

Sistemas de organização do conhecimento

Marcelo Schiessl

Milton Shintaku

Sistemas de Organização do Conhecimento (SOC) são representações do conhecimento, que por sua vez são modelos de abstração do mundo real, construídos para determinada finalidade. São sistemas de classificação, dicionários, ontologias, taxonomias, tesouros, redes semânticas, entre outros.

Carlan (2010) define SOC como

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. (CARLAN, 2010, p. 16)

Hjørland (2008) define Sistemas de Organização do Conhecimento como ferramentas que apresentam a interpretação organizada de estruturas do conhecimento, também chamadas de ferramentas semânticas. Contemplam, essencialmente, conceitos e suas relações semânticas. Para a *Simple Knowledge Organization System* (SKOS, 2004), é um conjunto de elementos geralmente estruturado e controlado que pode ser usado para descrever objetos, navegar em coleções etc. Para Broughton *et al.* (2005), são ferramentas semânticas que consistem em conceitos e relações semânticas definidas e selecionadas.

Os principais objetivos dos sistemas variam de autor para autor. Para Soergel (1999), o destaque está, sobretudo, no auxílio à recuperação da informação. Concretamente, fornecem uma base de conhecimento que apoia a busca e recuperação de informação pelo usuário final.

Para Vickery (2008), os sistemas explicitam o conhecimento de um domínio e possibilitam a organização e recuperação de informações ou de outros tipos de objetos. Também auxiliam a construção de mapas de conhecimento, a navegação para busca de informação, a criação de novo conhecimento a partir do existente, entre outras possibilidades.

No entendimento de Shiri e Molberg (2005), apresentam várias perspectivas de uma coleção, pois oferecem ao usuário uma estrutura conceitual rica e uma estrutura semântica que facilita a seleção de termos para a formulação, reformulação e expansão da busca.

Hodge (2000) aponta relevante função dos SOC em bibliotecas digitais: a de estabelecer elos entre recursos digitais e entre esses e objetos físicos. Os registros podem ser localizados diretamente por meio da navegação nos SOC ou indiretamente, por meio de uma chave de busca que pode ser usada para acessar outra fonte de informação.

O mapeamento terminológico entre diferentes SOC também traduz importante objetivo para Macgregor e McCulloch (2006), já que facilita aos usuários consultar e navegar simultaneamente em múltiplos repositórios distribuídos.

Atualmente, os sistemas informatizados com recuperação do conteúdo integral têm-se expandido, principalmente na web. A disponibilidade de documentos digitais revela certa unidade, pois todos os documentos estão no formato digital, independentemente da tipologia. Além disso, traz consigo diferenças básicas relacionada à própria tipologia, visto que imagem e texto, mesmo que no mesmo formato digital, apresentam diferenças.

Nesse contexto, os SOC tornam-se imprescindíveis na era da internet. Em ambiente digital, são responsáveis pela organização de um acervo de documentos eletrônicos. Diferentemente de acervos físicos, em que cada documento ocupa um lugar no espaço, em ambientes digitais todos os documentos são armazenados de forma não contígua. Nesse caso, a implementação dos SOC oferece diversos tipos de organização.

A organização de documentos eletrônicos, de mesma tipologia ou não, permite preservar as divergências, ao mesmo tempo em que ajusta as afinidades. Dessa maneira, permite que a forma de organização ora seja pelas semelhanças, ora pelas diferenças. Entretanto, todas essas facilidades requerem padronização na terminologia e nas relações entre termos.

Nesse sentido, as linguagens documentais, as terminologias, os vocabulários controlados, entre outros, possuem papel relevante na normalização dos SOC. Essa padronização pode, em alguns casos, restringir a riqueza oferecida pela língua natural, com todas as suas possibilidades, pela utilização de linguagens mais precisas. Contudo, permite a visualização geral dos temas, que leva à organização sistemática.

Classificação dos SOC

Para Vickery (2008), os SOC cumprem diferentes funções, de acordo com o momento da evolução dos sistemas de informação. Portanto, cronologicamente podem ser assim classificados:

- a. era da pré-coordenação: estruturas estáticas que atendem às necessidades dos sistemas manuais de organização e recuperação da informação, como índices e catálogos;
- b. era da pós-coordenação: mais dinâmicos, possibilitam que cada um de seus elementos seja manipulado de forma independente para representar os assuntos de cada documento, como vocabulários controlados e tesauros;
- c. era da internet: classificações hierárquicas que orientam o usuário na escolha do termo que melhor expressa sua questão de busca, como taxonomias;
- d. era da web semântica: são projetados por agentes inteligentes, como ontologias.

Para Hodge (2000), a classificação segue a evolução da linguagem natural à linguagem controlada, ao mesmo tempo em que a estrutura fraca caminha em direção a uma estrutura forte. São exemplos:

- a. listas de termos, como lista de autoridades, glossários, dicionários e *gazetteers*;
- b. classificações e categorização: enfatizam a criação de classes de assuntos, como cabeçalhos de assunto, classificação, taxonomias e esquemas de categorização;
- c. grupos de relacionamentos: conexão entre termos e conceitos, como tesouros, redes semânticas e ontologia.

Breves definições contextuais

Língua e linguagem

Preliminarmente, faz-se necessário distinguir dois termos de conceitos próximos, mas distintos em suas definições: língua e linguagem. Nota-se que em idiomas germânicos, por exemplo, existe somente um termo (*language* no inglês e *sprache* em alemão), enquanto nas línguas românicas existem dois termos (*langue* e *langage* em francês e *lengua* e *linguagem* em espanhol). Para Saussure (1975),

Linguagem é um fenômeno; é o exercício de uma faculdade que existe no homem. A língua é o conjunto de formas concordantes que esse fenômeno assume numa coletividade de indivíduos e numa época determinada. (SAUSSURE, 1975, p. 115)

A língua é uma convenção social de uma comunidade, um fenômeno social estabelecido em determinado espaço de tempo. A linguagem, por sua vez, torna-se mais ampla, é a capacidade de representar conceitos independentes da forma, por isso, torna-se mais flexível nas suas representações, podendo ser por meio da língua ou de outro sistema de representação.

Nesse sentido, língua refere-se aos idiomas, que são sistemas sociais, convenção por uma comunidade de usuários e em constante mutação. A linguagem, no entanto, pode representar sistemas de representação de conceitos com formatos diversos, como sons, sinais e outros, tais como as linguagens de programação de computadores ou a linguagem documentária.

Pela tradição nos textos da ciência da informação, usa-se aqui o termo linguagem sem grandes distinções, apesar de essas diferenças de cunho linguístico serem relevantes.

Linguagem documentária e vocabulário controlado

Um dos objetivos fundamentais da ciência da informação é representar o conteúdo de documentos com a menor perda de informação possível. Esse procedimento visa organizar a forma de armazenar e recuperar as informações neles contidas. Para tanto, utilizam-se linguagens convencionadas que traduzem os conceitos-chave dos documentos para linguagens reconhecidas por sistemas de informação ou de bibliotecas, conhecidas como linguagens documentárias ou linguagens de indexação.

Para se entender melhor o conceito de linguagem documentária, é necessário que se compreenda a linguagem natural, que significa o discurso comum com o qual nos comunicamos no dia a dia. O leitor pode encontrar também a expressão texto livre, que é traduzida do termo *free text*, frequentemente utilizado em textos em inglês para se referir à linguagem natural escrita. A linguagem natural é adaptada às formas de comunicação oral e escrita, que pressupõem sutileza e riqueza de detalhes que nos permitem expressar valores, emoções, associação de ideias, enfim, comunicar efetivamente.

Essa vertente pródiga de características que privilegia as linguagens naturais é também o maior obstáculo para utilizá-las no tratamento de informações. De modo a reduzir o espectro fornecido pela linguagem natural, buscam-se alternativas nas linguagens documentárias que simplificam e condensam o contexto da linguagem natural para que seja preservada a riqueza contida nos documentos originais. Elas se caracterizam pela eliminação da ambiguidade e pelo controle e padronização do vocabulário. Outra característica das linguagens documentárias é a capacidade de explicitar os relacionamentos entre termos.

De acordo com Guinchat e Menou (1994), ela é um instrumento destinado à realização de operações precisas, em condições precisas, para atender necessidades precisas. Também é utilizada no tratamento intelectual de