

---

# Evolução teórica sobre tesouro: estudo diacrônico das publicações de Dagobert Soergel (1972-2017)

*Evolución teórica del pensamiento de Dagobert Soergel sobre los tesauros (1972-2017)*

*Theoretical evolution of thesauri: a diachronic study of Dagobert Soergel publications (1972-2017)*

---

**Mariângela Spotti Lopes FUJITA, Luciana Beatriz Piovezan RIO BRANCO,  
Roberta Caroline Vesú ALVES**

Departamento de Ciência da Informação da Faculdade de Filosofia e Ciências da UNESP, Campus de Marília, Universidade Estadual Paulista – UNESP, Brasil, mariangela.fujita@unesp.br, lbpiovezan@gmail.com, robertavesu@gmail.com

## Resumen

Estudio de la evolución de la teoría acerca de los tesauros en el pensamiento del investigador Dagobert Soergel. Se analizó el discurso de Soergel en artículos que publicó sobre tesauros mediante la metodología del análisis de contenido de Bardin en muestra diacrónica. En particular se analizaron las categorías de construcción, funciones, mantenimiento, tecnologías y usos de los tesauros. Los artículos analizados muestran que Soergel se preocupa de los avances científicos y metodológicos de los tesauros en diferentes entornos tecnológicos. El investigador es innovador y sigue las tendencias sobre construcción, funciones, mantenimiento, tecnologías y usos de los tesauros, que apuntan a su desarrollo estructurado y su uso en nuevos ambientes tecnológicos y con interoperabilidad.

**Palabras clave:** Soergel, Dagobert. Tesauros. Evolución. Teoría.

## 1. Introdução

Os tesauros possuem uma evolução influenciada pelo contexto de uso, pelas tecnologias, como também pela diversidade cultural que se expressa pela linguagem.

As pesquisas sobre tesouro necessitam de investigações sobre seus fundamentos teóricos e metodológicos, para observar evolução técnica ou teórica e verificar aspectos negligenciados. Estudos de revisão de literatura temáticos são imprescindíveis pela análise evolutiva que se obtém e destaque de autores mais influentes. Por outro lado, a análise evolutiva de uma temática pode ser investigada pela trajetória de autores influentes na revisão de literatura e, sobretudo, na própria evolução do tema em questão.

Aspectos que fazem a diferença na evolução de tesauros podem ser explicados ou antecipados em publicações de Soergel, como por exemplo,

## Abstract

Study on the evolution of thesaurus theory in Dagobert Soergel's discourse. The articles he published on the subject were analyzed using Bardin's content analysis with a diachronic sample. The following categories were considered: construction, functions, maintenance, technologies and uses of thesaurus in the researcher's discourse. The articles showed that Soergel is concerned with scientific and methodological improvement of thesauri in different technological environments. The researcher is innovative and follows the more innovative trends on the construction, functions, maintenance, technologies and uses of thesauri, which and points to their structured development and use in new interoperable technological environments.

**Keywords:** Soergel, Dagobert. Thesauri. Evolution. Theory.

relacionamento entre termos, software de construção e manutenção, o uso em ambientes tecnológicos e digitais, a interoperabilidade entre diferentes vocabulários, entre outros. Consideramos que as pesquisas e publicações do pesquisador Dagobert Soergel têm influência na elaboração de normas sobre tesauros, como por exemplo, na norma ANSI/NISO Z39.19-2005.

Dagobert Soergel nasceu em 7 de janeiro de 1940, em *Freiburg*, Alemanha, se formou em Física, em 1960, pela *University of Freiburg* e se tornou Mestre em Matemática e Física pela mesma universidade em 1964, depois se tornou Ph.D. em Ciências Políticas, também pela *University of Freiburg*. Atualmente é professor da área de Ciência da Informação no *Department of Library and Information Studies*, da *University of Buffalo*, em *New York, US*. Atua, também, como Professor Emérito em estudos da informação no *College of Information Studies*, na *University of*

Maryland. É membro da sociedade científica *International Society of Knowledge Organization (ISKO)* desde sua fundação.

Vale destacar que Soergel apresenta como interesse de pesquisa os temas: Armazenamento e recuperação de informação; Ontologias; Classificações; Taxonomias; Tesouros; e Design orientado para o usuário (Soergel, 2005).

A análise da evolução científica e metodológica de tesouros pela visão privilegiada de Dagobert Soergel é a proposta deste artigo com o objetivo de analisar o discurso do pesquisador em suas publicações. Para isso, foi realizada análise de conteúdo de Bardin em uma amostra diacrônica de suas publicações relacionadas à tesouros.

## 2. Tesouros

Os tesouros são instrumentos de controle de vocabulário com objetivo de representação de dupla função: explicitar o conteúdo de documentos na indexação e expressá-los em sistemas de informação na recuperação. Essa segunda função atua na expressão das necessidades de pesquisa de usuários para formulação de estratégias de busca (American National Standards Institute, 2010; International Organization for Standardization, 2011, 2013).

A norma ISO 25964-1 define o tesouro como um “vocabulário controlado e estruturado no qual conceitos são representados por termos, organizados para que as relações entre conceitos sejam explícitas e termos preferidos estejam acompanhados por entradas para sinônimos ou quase-sinônimos” (International Organization for Standardization, 2011, p. 12, tradução nossa).

Gollub (2011 citado por Mazzochi, 2018) considera que vocabulário controlado tem um significado mais restrito a qualquer conjunto de termos ou lista de termos controlados, enquanto que linguagens de indexação “são tipos específicos de vocabulários controlados que representam linguagens formalizadas projetadas e usadas para descrever o conteúdo do assunto dos documentos para fins de recuperação da informação”. O termo “vocabulário controlado” tem sido utilizado em normas internacionais para se referir a outros tipos de vocabulários que não o tesouro e que se relacionam entre si, seja para interoperabilidade semântica ou tecnológica.

Fujita et al. (2018) analisaram a importância do termo linguagem de indexação em Organização do Conhecimento e constataram que na Enciclopédia de Organização do Conhecimento da ISKO, o termo é considerado como parte da terminologia de organização do conhecimento (OC) porque é “Ainda empregado em um ambiente de

Biblioteconomia e Ciência da Informação/ Organização do Conhecimento” (Mazzochi, 2018).

Diferentes Sistemas de Organização do Conhecimento que “abrangem todos os tipos de esquemas que organizam e representam o conhecimento” (Carlan e Medeiros, 2011, p. 54), são utilizados em ambiente digital.

Segundo Berti Junior et al. (2017), exemplos desses Sistemas de Organização do Conhecimento tais como, vocabulários, classificações, taxonomias e tesouros, entre outros, são usados em serviços de informação de banco de dados e de bibliotecas digitais para garantir a correta interpretação semântica e interoperabilidade das terminologias e registros de informações.

Currás (2005) explica que em meados de 1940 as bibliotecas e centros de documentação especializados encontravam dificuldades de identificação e localização de informações, tornando-os obsoletos. A resposta para este problema foi a flexibilização da representação dos assuntos pelo surgimento em fins da década de 1950 do tesouro com sua estrutura sistemática. Neste período os tesouros experimentaram grande evolução teórica e metodológica e passaram a ser utilizados pela Ciência da Informação nos processos de indexação e de recuperação da informação, como o instrumento que reúne um conjunto controlado de termos de um determinado domínio. Dotado de relações sintáticas e semânticas entre termos, o tesouro permite que a linguagem utilizada nos documentos seja traduzida em uma linguagem mais estrita.

O controle de vocabulário exercido no tesouro colabora com a redução de ambiguidades, controle de sinônimos, relacionamentos entre termos, entre outros aspectos, para a recuperação precisa da informação. A organização dos termos é predominantemente por ordem alfabética e de relações hierárquicas, (Martins, 2013).

O tesouro, portanto, deve ter uma linguagem controlada, com base na linguagem dos documentos e na terminologia de uma área do conhecimento, como também se basear em necessidades de informação dos usuários, expressas em sua linguagem de busca no sistema de informação conforme seus perfis de interesse. Também, consiste no instrumento para organização do conhecimento que melhor se desenvolveu com relação à sua estrutura e funções e, por ser dotada de precisão e relevância, pode ser considerada a mais adequada para a representação e recuperação da informação.

A evolução dos tesouros acompanha as tendências teóricas para sua construção, também de acordo com tecnologias em face de cada tempo

e meio de recuperação da informação. Atualmente, a tendência é o uso dos tesouros para organização semântica e recuperação da informação com interoperabilidade entre tesouros, bem como com outros sistemas de organização do conhecimento.

Diante disso, serão destacados a seguir aspectos de uso, finalidade, construção e avaliação dos tesouros.

### 2.1. Usos e finalidades dos tesouros: aspectos teóricos e recursos tecnológicos

O tesouro tem a finalidade de representar de forma mais estrita e normalizada a linguagem dos documentos durante os processos de indexação e de pesquisa. Assim, se mantém a comunicação nos sistemas de unidades de informação para a recuperação de informação. Outra função é de tradução da linguagem natural dos textos para linguagem controlada, que exerce controle de vocabulário e evita dispersões semânticas e sintáticas. De acordo com Soler Monreal e Gil Leiva (2010, p. 310) o tesouro “[...] se fundamenta em terminologia consensual, normalizada e controlada”.

Além disso, orienta, tanto o indexador sobre os melhores termos para representar o assunto, quanto o pesquisador sobre qual termo escolher que corresponda à representação do assunto procurado. Tem como atribuição a organização de conceitos e termos de uma área do conhecimento para adequar a relação desses termos com a linguagem do usuário, o que contribui para comunicar ao indexador os critérios de busca do usuário (Dextre Clarke; Zeng, 2012; Soergel, 1975).

Para objetivos mais contemporâneos, os tesouros passam a interagir nos ambientes *web*, oferecendo maiores graus de precisão à organização do conhecimento em rede (Marcondes, 2011), e atuando de forma interoperável e intercambiável (Mendes; Reis; Maculan, 2015), o que exige mais qualidade na escolha e organização dos termos.

Em ambiente *web* de mecanismos de pesquisa, segundo García-Marco (2016a), o tesouro pode ser usado para melhorar a recuperação e se tornará cada vez mais relevante nesse meio global, devido à precisão que proporciona ao usuário. O tesouro garante a interoperabilidade e possibilita a alimentação de novos mecanismos semânticos da Internet, que está se tornando a Web Semântica. Também, a interconexão de sistemas de organização do conhecimento consiste em uma ótima abordagem para melhorar a recuperação na Internet (García-Marco, 2016d).

Conforme analisa García-Marco (2017), a interoperabilidade dos tesouros ocorre com outros sistemas de organização do conhecimento no contexto da Internet e Web Semântica. A interoperabilidade, segundo García-Marco (2017, p. 306, tradução nossa),

[...] surge na estrutura da rede semântica e no movimento de dados abertos, uma orientação que não é estranha, mas sim, junto ao trabalho íntegro realizado pelas equipes de desenvolvimento da ISO 25964 e do padrão W3C SKOS.

García-Marco (2017, p. 306, tradução nossa), se refere à os três modelos de mapeamento estrutural para interoperabilidade da norma ISO 25964-2,

- de unidade estrutural, na qual os conceitos e estruturas relacionais de ambos os sistemas são equivalentes, por exemplo, entre dois tesouros;
- *link* direto, no qual todos os vocabulários estão ligados;
- de *'hub'* ou eixo central, no qual os vocabulários são mapeados em ambas as direções em relação àquela que funciona como o núcleo do sistema de vocabulários estruturados.

Para a interoperabilidade de relacionamentos semânticos, segundo do mesmo autor, são importantes em três tipos de mapeamentos, os de equivalência, os hierárquicos e os associativos, entre outros mais específicos. Ainda, a interoperabilidade é recomendada pela norma ISO 25964-2 entre tesouros e os principais sistemas de organização do conhecimento, portanto, que pode ocorrer com: “classificações bibliográficas e de arquivo; taxonomias; sistemas de cabeçalho de assunto; ontologias; terminologias; listas de controle de autoridade; anéis de sinônimos” (García-Marco, 2017, p. 307, tradução nossa).

A norma ISO 25964-2 não normaliza os diferentes tipos de sistemas de organização do conhecimento, mas é um avanço na questão da interoperabilidade entre esses sistemas.

### 2.2. Construção de tesouros

Para a elaboração dos tesouros, entre outros aspectos, deve-se determinar o domínio ou campo do conhecimento. Conhecer o campo conceitual a ser representado, bem como estabelecer a terminologia e o sistema nocional que decorre da relação entre os termos e conceitos, pois são aspectos que resultam da metodologia de elaboração de tesouro (Cervantes, 2009).

O método indutivo para elaboração de tesouros permite obter a terminologia mediante a literatura corrente de uma área, elegendo os termos e conceitos de maior frequência ou ocorrência, que

também correspondam às perguntas dos usuários e após a coleta e organização dos termos é realizada a estruturação do domínio. O método dedutivo permite obter a terminologia por meio da consulta aos especialistas de um domínio, que fornecerão os termos para compor o tesouro, portanto, a escolha dos termos ocorre pela garantia dos produtores e utilizadores do sistema de informação, (Cervantes, 2009).

Em geral, as diretrizes normativas indicam a possibilidade de combinação empregando cada método em um estágio da construção (Cervantes, 2009). Assim, os termos podem ser revisados posteriormente para assegurar a consistência de sua escolha e do estabelecimento de relações.

As abordagens indutiva e dedutiva de construção de tesouros são denominadas por Garcia (2016c) como bottom-up ou ascendente e top-down ou descendente, respectivamente. Esta última oferece mais autonomia para a construção ao considerar necessidades de usuário e interoperabilidade em diferentes ambientes digitais (García-Marco, 2016c)

A elaboração de tesouros se relaciona com dois princípios para o levantamento de termos, a garantia literária e garantia de uso ou do usuário. A garantia literária assegura que os assuntos dos textos ou documentos de uma área sejam representados do modo mais preciso possível e a garantia de uso permite que a representação ocorra correspondendo ao que é importante para a recuperação da informação dos usuários (Barité et al., 2010).

Nesse contexto, o tesouro é formado, segundo Soergel (1975), por um conjunto de termos descritores e outros símbolos. Estes não fazem parte, mas organizam sua estrutura, relacionando o descritor apropriado a ser usado, por exemplo, para relacionar dois termos sinônimos. Ainda, apresenta como item obrigatório: “a lista de descritores, o vocabulário ou léxico do sistema, as relações entre os descritores (hierárquicas e associativas)” (Soergel, 1975, tradução nossa); e como itens opcionais, que pertencem à sintaxe do tesouro, a lista de indicadores como descritores relacionais, relacionamentos hierárquicos ou associativos entre indicadores de função ou entre descritores relacionais de função sintática (Soergel, 1975).

Além da seleção e normalização dos conceitos, são estabelecidas relações entre termos de tipo lógico-hierárquicas, gênero/espécie e todo/parte (entre conceitos) e relações semânticas (entre termos), de equivalência e associativas (entre conceitos) (Dextre Clarke, 2017).

As normas sobre tesouros, de modo geral, são derivadas das primeiras diretrizes de construção de tesouro, do *Committee On Scientific And Technical Information* (COSATI) em 1967 e da UNESCO em 1974 (Lancaster, 2002, p.51).

O processo de construção dos tesouros é guiado pelas normas internacionais, cuja evolução levou à atual ANSI/NISO Z39.19-2005 (American National Standards Institute, 2010) e, na sequência, as normas ISO 25964-1 “*Information and documentation. Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval*” (International Organization for Standardization, 2011) e ISO 25964-2 “*Information and documentation. Thesauri and interoperability with other vocabularies - part 2: Interoperability with other vocabularies*” (International Organization for Standardization, 2013).

Fujita, et al. (2016), com base em Martins (2013), informa que a norma ANSI/NISO Z39.19 (2005) é a quarta edição da norma ANSI Z39.19-1974, cujo fator determinante de mudança foi a aplicação de tecnologias de informação para desenvolvimento de tesouros eletrônicos em função dos avanços da informação eletrônica e da Internet.

Esta norma se baseou em diferentes autores, incluindo Soergel, que contribuiu principalmente com critérios de avaliação e manutenção de vocabulários controlados, porque demonstram o desenvolvimento do tesouro no meio digital, entre outros aspectos.

Garcia Marco (2016b, p.36) analisa que a quarta edição da norma Z39.19-2005 foi precursora em realizar três principais inovações determinantes para a evolução dos tesouros atuais: abarcar todo tipo de recursos de informação, dos tradicionais aos eletrônicos; ampliar seu alcance para todos os vocabulários controlados, listas de termos controlados, anéis de sinônimos, taxonomias e tesouros; e, abordar o problema de interoperabilidade de vocabulários na internet.

As evoluções seguintes foram desenvolvidas pela norma britânica BS8723, publicada em cinco partes (BSI, 2005, 2007, 2008). Os principais avanços incluem: a interoperabilidade linguística para tesouros multilíngues, a distinção entre termos e conceitos, o desenvolvimento de um modelo de dados, a análise em facetas, o tratamento dos sistemas de classificação como ferramentas auxiliares do tesouro, uma estratégia de interoperabilidade entre linguagens documentárias e a abordagem de outras linguagens (listas de cabeçalhos de assuntos) e alguns outros sistemas de organização do conhecimento utilizados na internet tais como ontologias, taxonomias e anéis de sinônimos (García Marco, 2016b)

Portanto, as normas ISO 25964-1 e ISO 25964-2 apresentam aspectos evolutivos importantes advindos, da norma ANSI/NISO Z39.19-2005, “*Guidelines for the construction, format, and management of monolingual controlled vocabularies*” e, principalmente, da Norma BS8723, “*Structured vocabularies for information retrieval*”. A parte 1 reúne o tratamento dos tesouros monolíngues e multilíngues e a parte 2 é mais dedicada à interoperabilidade com outros vocabulários em contexto de web semântica e de dados abertos.

### 2.3. Avaliação e manutenção de tesouros

De acordo com a norma ANSI/NISO Z39.19-2005 “[...] os desenvolvedores de vocabulário controlado devem avaliar se os metadados são relevantes para sua comunidade e para o domínio do conhecimento, isso antes de desenvolver novos conjuntos ou esquemas com elementos de metadados.” (American National Standards Institute, 2010, p. 3, tradução nossa).

As razões para testar e avaliar, conforme a norma ANSI/NISO Z39.19-2005, são para determinar se o tesouro descreve o conteúdo e apresenta resultados de pesquisa adequados para os usuários (relevância e revocação) (American National Standards Institute, 2010).

Os seguintes métodos podem ser usados para medir a qualidade ou a eficácia do tesouro, segundo a mesma norma: a) Avaliação Heurística (especialistas avaliam em aspectos qualitativos e quantitativos); b) Modelagem de Afinidade (alguns usuários agrupam cartões com termos para que os resultados sejam analisados em relação à hierarquia dos termos existentes, entre outros aspectos); c) Teste de Usabilidade (conforme a literatura nas áreas de computação e arquitetura da informação sobre usabilidade e métodos de avaliação).

Os critérios de avaliação, expressos pelas perguntas abaixo conforme apontados na norma ANSI/NISO Z39.19-2005 e já mencionados por Soergel (2002), demonstram alguns dos cuidados para elaboração de tesouros (American National Standards Institute, 2010, p. 95, tradução nossa):

O vocabulário controlado contém termos sinônimos ou equivalentes sem indicar o relacionamento? A apresentação reflete a análise conceitual, por exemplo, a sequência de conceitos em nível hierárquico? As notas são fornecidas quando necessário? Todas as facetas necessárias estão incluídas? Os termos estão atualizados e são usáveis? (pontualidade dos termos). Os sinônimos de termos de entrada incluídos na lista alfabética são completos?

Outro fator de desenvolvimento do tesouro com qualidade consiste na manutenção, pois, a linguagem dinâmica do usuário e atualização constante dos documentos leva à necessidade de revisão e manutenção.

A manutenção do tesouro consiste em regulamentos para adicionar, modificar e excluir termos, de modo que cada termo tenha um responsável por sua elaboração e que um histórico fique registrado até a eliminação do termo. Vale destacar que, segundo a American National Standards Institute (2010, p. 96, tradução nossa) “[...] os vocabulários controlados refletem a linguagem e são, portanto, instrumentos dinâmicos [...]”. Segundo a norma ANSI/NISO Z39.19-2005, devem ocorrer revisões periódicas da terminologia por meio de políticas e procedimentos da instituição. Destaca-se que as versões eletrônicas têm a vantagem de serem acessadas e analisadas mais rapidamente.

### 3. Metodologia

A metodologia de análise de conteúdo de Bardin (2011) foi aplicada, conforme Amorin e Café (2016), em amostra de publicações científicas de Dagobert Soergel. A análise de conteúdo constituiu-se de três etapas metodológicas: a) pré-análise; b) exploração do material; c) tratamento de resultados com inferência e interpretação de aspectos sobre o tesouro apresentados pelo pesquisador. Assim, o pensamento de Soergel foi analisado, inferido e elucidado no que tange aos tesouros.

A etapa de pré-análise contribuiu para a seleção das obras do pesquisador, apresentando os seguintes passos e critérios:

a) *identificação da produção bibliográfica de artigos científicos do autor*: verificando o currículo do pesquisador Soergel disponibilizado em sua página web (<http://www.dsoergel.com/>), bem como suas publicações apresentadas no Google Acadêmico (<https://scholar.google.com/citations?user=4bJ6HcUAAA&hl=en>) para observação de suas características e estabelecimento da amostra e categorias;

b) *coleta de informações sobre os artigos publicados, como*: ano de publicação, título do artigo, título do periódico, indicação de coautoria e fonte para localização da publicação no período de 1972 a 2017;

c) *identificação e seleção de periódicos e eventos de maior relevância em que o autor publicou seus artigos*: Foi aplicado o critério de seleção de artigos com autoria única para evidenciar o pensamento de Soergel. O critério de seleção consistiu em considerar os periódicos da área de Ciência

da Informação, ativos até o presente com avaliação por pares às cegas, com destaque para *Knowledge Organization* (prev. *International Classification*) e *Journal of the American Society for Information Science and Technology*. Não foram considerados periódicos de áreas de assunto diversificadas ou amplas. Além disso, foram selecionados trabalhos publicados de conferências organizadas por sociedades de pesquisadores da área de Ciência da Informação e Organização do Conhecimento com destaque para *American Society for Information Science and Technology* (ASIS&T) e *International Society for Knowledge Organization* (ISKO). Sobre os artigos, quatro são artigos de periódicos, dois deles publicados no *Journal of the American Society for Information Science and Technology* (JASIS&T), dois no *Knowledge Organization* (KO) e três são trabalhos publicados em conferências, sendo estes, um do Workshop do *Special Interest Group – Classification Research* da *American Society for Information Science and Technology* (ASIS&T), outro proveniente da *ASIS&T Annual Meeting* e o terceiro do *International ISKO Conference*.

d) *identificação e seleção do tema a ser analisado nos artigos*: por meio de leitura rápida dos textos, verificando principalmente título, resumo e palavras-chaves para identificação do tema principal do artigo buscando entender termos tais como tesouro, linguagem de indexação, ontologia, esquemas de classificação, sistema de organização do conhecimento, terminologia, conceitos, para relacioná-los com linguagem de indexação e tesouro. Com base na fundamentação teórica desse estudo foram, principalmente, procurados os seguintes temas: elementos de construção, funções, manutenção, tecnologias e usos de tesouros e linguagens de indexação, utilizados, posteriormente, como categorias para análise do pensamento do autor nos artigos selecionados.

A etapa de exploração do material consistiu na identificação e análise, em sete trabalhos científicos de Soergel, de aspectos relacionados às cinco categorias destacadas na etapa anterior.

A etapa de tratamento dos resultados teve como enfoque a análise de inferência e interpretação de aspectos qualitativos segundo o pensamento de Soergel.

Foram definidas cinco categorias de análise (elementos de construção; funções; manutenção; tecnologias; e usos) após consulta às publicações com leitura flutuante, por serem aspectos recorrentes e abordados pelo autor nos trabalhos analisados. A etapa de exploração da amostra de 7 publicações possibilitou a identificação e análise

das categorias. Nesta etapa foram feitas análises, inferências e observações sobre os trabalhos publicados e coletadas citações, sempre com enfoque nas categorias de análise.

#### 4. Discussão dos resultados

A pesquisa abrangeu todo o período de publicação do autor disponível para ser consultado em seu currículo online e em sua página de autor na ferramenta Google Acadêmico (2018), compreendendo um total 202 publicações entre os anos de 1972 a 2017. Porém, o período situado entre os anos de 1975 e 1990 e entre 2000 e 2014 não foram representados no estudo, dado que não foi identificado nenhum artigo de periódico ou conferência publicado pelo autor que tenha se adequadado aos critérios de seleção.

Os resultados da análise de conteúdo demonstram a inferência e interpretação sobre aspectos da linguagem de indexação e tesouro, relativos à construção, uso, avaliação e manutenção encontrados em sete artigos publicados por Dagobert Soergel. As análises são:

*Primeiro Artigo – A Universal Source Thesaurus as a Classification Generator* (1972): Estimulado pelo avanço das ações de cooperação no âmbito dos serviços de informação da segunda metade da década de 1960, apresenta a ideia de um *Universal Source Thesaurus* (UST) – tesouro de fonte universal, ou seja, um tesouro para ser uma fonte de informação universal, que seria utilizado como fonte para criação de linguagens especializadas – classificações, tesouros e demais linguagens de indexação e como linguagem de intercâmbio (*switching language*), ferramenta de integração entre diferentes linguagens para facilitar o trabalho cooperativo.

Para isso o tesouro deve apresentar as seguintes características segundo Soergel (1972, p.300, tradução nossa):

- (1) A estrutura de um tesouro universal.
- (2) O uso de um tesouro universal na construção de esquemas especiais.
- (3) O uso de um tesouro universal para concordância ou "troca de linguagem" entre diferentes esquemas de classificação, como base para catalogação cooperativa.
- (4) Organização e gestão de um dicionário de sinônimos de fonte universal.

Além disso, apresenta bases para o pensamento pró-interoperabilidade de sistemas no âmbito da interação de linguagens. Percebe-se a consonância com o pensamento de Dahlberg com relação à sua visão ontológica da organização do

conhecimento, porém o autor apresenta um pensamento mais transigente, ao adotar esta visão de uma linguagem de intercâmbio ou de reconciliação, que foi posteriormente retomada por Gnoli (2011) em sua “defesa” da herança ontológica de Dahlberg.

A noção do conceito como elemento central da linguagem de indexação também está presente no artigo, especialmente perceptível na afirmação de que a base para a construção de um UST e seu uso como linguagem de intercâmbio é a existência de compatibilidade conceitual. Assim “em outras palavras, a indexação cooperativa não requer uma classificação universal; a compatibilidade conceitual entre diferentes esquemas e a disponibilidade de uma concordância entre eles são suficientes” (Soergel, 1972, p. 7, tradução nossa).

*Segundo Artigo – Automatic and semi-automatic methods as an aid in the construction of indexing languages and thesauri* (1974): Apresenta um método automático baseado em estatística para determinação de termos e conceitos e o estabelecimento de relacionamentos terminológicos e classificatórios entre eles com base na frequência de ocorrência e coocorrência de termos e conceitos para construção e manutenção de tesouros. O autor apresenta este método como semi-automático, destacando que embora seja interessante, à época um método completamente automatizado para construção de tesouros não era viável.

Segundo o autor, a frequência coocorrência de termos pode ser utilizada para a determinação de estruturas terminológicas, como a definição dos relacionamentos de equivalência. Já a frequência coocorrência de conceitos, pode ser usada para a construção da estrutura classificatória. Uma das dificuldades apresentadas pelo autor seria a de que os softwares disponíveis não seriam capazes de identificar os conceitos complexos, formados por mais de uma palavra.

Para demonstrar as palavras do autor,

As seções anteriores estavam relacionadas (a) com a identificação de termos e suas inter-relações entre pares e (b) com a identificação de conceitos e suas inter-relações entre pares - as últimas poderiam ser chamadas de informações classificatórias “locais”. O próximo passo é a derivação automática de uma estrutura “global” (um esquema de classificação) para obter a imagem geral. Há duas tarefas inter-relacionadas:

- (a) Encontrar agrupamentos úteis de conceitos.
- (b) Encontrar um padrão de subdivisão do conjunto de documentos em classes não sobrepostas.

Existem dois métodos principais para executar estas tarefas:

- (1) métodos de agrupamento.
- (2) métodos teóricos de gráfico.

Até certo ponto, esses métodos se sobrepõem; pode-se até dizer que o método de *clustering* é um caso especial de métodos baseado na teoria dos gráficos. (Soergel, 1974, p. 38-39, tradução nossa).

Nesse sentido, Soergel explicou a derivação automática de esquemas de classificação para métodos automáticos de construção de tesouros.

*Terceiro Artigo – Beyond facets: semantic roots and modifiers as elements of a conceptual morphology* (1991): O autor expõe a proposta de uma premissa para a construção de linguagens de indexação baseada na análise facetada, em que por meio de uma morfologia conceitual, que parte da raiz semântica de cadeias de conceitos e de um sistema de modificadores para promoção de indexação mais flexível. Ou nas palavras do autor, sua sugestão é de “[...] uma classificação que consiste de um esqueleto bem-estruturado de conceitos centrais ou raízes semânticas e um sistema de modificadores cuidadosamente realizado. Muitos conceitos podem ser formados por meio da combinação destes elementos” (Soergel, 1991, p. 2, tradução nossa).

Segundo o autor, o sistema alcança a flexibilização da indexação ao permitir a pesquisa genérica por meio da raiz semântica, bem como a pesquisa mais específica por meio da combinação de raiz semântica acrescida de modificadores. Contudo, é um sistema ainda em estágio inicial.

*Quarto Artigo – Software Support for Thesaurus Construction and Display* (1994): Apresenta uma proposta de software de construção e visualização de tesouros TermMaster utilizado na construção do tesouro “Alcohol and Other Drugs”. Entre as funcionalidades que são atribuídas ao software estão a de criação de hierarquias de termos, o suporte para atribuição de notas aos termos, criação de variados tipos de relações semânticas, bem como, a atribuição de níveis de “importância” aos relacionamentos entre os termos, a possibilidade de customizar os relacionamentos para incluir tipos definidos pelo usuário e a combinação de diversos tesouros em um. Segundo Soergel (1994, p. 2, tradução nossa),

Um termo do tesouro está vinculado a outros termos por meio de diversos relacionamentos. TermMaster permite um grande número de desses relacionamentos, tornando possível, por exemplo, distinguir no banco de dados (não necessariamente na versão do usuário) entre Termo Sinônimo, e Termo Equivalente (quase-sinônimo) [...]. Notas de escopo são tratadas como relações com o texto; assim, um termo pode ter várias notas de escopo e pode haver

diferentes tipos de notas de escopo, como Notas de Histórico ou Notas Internas [...].

Alguns relacionamentos são mais importantes que outros. O TermMaster permite (mas não exige) as especificações de relacionamento com três níveis de importância que podem ser usados para orientar a inclusão de relacionamentos [...]. O programa pode ser facilmente customizado para incluir um conjunto de relacionamentos definido pelo usuário.

De acordo com o autor “o programa aproxima todas as relações que são conceitualmente iguais, mas expressas em diferentes termos enquanto estes termos forem mostrados como sinônimos nas fontes utilizadas” (Soergel, 1994, p. 3, tradução nossa).

A proposta de combinação de elementos de outros tesouros, especialmente a inclusão de relacionamentos entre conceitos de diferentes tesouros que estão em acordo com as propostas das diretrizes atuais para construção de tesouros (ISO, 25964-2011/2013), no tocante ao mapeamento de vocabulários controlados, base das atuais discussões sobre interoperabilidade semântica com interligação de vocabulários a partir da existência de relacionamentos inter-linguagens.

Quinto Artigo – SemWeb: proposal for an open, multifunctional, multilingual system for integrated access to knowledge about concepts and terminology (1996): O autor parte das recomendações feitas no Congresso Internacional da ISKO realizado na Polônia em 1995, que se dirigia a infraestrutura conceitual referente à organização da informação e do conhecimento na sociedade da informação do então eminente século XXI, para propor um sistema de acesso à base de conhecimento sobre conceitos e terminologia. De acordo com a proposta, o sistema faria a pesquisa em diversas bases, coletaria os dados e apresentaria ao usuário em uma interface comum integrada. Assim, o sistema seria projetado para ser usado por muitos níveis de usuários para melhorar o intercâmbio de informações. De acordo com Soergel (1996, p. 2, tradução nossa),

Este artigo trata da infraestrutura intelectual. Apresenta a visão e arquitetura básica de um sistema para acesso integrado a dados sobre conceitos e terminologia. O sistema reuniria os dados do usuário de uma variedade de fontes que até agora existem em grande parte em mundos separados, incluindo dicionários, tesouros e esquemas de classificação; ele utilizaria bases de conhecimento existentes que são acessíveis através da Internet ou em CD-ROM e em uma base de conhecimento distribuída integrada e comum que cresceria de forma incremental ao longo do tempo. As bases de conhecimento existentes seriam acessadas por meio de uma interface comum que procuraria várias bases

de conhecimento, agruparia os dados em um formato comum e os apresentaria ao usuário. A base de conhecimento distribuída integrada comum forneceria um ambiente no qual muitos colaboradores poderiam realizar projetos classificatórios e terminológicos de forma mais eficiente, com os resultados disponíveis em um formato comum. Com o tempo, os dados de outras bases de conhecimento poderiam ser incorporados na base de conhecimento comum, seja por transferência real (desde que os produtores da base de conhecimento estejam dispostos) ou por referência através de um link. De qualquer forma, tal incorporação requer trabalho intelectual, mas permite uma integração mais forte do que o acesso comum à interface com várias bases de conhecimento. Cada informação na base de conhecimento comum terá todas as suas fontes anexadas, fornecendo um mecanismo de reconhecimento que dá o devido crédito a todos os colaboradores. Todo o sistema seria projetado para ser usado por muitos níveis de usuários para melhorar a troca de informações.

As funções dessa base de conhecimentos, contendo conceitos e terminologia, segundo Soergel (1996, p. 4, tradução nossa), consistem em “[...] mapear um campo conceitual, relacionar os conceitos aos termos e fornecer definições, também para orientar e servir como ferramenta de referência”.

Além disso, as potencialidades do *SemWeb template*, está em organizar as informações dos conceitos e termos.

O modelo se concentra em informações sobre conceitos e termos individuais. [...] o sistema também deve mostrar estruturas conceituais e classificatórias de vários formatos (listas lineares, mapas bidimensionais, etc.), com navegação e navegação adequada ao usuário para que possa passar de uma visão geral para classificações detalhadas. (Soergel, 1996, p. 9, tradução nossa).

Sexto artigo – *The rise of ontologies or the reinvention of classification* (1999): O autor faz uma retomada dos objetivos e propósitos das classificações usadas em bibliotecas e cuja importância foi recentemente descoberta por outras áreas, tais como inteligência artificial, processamento de linguagens naturais, engenharia de softwares, levando à ascensão das ontologias. Aborda o processo de adoção por tais áreas de pequenas classificações de objetos que, ao passo que foram tornando-se mais complexas, denominaram-se ontologias. Soergel (1999) argumenta aqui em defesa da classificação, independente do nome que receba, porém, ressaltando que o desenvolvimento de instrumentos semelhantes em áreas diversas e sem comunicação, considerando especialmente o vasto conhecimento desenvolvido pela Ciência da Informação sobre Classificação,



poderia causar não apenas a reinvenção de ferramentas, mas também a duplicação de trabalho e a elaboração de ferramentas pobremente construídas. De acordo com Soergel (1999, p. 1, tradução nossa),

A classificação desempenha muitas funções, portanto, é reivindicada por muitos campos, mas a comunicação entre esses campos é fraca, levando a uma abordagem marcada por uma reinvenção fragmentada e dispendiosa.

Estruturas ontológicas e lexicais refletem a base do trabalho científico e acadêmico, da aprendizagem e da inteligência das máquinas. Elas servem para muitas funções críticas de pensamento, comunicação, organização e recuperação de informações por pessoas e máquinas. As funções das ferramentas que fornecem tais estruturas (dicionários, tesouros, ontologias / classificações) incluem o seguinte:

- Fornecer um roteiro semântico para campos individuais e os relacionamentos entre campos, fornecendo orientação e servindo como uma ferramenta de referência. Isso inclui as seguintes funções específicas: relacionar conceitos a termos e fornecer definições; esclarecer conceitos colocando-os no contexto de uma classificação / ontologia; relacionar conceitos e termos ou ícones em disciplinas, idiomas e culturas. [...]
- Recuperação de informações de suporte: fornecer suporte baseado em conhecimento de pesquisa do usuário final (árvores de menu, análise de faceta orientada de um tópico de pesquisa, navegação em uma hierarquia ou mapa conceitual para identificar conceitos de pesquisa, mapeamento de termos de consulta do usuário para descritores usados em um ou mais bases de dados ou para as múltiplas expressões de linguagem natural para pesquisa de texto livre); suporte a pesquisa expandida hierarquicamente; suporte a exibições bem estruturadas de resultados de pesquisa; fornecer uma ferramenta para indexação (controle de vocabulário, indexação centrada no usuário ou orientada a problemas).
- Fornecer a base conceitual para sistemas baseados em conhecimento.
- Fornecer a base conceitual para definição de elementos de dados e hierarquias de objetos em sistemas de software.
- Fazer tudo isso em disciplinas, idiomas e culturas.
- Servir como dicionário mono-, bi- ou multilíngue para uso humano e como base de dicionário / conhecimento para processamento de linguagem natural - tradução automática e compreensão de linguagem natural para extração de dados e abstração / indexação automática.

Além disso, anda segundo o autor,

A classificação tem sido usada há muito tempo em bibliotecas e sistemas de informação para fornecer orientação ao usuário, esclarecendo aspectos de sua necessidade de informação e estruturando re-

sultados de busca para navegação, funções amplamente ignoradas pela comunidade de recuperação da informação textual, mas agora recebe atenção crescente no contexto de buscar ajudar os usuários a lidarem com a vasta quantidade de informação na web. Recentemente, outros campos como IA [Inteligência Artificial], o processamento de linguagem natural e a engenharia de software descobriram a necessidade de utilizar a classificação, levando à ascensão do que esses campos chamam de ontologias. [...] uma vez que essas comunidades aumentaram sua consciência de que não existe apenas um problema de classificação, mas também de terminologia, as “ontologias” incluíam também vocabulários e se tornaram verdadeiros tesouros. (Soergel, 1999, p. 2, tradução nossa).

Para o autor, a ontologia é vista, de modo geral, como uma classificação de categorias básicas, entre outros aspectos. Além disso, afirma que existem muitos tipos de bases de conhecimento que utilizam conceitos e terminologia, que embora sejam desenvolvidos para diferentes propósitos, apresentam princípios e métodos muito semelhantes para sua construção, sendo que para essa reflexão o autor cita o próprio trabalho de 1996, que foi analisado nesta pesquisa no quinto artigo.

*Sétimo artigo – Unleashing the power of data through organization: structure and connections for meaning, learning and discovery (2015):* Neste artigo o autor endereça uma série de problemas em que a Organização do Conhecimento pode ser útil, desenhando uma possível ampliação do escopo da área, entre estas questões destacam-se, tais como, computação cognitiva, gestão de dados organizacionais, gestão de prontuários eletrônicos de pacientes, *linked data*, *big data*, interoperabilidade de dados, extração automática de informação, entre outras.

É destacado o fato de que os Sistemas de Organização do Conhecimento (SOC) permeiam as diversas problemáticas abordadas, emergindo a noção dos SOC como elementos centrais da Organização do Conhecimento. Além disso, segundo o pesquisador,

Resumindo (Soergel 2009, 3), “sistemas de organização do conhecimento (KOS) cobrem uma ampla gama de sistemas em camadas, atendendo a uma ampla gama de propósitos. Eles são conhecidos com nomes como ontologia, esquema de metadados, taxonomia, classificação, estrutura de diretórios da Web, plano de arquivamento, tesouros, dicionários, folksonomia e muito mais”. (Soergel, 2015, p. 402, tradução nossa).

Quanto aos SOC especificamente o autor discute a existência e a dispersão dessas ferramentas, que existem em diferentes tipos – como dicionários, sistemas de classificações, tesouros, ontologias, que possuem não apenas objetivos,

mas também características, padrões normativos e softwares (para criação e uso) diferentes, apresentando a unificação de SOC como requisito para a melhoria destas ferramentas e consequentemente para que a própria área de Organização do Conhecimento seja beneficiada.

Soergel (2015) coloca duas modalidades de unificação, a formal e a flexível, com o objetivo da integração dos dados de SOC com outros SOC, mas também com outros tipos de dados estruturados. Para o autor a unificação formal caracteriza-se por SOC usarem “[...] os mesmos princípios de organização, estrutura interna e apresentação para sistemas com diferentes tipos de entidades” e requer o estabelecimento de um único esquema padronizado e abrangente para representar todos os tipos de dados em qualquer SOC a partir de qualquer software (Soergel, 2015, p.419, tradução nossa). A unificação flexível, segundo o autor, de dois ou mais SOC,

[...] não requer fusão total e perda de identidade. Ela pode ser alcançada pela análise dos sistemas, descobrindo semelhanças e concordâncias onde existem, criando acordos onde for possível e explicando as diferenças que devem ser mantidas embora ainda estabelecendo relações de concordância parcial. Um único SOC pode e deve ser hospitaleiro para muitos pontos de vista (Soergel, 2015, p. 420, tradução nossa).

Para Soergel (2015) este processo de unificação, muito identificável com o que tem sido denominado como interoperabilidade, é o foco central do trabalho, sendo nas palavras do autor “a chave para liberar o poder dos dados pela integração de conjuntos de dados” (Soergel, 2015, p.419, tradução nossa).

A síntese das análises realizadas com os sete artigos, aponta que, em geral, as publicações analisadas de Soergel são marcadas pela proposição de novas metodologias e ferramentas de representação de conteúdo, destacando-se, por exemplo, o tesouro universal que servem como base para sistemas especialistas e de classificação (Soergel, 1972), métodos automático e semi-automático de construção e manutenção de tesouros (Soergel, 1974), elementos de construção de tesouros bem estruturados que contribuem para que sejam realizadas as funções de recuperação e uso (Soergel, 1991, 1994), o tesouro que passa a ser um elemento de construção de outros sistemas (Soergel, 1996, 1999, 2015). Esses fatores fortemente embasados nos desenvolvimentos computacionais e tecnológicos mais recentes, mas segundo cada época.

Vale destacar que o autor é muito cuidadoso ao apresentar esses aspectos em seus textos,

exemplificando-os quando parece ser mais pertinente, principalmente, quando explica os relacionamentos entre termos. Assim, apresenta uma escrita bem elucidativa sobre a construção dos tesouros e outros aspectos que abordou.

Especificamente, quanto às categorias analisadas destacam-se:

*Os elementos de construção dos tesouros* visam uma estrutura capaz de ser interoperável com outros sistemas de organização do conhecimento (Soergel, 1972, 2015). Além disso, apresenta diferentes metodologias de estruturação pelos relacionamentos entre termos e uso de dicionários (Soergel, 1991, 1994, 1996, 1999). Por exemplo, os tesouros em meio digital, segundo Soergel (1999), podem fornecer: relacionamentos mais ricos e definições, além de notas; e, apresentar links para dicionários linguísticos e de sinônimos, e outros documentos que embasam o conceito. Esses elementos, de modo geral, contribuem para eliminar ambiguidades e controle de sinonímia, segundo Souto (2003), além de contribuírem com a estrutura sintática e a semântica do tesouro (Soergel, 1975). Nesse contexto, as abordagens para desenvolvimento de tesouros segundo García-Marco (2016c) e outras, conforme Cervantes (2009), são imprescindíveis para demonstrar a representação do domínio de modo a expressar também as necessidades ou perguntas dos usuários, destinatários desse instrumento de organização elaborado para fins de recuperação da informação.

*A manutenção do tesouro* (Soergel, 1974), mostra que pode colaborar com a melhoria de relacionamentos terminológicos e classificatórios, entre outros aspectos. Nesse sentido, a manutenção regulamenta o desenvolvimento do tesouro diante de mudanças da linguagem do usuário (American National Standards Institute 2010; International Organization for Standardization, 2011). Além disso, a manutenção não pode ser mais vista à parte do processo de construção dos tesouros, pois é um instrumento de representação em constante desenvolvimento, principalmente quando seu uso apoia ambientes tecnológicos na Web, como nos mecanismos de pesquisa que, segundo García-Marco (2016a), melhora a precisão na recuperação.

*O aspecto das funções* demonstrou que tesouros contribuem para mapear campos de conceitos e servir de fonte de pesquisa, além de interoperar com outros sistemas de organização do conhecimento (Soergel, 1996, 1999, 2015). Nesse sentido, outras funções dos tesouros, propiciadas por novas tecnologias são, segundo Soergel (1999, p. 5, tradução nossa),

Transmitir significado, orientação e estrutura: auxiliando ao usuário a pensar em um problema e ajudar a formular a consulta, o que requer bons métodos para exibir uma boa estrutura de interface, mostrando a hierarquia, gráfico conceitual e facetas para consulta.

O pesquisador, sobretudo, enfatiza a interoperabilidade como aspecto importante na relação do tesouro com outros sistemas de organização do conhecimento que, de modo geral, segundo Mendes, Reis e Maculan (2015), demanda mais cuidado no relacionamento entre termos. Por isso, a preocupação do pesquisador com os relacionamentos adequados entre termos quando se utilizam diferentes métodos, padrões e tecnologias para garantir a qualidade e integridade do tesouro na interoperabilidade. Esse aspecto da interoperabilidade demanda muita interação entre tecnologias que propiciem a troca de informações de modo estruturado, com vocabulários interligados (García-Marco, 2017).

A categoria de uso dos tesouros apresentou a preocupação em oferecer elementos ricos para a recuperação da informação mediante as diferentes modalidades de pesquisas: genérica, específica e pesquisa expandida hierarquicamente, além da possibilidade de customização da busca para incluir relacionamentos definidos pelo usuário que possam ser utilizados na indexação e na construção de tesouros (Soergel, 1991, 1994, 1996, 1999). Esses aspectos contribuem para orientar a busca e recuperação da informação pelo usuário, como também para o indexador usar o tesouro na indexação conforme as necessidades dos usuários (Dextre Clarke; Zeng, 2012; Soergel, 1975). Ainda, os tesouros são contruídos para os usuários e segundo necessidades e perguntas dos usuários, nesse sentido, a construção do tesouro também é voltada para o seu uso (García-Marco, 2016c).

As tecnologias utilizadas para desenvolvimento dos tesouros evoluíram de sistemas automáticos e semiautomáticos, passando por *softwares* específicos, para sistemas que podem ser utilizados em conjunto na *web*, como as ontologias, além das tecnologias de inteligência artificial, *linked data*, entre outras visando a interoperabilidade (Mendes, Reis e Maculan, 2015; Soergel, 1974, 1994, 1996, 1999, 2015). Nesse contexto, a interoperabilidade entre instrumentos de representação como o tesouro ocorre em ambiente de Internet e Web Semântica, por oferecer condições de alimentação de novos mecanismos semânticos desses ambientes, por exemplo, com o padrão W3C SKOS (García-Marco, 2016d, 2017).

## 5. Considerações finais

A evolução dos tesouros no discurso de Dagobert Soergel mostra o aprimoramento científico e metodológico desse instrumento de representação, considerando os elementos de construção, funções, manutenção e tecnologias.

As publicações analisadas foram de grande importância para o entendimento desses aspectos porque demonstram as apropriações contemporâneas das contribuições de Soergel para o desenvolvimento dos tesouros, como por exemplo, o tesouro universal que serve de fonte para a construção de outros tesouros e antecipa a idéia de interoperabilidade, ou a ênfase sempre presente do emprego de métodos automáticos e semiautomáticos na construção de tesouros.

O aspecto das funções demonstrou que tesouros contribuem com sistemas especialistas e classificação, que permitem a pesquisa genérica por meio da raiz semântica, como também a pesquisa mais específica. As funções da base de conhecimentos sobre conceitos e terminologia consistem em mapear campo conceitual, fornecer definições, e servir de ferramenta de referência.

As funções dos sistemas de organização do conhecimento, conhecidos como ontologias, esquema de metadados, taxonomia, classificação, tesouros, entre outros, tem o objetivo da integração dos dados com outros sistemas de organização do conhecimento para ampliar e melhorar a organização do conhecimento e sua área científica, também deve ser hospitaleiro. Portanto, abrangem a contribuição para construção de outros sistemas de organização do conhecimento, como também para apresentar bases de conhecimentos, entre outros aspectos. Em geral, os aspectos da construção dos tesouros contribuem para a realização das funções de recuperação da informação, contribuição da elaboração de outros instrumentos de representação, entre outros aspectos.

A manutenção do tesouro é apontada em apenas um trabalho, mostrando que permite atualização do tesouro nos relacionamentos terminológicos e classificatórios, também na frequência de ocorrência e coocorrência de termos e conceitos, mostrando que assim o tesouro pode ser atualizado.

As tecnologias utilizadas para desenvolvimento dos tesouros e sistemas de organização do conhecimento evoluíram de sistemas automáticos e semiautomáticos, passando por *softwares* (sistemas ou programas) específicos, para sistemas que podem ser utilizados na *Web*, ainda aborda

as tecnologias apropriadas para inteligência artificial e ontologias, entre outras.

A categoria de uso dos tesouros demonstrou a preocupação com a recuperação da informação por meio de pesquisas (genérica e específica), a facilidade de ser customizado para incluir relacionamentos definidos pelo usuário, a intenção de ser projetado para muitos níveis de usuários, visando a melhora de intercâmbio de informações, além de mostrar que o uso pode ser a partir de recuperação por meio de pesquisa expandida hierarquicamente, ou com exibições bem estruturadas em resultados de pesquisa. Também, pode ser usado como ferramenta para indexação (controle de vocabulário, indexação centrada no usuário ou orientada a problemas).

Nesse sentido, o pesquisador Dagobert Soergel demonstra uma visão abrangente das linguagens de indexação como parte de um sistema com o qual troca insumos com outros sistemas de organização do conhecimento para o seu desenvolvimento.

Percebemos que as temáticas abordadas pelo autor vêm sendo de interesse da área com os recentes desenvolvimentos das ferramentas computacionais, o que também deu vida nova às discussões conceituais trazidas pelo autor, como a noção do conceito/termo como elemento central da linguagem de indexação, a construção de tesouros por meio de fusão de tesouros existentes e a importância dos dados semânticos não apenas para a recuperação da informação online, mas como para o desenvolvimento da própria web em suas funções.

É possível notar que o seu pensamento pode ser inovador em muitos aspectos, como por exemplo, quanto à interoperabilidade entre linguagens de indexação, que na visão do autor pode ser alcançada sem que seja necessário um apagamento das diferenças culturais de cada sistema envolvido. Também, na publicação mais recente, percebe-se que o autor passou a considerar a noção de sistemas de organização do conhecimento enquanto um termo abrangente que abarca os diversos instrumentos de representação do conhecimento, como o tesouro, aqui observado em especial, assim como os sistemas de classificação e as ontologias, entre outros.

Os elementos de construção dos tesouros são fundamentados pela necessidade de controle terminológico e de vocabulário adequados, em face às tecnologias de cada época, também visando a interoperabilidade.

As funções e uso dos tesouros demonstram, principalmente, a sua importância para a recuperação da informação e seu uso na construção de

outros sistemas de organização do conhecimento.

Em suma, os artigos analisados mostraram que Soergel apresenta inovações e acompanha tendências para elaboração, utilização e melhoria dos tesouros.

## Referências

- American National Standards Institute; National Information Standards Organization. (2010). ANSI/NISO Z39.19-2005. Guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2010. <https://www.niso.org/> (2017.10.16)
- Amorin, I. S.; Café, L. M. A. (2016). Os conceitos de comunidade discursiva, domínio e linguagem na análise de domínio Hjørlandiana // 17 Encontro Nacional de Pesquisa em Ciência da Informação. Salvador: UFBA, 2016. <http://www.ufpb.br/evento/lti/ocs/index.php/enancib2016/enancib2016/paper/view/3580> (2018.09.15).
- Bardin, L. (2011). Análise de conteúdo. São Paulo: Edições 70, 2011.
- Barité, M. et al. (2010). Garantia literária: elementos para uma revisão crítica após um século // *TransInformação*, 22:2, (maio/ago., 2010) 123-138 <http://www.scielo.br/pdf/tinf/v22n2/a03v22n2.pdf> (2018.07.26).
- Berti Junior, D. W. et al. (2017). Semiautomação de relações em tesouros: uma proposta para refinamento de relacionamentos semânticos a partir do tesouro Agrovoc. // *Informação & Informação*, 22:3 (set./out. 2017) 377-404. <http://www.uel.br/revistas/uel/index.php/informacao/article/view/28724> (2018.07.26).
- 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 (ago./dez.2011) 53-73. <http://periodicos.unb.br/index.php/RIC/article/view/6209> (2018.07.26).
- Cervantes, B. M. N. (2009). A construção de tesouros com a integração de procedimentos terminográficos. Universidade Estadual Paulista: Marília. Tese de Doutorado.
- Currás, E. (2005). Ontologías, taxonomía y tesouros: manual de construcción y uso. Gijón: Trea, 2005.
- Dextre Clarke, S. G. (2017). Thesaurus (for information retrieval) // Hjørland, B.; Gnoli, C. (Eds.). *Encyclopedia of knowledge organization*, 2017 <http://www.isko.org/cyclo/thesaurus#top> (2019.02.28).
- Dextre Clarke, S. G.; Zeng, M. L. (2012). From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modelling. // *Information Standards Quarterly*, 24:1 (2012). <http://eprints.rclis.org/16818/> (2017.10.16).
- Fujita, M. S. L. et al. (2016). A gestão de linguagens de indexação por softwares: avaliação para uso no controle de registros de autoridade em catálogos online de bibliotecas. // *Scire*, 23:2 9 (jul.-dic. 2016). 63-73. <https://www.ibersid.eu/ojs/index.php/scire/article/view/4430> (2018.07.30).
- Fujita, M. S. L. et al. (2018) Construction and evaluation of hierarchical structures of indexing languages for online catalogs of libraries: an experience of the São Paulo State University (UNESP). // *Knowledge Organization*, 45 (2018) 220-231.
- García-Marco, F. J. (2016a) Enhancing the Visibility and Relevance of Thesauri in the Web: Searching for a Hub in the Linked Data Environment. // *Knowledge Organization*, 43 (2016) 193-202.

- García-Marco, F. J. (2016b). Normas y estándares para la elaboración de tesauros de patrimonio cultural. // El lenguaje sobre el patrimonio: estándares documentales para la descripción y gestión de colecciones. Ministerio de Educación, Cultura y Deporte, Editors: Secretaría General Técnica. Centro de Publicaciones, Ministerio de Educación, Cultura y Deporte, 29-46.
- García-Marco, F. J. (2016c). Teaching Thesaurus Construction: A Top-Down Approach for LIS Undergraduate Programmes. // AIKO, 5 (2016) 546-554.
- García-Marco, F. J. (2016d). The evolution of thesauri and the history of knowledge organization: between the sword of mapping knowledge and the wall of keeping it simple. // Brazilian Journal of Information Studies: Research Trends. 10 (2016) 1-11.
- García-Marco, F. J. (2017). ISO 25964-2: interoperabilidad de los tesauros con otros sistemas de organización del conocimiento en contextos semánticos. // Anuario ThinkEPI, 11, (2017) 305-309.
- Gnoli, C. 2011. Ontological foundations in knowledge organization: the theory of integrative levels applied in citation order. // Scire 17:1: (2011) 29-34.
- Golub, K. 2011. Knowledge Organisation Systems. *Technical Foundations*, UK. 6. <http://technicalfoundations.ukoln.ac.uk/subject/knowledge-organisation-systems.html>
- Google Acadêmico (2018). Dagobert Soergel. <https://scholar.google.com.br/citations?hl=pt-BR&user=4bJ6HcUAAAAJ> (2018.10.03).
- International Organization for Standardization. (2011). ISO/FDIS 25964-1: Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval. Geneva: International Organization for Standardization, 2011.
- International Organization for Standardization. (2013). ISO/FDIS 25964-2: Information and documentation – thesauri and interoperability with other vocabularies - part 2: Interoperability with other vocabularies. Geneva: International Organization for Standardization, 2013.
- Lancaster, F. W. (2002). El control del vocabulario en la recuperación de información. Valencia: Universidad de Valencia, 2002.
- Marcondes, C. H. (2011) Um modelo semântico de publicações eletrônicas. Liinc em Revista, 7 (2011) 82-103. <http://revista.ibict.br/liinc/article/view/3290>
- Martins, M. C. B. (2013). Indexação e controlo da terminologia em bibliotecas do ensino superior politécnico em Portugal: o sistema no Instituto Politécnico de Portalegre. Universidad de Salamanca: Salamanca. Tese de Doutorado. <http://comum.rcaap.pt/handle/10400.26/7171> (2017.03.25).
- Mazzocchi, F. (2018). Knowledge organization systems (KOS). // Knowledge Organization, 45 (2018). 54-78.
- Mendes, P. R.; Reis, R. M. dos; Maculan, B. C. M. S. (2015). Tesauros no acesso à informação: uma retrospectiva. // Revista ACB: Biblioteconomia em Santa Catarina, 20:1 (2015) 49-66. <https://bit.ly/2Sxw0eb>. (2016.02.28).
- Soergel, D. (1972). A universal source thesaurus as a classification generator. // Journal of the American Society for Information Science, 23:5 (1972) 299-305. <http://www.dsoergel.com/cv/B11.pdf> (2018.07.19).
- Soergel, D. (1974). Automatic and semi-automatic methods as an aid in construction of indexing languages and thesauri. // International Classification, 1:1 (1974) 34-38.
- Soergel, D. (1975) Theoretical problems of thesaurus building with particular reference to concept formation. 1975. <http://www.dsoergel.com/cv/B12.pdf> (2018.06.26).
- Soergel, D. (1991). Beyond facets: Semantic roots and modifiers as elements of a conceptual morphology. // Advances in Classification Research Online, 2:1 (1991) 149-166. <http://www.dsoergel.com/cv/B36.pdf> (2018.07.19).
- Soergel, D. (1994). Software support for thesaurus construction and display. // Advances in Classification Research Online, 5:1 (1994) 157-184. <http://journals.lib.washington.edu/index.php/acro/article/viewFile/13784/11898> (2018.07.198).
- Soergel, D. (1996). SemWeb: Proposal for an open, multi-functional, multilingual system for integrated access to knowledge about concepts and terminology. // Advances in Knowledge Organization, 5 (1996) 165-173.
- Soergel, D. (1999). The rise of ontologies or the reinvention of classification. // Journal of the American Society for Information Science, 50:12 (1999) 1119-1120. <http://www.dsoergel.com/cv/B70.pdf> (2018.07.19).
- Soergel, D. (2002). Thesauri and ontologies in digital libraries: Tutorial. <http://www.dsoergel.com/cv/B63.pdf> (2018.06.26).
- Soergel, D. (2005). <http://www.dsoergel.com/> (2018.10.03).
- Soergel, D. (2015). Unleashing the power of data through organization: structure and connections for meaning, learning, and discovery. // Knowledge Organization, 42:6 (2015) 401-427.
- Soler Monreal, C.; Gil Leiva, I. (2010). Posibilidades y límites de los tesauros frente a otros sistemas de organización del conocimiento: folksonomías, taxonomías y ontologías. // Revista Interamericana de Bibliotecología, 33:2 (2010). <https://bit.ly/2NoVNrd> (2019.02.28).
- Souto, L. F. (2003). Recuperação de informações em bases de dados: usos de tesouro. // Transinformação, 15:1 (jan./abr.,2003) 73-81. <http://www.scielo.br/pdf/tinf/v15n1/06.pdf> (2019.02.28).

---

Enviado: 2019-04-01. Segunda versión: 2020-01-13.  
Aceptado: 2020-06-04.

---

