

Sistemas de Organização do Conhecimento na visão da Ciência da Informação

Knowledge Organization Systems in the Information Science's approach

MSc. Eliana Carlan¹

Dra. Marisa Bräscher Basílio Medeiros²

Resumo – o artigo apresenta os sistemas de organização do conhecimento e a relação com as bases teóricas utilizadas na construção de tesouros, taxonomias, ontologias e sistemas de classificação na literatura da área de Ciência da Informação. O estudo utiliza a metodologia de revisão de literatura na área de organização e representação do conhecimento, além de pesquisa nas bases de dados, a fim de investigar a produção bibliográfica sobre o tema, entre 1998 e 2009. Identifica um caminho teórico percorrido na construção desses sistemas passando pela teoria da classificação, teoria do conceito, relacionamentos entre conceitos e princípios da Linguística e Terminologia. A partir de amostra representativa, faz-se análise nas suas características extrínsecas e intrínsecas. Conclui-se que tesouros e sistemas de classificação são mais citados na literatura, inclusive servindo de referência teórica para o desenvolvimento desses sistemas. Apesar dessa base teórica proveniente dos estudos sobre tesouros e classificações com normas e padrões reconhecidos internacionalmente, evidencia-se a necessidade da consolidação de padrões comuns para o desenvolvimento dos diferentes tipos de SOC. Aponta-se a necessidade de congregar interesses multidisciplinares pela convergência de objetivos e, com isso, gerar melhores práticas de organização e representação do conhecimento.

Palavras-chave: Sistemas de Organização do Conhecimento. Tesouros. Taxonomias. Ontologias. Sistemas de Classificação.

53

Abstract - This article presents the knowledge organization systems related to theories to build thesaurus, taxonomies, ontologies and classification systems in the literature field of Information Science. It uses the methodology of literature review and a research on the same field databases in order to investigate the bibliographic production about the theme, from 1998 up to July 2009. It identifies the same theoretical way to build KOS through the classification theory, concept theory, the relationship between the concepts and the foundation of Linguistics and Terminology. Extrinsic and intrinsic characteristics were analysed from the representative sample of the bibliographic production about KOS. The last chapter verifies that the thesaurus and classification systems are the most quoted in the literature about KOS, being a theoretical reference to the development of these systems based on the international standards and rules. It highlights the importance of consolidating common standards to build different types of KOS and shows the need of gathering the multidisciplinary interests linked by the same goals and also getting better practices in the knowledge organization and representation.

Keywords: Knowledge Organization System. Thesaurus. Taxonomy. Ontology. Classification Systems

Resumem - El artículo presenta los sistemas de organización del conocimiento y la relación con os fundamentos teóricos utilizados en la construcción de tesouros, taxonomias, ontologias y sistemas de clasificación en la literatura de Ciencia de la Información. El estudio utiliza la metodología de revisión de la literatura en el área de organización y representación del conocimiento, y la búsqueda de las bases de datos con el fin de investigar la producción de la literatura sobre el tema entre 1998 y 2009. Identifica un camino recorrido teórico en la construcción de estos sistemas a través de la teoría de la clasificación, la teoría del concepto, las relaciones entre los conceptos y principios de la Lingüística y

¹ Mestre em Ciência da Informação. Consultora OPAS/OMS - área de disseminação da informação.

E-mail: eliana.carlan@gmail.com

² Doutora em Ciência da Informação. Professora da Faculdade de Ciência da Informação da Universidade e Brasília. E-mail: marisab@unb.br

Terminología. De una muestra representativa, es el análisis de sus características intrínsecas y extrínsecas. Se concluye que los sistemas de tesauros y la clasificación son los más citados, incluyendo servir como marco teórico para el desarrollo de tales sistemas. A pesar de esta base teórica de los estudios de las normas de tesauros y la clasificación y con las normas reconocidas internacionalmente, pone de relieve la necesidad de consolidación de las normas comunes para el desarrollo de diferentes tipos de SOC. Este artículo pone de relieve la necesidad de poner en común la convergencia de intereses y objetivos multidisciplinario, lo que genera las mejores prácticas en la organización y representación del conocimiento.

Palabra-clave: Sistemas de Organización del Conocimiento, Tesouro, Taxonomía, Ontología, Sistemas de Classificación.

1. Introdução

Este trabalho³ apresenta a representação do conhecimento por meio dos Sistemas de Organização do Conhecimento (SOC)⁴, os quais abrangem todos os tipos de esquemas que organizam e representam o conhecimento, por exemplo, as classificações, taxonomias, tesauros e ontologias. SOC são sistemas conceituais semanticamente estruturados que contemplam termos, definições, relacionamentos e propriedades dos conceitos. Na organização e recuperação da informação, os SOC cumprem o objetivo de padronização terminológica para facilitar e orientar a indexação e os usuários. Quanto à estrutura variam de um esquema simples até o multidimensional, enquanto que suas funções incluem a eliminação da ambiguidade, controle de sinônimos ou equivalentes e estabelecimento de relacionamentos semânticos entre conceitos. A pesquisa foi realizada por meio de revisão de literatura sobre SOC observando seus progressos científicos e os aspectos conceituais e estruturais de elaboração desses sistemas. O objetivo da investigação foi relacionar os fundamentos teóricos indicados na literatura da área e os princípios teóricos encontrados nas publicações sobre o tema, desde seu surgimento até a realização deste trabalho, por meio das análises bibliométrica e de conteúdo. Contemplando também a relação dos SOC, no âmbito da Ciência da Informação (CI), e os princípios teóricos que orientam a construção de tesauros, taxonomias, ontologias e sistemas de classificação.

Para atingir os objetivos foi feito um levantamento bibliográfico na área com a opinião dos diversos autores percorrendo um caminho teórico fundamentado na Teoria da Classificação, Teoria do Conceito, nos relacionamentos entre conceitos, nos princípios da Linguística e da

3 Artigo decorrente da dissertação de mestrado: "Sistemas de Organização do Conhecimento: uma reflexão no contexto da Ciência da Informação", de Eliana Carlan sob a orientação da Dra. Marisa Bräscher, Universidade de Brasília, Fev. de 2010. Disponível em: http://bdtd.bce.unb.br/tesdesimplificado/tde_busca/processaPesquisa.php?pesqExecutada=1&id=5381

4 O termo Sistema de Organização do Conhecimento é a tradução para o português do original inglês "Knowledge Organization System" (KOS), proposto pelo *Networked Knowledge Organization Systems Working Group* na 1ª Conferência da ACM *Digital Libraries* em 1998, Pittsburgh, Pennsylvania. Assim como a sigla KOS, adotou-se o correspondente SOC em português.

Terminologia. Dos diferentes tipos de SOC, deu-se destaque aos **tesauros**, pelo uso ainda frequente na área de organização de informação e por sua existência consolidada e padronizada por normas internacionais (ISO, ANSI/NISO); as **taxonomias**, por sua importância na organização de informações em empresas e instituições, principalmente, no desenvolvimento de portais no ambiente web; as **ontologias**, pelo interesse da comunidade de pesquisa na área, em função das promessas da web semântica e, ainda pelo potencial que oferecem em relação à capacidade de representação do conhecimento de forma complexa e completa e os **sistemas de classificação**, pela ampla utilização na organização da informação em bibliotecas, onde são empregadas até hoje.

2. Sistemas de Organização do Conhecimento

O conhecimento está em “eterno” crescimento, transformando-se e acumulando-se. Quando se adquire conhecimento relaciona-se com algo já existente, e sobre o qual somos capazes de raciocinar e chegar a conclusões. A criação de conhecimento sobre os objetos que nos cercam constitui uma prerrogativa essencial da racionalidade humana. Desenvolve-se, por instinto, num processo cognitivo que leva a identificar características do objeto percebido e comparar com características identificadas em outros objetos já conhecidos. A partir desse conhecimento, inicia-se um processo classificatório do objeto.

Para Dahlberg (2006), a organização do conhecimento é a ciência que ordena a estruturação e sistematização dos conceitos, de acordo com suas características, que podem ser definidas como elementos de herança do objeto, e a aplicação dos conceitos e classes dos conceitos ordenados pela indicação de valores, dos referentes conteúdos dos objetos ou assuntos. A partir dessa organização do conhecimento criam-se ferramentas que apresentam a interpretação organizada e estruturada do objeto, chamados de SOC. No contexto da Ciência da Informação, os SOC ou esquemas de representação do conhecimento são instrumentos que fazem a tradução dos conteúdos dos documentos originais e completos, para um esquema estruturado sistematicamente, que representa esse conteúdo, com a finalidade principal de organizar a informação e o conhecimento e, conseqüentemente, facilitar a recuperação das informações contidas nos documentos. A infraestrutura que dá suporte ao desenvolvimento dos SOC requer, antes de tudo, uma análise das necessidades dos usuários dos sistemas; a identificação do tipo de SOC apropriado e o desenvolvimento do *hardware* e do *software* adequado à arquitetura de rede, sua integração e manutenção. Portanto, pode-se dizer que sistema de organização do conhecimento é uma denominação nova para as linguagens documentárias que agregam elementos incorporados nas inovações tecnológicas da era digital.

Os SOC são ferramentas semânticas com vocabulários estruturados e formalizados, usadas para o tratamento e a recuperação da informação, tanto no ambiente web como no tradicional, inclusive no desenvolvimento da web semântica. Para Hodge (2000) os SOC constituem o “coração” dos Sistemas de Recuperação da Informação das bibliotecas, museus e arquivos. Não existe um esquema de classificação do conhecimento sobre o qual todos concordem. Um SOC pode ser significativo e vantajoso para uma cultura, uma coleção ou um domínio e para outros pode não ser.

O *World Wide Web Consortium* (W3C) desenvolveu um modelo e especificação, denominado *Simple Knowledge Organization System* (SKOS), para dar suporte aos tradicionais SOC (tesauro, taxonomia, vocabulário controlado e sistemas de classificação) de forma que seja entendido pela máquina. SKOS usa uma sintaxe flexível de XML/RDF que fornece estrutura para publicação dos termos usados nos SOC e seus relacionamentos para dar suporte às buscas, mapeamentos e conexões entre os diferentes SOC.

No final da década de 50 do século XX, inicia-se o desenvolvimento e utilização dos sistemas de indexação e classificação por meio de palavras chave. Cavalcanti (1978, p. 27) define **tesauro** como “uma lista estruturada de termos associados empregada por analistas de informação e indexadores, para descrever um documento com a desejada especificidade, em nível de entrada, e para permitir aos pesquisadores a recuperação da informação que procuram”. No contexto da Documentação, o tesauro é considerado um instrumento de controle terminológico eficaz para a organização do conhecimento e importante ferramenta no tratamento e recuperação da informação. Os tesauros são estruturas consolidadas e padronizadas por normas internacionais, como a ISO 2788 (1986), a ISO 5964 (1985) e a ANSI/NISO Z39.19-2003 que tratam da construção, formatação e manutenção de tesauros monolíngues e multilíngues. Com base nessas normas pode-se sintetizar a estrutura de tesauros em duas partes: **Base teórica** é a essência e a fundamentação para a construção de tesauros: **conceito, termo, categorias e faceta**. **Base técnico-operacional** abordagem mais prática no desenvolvimento de tesauros: **planejamento, coleta de termo, controle terminológico, estabelecimento de relações entre conceitos e formas de divulgação e publicação**. A construção de tesauros é uma tarefa difícil, que exige um esforço coletivo, pois as linguagens construídas são únicas em cada domínio do conhecimento e, portanto, sofrem constantes modificações à medida que as línguas evoluem.

O ponto de partida das **taxonomias** é a classificação, por semelhanças e diferenças entre características do objeto num dado domínio, em que objetos e fenômenos são divididos em classes, essas subdivididas em subclasses, e em sub-subclasses e assim sucessivamente. Embora as taxonomias tenham suas raízes nos trabalhos de Aristóteles, Linnaeus e Darwin

(CONWAY e SLIGAR, 2002; TERRA, 2005) o conceito de taxonomia sofreu uma transformação nos tempos da informática, chegando ao âmbito da CI e da Documentação, no que se refere aos sistemas de classificação. As taxonomias vêm sendo usadas para a criação de metadados ou termos comuns para descrever um objeto, com foco na recuperação da informação e na categorização, como suporte de navegação e esquemas que organizam conteúdos das páginas na web e lista de controle de dados usados para suporte de mineração de dados. Uma taxonomia bem definida e bem construída racionaliza o processo de busca, reflete a necessidade dos usuários e o conteúdo que ela organiza.

As taxonomias não são documentos estáticos, pelo contrário, adaptam-se às influências do conteúdo e conhecimento dos trabalhadores que as utilizam. Quanto à estrutura, devem ser objetivas e estratégicas. O ponto de partida é a busca pela existência de taxonomias, assim como outros documentos já existentes, do mesmo domínio, resultando em economia de tempo e esforços. Conway e Sligar (2002) afirmam que uma taxonomia bem formada não só reflete a necessidade de seus usuários como o conteúdo que ela organiza.

Ao contrário do princípio dicotômico adotado por Aristóteles, pode-se, atualmente, construir taxonomias policotômicas, ou seja, um termo é associado a tantas classes e subclasses quantas se fizerem necessárias, em um domínio especializado. Com isso, evidencia-se a grandeza do problema de mapeamento multidimensional de qualquer área especializada (CAMPOS e GOMES, 2008). Conway e Sligar (2002), descrevem três tipos de taxonomias: **descritiva**, construída nos modelos de tesouros ou vocabulários controlados; **navegacional**⁵, inerente neste conceito é a ideia da relação gênero/espécie entre vários documentos; e **gerenciamento de dados**, que contém um pequeno conjunto de termos controlados rigidamente e tem particular significância enumerativa. Svenonious (2000) estabeleceu cinco metas centrais para a classificação de documentos em um sistema de informação: encontrar; identificar; selecionar; obter e navegar no acervo documental. As taxonomias criam uma rede semântica comum baseada na necessidade dos negócios e considera os bens intelectuais e a maneira pela qual os funcionários procuram pela informação.

O desenvolvimento de **ontologias** cresce e ganha papel específico em diversas comunidades científicas, como a Ciência da Computação, Ciência da Informação, Medicina, para citar algumas. Uma ontologia define os termos usados para descrever e representar uma área do conhecimento. Pode ser usada por pessoas, bancos de dados, em técnicas e aplicações de raciocínio indutivo e inferências que necessitam compartilhar informações dentro de um domínio. Assim, Robredo (2005, p. 321) define uma ontologia como “o resultado da

5 Sugere-se a leitura do artigo: Princípios classificatórios para a construção de taxonomias disponível em: <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/3626>

formulação, tão rigorosa e completa quanto possível, de um esquema conceitual sobre um domínio”. O fato é que existe uma multiplicidade de acepções e definições para ontologia, que refletem em diferentes tipos, pois são objetos de estudo de diferentes áreas do conhecimento. Noy e McGuinness (2001) apresentam cinco motivos pelos quais o desenvolvimento de ontologias torna-se algo importante e considerável: a) compartilhamento de conhecimento comum em estruturas de informação entre outros povos ou para os agentes de *software*; b) permite o reuso do conhecimento; c) realiza inferências em um domínio de conhecimento; d) separa o conhecimento de domínio do conhecimento operacional; e e) realiza a análise do conhecimento estruturado tendo como resultado respostas mais relevantes.

As ontologias promovem e facilitam a interoperabilidade entre sistemas de informação. Por meio de um processo “inteligente” dos computadores, é possível compartilhar e reutilizar o conhecimento entre os sistemas. Na literatura alguns autores definem ontologias como linguagens documentárias pelos seus elementos de formação: termos, definições e relações. Porém, apesar de possuírem elementos comuns que outros SOC, as ontologias são mais que linguagens documentárias, elas possuem funcionalidades que permitem que máquinas possam processar o “raciocínio” automatizado por meio de regras e inferências (SALES, CAMPOS e GOMES, 2008 p. 63). As ontologias são o elemento da web semântica que possibilita o nível de representação semântico. Por meio delas os *softwares* usados na web semântica, como agentes inteligentes e *web services*, são capazes de utilizar o conhecimento codificado para, ao menos parcialmente, entender e interpretar semanticamente, os documentos e objetos. Para isso, as ontologias são usadas com linguagens OWL (*Web Ontology Language*), XML (*eXtensible Markup Language*) e XML Scheme, RDF (*Resource Description Framework*) e RDF Scheme, que são tecnologias capazes de pesquisar e/ou captar informações de diferentes comunidades.

Os **sistemas de classificação** bibliográfica foram desenvolvidos com o objetivo de organizar os acervos de bibliotecas facilitando o acesso às informações pelos usuários. A Classificação Decimal de Dewey (CDD) foi o primeiro sistema de classificação elaborado e influenciou a construção de muitos outros sistemas. Surgiu em um determinado contexto histórico-científico de muitas décadas atrás, e vem sendo atualizado, revisado e mantido pelo comitê editorial da CDD, localizado na Biblioteca do Congresso dos Estados Unidos desde 1972. A CDD é atualizada a cada sete anos e a sua versão impressa está agora na 22ª edição.

Além da CDD existem outros sistemas de classificação citados na literatura e usados em bibliotecas de todo o mundo: Classificação da Biblioteca do Congresso dos Estados Unidos é pouco sistemática e o seu esquema enumerativo resulta em rigidez e falta de hospitalidade,

criada para atender especialmente as necessidades organizacionais da própria Biblioteca; Classificação Decimal Universal, criada a partir da CDD com alterações e adições, como o uso de um enfoque facetado para possibilitar a análise de assunto mais detalhada; Classificação de dois pontos, primeiro sistema de classificação bibliográfica com bases no princípio analítico-sintético ou análise por facetas, priorizando as categorias - *Personality, Matter, Energy, Space e Time* (PMEST); e Classificação de Bliss, primeiro a dizer que um esquema de classificação representava a “organização do conhecimento” e estabeleceu sua base filosófica e teórica tendo os seguintes princípios básicos: **consenso, arranjo de assuntos correlatos, gradação em especificidade, localização alternativa e brevidade da notação.**

Um grupo de informações pode ser organizado de várias formas, seguindo diferentes esquemas de organização. As notícias de um jornal, por exemplo, podem ser organizadas pelo seu assunto (política, economia, etc.), cronologicamente pela data de publicação ou alfabeticamente pelo nome do repórter. A classificação desse conjunto de informações em diferentes esquemas é chamada de classificação facetada ou multidimensional. Cada esquema representa uma dimensão na estrutura que organiza a informação. A web, por ser um ambiente virtual, permite facilmente apresentar a mesma informação organizada de várias formas. No mundo físico, ao contrário, isso é praticamente impossível. As longas notações criadas em esquemas multidimensionais não são fáceis de usar para fazer o arranjo de pacotes de informação física em prateleiras. Entretanto, em catálogos elas são muito úteis, uma vez que cada dimensão pode ser buscada independentemente.

A definição de classificação, ainda que possa variar de acordo com outros autores, tem elementos essenciais que caracterizam o processo de classificar, que é a formação metódica e sistemática de grupos onde se estabelecem critérios para a divisão. É, provavelmente, o método mais simples de descobrir ordem na múltipla e confusa diversidade da natureza. É usada como instrumento de representação do conhecimento com a finalidade de organizar e recuperar informações. Processo mental que está incorporado ao nosso cotidiano, desde quando temos consciência e armazenamos algum conhecimento. Enquanto fenômeno social, as pessoas classificam intuitivamente as coisas, o tempo todo. Com isso, a proposta inicial da classificação é facilitar as operações da mente quando se percebe e guarda na memória as características dos objetos em questão.

2.1 Princípios de hierarquia

Herança de Aristóteles que dividiu toda a natureza unificada em um todo e o todo em partes: classes naturais, subclasses, sub-subclasses e assim por diante, diz respeito à divisão dicotômica dos objetos em gênero e espécie. Esse processo segue uma série de regras de

associação e distinção ordenadas e sistematizadas. De acordo com Aristóteles, somente uma observação exaustiva pode revelar cada atributo verdadeiro de uma entidade e somente a filosofia pode nos guiar em determinar os atributos necessários e suficientes para os membros de uma classe. Somente quando a entidade é classificada corretamente e suas propriedades essenciais são identificadas que podemos dizer que a conhecemos verdadeiramente.

Kwasnik (1999) acrescenta que o legado de Aristóteles sobrevive no espírito das modernas aplicações de classificação e sugere alguns requisitos para uma estrutura hierárquica adequada: inclusividade, gênero/espécie, herança, transitividade, regras sistemáticas e predeterminadas para associação e distinção, mútua exclusividade, critério necessário e suficiente, informação completa e compreensível, herança e economia nas notações, inferência, definições reais e visão de nível elevado e perspectiva holística. Em resumo, hierarquias são excelentes representações para o conhecimento de domínios em que a natureza das entidades e a natureza do significado dos relacionamentos são conhecidas com classes de limites claros.

2.2 Teoria do Conceito e Relações entre Conceitos

A Teoria do Conceito foi desenvolvida por Ingertraut Dahlberg na década de 1970 que viabiliza uma fundamentação sólida para a determinação e o entendimento dos conceitos, assim denominados, com a finalidade de representação e recuperação da informação. Dahlberg (1978) define “conceito” como unidades do conhecimento, identificadas por meio de enunciados verdadeiros de um determinado objeto representado na forma verbal. Interpretando o triângulo conceitual de Dahlberg, temos no ápice o referente (aquilo que se quer conceituar), as características (predicação verdadeira sobre o referente) e a forma verbal (denominação do referente) representado por um termo. O **conceito individual** acontece basicamente por meio de analogia. Quando nos deparamos com algo desconhecido tentamos aproximar com algum conteúdo já presente em nossos registros mentais. Dahlberg (1978), define como algo único, diferente dos demais, constituindo uma unidade inconfundível e caracterizado pela presença de tempo e espaço, por exemplo, explosão da bomba atômica.

Enquanto que o **conceito geral** se situa fora do tempo e do espaço e se refere a uma multiplicidade de coisas, podendo também ser algo abstrato. Para cada item de referência são elaborados enunciados verdadeiros que formarão o conceito desse objeto, representado por um signo linguístico (termo). A partir disso, atribuem-se predicados ao referente denominados de características, que são propriedades dos objetos e, no nível do conceito, passam a ser características do conceito. Essas características podem ser de dois tipos: as essenciais e as acidentais. As características essenciais definem os conceitos gerais e os conceitos individuais

são determinados a partir das características essenciais acrescentando as acidentais. Para Dahlberg (1978, p. 24-25), a **intensão** de um conceito é a soma total de suas características, bem como a soma total de seus conceitos genéricos e das características especificadoras. Numa definição, nem todos os conceitos genéricos precisam ser mencionados, para representar a intensão do conceito. Apenas o mais próximo é suficiente, uma vez que este necessariamente contém os demais. Por ex.: a intensão do conceito “casa” é a seguinte: - construção; - feita de tijolo ou madeira; - contém sala e dormitórios; etc.

A **extensão** do conceito pode ser entendida como a soma dos conceitos mais específicos para os quais a intensão é verdadeira. Isto é, a classe dos conceitos de tais objetos dos quais se pode dizer que possuem aquelas características em comum que se encontram na intensão do mesmo conceito. A extensão do conceito pode ser de dois tipos: a) extensão de um conceito genérico em relação com os conceitos específicos. Por ex.: Casa - casa de pedra; b) extensão dos possíveis conceitos individuais. Compreende os indivíduos para os quais é válida a predicação genérica do conceito. Por ex.: Casa - casa do Presidente da República.

Com a evolução tecnológica e a demanda por formas de organização e recuperação de informações mais complexas, a **categorização e as relações entre conceitos** são fundamentais e encontram-se cada vez mais detalhadas nos SOC. Considerando que o conceito, geral ou individual, é uma síntese de características, Dahlberg (1978, p. 18) afirma que há relação entre conceitos quando conceitos diferentes possuem uma ou mais características em comum. Para tal comparação pode-se valer de dois tipos de abordagem: quantitativa (ponto de vista da lógica formal) e qualitativa (aspectos formais e materiais).

As relações quantitativas medem as similaridades das características de diferentes conceitos, por meio de uma visão lógica. Na relação de identidade conceitual as características de dois conceitos são as mesmas; na relação de inclusão conceitual as características de um conceito A estão contidas em um conceito B, acrescida de uma ou mais característica; na relação de interseção conceitual as características de dois conceitos se sobrepõem; e na disjunção conceitual as características de dois conceitos não têm nada em comum.

As relações qualitativas consideram os aspectos formais e materiais e está subdividida em: **relação formal/categorial** acontece entre o tipo de item de referência, que está sendo analisado e uma categoria. As características podem, às vezes, corresponder a uma hierarquia de características, sendo que o predicado de um enunciado pode tornar-se sujeito de um novo enunciado, e assim sucessivamente até atingir uma característica tão geral, na sua mais ampla extensão, tornando-se uma categoria. A **relação material/paradigmática** acontece entre conceitos que têm características da mesma natureza no interior de uma categoria. Pode ser de três tipos: **hierárquica**, numa relação lógica de implicação. Quando o conceito mais

específico possui características adicionais em relação ao conceito mais genérico do tipo gênero-espécie ou relação de abstração ou de subordinação. Quando dois conceitos diferem em uma característica, e são especializações de um mesmo conceito mais genérico, a relação hierárquica é dita relação lateral ou relação em renques ou de coordenação. A relação **hierárquica partitiva** existe entre o todo e suas partes, sendo que as partes também podem estar relacionadas entre si. Podendo ocorrer em: a) sistema natural - um organismo vivo - animal ou vegetal (todo) e as partes desse organismo (partes); b) sistema artificial - uma máquina (todo) e as partes da máquina (partes); c) organização humana - um país (todo) e os estados, municípios (partes); e d) assunto ou área do conhecimento - uma disciplina (todo) e os diferentes sub-campos (partes). A relação **de oposição** pode ser de três tipos: relação de oposição contraditória (numérico - não numérico), relação de oposição contrária (preto - branco), relação de oposição positivo - neutro - negativo (favorável - neutro - desfavorável).

As **relações funcionais/sintagmáticas** ocorrem entre conceitos de diferentes categorias. As relações funcionais aplicam-se, sobretudo, a conceitos que expressam processo ou operação. No plano da linguagem, os conceitos podem ser expressos por verbos ou pela derivação de verbos. Num primeiro nível de relacionamento pode-se trabalhar com as perguntas Quem? Fez o que? Num segundo nível as relações envolvem também, proposta, condição, tempo, lugar, pessoa, objeto e correspondem às questões: Porque? Como? Quando? Onde? Quem? e O que? por ex.: o conceito “medição” (verbo medir), relaciona-se com “objeto medido”, “propósito da medição”, “instrumento de medição”, “unidade de medição”, etc. A grande dificuldade de definir as relações funcionais/sintagmáticas deve-se ao fato de que as relações entre os conceitos ocorrem em diferentes categorias e podem se relacionar entre si em algum momento. Isso, porque as associações dependem, em larga escala, do universo de referência considerado e da organização dos domínios de especificidade. O Quadro 1 reúne exemplos passíveis de esclarecer as complexas relações entre termos. Sobre as relações associativas Saussure (1977, p. 145) aponta que os grupos formados por associação mental, não se limitam a aproximar os termos que apresentam algo em comum, mas o “espírito” capta a natureza das relações que os unem em cada caso e cria, com isso, tantas séries associativas quantas relações diversas existam.

Quadro 1. Relações associativas entre termos

Relação	Exemplos
Atributiva	economia / nível de atividade econômica
Disciplina / Objeto	entomologia / insetos
Processo / instrumento	iluminação / lâmpadas
Influência	política monetária / inflação
Matéria-prima / produto	bauxita / alumínio
Coisa / aplicação	abastecimento de água / irrigação
Ação / resultado da ação	tecelagem / tecido
Causalidade ou causa / consequência	crescimento econômico / desenvolvimento econômico
Efeito / causa	medicamento / cura da doença
Dependência causal	doenças patogênicas / agentes
Atividade / agente	tabagismo / fumo
Produto / propriedade	veneno / toxicidade
Atividades complementares	compra / venda
Opostos	vida / morte
Ação / seu paciente	prisão / criminoso
Coisa / seu contra-agente	insetos / inseticidas
Atividade / produto	tecelagem / tecido
Pessoas ou coisas / suas origens	brasileiro / Brasil

Fonte: Cintra e co-autores, 2000, p. 48-51, com adaptação

Existe ainda os sistemas de conceitos multidimensionais ou polihierarquia em que um assunto pode ser classificado em hierarquias diferentes. Campos (2001, p. 88) afirma que nesse tipo de sistema não pode haver simplificação, a maneira de representação reflete que a realidade pode ser percebida e concebida em diferentes contextos conceituais. Contudo, quando as polihierarquias são complexas, sugere-se que sejam divididas em várias monohierarquias, com referências cruzadas entre elas. A polihierarquia permite que um mesmo termo específico possa estar subordinado a mais de um termo genérico, podendo ser representadas nos diagramas em linha e diagramas em linha bidirecional, chamados de Árvore de Busca. Esse tipo de relação é comum em taxonomias navegacionais, pois oferecem ao usuário mais de um caminho para a localização de um mesmo produto, considerando que esse produto tenha caráter multidimensional. Com isso, facilitando a recuperação da informação.

2.3 Linguística, Terminologia e os SOC

Os SOC são constituídos de elementos que delimitam uma determinada área do conhecimento, sendo o termo, um dos elementos. Esses termos são subordinados a uma terminologia contextualizada pelo conteúdo informacional dentro de um domínio específico do conhecimento. Não obstante, a noção semiótica de signo, significado e significante, postulada por Ferdinand de Saussure são fatores relevantes na elaboração dos SOC. O signo linguístico dito por Saussure (1977, p. 80-87) não une uma coisa e uma palavra, mas um conceito e uma imagem acústica. Esta imagem acústica não é o som material e sim a representação deste na nossa mente e nos nossos sentidos. O conceito e a imagem acústica podem ser substituídos por significado e significante, respectivamente. Daí, utilizou-se a palavra símbolo para designar o signo linguístico ou significante que tem uma relação racional com o significado.

Quando se fala em significado, no contexto da organização e representação do conhecimento, logo pensamos em semântica. Na Linguística, semântica é o estudo dos significados. Aqui, toma-se como definição para significado a descrita por Lyons (1987, p. 133) onde diz que “os significados são ideias ou conceitos que se podem transferir da mente do falante para a do ouvinte por encarnar-se, por assim dizer, nas formas de uma ou outra língua”. Assim, os signos para os SOC são descritores resultantes de uma construção pertinente para a significação, ou seja, formas linguísticas que expressam conceitos que representam o conteúdo de documentos. Portanto, a Linguística fornece o aporte teórico que possibilita o desenvolvimento desses instrumentos e permite que se estabeleça uma ponte entre, pelo menos, duas linguagens: a linguagem do sistema e a linguagem do usuário. Os SOC são constituídos de termos, conceitos e relações semânticas e uma teoria para os SOC deve explicar como esses elementos semânticos devem ser selecionados e relacionados.

A Terminologia nasceu na década de 1930 com Eugène Wüster, a quem se atribui o papel de fundador da Teoria Geral da Terminologia (TGT). Wüster tem o foco na precisão da linguagem, com a criação de uma teoria e de uma metodologia da terminologia a fim de conseguir uma comunicação inequívoca e sem ambiguidade sobre os assuntos especializados. O próprio termo “terminologia” é polissêmico, refere-se tanto à Terminologia⁶ teórica quanto à terminologia concreta. Na primeira acepção, a Terminologia trata de uma área do conhecimento, que denomina conceitos, ou seja, conjunto de metodologias e diretrizes que regem a descrição, formação dos termos e estruturação de campos conceituais. Na segunda acepção, refere-se a um conjunto de termos de um campo específico do conhecimento relacionados com uma língua de especialidade, são produtos gerados pela prática. Nesse

6 Para distinguir, optou-se pela grafia de Terminologia, com T maiúsculo, para a Terminologia teórico-metodológica, e com t minúsculo, a terminologia concreta ou produto.

contexto, os termos estão relacionados e definidos rigorosamente para designar os conceitos que lhe são úteis, usados na construção de linguagens especializadas, linguagens documentárias e na elaboração dos SOC. A Terminologia como campo inter e transdisciplinar trabalha com a descrição e ordenamento do conhecimento no nível cognitivo, com a transferência do conhecimento em nível comunicacional e com os elementos centrais, conceitos, termos e áreas de experiências no nível de normas internacionais (ISO).

De fato, a transferência de informação, requer um sistema de significação claro e explícito, pois um sistema de comunicação não existe sem um repertório de signos e sem um corpo de regras que definam como se selecionam e se combinam esses signos para formar as mensagens transmissíveis. A terminologia de uma área específica, como se sabe, é constituída pelo conjunto de palavras especializadas de uma determinada disciplina. Os termos, que são as unidades de base da terminologia, designam os conceitos próprios de cada disciplina especializada (CABRÉ, 1993, p. 169). Com isso, proporcionando um maior entendimento entre pesquisadores e demais usuário dos sistemas de comunicação nas ciências. Um princípio fundamental da comunicação relaciona-se entre o entendimento da mensagem e o receptor, ou seja, não será possível estabelecer comunicação sem a compreensão mútua. A CI manifesta preocupação quando toma a informação como objeto de estudo em seu aspecto de construção, comunicação e uso da informação. Nos contextos técnico-científicos, a terminologia assume papel fundamental por trabalhar com as línguas de especialidades, responsáveis por garantir a comunicação rápida e precisa entre os interlocutores.

Estudos atuais na organização e representação do conhecimento recomendam a escolha de uma das teorias da Terminologia para a construção de SOC, de acordo com o campo do conhecimento a ser trabalhado. Algumas teorias da Terminologia merecem destaque: a TGT, pioneira e responsável pela base do trabalho terminológico, possui uma perspectiva prescritiva e normativa que visa rotular e padronizar definitivamente o uso de termos e conceitos, sem considerar as inúmeras variações possíveis em um contexto comunicacional. A Teoria da Socioterminologia de François Gaudin, sob uma perspectiva descritiva com ênfase ao uso social da língua, aceita a variação e a flexibilidade lexical e conceitual, proporcionando uma aproximação entre mecanismos de informação e usuários. E a Teoria Comunicativa da Terminologia de Maria Teresa Cabré fundamenta aspectos comunicativos das línguas naturais para melhor conduzir a comunicação entre especialistas.

3. Procedimentos Metodológicos

O processo de análise valeu-se de métodos quantitativos e qualitativos. O período definido para a pesquisa foi entre 1998, ano em que foi sugerida a adoção da expressão

“sistema de organização do conhecimento” e o mês de julho de 2009. O caminho metodológico teve início com uma revisão bibliográfica, nas áreas de organização e representação da informação e do conhecimento, fundamentada nos conhecimentos de Biblioteconomia, Documentação e CI. Dentre as diversas bases de dados da Ciência da Informação, optou-se por quatro bases específicas: LISA, LISTA, ISTA e WILSON⁷. Além das bases de dados, foi incluído um repositório de acesso livre, na área de CI: o E-LIS e a base de dados de trabalhos acadêmicos, Base de Dados de teses e Dissertações (BDTD).

A estratégia de busca, nas bases de dados internacionais foi pela expressão “*Knowledge Organization System*” e “*Knowledge Organization Systems*”. Na base BDTD decidiu-se pela expressão “Sistema de Organização do Conhecimento”. Apenas para obtenção de dados complementares ou busca de textos completos, não incluídos nas bases citadas, optou-se por consultar o google acadêmico. A delimitação das bases de dados, o período da busca, as estratégias de busca, a seleção dos indicadores para a análise bibliométrica e de conteúdo e a determinação das inconsistências nos registros, constituíram os elementos que estabeleceram os critérios para a definição do escopo das amostras representativas.

Para a análise dos dados, foram observados dois conjuntos de dados: o primeiro relativo aos aspectos de forma ou extrínsecos numa análise bibliométrica e privilegiou os indicadores: ano de publicação, autoria, título do documento, publicação, idioma, país de publicação, palavra-chave e resumo. O segundo é relativo aos aspectos de conteúdo ou intrínsecos, numa análise de conteúdo. As análises bibliométricas e de conteúdo são realizadas a partir de levantamento estatístico e frequência. A escolha do método da análise de conteúdo se deve a dois motivos principais: a) os procedimentos possibilitam uma análise com base em interpretações dos indicadores; e b) por ser uma análise para observar a relação entre a produção técnico-científica sobre SOC e os princípios teóricos para a construção dos SOC, em que os documentos têm características semelhantes e homogêneas. Dividida em três etapas: pré-análise, especulação e análise do material e tratamento dos resultados obtidos e interpretações.

Na coleta de dados nas bases LISA, ISTA, LISTA e H. W. Wilson ocorreu os seguintes resultados: LISA 30 registros; ISTA 33; LISTA 49; WILSON 13; E-LIS 70 e BDTD zero, totalizando o universo de 195 registros. Do resultado obtido da busca verificou-se baixa revocação⁸. É

7 LISA – Library and Information Science Abstracts; H. W. WILSON – Library Literature and Information Science Full Text; LISTA – Library, Information Science & Technology Abstracts with Full Text; ISTA – Information Science & Technology Abstracts.

8 “Habilidade de recuperar informações relevantes para um assunto específico, quando uma solicitação sobre esse assunto é feita ao sistema. (LANCASTER, 1977)”

oportuno ressaltar que as expressões usadas na estratégia de busca, não constam em nenhum tesouro, índice ou glossário das bases de dados.

3.1 Considerações acerca dos resultados encontrados nas bases de dados

A partir da primeira análise para a seleção da amostra significativa observou-se algumas inconsistências como: trabalhos com apenas o nome do autor, afiliação e país; artigos sem autoria; programação de eventos; trabalhos repetidos; dados insuficientes para análise; documentos não publicados; apresentações de trabalhos; um manual e um livro que não foi possível acessar; um estudo de caso; artigos identificados em línguas estrangeiras diferentes do inglês. A BDTD foi descartada para o estudo por obter resultado zero ou busca vazia na pesquisa. Pela falta de elementos, esses itens não foram computados. O total de registros recuperados foi de 195. No entanto, após este levantamento inicial e descartados os registros considerados inconsistentes, delimitou-se a amostra representativa da pesquisa em 90 registros para a análise.

4 Análise dos Dados

Para a análise dos **aspectos extrínsecos**, foram levantados os itens: base de dados, ano de publicação, autor, título do documento, publicação, palavra-chave, idioma e país de publicação. Porém, no decorrer da pesquisa observou-se que nem todas as bases de dados indexam o idioma do documento e o país de publicação. Das bases consideradas para análise apenas a LISA indexa o idioma e o país de publicação. As bases E-LIS e Wilson indexam, somente, o idioma e as bases ISTA e LISTA não indexam idioma nem país de publicação. Pela falta de dados para analisar o conjunto, esses itens não foram analisados. Embora o idioma para a indexação dos documentos nas bases de dados seja, predominantemente, o inglês.

Ano de publicação - no período de 1998 até julho de 2009 observa-se que nos primeiros anos houve modesta contribuição de publicações sobre o tema. Mas, a partir de 2002 existe uma tendência crescente de publicações sobre SOC. Os anos 2006 e 2007 apresentam maior número de publicações, que somados atingem 41% da produção. Em 2008, a produção foi de 13%, porém até meados de 2009 houve diminuição no número de publicações.

Publicação - foram selecionados e examinados 53 diferentes tipos de publicações. As publicações sobre o tema encontram-se distribuídas em vários tipos, o que aponta para a dispersão de publicações. Sendo que o menor número de trabalhos produzidos estava em 35 publicações diferentes e o de maior número de trabalhos ocorreu em duas publicações. Quanto maior o número de publicação, menor é o número de produção e vice-versa. Devido à diversidade de publicações encontrada, optou-se por dividir em três categorias, para facilitar a

visualização. As categorias são: **periódicos** (67,92%), que fazem parte os jornais, boletins e revistas; **eventos** (26,41%) englobam as conferências, congressos, seminários, anais, fóruns e *workshop* e **acadêmicos** (5,66%) são as teses, dissertações e monografias.

Autores - houve a ocorrência de 141 autores. Desse total, 125 autores aparecem com uma única produção sobre o tema, sendo 62 citados como autor principal e 63 como autores secundários. Os autores com mais produção são Marcia Lei Zeng, Katrin Weller, Núñez Paula e Israel Adrián, Claudio Gnoli e Douglas Tudhope.

Palavra-chave - buscou-se identificar os termos mais frequentes no campo palavra-chave dos registros recuperados e, também identificar as áreas às quais o termo "*Knowledge Organization System*" pode estar relacionado. No total foram computadas 404 palavras-chave, sendo 225 termos diferentes. Dentre os artigos recuperados o termo "*Knowledge*" sozinho ou associado a outros termos aparece como palavra-chave desde 1998, mas somente em 1999 a expressão "*knowledge organization system*" aparece como palavra-chave. "*Information retrieval*" e "*digital libraries*" apareceram pela primeira vez em 1998 e 2002 respectivamente. O primeiro artigo com o termo "*semantic*" data de 2001, sendo que a expressão "*web semantic*" aparece apenas em 2005 e em 2006 observa-se a expressão "*semantic relation*".

A análise dos **aspectos intrínsecos** ou de conteúdo foi desenvolvida em:

Pré-análise - foram selecionados os trabalhos que mencionam em seus conteúdos, algum tipo de relação com a base teórica apresentada na revisão de literatura. Esta seleção foi feita por meio de leituras dos campos: título, palavra-chave e resumo. Com base nesse critério foram identificados 43 documentos para a análise temática e selecionados para a leitura na íntegra. Exceto algumas inconsistências que foram apontadas: artigos não localizados na íntegra; artigos completos no idioma desconhecimento da autora. Portanto, esses quatro trabalhos não fizeram parte do cômputo da análise de conteúdo totalizando para análise temática 39 documentos.

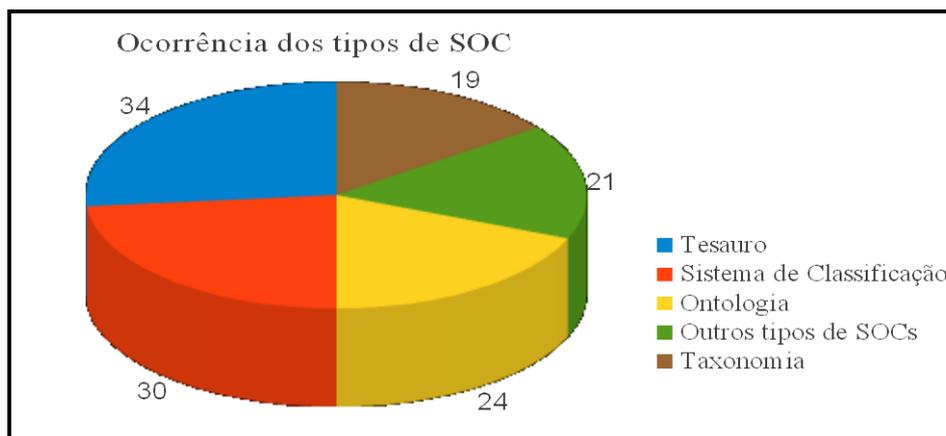
Especulação e análise do material - leitura na íntegra dos documentos relevantes. Os diferentes tipos de SOC foram classificados de acordo com o número de vezes que foram abordados nos documentos com relação ao desenvolvimento de tesauro e/ou taxonomia e/ou ontologia e/ou sistemas de classificação e/ou outro tipo de SOC⁹ e as categorias pré-estabelecidas: princípios da classificação; conceitos e relações entre conceitos; linguística e terminologia. No número de publicações por ano verificou-se apenas uma publicação nos anos

9 No decorrer da pesquisa observou-se que muitos autores não especificam o tipo de SOC, por isso a necessidade de criar o grupo "outro tipo de SOC".

de 1999 e 2000. Uma modesta contribuição até 2004 e obtendo um crescimento na produção com o auge em 2006, seguindo em 2008 e 2007.

Tratamento dos resultados obtidos e interpretações - análise e discussão de resultados, desdobradas em dois conjuntos principais que tratam em seus conteúdos, sobre os tipos de SOC e as categorias abordadas nos trabalhos. As categorias foram pré-determinadas na metodologia e são baseadas na fundamentação teórica desenvolvida na revisão de literatura.

Gráfico 2. Número de citação dos tipos de SOC

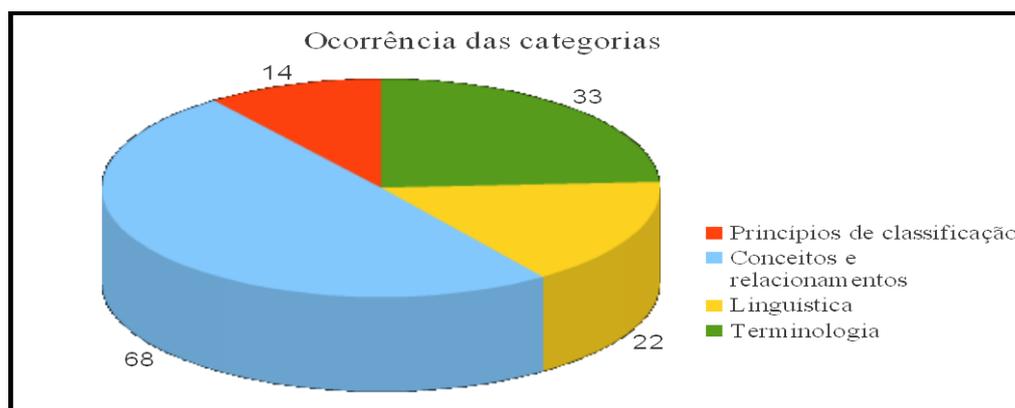


Fonte: elaboração própria

O gráfico mostra que os tesouros são os mais citados com 87% de frequência dos documentos analisados. Seguida dos sistemas de classificação com 77%, as ontologias com 61%, os outros tipos de SOC com 54% e as taxonomias com o menor índice de frequência 49%. O fato dos tesouros e os sistemas de classificação estarem nos tipos de SOC que apresentam maior índice de ocorrência permitem refletir, que tal fato pode ser justificado por esses dois sistemas possuírem estruturas de construção consolidadas por normas e padrões internacionais, enquanto que as taxonomias e ontologias ainda não. Além disso, não se pode deixar de mencionar a preocupação com a própria área de CI em refletir sobre a necessidade da consolidação de uma base teórica comum para a construção de tais sistemas.

Dentre os sistemas de classificação os que mais se destacam são CDD e CDU com 14 citações, LCC com 9, classificação facetada 7 e classificação de Bliss uma. Sobre os outros tipos de SOC identificados na análise ressaltam os vocabulários controlados, lista de cabeçalho de assunto, *gazetteer*, glossário, dicionário, mapa conceitual e folksonomia. Na análise por categorias, foram evidenciadas as influências teóricas quando a temática é SOC, de acordo com o gráfico a seguir.

Gráfico 2. Número de citação por categorias da base teórica



A categoria que representa a base teórica com maior frequência nos artigos analisados foi “teoria do conceito e relações entre conceitos”. Como parte dessa categoria, os temas específicos mencionados são: teoria do conceito (frequência 20), relação hierárquica (17), relação associativa (12) e categorização (12). No entanto, alguns autores referem-se a “relações entre conceitos” (7) sem especificar qual o tipo de relacionamento. Os resultados apontam com clareza a influência da fundamentação teórica para o melhor entendimento e precisão do conceito do objeto a ser representado, fornecendo bases seguras tanto para estabelecer relacionamentos quanto para a determinação dos termos. A “terminologia” representa o aporte teórico com o segundo maior número de ocorrências. Conforme constatado, o tema predominante relatado nos documentos é o controle terminológico (19), que engloba o controle da polissemia, as estruturas sintáticas e semânticas e os termos de origem mais complexa. Observou-se em muitos trabalhos o interesse em identificar e discutir os desafios para o desenvolvimento geral de estruturas de conceitualização, que abrange os conceitos e as terminologias a fim de melhorar as informações cruzadas e a representação de conceitos entre diferentes sistemas usados para fins da recuperação de informações.

Na categoria “linguística” identificou-se os temas sobre significado e significante, relação de sinonímia, hiponímia, homonímia, hiperonímia, meronímia e equivalência. Verificou-se, também, que diversos autores mencionam apenas relações semânticas (23) sem especificar o tipo de relacionamento. Por outro lado, o número de ocorrências é bastante elevado, merecendo ser destacado, inclusive, se considerados os relacionamentos especificados nas categorias “conceitos e relações entre conceitos” e “linguística”. Percebe-se que a natureza de diferentes relações semânticas, seu uso apropriado e bem definido podem ajudar na construção dos SOC consistentes. Finalmente, “princípios de classificação” foi a categoria com o menor índice de ocorrência, porém não menos significativo. Durante a análise, observou-se os fundamentos teóricos da teoria da classificação, os princípios da divisão de classes, das facetas e das hierarquias. O estudo evidenciou a baixa frequência com

que os autores tratam o assunto, mas ao mesmo tempo, alguns autores concordam que o núcleo das relações semânticas para todos os SOC é a hierarquia.

A análise dos temas associados aos SOC sugere a tendência de estudos sobre o assunto no campo da CI. Web semântica e interoperabilidade aparecem como os temas mais frequentes, juntamente com as bibliotecas digitais para fins de recuperação de informação. A folksonomia, por ser tema novo, surpreende pela frequência de ocorrência desse tema relacionado com os SOC. Porém, com a ocorrência mais baixa, destacam-se ciência da informação, teoria da informação, *softwares* para construção de SOC, construção automática e semi-automática de tesouros, rede semântica, bibliometria, abstração automática de termos, mapeamento de diferentes SOC e e-comércio. As palavras-chave mais usadas pelos autores são *knowledge organization*, *thesauri*, *knowledge organization system*, *knowledge management e information retrieval*. E o caminho teórico percorrido para o desenvolvimento dos SOC passa pela teoria do conceitos, da terminologia, da linguística, dos princípios de classificação e com destaque os relacionamentos entre conceitos. Os SOC estão diretamente relacionados com os temas da web semântica, interoperabilidade entre diferentes sistemas, recuperação da informação e bibliotecas digitais.

5 Considerações finais

Quanto ao objetivo geral proposto pela pesquisa a investigação realizada evidenciou que os SOC são estruturas sistematizadas que têm no seu desenvolvimento uma relação muito estreita com os princípios teóricos que subsidiam a construção de taxonomias, ontologias e, principalmente, tesouros e sistemas de classificação. As mudanças são inerentes aos processos de evolução. Alguns elementos são agregados enquanto outros se tornam obsoletos, substituídos ou transformados. Na organização e representação do conhecimento, não é diferente. Com base no que foi verificado na pesquisa, principalmente, na revisão de literatura, estima-se que haja uma probabilidade para a tendência da evolução, que acontece naturalmente das linguagens documentárias para os SOC, acrescidos de elementos próprios da tecnologia moderna para atuar no ambiente web. Outra observação significativa é que os SOC, mesmo empregados no ambiente informatizado, herdaram características no desenvolvimento de suas estruturas que estão fundamentadas em teorias que os profissionais da informação testemunham há pelo menos um século. Como a influência dos sistemas de classificação na construção de outros sistemas, apontada na análise de conteúdo. Porém, não é simplesmente uma repetição do passado, é uma versão nova, incorporada de linguagens novas, *softwares* e ferramentas tecnológicas, que criam estruturas de elevada representatividade conceitual e permitem estabelecer ricas e exaustivas redes semânticas. Os SOC são instrumentos que

dependem de padronização para melhor interagirem, integrarem e interoperarem entre eles e entre humanos, sejam usuários, provedores de informações ou máquinas. Esses sistemas contribuem para diminuir o abismo na comunicação entre pesquisadores, bancos de dados e documentos armazenados, criados pelas dificuldades na organização das informações e do conhecimento.

Na análise de conteúdo ficou evidente que existe um caminho teórico a ser percorrido para a construção de SOC que também fundamenta a construção de taxonomias e ontologias. Esse caminho passa pela teoria da classificação, teoria do conceito, os relacionamentos entre conceitos e os princípios da Linguística e da Terminologia para o tratamento dos termos. Esse percurso teórico fundamenta e garante a consistência nas estruturas dos SOC e nas relações semânticas para o melhor desempenho de suas funções.

É importante destacar alguns problemas detectados no decorrer da pesquisa que, no entendimento da autora, prejudicaram o seu desenvolvimento, como a baixa revocação no resultado das buscas nas bases de dados. A ausência de recuperação de publicações nacionais é outra limitação do estudo, uma vez que se obteve resultado “zero” nas buscas da BDTD. A falta de uma base de dados atualizada que indexe e recupere artigos de periódicos brasileiros na área de CI também contribuiu para ausência de análise de publicações brasileiras.

Referências

American National Standard Institute / National Information Standards Organization. ANSI/NISO Z39.19. *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. 2005. Disponível em: <<http://www.niso.org/kst/reports/standards/>>. Acesso em: out. 2009.

CABRÉ, M. T. **La Terminología: Teoría, metodología, aplicaciones**. Barcelona: Editorial Antártida/Empúries, 1993. 529 p.

CAMPOS, M. L. de A. **A Organização de Unidades do Conhecimento em Hipertextos: o modelo conceitual como um espaço comunicacional para a realização da autoria**. 2001, 198 fl. Tese. (Doutorado em Ciência da Informação) Programa de Pós-Graduação e Ciência da Informação da Escola de Comunicação da Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2001.

CAMPOS, M. L. de A. ; GOMES, H. E. Taxonomia e Classificação: o princípio de categorização. **DataGramZero – Revista de Ciência da Informação**, v. 9, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/F_I_art.htm> Acesso em: abril 2009.

CAVALCANTI, C.R. **Indexação e tesouro: metodologia e técnicas**. Brasília, ABDF, 1978.

CINTRA, A.M.M.; et al. **Para entender as linguagens documentárias**. São Paulo: Polis, 2002.

CONWAY, S.; SLIGAR, C. Building Taxonomy. In: **Unlocking Knowledge Assets**. Washington, Microsoft Press, 2002. 256 p.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.101-107, 1978.

- DAHLBERG, I. **Knowledge Organization**. 2006. Disponível em:
http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_Dahlberg.htm> Acesso em: mar. 2008.
- HODGE, G. **Systems of Knowledge Organization for Digital Libraries: beyond traditional authorities files**. Washington, DC, the Council on Library and Information Resources. 2000. Disponível em: <<http://www.clir.org/pubs/reports/pub91/contents.htm>> Acesso em: 24 maio 2008.
- KWASNIK, B. H. **The Role of Classification in Knowledge Representation and Discovery – 1**. Library Trends. FindArticles.com. 1999. Disponível em:
<http://findarticles.com/p/articles/mi_m1387/is_1_48/ai_57046525>. Acesso em: mar 2009.
- LYON, J. **Linguagem e Linguística: uma introdução**. Rio de Janeiro: Guanabara Koogan, 1987. 322 p.
- NOY, N. F.; MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Stanford University: Stanford, 2001. Disponível em:
<<http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>> Acesso em: maio 2008.
- ROBREDO, J. **Documentação de Hoje e de Amanhã**. 4ª ed. rev. e ampl. Brasília: Edição de autor, 2005, 409 p.
- SALES, L. F.; CAMPOS, M. L. de A.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais. **Perspectivas em Ciência da Informação**, v. 13, n. 2, p. 62-76, maio/ago. 2008.
- SAUSSURE, F. 1977. **Curso de Linguística Geral**. São Paulo: Cultrix/USP, 1977.
- SVENONIOUS, E. **The Intellectual Foundation of Information Organization**. London: The MIT Press, 2000. 254 p.
- W3C – World Wide Web Consortium. **KOS**. Disponível em:
<<http://www.w3.org/2004/02/skos/>> Acesso em: jan. 2009.