

**ENANCIB 2022**

PORTO ALEGRE | UFRGS | PPGCIN

XXII Encontro Nacional de Pesquisa em Ciência da Informação • ENANCIB
Porto Alegre • 07 a 11 de novembro de 2022**XXII Encontro Nacional de Pesquisa em Ciência da Informação – XXII ENANCIB****ISSN 2177-3688****GT-4 – Gestão da Informação e do Conhecimento****TESAURO DE GESTÃO DO CONHECIMENTO: OS PRIMEIROS PASSOS EM SEU DESENVOLVIMENTO*****KNOWLEDGE MANAGEMENT THESAURUS: THE INITIAL STEPS*****Sandro Rautenberg.** UNICENTRO.**Marcos Roberto Kuhl.** UNICENTRO.**Zoraide da Fonseca Costa.** UNICENTRO.**Modalidade: Resumo Expandido**

Resumo: A organização e a representação de conhecimento são atividades interdisciplinares. Envolve o desenvolvimento de vocabulários controlados, como os tesouros, para promover a classificação e a comunicação de elementos informacionais em diversos domínios do conhecimento. Especificamente, este trabalho se insere na organização e na representação de elementos de um domínio interdisciplinar de interesse, a Gestão do Conhecimento. Deste modo, objetiva-se o desenvolvimento do Tesouro da Gestão do Conhecimento. No sentido a contribuir à discussão, apresenta-se as escolhas e os esforços empreendidos no desenvolvimento inicial do referido vocabulário controlado. Como uma pesquisa aplicada, um processo de desenvolvimento de ontologias é utilizado, ao considerar como premissas que ontologias e tesouros são sistemas de organização de conhecimento e, atualmente, são a base conceitual para utilização do *Simple Knowledge Organization System* – SKOS – na comunicação de conhecimento em ambientes da web. Como resultado parcial da pesquisa, especifica-se as fontes de conhecimento aderentes à aquisição de palavras-chave, aferindo o conjunto de termos, definições e relações conceituais relevantes do domínio Gestão do Conhecimento. Em face dos primeiros passos no desenvolvimento, como trabalhos futuros, almeja-se a finalização da implementação e a publicação do tesouro na Web de Dados para sua livre utilização por pesquisadores e profissionais da área.

Palavras-Chave: Tesouro. Gestão do Conhecimento. Organização do Conhecimento. Representação do Conhecimento. Vocabulários Controlados.

Abstract: Knowledge Organization and Representation are interdisciplinary activities. They involve the development of controlled vocabularies, such as thesauri, for promoting the classification and communication of informational units in different knowledge domains. Specifically, this work argues about the organization and representation of elements of an interdisciplinary domain of interest, Knowledge Management, aiming to develop the Knowledge Management Thesaurus. Contributing to the discussion, the choices and the first efforts undertaken for developing the referred controlled vocabulary are presented. As an applied research, an ontology development process is used, considering as the premises that: ontologies and thesauri are Knowledge Organization Systems and, currently, are the conceptual basis for using the Simple Knowledge Organization System in web environments. As a partial result, a set of knowledge sources are specified for acquiring relevant terms, definitions, and conceptual relationships in the Knowledge Management domain. Considering the first



steps in the development, as future work, we aim to implement and share the thesaurus on the Web of Data for free use.

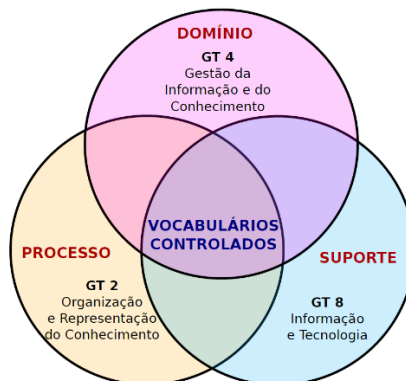
Keywords: Thesaurus. Knowledge Management. Knowledge Organization. Knowledge Representation. Controlled Vocabularies.

1 INTRODUÇÃO

É notória que a massificação da utilização da web como meio de socialização da informação trouxe, como consequência, a possibilidade da construção interdisciplinar de novos conhecimentos em diversas atividades humanas. Isso sustenta o surgimento e a sedimentação de novos domínios conceituais, dentre os quais a Gestão do Conhecimento como um dos constructos voltados aos estudos para as atividades de criação, organização, representação, compartilhamento, aplicação e refinamento de conhecimento.

Dentre as atividades anteriormente relatadas, no âmbito interdisciplinar da Gestão do Conhecimento e da Ciência da Informação, destaca-se que a organização e a representação de conhecimento têm suas fronteiras expandidas pelos avanços conceituais e tecnológicos em padrões, modelos e práticas (W3C, 2022). Conforme ilustrado na Figura 1, na Ciência da Informação, isso enseja-se um campo fértil de pesquisa para melhorar a comunicação de conhecimentos com o desenvolvimento vocabulários controlados.

Figura 1 – Desenvolvimento de Vocabulários Controlados no contexto da Ciência da Informação.



Fonte: Dados da pesquisa.

Já em relação à Gestão do Conhecimento como o domínio de aplicação, pode-se aderir à premissa da utilização de vocabulários controlados para sistematizar e explorar recursos informacionais em pesquisas interdisciplinares.



Em face dessas assertivas, como uma pesquisa aplicada, este trabalho delimita-se à investigação de uma classe especializada de vocabulários controlados, abordando o desenvolvimento de um tesauro para o domínio da Gestão do Conhecimento.

Para versar sobre os esforços iniciais realizados em direção ao desenvolvimento do referido tesauro, além dessa seção introdutória, este resumo compreende: **i)** a fundamentação teórica, abordando os conceitos Gestão do Conhecimento, vocabulários controlados e tesouros; **ii)** o procedimento metodológico que ampara o desenvolvimento de tesouros; **iii)** os resultados parciais alcançados no desenrolar do percurso metodológico; e **vi)** as considerações parciais, com o direcionamento de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este resumo é circunscrito ao desenvolvimento de um tipo especializado de vocabulário controlado, um tesauro para o domínio denominado Gestão do Conhecimento.

Academicamente, a Gestão do Conhecimento é um campo multidisciplinar com contribuições e visões de diferentes autores ou escolas. Por isso, não há um consenso sobre suas abordagens, mas a constatação de sua importância para a perenidade das organizações (LLORIA, 2008). Em uma visão preliminar, baseando-se nas perspectivas de Maier (2007) e Nissen (2006), entende-se que a Gestão do Conhecimento como a gestão formal de elementos de conhecimento (por exemplo: conceitos, regras, modelos, lições aprendidas, objetos de aprendizagem, repositórios, entre outros) que pode ser mediada por tecnologias da informação para conectar pessoas a pessoas ou pessoas a conteúdo, de modo a auxiliar o indivíduo nas atividades organizacionais. Diante essa perspectiva, as pessoas, as atividades organizacionais, os elementos de conhecimento e as tecnologias da informação e comunicação são conceitos importantes de discurso que necessitam ser mais bem compreendidos para sustentar as práticas da Gestão do Conhecimento. Assim, uma das formas de contribuir à Gestão do Conhecimento é o desenvolvimento de vocabulários controlados para organizar e representar os elementos de conhecimento, suportando a comunicação desses elementos entre os atores no referido domínio.

Segundo a NISO (2010, p. 5), um vocabulário controlado constitui uma representação de um conjunto finito de termos em consonância ao entendimento conceitual e limite de um domínio de interesse. Ressalta-se que, para cada termo presente no domínio em investigação,



deve-se ter um entendimento inequívoco e não redundante para: **i)** evitar a ambiguidade; **ii)** controlar a sinonímia existente; **iii)** estabelecer as relações entre os termos; e **iv)** realizar as validações do vocabulário perante o domínio de conhecimento. Em sua essência, os vocabulários controlados listam e organizam conceitos de um domínio, ao associar termos significantes e semântica aos termos representados. Dito isso, os vocabulários controlados servem a cinco propósitos (NISO, 2010, p. 11):

- **Tradução.** Fornece um meio para sistematizar a linguagem natural de um domínio de conhecimento em um vocabulário adequado para indexação e recuperação de recursos em sistemas de informação.
- **Consistência.** Promove a uniformidade no formato e na atribuição dos termos ao organizar, classificar e representar o conhecimento de um domínio.
- **Indicação de associações.** Indica as relações semânticas (hierárquicas ou transitivas) entre os termos.
- **Rotulagem e Navegação.** Auxilia os usuários a localizar os recursos informacionais desejados, ao fornecer uma organização adequada em um sistema de navegação.
- **Recuperação.** Computacionalmente, auxilia na localização de recursos informacionais em bases de dados.

Cabe ressaltar que a concepção de um vocabulário controlado envolve níveis de distintos de abstração e complexidade quanto ao propósito de utilização e à expressividade para o entendimento dos termos circunscritos ao domínio (Figura 2).

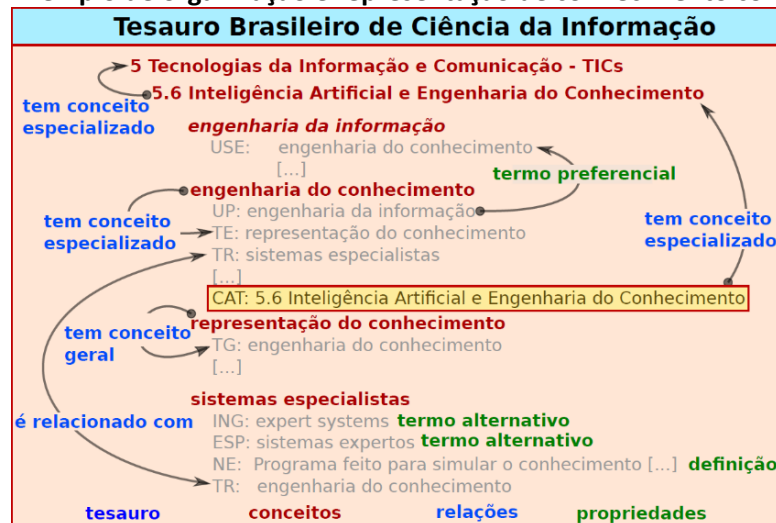
Figura 2 - Complexidade das classes de vocabulários controlados.



Fonte: Traduzida de NISO (2010, p. 17).

De acordo a Figura 2, a classe de vocabulário controlado de maior expressividade conceitual é denominada tesouro. Pontualmente, um tesouro é um sistema especializado de organização de conhecimento (PINHEIRO; FERREZ, 2014), podendo ser percebido como uma “lista de termos em linguagem natural, normalizados, preferenciais e organizados de modo conceitual, de acordo com regras terminológicas próprias e ligados entre si por relações hierárquicas ou semânticas” (CUNHA; CAVALCANTI 2008, p. 362).

Figura 3 – Exemplo de organização e representação de conhecimento com tesouros.



Fonte: Extraído de PINHEIRO e FERREZ (2014).

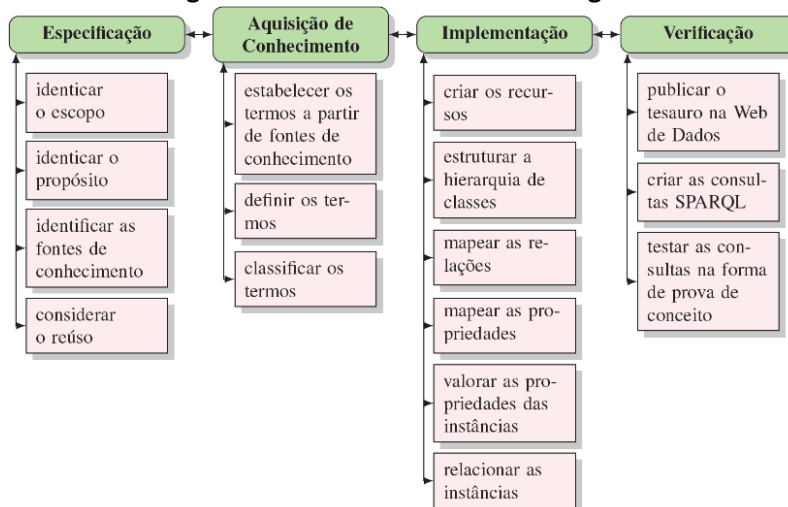
Para melhor entendimento sobre a estruturação de um tesauro, a Figura 3 exemplifica como este tipo especializado de vocabulário controlado é utilizado para organizar e representar elementos conceituais em um domínio particular. No exemplo, alguns termos do Tesauro Brasileiro de Ciência da Informação são apresentados em linguagem natural, destacando a definição e os termos preferenciais e alternativos de alguns conceitos (em verde). Para manter uma organização coesa, as relações entre os conceitos são ilustradas em azul, expressando uma representação quanto ao: **i)** entendimento conceitual especializado em direção ao mais geral (a relação de Termo Geral – TG – rotulada como: tem conceito geral); **ii)** entendimento geral em direção ao mais específico (semanticamente, a relação de Termo Específico – TE - denominada: tem conceito especializado); ou **iii)** a transitividade intraconceitos (com a relação Termo Relacionado – TR – expressa por: é relacionado com).

Cabe ressaltar ainda que as relações suportadas caracterizam os tesouros como sistemas de organização de conhecimento dinâmicos, com a capacidade de adaptar sua estrutura e seus elementos frente às mudanças e ressignificados no domínio de conhecimento em representação (STOCK; STOCK, 2015, p. 691). Essa característica é um diferencial na organização e na representação do conhecimento circunscrito ao domínio da Gestão do Conhecimento, uma vez que o referido domínio tem evoluído sua base conceitual em virtude das contribuições de suas diferentes escolas e abordagens. No sentido de desenvolver o Tesauro da Gestão do Conhecimento, na próxima seção é apresentado o procedimento metodológico que sustenta tal desenvolvimento.

3 PROCEDIMENTO METODOLÓGICO

De modo geral, organizar e representar conhecimento mediante tesouros são atividades próximas ao desenvolvimento de ontologias por oportunizar o inter-relacionamento semântico de conceitos de um domínio (MACULAN; AGANETTE, 2017; MOREIRA, 2019). Por isso, neste trabalho, inspira-se no processo iterativo de desenvolvimento de ontologias, adequando as macroatividades (Figura 4):

Figura 4 – Procedimento metodológico.



Fonte: Extraído de RAUTENBERG, SOUZA e KELNIAR (2021).

- **Especificação.** Discerne-se sobre os custos do desenvolvimento do tesouro ao: **a)** identificar o propósito - verificar em que ambiente o tesouro se insere, definir as circunstâncias do porquê deve ser desenvolvido, entre outros; **b)** identificar o escopo – observar os requisitos de “quem são os usuários”, “quais são as intenções de uso”, entre outros; **c)** identificar as fontes de conhecimento - procurar por livros, artigos, dicionários, entre outras fontes, dos quais pode-se abstrair termos e conceitualizações; e **d)** considerar o reúso - verificar a existência de ontologias ou vocabulários já estabelecidos para reutilizar conceitos.
- **Aquisição do conhecimento.** Identifica-se os elementos de conhecimento aderentes ao contexto do tesouro. São consideradas as tarefas de: **a)** estabelecer os termos a partir das fontes de conhecimento - listar termos comumente utilizados no domínio, agregando os elementos interdisciplinares previamente estabelecidos em outros domínios; **b)** classificar os termos - verificar na lista de termos se o entendimento de



determinado recurso remete: a um conceito geral do domínio; a uma associação entre os conceitos do domínio; ou a um exemplo de um conceito do domínio; e **c)** definir os termos – atribuir uma nota explicativa aos termos selecionados.

- **Implementação.** Utiliza-se as ferramentas computacionais para: **a)** criar os recursos – a partir da lista de termos, os itens conceituais são criados no arquivo inerente ao tesauro; **b)** estruturar a hierarquia conceitual – a partir dos itens criados, numa visão *top-down*, os conceitos gerais são vinculados aos conceitos semanticamente específicos; **c)** mapear as relações - para cada conceito, agregam-se os itens conceituais que evidenciam explicitamente o inter-relacionamento entre os conceitos do domínio; **d)** mapear as propriedades - para cada classe, agregam-se os itens são percebidos explicitamente como uma característica do conceito em questão (por exemplo, a tradução do termo para outros idiomas); **e)** valorar as propriedades - definir os valores a cada propriedade identificada (por exemplo, o rótulo, os nomes alternativos e a definição de um conceito); e **f)** relacionar os conceitos – inter-relacionar as instâncias conceituais do tesauro.
- **Verificação.** Trata-se de uma atividade para averiguar se o tesauro atende aos requisitos estabelecidos na atividade “1. Especificação”, sendo que se necessita: **a)** publicar o tesauro na Web de Dados – para privilegiar a execução da prova de conceito pelos especialistas, o tesauro pode ser compartilhado na web; **b)** criar as consultas SPARQL - desenvolver consultas em linguagem SPARQL para recuperar os recursos codificados no tesauro em consonância aos requisitos previamente estabelecidos; e **c)** testar as consultas na forma de prova de conceito – averiguar se as consultas SPARQL retornam os elementos requisitados.

4 ESTÁGIO ATUAL DO DESENVOLVIMENTO

Em consonância ao procedimento metodológico, o desenvolvimento do Tesauro de Gestão do Conhecimento encontra-se na atividade aquisição do conhecimento, mais precisamente na tarefa de estabelecer os termos a partir das fontes de conhecimento. Neste sentido, o Quadro 1 lista as fontes de conhecimento identificadas previamente para o estudo. Cabe ressaltar que das referidas fontes é extraído um conjunto de palavras-chave no idioma



inglês que apontam aos termos pertinentes ao tesouro considerando o domínio interdisciplinar da Gestão do Conhecimento.

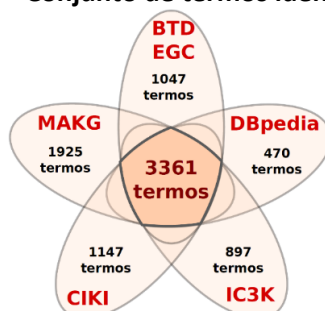
Quadro 1 – Identificação das fontes de conhecimento.

Fonte	Descrição
MAKG	Descrição: acrônimo de <i>Microsoft Academic Knowledge Graph</i> , é uma base de conhecimento que contempla mais de oito bilhões de registros (triplos descritas no formato RDF - <i>Resource Description Framework</i>) com informações acerca de publicações científicas e entidades como: autores, instituições, periódicos, palavras-chave e áreas de conhecimento (FÄRBER, 2019)
	Origem: https://makg.org/
	Data de consulta: 23 abr. 2021
BTD EGC	Descrição: banco de teses e dissertações do EGC do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento. Disponibiliza cerca de 280 teses e 312 dissertações nas linhas de pesquisa Engenharia do Conhecimento, Gestão do Conhecimento e Mídia do Conhecimento (EGC, 2022).
	Origem: http://btd.egc.ufsc.br/
	Data de consulta: 12 out. 2021
DBpedia	Descrição: Base de conhecimento multilíngue extraída a partir dos dados estruturados da Wikipedia (BIZER <i>et al.</i> , 2009).
	Origem: https://dbpedia.org/sparql
	Data de consulta: 27 abr. 2021
I3CK	Descrição: anais das edições da <i>International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management</i> .
	Origem: https://link.springer.com/
	Data de consulta: 1º mai. 2021
CIKI	Descrição: anais das edições do Congresso Internacional de Conhecimento e Inovação
	Origem: https://proceeding.ciki.ufsc.br/index.php/ciki
	Data de consulta: 13 out. 2021

Fonte: Dados da pesquisa.

Das fontes de conhecimento listadas, cabe ressaltar a disponibilidade das bases de conhecimento MAKG e DBpedia, das quais futuramente pode-se abstrair outros elementos constituintes do tesouro com consultas customizadas na Web de Dados.

Figura 5 – Conjunto de termos identificados.



Fonte: Dados da pesquisa.



A Figura 5 ilustra o resultado do processo de eliciação do conjunto de termos pertinentes. Ressalta-se que durante o processo foram identificadas 5.486 palavras-chave, as quais expressam 3.361 termos exclusivos que podem compor uma versão estável do Tesouro de Gestão do Conhecimento.

5 CONSIDERAÇÕES PARCIAIS

Este resumo expressa o esforço inicial empreendido para desenvolver um tesouro. Ao utilizar a base teórica de Sistemas de Organização de Conhecimento, pretende-se desenvolver o Tesouro da Gestão do Conhecimento, abarcando conceitos sobre indivíduos, atividades organizacionais, os elementos de conhecimento e tecnologias da informação e comunicação.

No estágio atual do desenvolvimento, foram identificadas cinco bases de conhecimento que contribuem na definição do conjunto de termos (palavras-chave) que pode constituir a estrutura do referido tesouro. Neste sentido, foram elicitados 3.361 termos os quais necessitam ser organizados e representados mediante um Sistema de Organização do Conhecimento. Em face disso, como avanços futuros, são traçadas como atividades:

- a evolução dos termos quanto à atribuição de definições multilíngues, à classificação hierárquica e o estabelecimento das relações semânticas intratermos em consonância aos preceitos do desenvolvimento de tesouros e ao domínio Gestão do Conhecimento; e
- a sistematização dos conceitos de acordo as tecnologias atuais da Web Semântica, pontualmente, utilizando a recomendação do *World Wide Web Consortium* para Sistemas de Organização de Conhecimento denominado SKOS (W3C, 2022), com vista ao compartilhamento e utilização do tesouro em sistemas da web.

AGRADECIMENTOS

À Fundação Araucária pelo suporte ao projeto de pesquisa “Desenvolvimento do Tesouro Brasileiro de Gestão do Conhecimento” (FA - Convênio 301/2022FA).

REFERÊNCIAS

BIZER, C.; LEHMANN, J.; KOBILAROV, G.; AUER, S.; BECKER, C.; CYGANIAK, R.; HELLMANN, S. DBpedia - A crystallization point for the Web of Data. **Web Semantics: Science, Services and Agents on the World Wide Web**, Amsterdam, v. 7, p. 154 – 165, 2009.



CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

EGC. **Banco de Teses e Dissertações do EGC**. Disponível em: <http://btd.egc.ufsc.br>. Acessado em: 11 mai. 2022.

FÄRBER, M. The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data. In: International Semantic Web Conference – ISWC, 18., 2019, Auckland. **Anais [...]**. Auckland: Springer, 2017.

ISKE, P.; BOERSMA, W. Connected brains: Question and answer systems for knowledge sharing: concepts, implementation and return on investment. **Journal of Knowledge Management**, Bingley, v. 9, n. 1, p. 126-145, 2005.

LLORIA, M. B. A review of the main approaches to knowledge management. **Knowledge Management Research & Practice**, Heidelberg, v. 6, n. 1, p. 77-89, 2008.

MACULAN, B. C. M. dos S.; AGANETTE, E. C. Definições: Convergência na Construção de Tesouros e Ontologias. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB, 18., 2017, Marília. **Anais [...]**. Marília: UNESP, 2017.

MAIER, R. **Knowledge Management Systems Information and Communication Technologies for Knowledge Management**, 3° ed, Springer, 2007.

MOREIRA, W. Tesouros e ontologias como modelos de sistemas de organização do conhecimento. **Brazilian Journal of Information Studies: Research Trends**, Marília, v. 13, n. 1, p.15-20, 2019.

NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19-2005 (R2010): guidelines for the construction, format, and management of monolingual thesauri**. American National Standards Institute, 2010.

NISSEN, M. E. **Harnessing knowledge dynamics**. Idea Group Inc., 2006.

PINHEIRO, L. V. R.; FERREZ H, D. **Tesouro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), 2014. E-book. Disponível em: https://ibict.br/images/internas/TESAURO-COMPLETO-FINAL-COM-CAPA-_24102014.pdf. Acesso em: 17 mai. 2021.

RAUTENBERG, S.; SOUZA, L. de; KELNIAR, J. P. Uma proposta de tesouro da Ciência da Computação: organização de elementos de conhecimento com SKOS. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 26, n. 1, p. 1-15, 2021.

STOCK, W. G.; STOCK, M. Handbook of Information Science. Boston: De Gruyter Saur, 2015. W3C – WORLD WIDE WEB CONSORTIUM. **SKOS Simple Knowledge Organization System Reference**. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818> Acesso em: 10 mai. 2022.