistemas de organização do conhecimento teve o seu período mais produtivo entre o início da década de 1960 e o final da de 1980, impulsionado pelo crescimento das bases de dados online. Entre meados e o final dos anos 1950, com a rápida expansão das tecnologias da informação, bem como da produção científica, começaram a ser desenvolvidas teorias sobre informação e comunicação aplicadas à indexação e à recuperação da informação.

As primeiras bases de dados em linha eram sobretudo bases de dados de indexação e resumo ou fontes secundárias. A importância dos vocabulários controlados para aumentar a eficácia da recuperação de informação (medida basicamente pelos indicadores de

² É de referir que a IFLA, na edição de 2009 da *Declaração dos Princípios Internacionais de Catalogação (PIC)*, recomendou a utilização da expressão “ponto de acesso autorizado” ou “ponto de acesso controlado” em substituição de “cabeçalho” (“heading”), no entanto o termo permanece no título de vários vocabulários controlados, como os Medical Subject Headings, ou os Library of Congress Subject Headings.

³ Ver, a este respeito, por exemplo, as abordagens de Mario Barité (2016), Robert Fugman (1993) ou Claudio Gnoli (2020) e, ainda, o contributo da discussão apresentada por Armando Malheiro da Silva (2006) no capítulo intitulado “Informação e Conhecimento”, bem como por Fulvio Mazzocchi (2018).

revocação e de precisão) foi confirmada por experiências de recuperação de informação, como os testes de Cranfield (Cleverdon, 1967) e a investigação MEDLARS (Lancaster, 1969). Esta última reveste-se de elevada relevância, dado que resulta na consolidação do sistema que deu origem à base de dados MEDLINE, ou seja, MEDLARS online. Um estudo mais recente, desenvolvido pela autora desta lição, permitiu, justamente, avaliar a recuperação da informação na base de dados MEDLINE, através da plataforma PubMed, recorrendo aos indicadores de revocação e de precisão e comparando os resultados de pesquisas por linguagem natural vs. vocabulário controlado, tendo incidido em dois assuntos de uma área especializada da prestação de cuidados de saúde, os cuidados intensivos (Pestana, 2014).

A abordagem pós-coordenada, que se caracteriza pela combinação explícita, no momento da pesquisa, de termos conceptualmente simples, surgiu em resposta ao crescimento da literatura, especialmente dos artigos de publicações periódicas científicas e à necessidade de uma maior especificação do conteúdo. Os tesauros orientados para o domínio, devido à sua utilização de termos com superior grau de especificidade, foram frequentemente considerados mais eficazes para a recuperação da informação do que os esquemas de classificação.

Os recursos de informação fornecidos por bases de dados em linha e disponíveis na Internet são indexados de muitas formas diferentes, muitas vezes utilizando vocabulários temáticos distintos e organizados de acordo com esquemas diversos, daí que os profissionais da informação tenham, desde há algum tempo, procurado formas de facilitar a pesquisa entre domínios (Zeng e Chan, 2004).

Os pontos de acesso são atribuídos no âmbito do processo de indexação de assuntos, que inclui a análise de conteúdo, a seleção dos conceitos importantes e a sua posterior representação com base na utilização de um vocabulário controlado. A utilização deste vocabulário permitirá, depois, proceder-se ao exercício do controlo da autoridade. Para garantir a coerência na aplicação do processo em causa pelos vários indexadores, a NLM estabeleceu uma política interna aplicável a todos os tipos de materiais processados: NLM Policy on Subject Analysis and Classification⁴ (Política da NLM sobre análise e classificação de assuntos). Desde 1879 que a NLM indexa a literatura biomédica com

⁴ Disponível em: <http://www.nlm.nih.gov/tsd/cataloging/subjclasspolicy.html>. Consultado em 30 de junho de 2025.

vocabulário controlado, permitindo, aos profissionais de saúde e da área da biomedicina, aceder à informação de que necessitam para a investigação, os cuidados de saúde e a formação através da linguagem desenvolvida para o efeito. A partir da década de 1960, com o desenvolvimento do sistema de análise e recuperação da literatura, a NLM passou a permitir o acesso rápido a uma vasta gama de referências de informação biomédica. Em 1965, foi iniciado o serviço de pesquisa por pedido direto à NLM ou a um dos seus centros e, em 1968, foi inaugurado em Nova Iorque o primeiro sistema de recuperação de informação bibliográfica em linha: SUNY Biomedical Communication Network (Egeland, 1975; McClure, 2013). Tratava-se de um sistema de telex através do qual as pesquisas bibliográficas eram enviadas para a base de dados MEDLARS sem ser necessário aguardar pelo envio dos resultados impressos. Após o estudo de Lancaster (1969), já referido, a NLM lançou, em 1971, uma versão mais avançada do software, disponível através de uma rede de telecomunicações de valor acrescentado: o MEDLARS Online ou MEDLINE (Coletti e Bleich, 2001). Duas décadas mais tarde, em julho de 1992, foi anunciado um serviço de acesso à MEDLINE através da Internet. Este utilizava a capacidade Telnet e, em 1996, devido ao aumento da utilização dos serviços, a NLM lançou uma interface de pesquisa amigável: o Internet Grateful Med. Esta foi a primeira aplicação de software através da World Wide Web. A partir de 1997, o acesso à base de dados passou a ser gratuito, também através da World Wide Web, com a PubMed a substituir a antiga interface, permitindo ao utilizador final a possibilidade de percorrer as diversas modalidades de pesquisa, incluindo por pontos de acesso.

Para a formulação dos pontos de acesso é, então, necessário ter medidas estabelecidas para exercer o chamado controlo de autoridade, ou seja, garantir a uniformidade de modo a que os registos pertencentes à mesma entidade, publicação ou conceito não se dispersem por sinónimos ou variantes. O controlo de autoridade é o estabelecimento consistente e padronizado de um termo autorizado para um cabeçalho de assunto, de nome ou de título e as respetivas ligações dos termos variantes, facilitando a recuperação de modo a obter apenas o que é relevante e não um vasto resultado que também inclua o que é relevante (Garrett, 1997; Pisciotto, 2001).

Os pontos de acesso por assuntos constituem uma importante parte do sistema de informação com vista à recuperação da informação mais relevante. A sua formulação, para além de controlar a sinonímia de termos passíveis de representar o mesmo conceito, permite distinguir palavras homógrafas e, ainda, reunir num só termo as diversas formas

associáveis à representação de um conceito, como por exemplo, o singular vs. plural, as abreviaturas, as variações ortográficas ou os termos estrangeiros.

Como exemplo de um dos vocabulários controlados do domínio da biomedicina, saliente-se o tesauro Medical Subject Headings (MeSH), o qual consiste num vocabulário controlado produzido pela NLM para indexar e pesquisar artigos de publicações periódicas na base de dados MEDLINE, bem como publicações de outras bases de dados produzidas pela NLM. A cada artigo está associado um conjunto de cabeçalhos de assunto que melhor representam o seu conteúdo. Os termos que representam os principais conceitos são assinalados pelos indexadores com um asterisco (*). Tal tem um efeito direto na estratégia de pesquisa e, desde que os utilizadores conheçam esta característica, podem aplicar o limite "Major Topic Heading" no menu de pesquisa MeSH.

O MeSH, versão de 2025, contém cerca de 30.000 descritores, apresentados por ordem alfabética ou numa estrutura hierárquica que permite efetuar pesquisas a vários níveis de especificidade. Está organizado em 16 categorias denominadas "estruturas em árvore", devido à sua estrutura ramificada. Cada categoria está dividida em subcategorias, que contêm conjuntos de descritores organizados hierarquicamente, do mais geral para o mais específico. Os indexadores recebem instruções precisas para procurar e utilizar o descritor mais específico possível para representar cada conceito. A cada descritor é atribuído um número que indica a sua posição na hierarquia. Alguns termos têm vários números, dado poderem ocorrer em mais do que um lugar na hierarquia do vocabulário. Cada termo tem, também, um conjunto de qualificadores utilizados para especificar um determinado aspeto, por exemplo, "classificação" (CL), "diagnóstico" (DI) ou "terapia" (TH). Juntamente com cada descritor, podem existir informações sobre termos relacionados ("Ver também") e sobre termos sinónimos ("Termo de entrada"). Conclui-se, portanto, que a dimensão e a complexidade de um vocabulário como o MeSH o tornam como indispensável no estudo dos vocabulários controlados.

Ao serem publicados por ano mais de 1 milhão de artigos no domínio das ciências biomédicas e da vida, conforme já referido, o caminho na direção da total automatização do processo de indexação torna-se uma evidência, todavia, ainda sem ter atingido o estado de total processamento sem necessidade de correções. Amar-Zifkin et al. (2025), por exemplo, observaram, num estudo, que a indexação algorítmica na MEDLINE atribuía

frequentemente termos irrelevantes ou imprecisos às publicações, podendo, por isso, gerar os fatores de silêncio e de ruído na recuperação da informação.

Com o aparecimento de plataformas de agregação de recursos de informação, tornou-se necessário o estabelecimento da interoperabilidade entre os vocabulários existentes. Zeng e Chan (2004) elencaram as principais tendências e questões no estabelecimento da interoperabilidade e afirmaram que é uma questão inevitável no ambiente em rede. Todavia, trata-se de uma tarefa complexa, pois, conforme referido por Dextre Clarke (2001), as relações semânticas básicas existentes, por exemplo, nas estruturas tradicionais dos tesouros, nem sempre foram estabelecidas de forma consistente, o que resulta em estruturas semânticas imprevisíveis. Não obstante, a investigação em torno desta problemática está, já, consolidada e são múltiplos os estudos que abordam a interoperabilidade entre vocabulários controlados, como é o caso dos trabalhos desenvolvidos por Bekhuis, Demner-Fushman, & Crowley (2013), Andrade, & Lara (2016), ou, ainda, por Fung, K.W. et al. (2024). Do ponto de vista dos resultados práticos desta ação, refira-se, como exemplo, o trabalho que a NLM tem realizado com vista à promoção da interoperabilidade entre o MeSH e a NLM Classification, em que nove novos descritores da edição de 2025 do MESH foram adicionados ao índice da NLM Classification⁵.

Torna-se, portanto, permanentemente oportuno o ensino e o estudo não só da evolução dos vocabulários controlados quanto à sua estrutura e seleção terminológica, como, também, quanto à evolução tecnológica que vá de encontro a uma total automatização dos processos de indexação, de incremento dos indicadores de recuperação da informação e da possibilidade de interoperabilidade entre os diversos vocabulários controlados.

⁵ Cf. informação disponível em: <https://classification.nlm.nih.gov/scope> (Endereço consultado em 30 de junho de 2025)

Sumário da lição a apresentar

A lição a apresentar seguirá o seguinte plano:

1. O conceito de vocabulário controlado em Ciência da Informação
 - 1.1. Vocabulário controlado: origens de um conceito
 - 1.2. Conceito de sistema de organização do conhecimento
 - 1.3. Conceito de linguagem de indexação

2. O papel dos vocabulários controlados e o acesso à informação em biomedicina
 - 2.1. Controlo de autoridade nos serviços de informação
 - 2.2. Indexação humana, semiautomática e automática
 - 2.3. Acesso à informação por assunto
 - 2.3.1. Acesso através da utilização de linguagem natural
 - 2.3.2. Acesso através do uso de vocabulários controlados

3. Os vocabulários controlados no domínio da biomedicina e a sua utilização
 - 3.1. Caracterização dos vocabulários controlados de biomedicina
 - 3.2. Estruturas e delimitações terminológicas dos vocabulários controlados
 - 3.3. Utilização de vocabulários controlados para indexação de literatura científica

4. A interoperabilidade dos vocabulários controlados de biomedicina
 - 4.1. Modalidades de interoperabilidade entre os vocabulários controlados
 - 4.2. Orientações da normalização internacional
 - 4.3. Exemplos e estudos de interoperabilidade

5. O papel da inteligência artificial e o futuro dos vocabulários controlados: um ponto de interrogação
 - 5.1. Inteligência artificial no contexto da indexação de literatura biomédica
 - 5.2. Desafios e potenciais contributos tecnológicos para a evolução dos vocabulários controlados

Referências citadas neste texto

Amar-Zifkin, A., Ekmekjian, T., Paquet, V., & Landry, T. (2025). Algorithmic indexing in MEDLINE frequently overlooks important concepts and may compromise literature search results. *Journal of the Medical Library Association*, 113(1), 39–48. <https://doi.org/10.5195/jmla.2025.1936>

Andrade, J., & Lara, M. L. G. (2016). Interoperability and Mapping Between Knowledge Organization Systems: Metathesaurus—Unified Medical Language System of the National Library of Medicine. *Knowledge Organization*, 43(2), 107–112. <https://doi.org/10.5771/0943-7444-2016-2-107>

Barité, M. (2016). Control de vocabulario: orígenes, evolución y proyección. *Ciência da Informação*, 41(1), 95–119. <https://doi.org/10.18225/ci.inf.v43i1.1421>

Bekhuis, T., Demner-Fushman, D., & Crowley, R. S. (2013). Comparative effectiveness research designs: an analysis of terms and coverage in Medical Subject Headings (MeSH) and Emtree. *Journal of the Medical Library Association*, 101(2), 92–100. <https://doi.org/10.3163/1536-5050.101.2.004>

Chen, E., Bullard, J. & Giustini, D. (2023). Automated indexing using NLM's Medical Text Indexer (MTI) compared to human indexing in Medline: A pilot study. *Journal of the Medical Library Association*, 111(3), 684–695. <https://doi.org/10.5195/jmla.2023.1588>

Clack, D. H. (1990). *Authority control: principles, applications, and instruction*. American Library Association.

Cleverdon, C. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19, 173–192.

Coletti, M. H., & Bleich, H. L. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association*, 8(4), 317–323. <http://doi.org/10.1136/jamia.2001.0080317>

Dextre Clarke, S. G. (2001). Thesaural relationships. In Bean, C., & Green, R. (Eds), *Relationships in the Organization of Knowledge* (pp. 37–52). Kluwer. <https://doi.org/10.1007/978-94-015-9696-1>

Egeland J. (1975). The SUNY biomedical communication network: six years of progress in on-line bibliographic retrieval. *Bulletin of the Medical Library Association*, 63(2), 189–194.

Fugmann, R. (1993). *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice*. Indeks-Verlag.

Fung, K.W. et al. (2024). Promoting interoperability between SNOMED CT and ICD-11: lessons learned from the pilot project mapping between SNOMED CT and the ICD-11 Foundation. *Journal of the American Medical Informatics Association*, 31(8), 1631-1637. <https://doi.org/10.1093/jamia/ocae143>

Garrett, L. (1997). *Authority control and its influence on recall and precision in an online bibliographic catalog*. University of Hawaii. Tese de Doutorado. (acesso através da base de dados ProQuest Dissertations & Theses)

Gnoli, C. (2020). *Introduction to knowledge organization*. Facet Publishing.

González-Márquez, R. et al. (2024). The landscape of biomedical research. *Patterns*, 5, 100968. <https://doi.org/10.1016/j.patter.2024.100968>

Hjørland, B. (1997). Information seeking and subject representation: an activity-theoretical approach to information science. Greenwood Press.

Lancaster, F. W. (1969). MEDLARS: Report on the evaluation of its operating efficiency. *American Documentation*, 20, 119 –142.

Lopes, I. L. (2002). Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ciência da Informação*, 31(1), 41–52.

Mazzocchi, F. (2018). Knowledge organization system (KOS): an introductory critical account. *Knowledge Organization*, 45(1), 54-78. Também disponível em: ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli, <https://www.isko.org/cyclo/kos>

McClure L. W. (2013). When the librarian was the search engine: introduction to the special issue on new roles for health sciences librarians. *Journal of the Medical Library Association*, 101(4), 257–260. <https://doi.org/10.3163/1536-5050.101.4.006>

Pestana, O. (2014). Information value and quality for the health sector: a case study of search strategies for optimal information retrieval. In G. L. Jamil, A. M. Silva, & F. Ribeiro (Eds.),

Rethinking the Conceptual Base for New Practical Applications in Information Value and Quality (pp. 116-133). IGI Global. <http://doi.org/10.4018/978-1-4666-4562-2.ch006>

Pisciotta, Robert A. (2001) – Authority control in library catalogs. In L. L. Thompson, *Bibliographic management of information resources in health sciences libraries* (pp. 115-149). Medical Library Association and The Scarecrow Press.

Rae, A. R., Pritchard, D. O., Mork, J. G., & Demner-Fushman, D. (2020). Automatic MeSH Indexing: Revisiting the Subheading Attachment Problem. *AMIA. Annual Symposium Proceedings. AMIA Symposium, 2020*, 1031–1040.

Silva, A. M. (2006). *A informação: da compreensão do fenómeno e construção do objecto científico*. CETAC.com e Afrontamento.

Zeng, M. L., & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377-395. <https://doi.org/10.1002/asi.10387>

Referências consultadas para preparação da lição

- Al-Rawas, M., Qader, O. A. J. A., Othman, N. H., Ismail, N. H., Mamat, R., Halim, M. S., Abdullah, J. Y., & Noorani, T. Y. (2025). Identification of dental related ChatGPT generated abstracts by senior and young academicians versus artificial intelligence detectors and a similarity detector. *Scientific Reports*, 15(1), 11275. <https://doi.org/10.1038/s41598-025-95387-y>
- Amar-Zifkin, A., Ekmekjian, T., Paquet, V., & Landry, T. (2025). Algorithmic indexing in MEDLINE frequently overlooks important concepts and may compromise literature search results. *Journal of the Medical Library Association*, 113(1), 39–48. <https://doi.org/10.5195/jmla.2025.1936>
- Andrade, J., & Lara, M. L. G. (2016). Interoperability and Mapping Between Knowledge Organization Systems: Metathesaurus—Unified Medical Language System of the National Library of Medicine. *Knowledge Organization*, 43(2), 107-112. <https://doi.org/10.5771/0943-7444-2016-2-107>
- Barité, M. (2011). Sistemas de organização do conhecimento: uma tipologia atualizada. *Informação & Informação*, 16(2), 122–139. <https://doi.org/10.5433/1981-8920.2011v16n2p122>
- Barité, M. (2014). El control de vocabulario en la era digital: revisión conceptual. *Scire*. 20(1) 99-108.
- Barité, M. (2016). Control de vocabulario: orígenes, evolución y proyección. *Ciência da Informação*, 41(1), 95-119. <https://doi.org/10.18225/ci.inf.v43i1.1421>
- Barité, M. (2025). Knowledge Organization and Terminology: intersections, interlocutions and projections. *Informatio*, 30(1), 2025, e209. <http://doi.org/10.35643/Info.30.1.14>
- Barité, M. et al. (2015). *Diccionario de Organización del Conocimiento: Clasificación, Indización, Terminología*. 6.a ed. corregida y aumentada. CSIC.
- Bawden, D., & Robinson, L. (2022). *Introduction to information science* (2nd. ed.). Facet Publishing.
- Bekhuis, T., Demner-Fushman, D., & Crowley, R. S. (2013). Comparative effectiveness research designs: an analysis of terms and coverage in Medical Subject Headings (MeSH) and Emtree. *Journal of the Medical Library Association*, 101(2), 92-100. <https://doi.org/10.3163/1536-5050.101.2.004>

- Borko, H., & Bernier, C. L. (1978). *Indexing concepts and methods*. Academic Press.
- Boteram, F., Gödert, W., & Hubrich, J. (Eds.) (2011). *Concepts in context: Proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization : July 19th-20th, 2010*. Ergon Verlag.
- Buchanan, B. (1976). *A glossary of indexing terms*. Clive Bingley.
- Cavalié, E. (2019). *L'indexation matière en transition: de la réforme de Rameau à l'indexation automatique*. Éditions du Cercle de la Librairie.
- Chen, E., Bullard, J., & Giustini, D. (2023). Automated indexing using NLM's Medical Text Indexer (MTI) compared to human indexing in Medline: A pilot study. *Journal of the Medical Library Association*, 111(3), 684–694. <https://doi.org/10.5195/jmla.2023.1588>
- Clarck, D. H. (1990). *Authority control: principles, applications, and instruction*. American Library Association.
- Cleveland, A. D., & Cleveland, D. B. (2009). *Health informatics for medical librarians*. New York: Neal-Schuman Publishers.
- Cleverdon, C. (1967). The Cranfield tests on index language devices. *Aslib Proceedings*, 19(6), 173–192.
- Coletti, M. H., & Bleich, H. L. (2001). Medical subject headings used to search the biomedical literature. *Journal of the American Medical Informatics Association*, 8(4), 317–323. <http://doi.org/10.1136/jamia.2001.0080317>
- Dahlberg, I. (2014). *Wissensorganisation: Entwicklung, Aufgabe, Anwendung, Zukunft*. Ergon Verlag.
- Dai, S., You, R., Lu, Z., Huang, X., Mamitsuka, H., & Zhu, S. (2020). FullMeSH: improving large-scale MeSH indexing with full text. *Bioinformatics*, 36(5), 1533–1541. <https://doi.org/10.1093/bioinformatics/btz756>
- DeMars, M. M., & Perruso, C. (2022). MeSH and text-word search strategies: Precision, recall, and their implications for library instruction. *Journal of the Medical Library Association*, 110(1), 23–33. <https://doi.org/10.5195/jmla.2022.1283>

Dextre Clarke, S. G. (2001). Thesaural relationships. In Bean, C., & Green, R. (Eds), *Relationships in the Organization of Knowledge* (pp. 37-52). Kluwer. <https://doi.org/10.1007/978-94-015-9696-1>

Dusenne, M., Billey, K., Desgrippes, F., Benis, A., Darmoni, S. J., & Grosjean, J. (2022). WikiMeSH: Multi Lingual MeSH Translations via Wikipedia. *Studies in Health Technology and Informatics*, 294, 403–404. <https://doi.org/10.3233/SHTI220483>

Egeland J. (1975). The SUNY biomedical communication network: six years of progress in on-line bibliographic retrieval. *Bulletin of the Medical Library Association*, 63(2), 189–194.

Fernandez-Llimos, F., Negrão, L. G., Bond, C., & Stewart, D. (2024). Influence of automated indexing in Medical Subject Headings (MeSH) selection for pharmacy practice journals. *Research in Social & Administrative Pharmacy: RSAP*, 20(9), 911–917. <https://doi.org/10.1016/j.sapharm.2024.06.003>

Fugmann, R. (1993). *Subject Analysis and Indexing: Theoretical Foundation and Practical Advice*. Indeks-Verlag.

Fung, K.W. et al. (2024). Promoting interoperability between SNOMED CT and ICD-11: lessons learned from the pilot project mapping between SNOMED CT and the ICD-11 Foundation. *Journal of the American Medical Informatics Association*, 31(8), 1631-1637. <https://doi.org/10.1093/jamia/ocae143>

Garrett, L. (1997). *Authority control and its influence on recall and precision in an online bibliographic catalog*. University of Hawaii. Tese de Doutoramento. (acesso através da base de dados ProQuest Dissertations & Theses)

Gil Leiva, I. (1999). *La automatización de la indización de documentos*. EdicionesTrea.

Gil Leiva, I. (2008). *Manual de indización: teoría y práctica*. EdicionesTrea.

Gnoli, C. (2020). *Introduction to knowledge organization*. Facet Publishing.

Golub, K. (2014). *Subject access to information: An interdisciplinary approach*. Libraries Unlimited.

Golub, K. (2019). Automatic Subject Indexing of Text. *Knowledge Organization*, 46(2), 104–121.

- Golub, K. (2021). Automated Subject Indexing: An Overview. *Cataloging & Classification Quarterly*, 59(8), 702-719. <https://doi.org/10.1080/01639374.2021.2012311>
- González-Márquez, R. et al. (2024). The landscape of biomedical research. *Patterns*, 5, 100968. <https://doi.org/10.1016/j.patter.2024.100968>
- Hersh, W. R. (2009). *Information retrieval: a health and biomedical perspective*. Springer.
- Hider, P. (2018). *Information resource description : creating and managing metadata* (2nd. ed.). Facet Publishing.
- Hjørland, B. (1997). *Information seeking and subject representation: an activity-theoretical approach to information science*. Greenwood Press.
- Hjørland, B. (2008). What is Knowledge Organization (KO)? *Knowledge Organization*, 35(2/3), 86-101.
- Hjørland, B. (2016). Knowledge organization. *Knowledge Organization*, 43(6), 475-84. Também disponível in Hjørland, Birger, ed. ISKO Encyclopedia of Knowledge Organization, https://www.isko.org/cyclo/knowledge_organization
- Hjørland, B. (2018). Indexing: concepts and theory. *Knowledge Organization*, 45(7), 609-39. Também disponível em: ISKO Encyclopedia of Knowledge Organization, ed. Birger Hjørland, coed. Claudio Gnoli, <https://www.isko.org/cyclo/indexing>
- Hodge, G. (2000). *Systems of knowledge organization for digital libraries: beyond traditional authorities files*. CLIR. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf>. Acesso em: 30 junho 2025.
- Howard, F. M., Li, A., Riffon, M. F., Garrett-Mayer, E., & Pearson, A. T. (2024). Characterizing the Increase in Artificial Intelligence Content Detection in Oncology Scientific Abstracts From 2021 to 2023. *JCO Clinical Cancer Informatics*, 8, e2400077. <https://doi.org/10.1200/CCI.24.00077>
- International Organization for Standardization. (2013). *Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*. (ISO standard No. 25964-2:2013). <https://www.iso.org/standard/53658.html>
- Iyer, H. (2012). *Classificatory Structures: Concepts, Relations and Representation*. Ergon Verlag.

Kim H. J. et al. (2024). 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. *Journal of Medical Internet Research*, 26: e52001. <https://doi.org/10.2196/52001>

Kim, M. C., Nam, S., Wang, F., & Zhu, Y. (2020). Mapping scientific landscapes in UMLS research: A scientometric review. *Journal of the American Medical Informatics Association*, 27(10), 1612–1624. <https://doi.org/10.1093/jamia/ocaa107>

Koutsomitropoulos, D. A., & Andriopoulos, A. D. (2020). Automated MeSH indexing of biomedical literature using contextualized word representations. In Maglogiannis I., Iliadis L., & Pimenidis E. (Eds.), *IFIP Advances in Information and Communication Technology: Vol. 583 IFIP* (pp. 343–354). Springer. https://doi.org/10.1007/978-3-030-49161-1_29

Lancaster, F. W. (1969). MEDLARS: Report on the evaluation of its operating efficiency. *American Documentation*, 20, 119 –142.

Lancaster, F. W. (1986). *Vocabulary control for information retrieval*. Information Resources Press

Lancaster, F. W. (2003). *Indexing and Abstracting in Theory and Practice*. 3rd. ed. Facet Publishing.

Levin, G., Meyer, R., Guigue, P.-A., & Brezinov, Y. (2024). It takes one to know one-Machine learning for identifying OBGYN abstracts written by ChatGPT. *International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics*, 165(3), 1257–1260. <https://doi.org/10.1002/ijgo.15365>

Li, Z., Belkadi, S., Micheletti, N., Han, L., Shardlow, M., Nenadic, G., & IEEE Computer Soc. (2024). Investigating Large Language Models and Control Mechanisms to Improve Text Readability of Biomedical Abstracts. *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, 265–274. <https://doi.org/10.1109/ICHI61247.2024.00042>

Lippell, H. (Ed.) (2022). *Taxonomies: Practical Approaches to Developing and Managing Vocabularies for Digital Information*. Facet Publishing.

Lopes, I. L. (2002). Uso das linguagens controlada e natural em bases de dados: revisão da literatura. *Ciência da Informação*, 31(1), 41–52.

Mazzocchi, F. (2018). Knowledge organization system (KOS): an introductory critical account. *Knowledge Organization*, 45(1), 54-78. Também disponível em: ISKO Encyclopedia of Knowledge Organization, eds. Birger Hjørland and Claudio Gnoli, <https://www.isko.org/cyclo/kos>.

McClure L. W. (2013). When the librarian was the search engine: introduction to the special issue on new roles for health sciences librarians. *Journal of the Medical Library Association*, 101(4), 257–260. <https://doi.org/10.3163/1536-5050.101.4.006>

Miller, L., Bhattacharyya, D., Miller, V., & Bhattacharyya, M. (2023). Recent Trend in Artificial Intelligence-Assisted Biomedical Publishing: A Quantitative Bibliometric Analysis. *Cureus Journal of Medical Science*, 15(5) e39224. <https://doi.org/10.7759/cureus.39224>

Moore, D. A. Q., Yaqub, O., & Sampat, B. N. (2024). Manual versus machine: How accurately does the Medical Text Indexer (MTI) classify different document types into disease areas? *PloS One*, 19(3), e0297526. <https://doi.org/10.1371/journal.pone.0297526>

Myszewski, J., Klossowski, E., & Schroeder, K. (2022). Evaluation of language analysis to summarize the literature: A comparison to traditional meta-analysis in primary hip and knee surgery. *Regional Anesthesia and Pain Medicine*, 47(3), 151–154. <https://doi.org/10.1136/rapm-2021-103261>

Pestana, O. (2014). Information value and quality for the health sector: a case study of search strategies for optimal information retrieval. In G. L. Jamil, A. M. Silva, & F. Ribeiro (Eds.), *Rethinking the Conceptual Base for New Practical Applications in Information Value and Quality* (pp. 116-133). IGI Global. <http://doi.org/10.4018/978-1-4666-4562-2.ch006>

Pestana, O. (2015). Alignment in medical sciences: towards improvement of UDC. In A. Slavic, & M. I. Cordeiro, *Classification & authority control: expanding resource discovery: proceedings of the International UDC Seminar 2015* (pp. 135-150). UDCC.

Pisciotta, Robert A. (2001) – Authority control in library catalogs. In THOMPSON, Laurie L. – *Bibliographic management of information resources in health sciences libraries* (pp. 115-149). Medical Library Association and The Scarecrow Press.

Pita Costa, J., Rei, L., Stopar, L., Fuat, F., Grobelnik, M., Mladenčić, D., Novalija, I., Staines, A., Pääkkönen, J., Konttila, J., Bidaurrezaga, J., Belar, O., Henderson, C., Epelde, G., Gabaráin, M. A., Carlin, P., & Wallace, J. (2021). NewsMeSH: A new classifier designed to annotate health news with MeSH headings. *Artificial Intelligence in Medicine*, 114, 102053. <https://doi.org/10.1016/j.artmed.2021.102053>

Rae, A. R., Pritchard, D. O., Mork, J. G., & Demner-Fushman, D. (2020). Automatic MeSH Indexing: Revisiting the Subheading Attachment Problem. *AMIA Annual Symposium Proceedings, 2020*, 1031–1040.

Ramachandran, R., & Arutchelvan, K. (2022). ArRaNER: A novel named entity recognition model for biomedical literature documents. *Journal of Supercomputing*, 78(14), 16498–16511. <https://doi.org/10.1007/s11227-022-04527-y>

Ribeiro, F. (1996). *Indexação e controlo de autoridade em arquivos*. Câmara Municipal do Porto. Arquivo Histórico.

Ribeiro, F. (1996). Subject indexing and authority control in archives: The need for subject indexing in archives and for an indexing policy using controlled language. *Journal of the Society of Archivists*, 17(1), 27–54. <https://doi.org/10.1080/00379819609511787>

Rodríguez Bravo, B. (2011). *Apuntes sobre representación y organización de la información*. Ediciones Trea.

Rowley, J., & Hartley, R. (2008). *Organizing Knowledge: An Introduction to Managing Access to Information*. 4th ed. Ashgate.

Satija, M.P. (2004). *A Dictionary of Knowledge Organization*. Guru Nanak Dev University.

Scheible, R., Strecker, P., Yazijy, S., Thomczyk, F., Talpa, R., Puhl, A., & Boeker, M. (2022). A Multilingual Browser Platform for Medical Subject Headings. *Studies in Health Technology and Informatics*, 289, 384–387. <https://doi.org/10.3233/SHTI210939>

Simões, M. G. et al. (2000). *Do tratamento à organização da informação: reflexões sobre concepções, perspectivas e tendências*. Imprensa da Universidade de Coimbra. <https://doi.org/10.14195/978-989-26-1944-6>

Silva, A. M. (2006). *A informação: da compreensão do fenómeno e construção do objecto científico*. CETAC.com e Afrontamento.

Silva, A. M., & Paletta, F. C. (2019). *Ciência da informação: estudos de epistemologia e de ética*. CIC.Digital.

Silva, A. M., & Ribeiro, F. (2011). *Paradigmas serviços e mediações em ciência da informação*. Nectar.

- Smiraglia, R. P. (2014). *The Elements of Knowledge Organization*. Springer. <https://doi.org/10.1007/978-3-319-09357-4>
- Stock, W. G., & Stock, M. (2015). *Handbook of information science*. Walter de Gruyter.
- Szostak, R., Gnoli, C., & López-Huertas, M. (2016). *Interdisciplinary Knowledge Organization*. Springer.
- Tang, W., Klang, M., Medved, D., Nugues, P., & Nilsson, J. (2023). Automatic Annotation of PubMed Articles with MeSH Qualifiers. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2023*, 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340998>
- Taylor, A. G., & Joudrey, D. N. (2009). *The organization of information* (3rd ed.). Libraries Unlimited.
- Taylor, A. G., & Tillett, B. B. (2004). *Authority Control in Organizing and Accessing Information: Definition and International Experience*. Haworth Information Press.
- Turki, H., Dossou, B. F. P., Emezue, C. C., Owodunni, A. T., Hadj Taieb, M. A., Ben Aouicha, M., Ben Hassen, H., & Masmoudi, A. (2024). MeSH2Matrix: Combining MeSH keywords and machine learning for biomedical relation classification based on PubMed. *Journal of Biomedical Semantics*, 15(1), 18. <https://doi.org/10.1186/s13326-024-00319-w>
- Williamson, N.; & Mcilwaine, I. C. (2009). UDC Medical Sciences Project: progress and problems. *Extensions and Corrections to the UDC*, 31, pp. 33-36.
- Wyman, L. P. (1999). *Indexing specialties: medicine*. Information Today.
- You, R., Liu, Y., Mamitsuka, H., & Zhu, S. (2021). BERTMeSH: deep contextual representation learning for large-scale high-performance MeSH indexing with full text. *Bioinformatics*, 37(5), 684–692. <https://doi.org/10.1093/bioinformatics/btaa837>
- Zeng, M. L., & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377-395. <https://doi.org/10.1002/asi.10387>
- Zeng, M. L., & Qin, J. (2022). *Metadata* (3rd ed.). Facet Publishing.

Zheng, L., He, Z., Wei, D., Keloth, V., Fan, J.-W., Lindemann, L., Zhu, X., Cimino, J. J., & Perl, Y. (2020). A review of auditing techniques for the Unified Medical Language System. *Journal of the American Medical Informatics Association*, 27(10), 1625–1638. <https://doi.org/10.1093/jamia/ocaa108>