



## GT-4 – Gestão da Informação e do Conhecimento

ISSN 2177-3688

### DESENVOLVENDO UM TESAURO PARA GESTÃO DO CONHECIMENTO: UMA IMPLEMENTAÇÃO EM ANDAMENTO

### *DEVELOPING A KNOWLEDGE MANAGEMENT THESAURUS: AN IMPLEMENTATION IN PROGRESS*

**Sandro Rautenberg** - Universidade Estadual do Centro-Oeste (UNICENTRO)

#### **Modalidade: Resumo Expandido**

**Resumo:** padrões, modelos e práticas da Web Semântica são empregados em sistemas de organização do conhecimento. Tal fato inspira um desenvolvimento em curso denominado Tesouro Brasileiro de Gestão do Conhecimento. Como uma pesquisa aplicada, este desenvolvimento baseia-se num processo dividido nas atividades: especificação, aquisição de conhecimento, implementação e verificação. Contribuindo à discussão, apresenta-se as escolhas realizadas na atividade de implementação, as quais conduziram a definição dos conjuntos iniciais para: termos, notas explicativas, categorias gerais e relações do tesouro. Como trabalho futuro, aponta-se à versão estável do tesouro, compartilhando-a aos pesquisadores/profissionais com interesses no domínio da Gestão do Conhecimento.

**Palavras-chave:** tesouro; gestão do conhecimento; web semântica; sistemas de organização do conhecimento; vocabulários controlados.

**Abstract:** standards, models and practices of the Semantic Web have been used in knowledge organization systems. This inspires an ongoing development named Brazilian Knowledge Management Thesaurus. As applied research, the employed development process is based on the activities of specification, knowledge acquisition, implementation, and verification. Contributing to the discussion, we present the taken choices made in the implementation activity, which conducted the definition of initial sets of terms, explanatory notes, general categories and relations of the thesaurus. As future work, a stable version of the thesaurus is pointed out, sharing it with researchers/professionals of the Knowledge Management domain.

**Keywords:** thesaurus; knowledge management; semantic web; knowledge organization system; controlled vocabularies.

## **1 INTRODUÇÃO**

Como meio de socialização da informação, a utilização da web potencializa o desenvolvimento de novos conhecimentos. Isso sustenta o surgimento e a sedimentação de novos domínios conceituais, dentre os quais aponta-se para a Gestão do Conhecimento (GC) como o domínio interdisciplinar de pesquisa e aplicação de tecnologias voltado para criação,

organização, formalização, compartilhamento, aplicação e refinamento do conhecimento em organizações. Neste enredo, destaca-se o alinhamento dos preceitos da Web Semântica e da GC para a digitalização, a organização e a representação de elementos de conhecimento. À luz da Ciência da Informação, esse alinhamento também é uma oportunidade para conduzir novos estudos e aplicações, como por exemplo, o desenvolvimento de vocabulários controlados.

Em face das assertivas anteriores, uma pesquisa aplicada em curso com o objetivo do desenvolvimento do Tesouro Brasileiro de Gestão do Conhecimento é proposta (RAUTENBERG; KUHL; COSTA, 2022). Na gênese da referida pesquisa, como resultados parciais são apontados: **i)** a definição das fontes de conhecimento; e **ii)** o levantamento de palavras-chave do domínio da GC candidatas a evoluir aos termos e relações conceituais relevantes do referido tesouro. Considerando a continuidade da pesquisa, este resumo sintetiza os resultados recentes, os quais são alcançados na atividade em curso de implementação do tesouro. Por ora, os resultados revelam as opções sedimentares para: **i)** a definição das categorias gerais do tesouro; e **ii)** o uso de padrões e práticas da Web Semântica na organização e na representação dos recursos digitais, ao utilizar o padrão *Simple Knowledge Organization System* – SKOS (W3C, 2009) e o metadados *Dublin Core* (DCMI, 2020).

Para dissertar sobre o estágio atual do desenvolvimento, além dessa seção introdutória, este resumo compreende: **i)** a fundamentação teórica, abordando os conceitos GC e tesouros; **ii)** o procedimento metodológico que ampara a pesquisa aplicada; **iii)** a discussão dos resultados por ora alcançados; e **iv)** as considerações parciais, com o direcionamento aos trabalhos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este resumo é circunscrito ao desenvolvimento de um tipo especializado de vocabulário controlado para o domínio da GC. Em face disso, a seguir, algumas assertivas são pontuadas para versar sobre GC, tesouro e SKOS, acenando ao entendimento do projeto denominado Tesouro Brasileiro de Gestão do Conhecimento.

Academicamente, a GC é um domínio multidisciplinar com contribuições e visões de diferentes correntes e pesquisadores. Isso revela a dificuldade em estabelecer uma definição unicamente aceita. Sendo assim, para este trabalho, inspira-se nas perspectivas à GC de Maier (2007) e Nissen (2006). Complementarmente, os referidos autores pontuam a GC como a gestão formal de elementos de conhecimento (termos, regras, modelos, lições aprendidas, objetos de aprendizagem, repositórios, entre outros), mediada por tecnologias para conectar pessoas a

conteúdo e, por conseguinte, auxiliar o indivíduo em suas atividades organizacionais. Diante essa perspectiva, as pessoas, as atividades organizacionais, os elementos de conhecimento e as tecnologias são conceitos primordiais que necessitam ser melhor definidos, correlacionados e compreendidos por pesquisadores e profissionais para sustentar as práticas de GC. Nesta linha, uma ação afirmativa ao entendimento inequívoco dos conceitos da GC pode endereçar o desenvolvimento e a adoção de vocabulários controlados para suportar a comunicação interdisciplinar entre os atores. Neste sentido, destaca-se a utilização de tesouros.

Os tesouros são caracterizados como uma classe especializada de vocabulário controlado para sistematizar e expressar os elementos de conhecimento de um domínio. Conceitualmente, um tesouro é 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). Neste sentido, Austin e Dale (1993) pontuam que, estruturalmente, os tesouros organizam os termos de interesse mediante relações associativas, hierárquicas e de equivalência, enriquecendo o entendimento conceitual mediante notas explicativas.

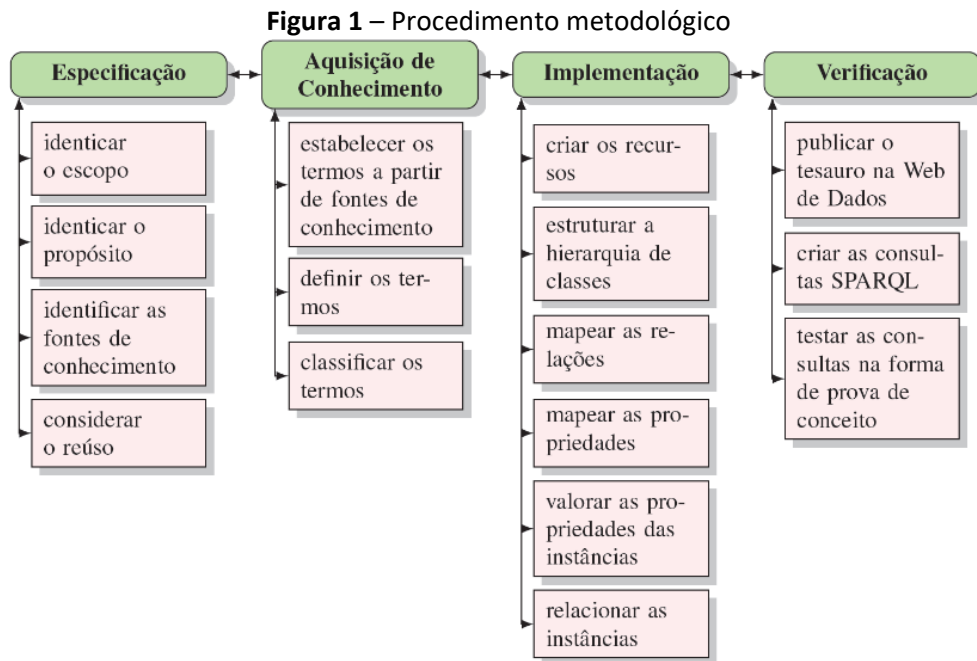
Alinhando os tesouros aos recentes avanços tecnológicos, principalmente para a web, cabe ressaltar que novos padrões e práticas de organização e representação de recursos digitais são propostos. Neste contexto, pode-se citar os padrões e práticas desenvolvidos à luz da Web Semântica para disponibilizar dados, informação e conhecimento em tecnologias computacionais. Em especial, destaca-se a utilização do SKOS para organizar, representar, evoluir e disponibilizar tesouros em ambientes digitais. Em síntese, o SKOS é um padrão referendado pelo *World Wide Web Consortium* que especifica modelos e práticas para manipular sistemas de organização do conhecimento (por exemplo, taxonomias e tesouros) no âmbito disponibilização de recursos digitais na web (W3C, 2009).

Considerando as ponderações anteriores sobre GC, tesouro e SKOS, a próxima seção disserta sobre o procedimento metodológico que sustenta o desenvolvimento do Tesouro da Gestão do Conhecimento.

### **3 PERCURSO METODOLÓGICO**

De modo geral, as atividades para organizar e representar conhecimento com tesouros encontram consonância no ciclo de desenvolvimento de ontologias, ao explicitar as relações semânticas de termos/conceitos aderentes em um domínio em investigação (MACULAN;

AGANETTE, 2017; MOREIRA, 2019). Por isso, neste trabalho, inspira-se no processo interativo de desenvolvimento de ontologias, adequando suas atividades gerais como segue (Figura 1):



Fonte: Extraído de Rautenberg, Souza e Kelnar (2021).

1. **Especificação.** Discerne-se sobre as oportunidades de construção de tesouros, principalmente, ao pontuar sobre os esforços/custos de desenvolvimento, o público-alvo, o propósito de utilização e as fontes de conhecimento disponíveis.
2. **Aquisição do conhecimento.** A partir das fontes de conhecimento, identifica-se o conjunto de termos aderentes ao domínio do tesouro, mediante as tarefas de triagem/classificação, entendimento e definição dos elementos percebidos.
3. **Implementação.** Utiliza-se as ferramentas computacionais para criar os recursos digitais que representam os termos do tesouro, estruturar a hierarquia conceitual e interrelacionar os termos representados. Adicionalmente, a cada termo, pode-se propor uma definição ou significado semântico.
4. **Verificação.** Trata-se de uma atividade para averiguar se os termos representados seguem o regramento dos requisitos pré-estabelecidos. Ao disponibilizar o tesouro em ambientes digitais, consultas semânticas são criadas e testadas na forma de prova de conceito para certificar a cobertura conceitual do tesouro desenvolvido.

Perante o procedimento metodológico, também cabe definir o suporte tecnológico adequado para o efetivo desenvolvimento. Neste trabalho, adota-se o ambiente Protégé

(MUSEN, 2015) para a edição de um arquivo digital referente ao tesouro. Pontualmente, com o Protégé, é possível criar os recursos semânticos correspondentes aos termos percebidos, estruturar a hierarquia dos termos, associar os intratermos, atribuir as definições aos termos e registrar os dados de proveniência com os metadados *Dublin Core* (DCMI, 2020).

#### 4 ESTÁGIO ATUAL E RESULTADOS PARCIAS DO DESENVOLVIMENTO

Considerando o percurso metodológico adotado, atualmente, realiza-se a atividade de implementação, executando iterativamente as tarefas: criar os recursos semânticos para os termos do tesouro; estruturar a hierarquia conceitual; relacionar os intratermos; atribuir as definições dos termos em consonância ao domínio da GC.

Em face disso, destaca-se a identificação e a edição em curso de 1.507 termos aderentes ao domínio da GC. Também foram previamente identificadas 3.102 relações intratermos (relação associativas - TR), 839 relações hierárquicas (relações todo e parte - TG ou TE) e 2.335 definições nos idiomas inglês e português (notas explicativas - NE).

**Quadro 1** – Termo Habilidades do Século 21 no Tesouro Brasileiro de Gestão do Conhecimento

##### **Habilidades do Século 21**

**ING** 21St Century Skills

**TE** Habilidades para o Aprendizado e a Inovação  
Habilidades para a Alfabetização Digital  
Habilidades para a Carreira e a Vida

**TR** Aprendizado  
Colaboração  
Formação Profissional  
Inovação  
Trabalho em Equipe

**NE** As Habilidades do Século 21 compreendem competências, habilidades e disposição à aprendizagem, identificadas como necessárias, considerando os aspectos sociais e laborais do século 21. Foram definidas por educadores, líderes empresariais, acadêmicos e agências governamentais. Muitas dessas habilidades estão associadas ao aprendizado, raciocínio analítico, resolução de problemas complexos, colaboração e trabalho em equipe. As habilidades foram agrupadas em três dimensões: habilidades para o aprendizado e a inovação; habilidades para a alfabetização digital; e habilidades para a carreira e vida.

21st Century Skills comprise skills, abilities, and learning dispositions that have been identified as being required for success in 21st century society and workplaces by educators, business leaders, academics, and governmental agencies. Many of these skills are also associated with deeper learning, which is based on mastering skills such as analytic reasoning, complex problem solving, and Teamwork. The skills have been grouped into three main areas: Learning and Innovation skills, Digital Literacy skills, and Career and Life skills.

**CAT** 02. Indivíduos, Papéis e Competências

**Fonte:** dados da pesquisa (2022).

O Quadro 1 exemplifica um termo em edição do Tesouro Brasileiro de Gestão do Conhecimento. Destaca-se o termo “Habilidades do Século 21”, apontando: **i)** o rótulo correspondente no idioma inglês (ING); **ii)** os termos específicos inerentes ao referido termo (TE); **iii)** outros termos que expressam alguma associação semântica (TR); **iv)** as definições do termo nos idiomas português e inglês (NE); e por fim **v)** a classificação do termo, segundo as categorias gerais estabelecidas.

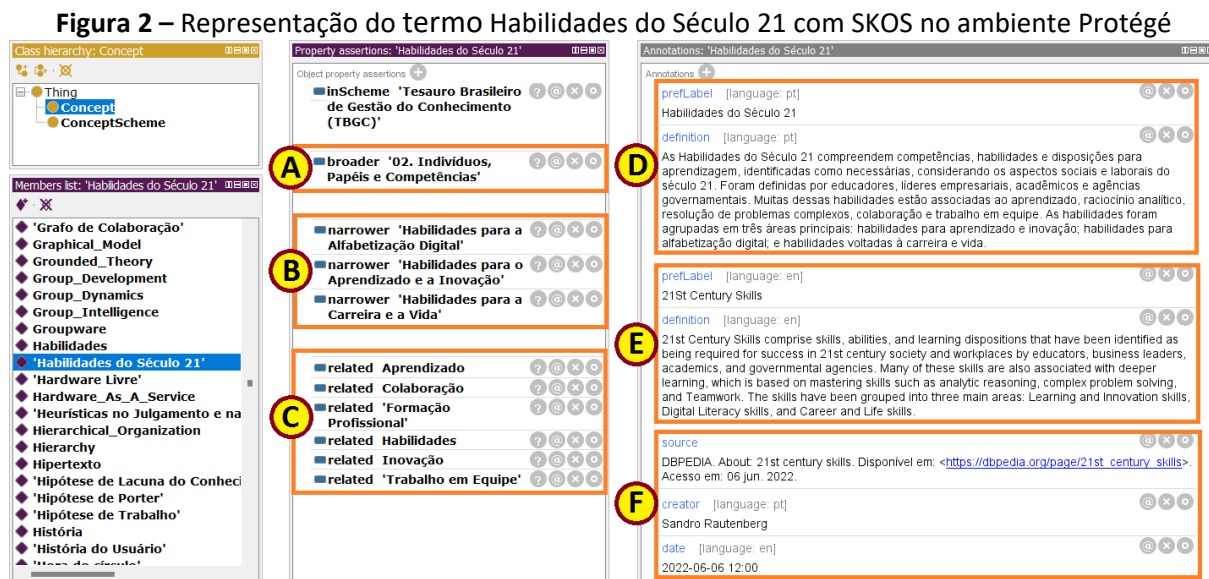
Em relação às categorias gerais, ressalta-se que estas são um resultado relevante adicional da atividade de implementação. Neste sentido, seis categorias são definidas:

- **01. Conceitos, Teorias e Modelos (*Concepts, Theories and Models*)**. Em consonância a sua origem, a GC é um domínio multidisciplinar e é influenciada por diversas bases conceituais. Como consequência, a tarefa de privilegiar definições, teorias e modelos em detrimento de outras torna-se complexa. Em face disso, opta-se por organizar e categorizar as contribuições interdisciplinares na dimensão “01. Conceitos, Teorias e Modelos”.
- **02. Indivíduos, Papéis e Competências (*People, Roles and Competences*)**. A GC deve manter e disponibilizar os elementos de conhecimento com vistas a sua utilização. Neste contexto, os indivíduos são sensibilizados por preceitos de aprendizagem e colaboração para desempenhar o papel fundamental nos processos de criação, refinamento e aplicação do conhecimento organizacional. Isso ocorre quando os indivíduos integram conteúdos (dados, informação e outros conhecimentos) mediante suas especialidades, competências e experiências. Tais assertivas sustentam categorizar alguns termos inerentes à GC em uma dimensão denominada “02. Indivíduos, Papéis e Competências”.
- **03. Processos, Atividades e Ações (*Processes, Activities and Actions*)**. A GC baseia-se em estratégias que consideram o conhecimento como recurso primordial no alcance dos objetivos organizacionais. Em outras palavras, a GC é envolta por processos e suas atividades que, como uma cadeia de valor, a partir dos dados (matéria-prima) evolui-se em direção ao conhecimento (produtos e/ou processos). Pontualmente, aquisição (de dados), triagem, indexação e organização (da informação), análise, síntese e raciocínio (com conhecimento) são alguns exemplos de atividades imprescindíveis nos processos de GC. Por isso, alguns os termos inerentes merecem ser explicitados na dimensão “Processos, Atividades e Ações”.

- **04. Sistemas, Ferramentas e Tecnologias (*Systems, Tools and Technologies*)**. Para contribuir ao sucesso organizacional, a GC tipicamente permeia por diversas tecnologias para preservar as bases de conhecimento institucionais, permitindo sua exploração nas atividades da organização. Neste sentido, alguns sistemas e ferramentas facilitam a instrução, a comunicação, o trabalho colaborativo e a construção de novos conhecimentos, por exemplo. Também conectam pessoas a pessoas, pessoas a conteúdos de conhecimento e auxiliam o indivíduo em suas atividades laborais. De acordo esta perspectiva, pontuar sobre os termos de “Sistemas, Ferramentas e Tecnologias” torna-se fundamental à GC.
- **05. Conteúdo, Linguagens e Estruturas (*Content, Languages and Structures*)**. Na GC, o conteúdo (os dados) deve ser rotulado, formalizado, estruturado, indexado, armazenado para, possivelmente, ser digitalizado, recuperado e manipulado, tornando-se parte da informação ou do conhecimento organizacional. Neste enredo, evoluir o conteúdo ao nível de insumo para o conhecimento organizacional ocorre com o emprego de linguagens e metadados de estruturação/representação. Devido à evolução das tecnologias computacionais, diversos metadados, linguagens e estruturas de representação voltados ao tratamento e ao processamento de conteúdos são propostos e são passíveis de serem utilizados em tecnologias voltadas à GC. Dito isto, é compreensível abarcar termos sobre “Conteúdo, Linguagens e Estruturas” em um tesouro voltado ao domínio da GC.
- **06. Organização e Cultura (*Organization and Culture*)**. A GC necessita pontuar os valores e comportamentos coletivos que circundam a construção e a preservação de bases de conhecimento em uma organização. Neste sentido, além dos indivíduos, processos, tecnologias e conteúdo, deve-se ponderar sobre outros fatores, como por exemplo, aqueles relativos à estrutura, à identidade, à cultura, ao regimento e ao clima de uma organização. Lidar preventivamente com os referidos fatores pode potencializar ou atenuar os fluxos informacionais, impactando sobremaneira nos processos de GC. Desta maneira, também é necessário considerar os termos de “Organização e Cultura” para a efetiva promoção da GC.

Por fim, como outro resultado importante alcançado, pontua-se sobre a utilização do padrão SKOS na implementação do tesouro. Neste sentido, com o referido padrão, é possível

organizar o tesouro e representar os termos inerentes de acordo os preceitos da Web Semântica. Em consonância ao exemplo expresso no Quadro 1, a Figura 2 ilustra como o termo “Habilidades do Século 21” tem sua representação orquestrada com o SKOS no ambiente Protégé, onde:



Fonte: dados da pesquisa (2023).

- (A) a propriedade *skos:broader* define que o termo em questão é subordinado à dimensão “02. Indivíduos, Papéis e Competência” do “Tesouro Brasileiro de Gestão do Conhecimento”;
- (B) a propriedade *skos:narrower* relaciona três termos derivados - “Habilidades para o Aprendizado Digital”, “Habilidades para o Aprendizado e a Inovação” e “Habilidades para a Carreira e a Vida”;
- (C) a propriedade *skos:related* vincula o termo “Habilidades do Século 21” a outros termos do tesouro, como por exemplo: “Aprendizado”, “Colaboração”, “Formação Profissional”, entre outros;
- (D) as propriedades *skos:prefLabel* e *skos:definition* apontam para o entendimento do termo, considerando o idioma português;
- (E) sob o mesmo prisma do item anterior, as propriedades *skos:prefLabel* e *skos:definition* associam o rótulo e a definição do termo no idioma inglês; e
- (F) as propriedades *dct:source*, *dct:creator* e *dct:date* são utilizadas para descrever a proveniência do termo “Habilidades do Século 21” como um recurso digital.

## 5 CONSIDERAÇÕES PARCIAIS

Este resumo disserta sobre as ações em curso para implementação do Tesouro Brasileiro de Gestão do Conhecimento. Como resultados parciais, destaca-se que a versão em construção do tesouro abarca 1.507 termos, organizados em seis categorias gerais: “01. Conceitos, Teorias e Modelos”, “02. Indivíduos, Papéis e Competências”, “03. Processos, Atividades e Ações”, “04. Sistemas, Ferramentas e Tecnologias”, “05. Conteúdo, Linguagens e Estruturas” e “06. Organização e Cultura”.

Salienta-se que, segundo o procedimento metodológico adotado, a implementação do tesouro é tecnologicamente suportada pelo ambiente Protégé para editar os recursos digitais aderentes ao domínio da GC. Vale destacar que o tesouro é editado baseando-se no padrão SKOS, a recomendação da *World Wide Web Consortium* para sistematizar vocabulários controlados.

Para a continuidade dos trabalhos, aponta-se como ações futuras:

- a evolução dos termos inerentes ao tesouro quanto à atribuição de definições multilíngues;
- o estabelecimento das relações semânticas intratermos percebidas a partir das definições atribuídas;
- a organização hierárquica dos termos representados; e
- a publicação de uma versão estável do tesouro na Web de Dados para utilização compartilhada por pesquisadores e sistemas web.

## AGRADECIMENTOS

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

## REFERÊNCIAS

AUSTIN, D.; DALE, P. **Diretrizes para o estabelecimento e desenvolvimento de tesouros monolíngües**. Brasília: IBICT/SENAI, 1993. E-book. Disponível em: [https://livroaberto.ibict.br/bitstream/1/731/7/Diretrizes\\_estabelecimento\\_tesouros.pdf](https://livroaberto.ibict.br/bitstream/1/731/7/Diretrizes_estabelecimento_tesouros.pdf). Acesso em: 3 jun. 2023.

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

DCMI. **DCMI Metadata Terms**. 2020. Disponível em: <http://dublincore.org/documents/dcmi-terms/>. Acesso em: 3 jun. 2023.

MACULAN, B. C. M. 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. Berlin: 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.

MUSEN, M. A. The Protégé project: A look back and a look forward. **AI Matters**. v. 1, n. 4, p. 4-12, 2015.

NISSEN, M. E. **Harnessing knowledge dynamics**. Hershey: IRM Press, 2006.

RAUTENBERG, S.; SOUZA, L.; 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.

RAUTENBERG, S.; KUHL, M. R.; COSTA, Z. F. Tesouro de gestão do conhecimento: os primeiros passos em seu desenvolvimento. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 22., Porto Alegre, RS, 2022. **Anais [...]** Porto Alegre, RS: UFRGS, 2022.

W3C. **SKOS Simple Knowledge Organization System Reference**. 2009. Disponível em: <https://www.w3.org/TR/2009/REC-skos-reference-20090818>. Acesso em: 3 jun. 2023.