

Resumo: A Ciência da Informação (CI) encontra-se no atual contexto diante de um paradoxo interessante. De um lado temos um grande *boom* de informações e de outro o desafio imenso que pode consolidar a CI e os Sistemas de Organização do Conhecimento (SOC) mediante esse cenário. Esse artigo busca apontar caminhos que os SOC podem tomar para fazerem frente aos buscadores da *web* como o Google, instrumentos amplamente utilizados e que cada vez mais levam as pessoas a questionarem os sistemas de biblioteca e o porquê de sua existência. Os buscadores da *web* trouxeram uma grande contribuição para a busca de informações no meio digital. Porém, estão imersos em visões de grandes corporações e limitados a estereótipos por vezes preconceituosos. Tais sistemas ainda carecem de competências que podem ser encontradas em SOC que sigam valores éticos e parâmetros de construção que busquem conhecer seus usuários e sua realidade sócio-histórica-cultural.

Palavras-chave: Organização e Representação do Conhecimento; Sistemas de Organização do Conhecimento; Vocabulários Controlados Multilíngue

Abstract: Information Science (IS) faces an interesting paradox in the current context. On the one hand we have a big boom of information and on the other hand the immense challenge that can consolidate IS and Knowledge Organization Systems (KOS) through this scenario. This article seeks to point out ways that KOS can follow to counter web searchers such as Google, widely used tools that increasingly lead people to question library systems and why they exist. The web searchers have given a great contribution to the search for information in the digital environment. However, they are immersed in visions of large corporations and limited to stereotypes, sometimes prejudiced. Such systems still lack skills that can be found in KOS which follow ethical values and construction parameters that seek to know their users and their socio-historical-cultural reality.

Keywords: Knowledge Organization and Representation; Knowledge Organization Systems; Multilingual Controlled Vocabulary

1. Introdução

A Ciência da Informação (CI) encontra-se no atual contexto diante de um paradoxo interessante. De um lado temos um grande *boom* de informações, advento principalmente da era digital, que trouxe a quebra de barreiras e acelerou a velocidade da produção e da também da obsolescência da informação. Isso visto de maneira simplista nos faz acreditar que a CI se encontra em um momento de tranquilidade pois seu objeto de pesquisa maior, a informação, se encontra cada vez mais presente em tudo.

Porém, como havia comentado, temos outro lado. Com essa super produção de informação em distintos meios, formatos e suportes, com o acesso a distância cada vez mais facilitado via *web*, temos um desafio imenso que pode consolidar a CI e os SOC, ou torná-los questionáveis ou até mesmo dispensáveis para os usuários.

Esse artigo busca apontar caminhos que os SOC podem tomar para fazerem frente aos buscadores da *web* como o Google, instrumentos amplamente utilizados e que cada vez

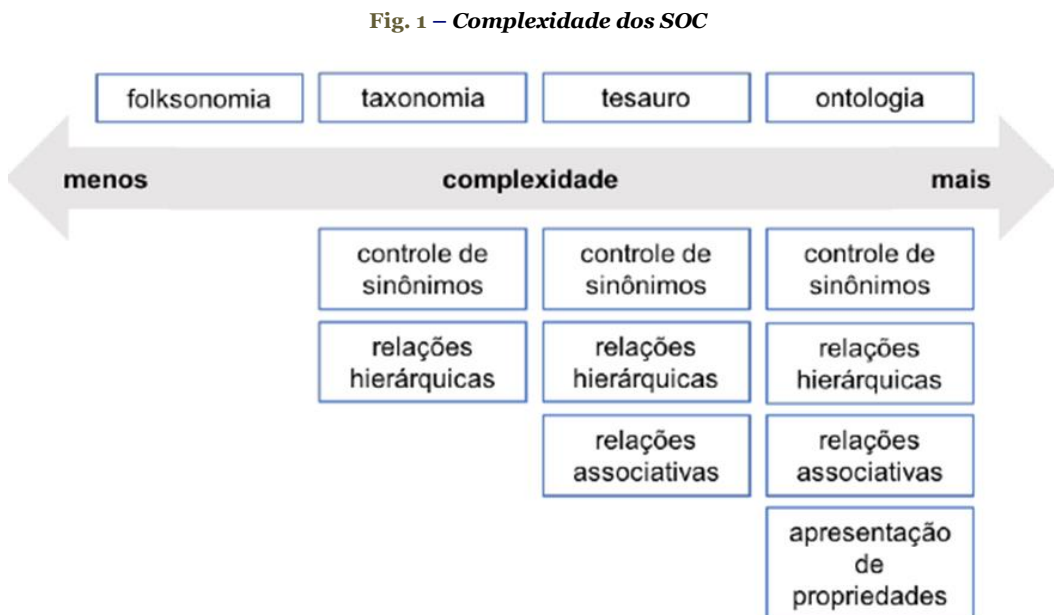
mais levam as pessoas a questionarem os sistemas de biblioteca e o porquê de sua existência.

2. Os Sistemas de Organização do Conhecimento (SOC)

Segundo Hodge (2000:1), Sistemas de Organização do Conhecimento são

todos os tipos de esquemas para organizar a informação e promover a gestão do conhecimento. [...] incluem esquemas de classificação e categorização que organizam materiais em um nível geral, cabeçalhos de assunto que fornecem acesso mais detalhado e arquivos de autoridade que controlam versões variantes de informações-chave, tais como nomes geográficos e nomes pessoais. [...] também incluem vocabulários altamente estruturados, como tesouros, e esquemas menos tradicionais, tais como redes semânticas e ontologias. Como os sistemas de organização do conhecimento são mecanismos para organizar a informação, eles estão no coração de cada biblioteca, museu e arquivo.

Os SOC consistem em estruturas terminológicas que apresentam relações conceituais por meio de termos. Seu objetivo é organizar e representar a informação para sua recuperação. Moreira (2018:102) apresenta a tipologia de SOC e seu nível de complexidade na figura abaixo:



Fonte: (MOREIRA, 2018:102)

Como elemento em comum podemos afirmar que os SOC consistem em conceitos representados por termos. As variações entre as diferentes tipologias consiste na

complexidade para a definição terminológica conceitual, partindo da ausência de controle terminológico (folksonomia) ao controle mais detalhado (ontologia).

2.1. Por que utilizar esses sistemas ao invés de buscadores da internet?

Atuando como bibliotecário, diversas vezes sou questionado sobre o porque e para que, existem os sistemas de classificação da informação. Todos crêem que os buscadores da *web* são imparciais e extremamente eficientes, portadores da verdade e capazes de trazer sempre a informação que precisam. Se ela não é encontrada, é porque não existe. Me questiono como essa confiança absoluta nesses sistemas surgiu, mas não pretendo discutir isso nesse trabalho.

A busca aqui é como criar SOC que atendam as demandas do usuário e demonstrem a ele o quanto esses sistemas são eficientes e necessários como instrumentos para o fomento da ciência.

Hjørland (2012:299) apresenta o seguinte questionamento

The outlook for the future is, however, fundamentally challenged by digital technologies at both the practical and at the theoretical levels. In hindsight, the question arises as to whether KO has ever had a sound theoretical basis because, as discussed below, questions such as: “How do we decide whether A is a kind of B?” have not been properly addressed in the field of KO.

De fato a CI carece de uma definição científica clara no tocante aos critérios para a seleção de um termo em detrimento do outro. Essa definição científica metodológica é extremamente importante para diferenciar os SOC de buscadores da internet.

Biscalchin (2013:112-113) apresentou por meio de quatro categorias de análise um conjunto de catorze parâmetros¹ para construção de vocabulários controlados multilíngue. Esse

-
- 1 1. empregar termos representativos da cultura de um indivíduo ou de uma coletividade deles;
 2. contemplar termos representativos de grupos sociais minoritários;
 3. promover a igualdade de termos das temáticas homens e mulheres, paz e guerra, religiões, correntes ideológicas e convicções político-sociais;
 4. evitar termos discriminatórios ou depreciativos;
 5. fomentar a imparcialidade na recolha de termos a partir dos princípios das garantias literária, de uso, organizacional e cultural;
 6. praticar a seleção ética de termos para a representação por assunto de áreas científicas especializadas;
 7. traduzir os termos da linguagem fonte para a(s) linguagem(ns) alvo, considerando-se a correspondência conceitual que deve ocorrer entre as terminologias das áreas científicas especializadas das linguagens/idiomas envolvidos;
 8. realizar a compatibilidade semântica entre os termos representativos de diferentes culturas nas perspectivas dos preceitos da multiculturalidade e da transculturalidade;
 9. reconhecer a variação formal e conceitual dos termos, considerando-se o desenvolvimento de estruturas poli-hierárquicas de áreas científicas especializadas;
 10. selecionar termos técnico-científicos representativos da cultura e da linguagem de especialidade de áreas do conhecimento;
 11. estabelecer relações sintático-semânticas consistentes entre os termos de áreas científicas especializadas de ordens hierárquica, equivalente e associativa;

conjunto de parâmetros busca otimizar esses SOC de modo que todos atendam critérios para a definição terminológica e para a representação conceitual da informação.

Essa pesquisa apresenta um possível caminho a ser seguido na busca por uma definição de critérios científicos na construção de SOC. O acesso global e facilitado é a chave para o sucesso de um sistema de organização do conhecimento. O modelo tradicional leva grande desvantagem em relação a esse modelo digital que possibilita acesso remoto ao usuário.

Given the significant reliance on centralized classification and the challenges posed by smart technologies, it follows that there is a need for control mechanisms to check the quality of the classifications made, for example, by LC. There is far too little research about the quality of indexing and classification today, and particular decisions are not documented and made available for research. If the centralized classifications are not of a very high quality, users may not find them to be viable, given the many alternatives. In his study, Larson (1991) indicated that many users did not find Library of Congress Subject Headings (LCSH) useful, and even found them to be harmful in the online context. (HJØRLAND, 2012:300)

O modo como a classificação é realizada não é documentada, e isso dificulta os estudos. Desse modo os parâmetros apresentados por (BISCALCHIN:2013) com a aporte da terminologia, da garantia cultural e da multiculturalidade podem ser norteadores para o aprimoramento na classificação de acervos para os usuários, levando em conta seu contexto sócio-histórico-cultural.

Estudos com o de Larson (1991) e de Villen-Rueda, Senso e Moya-Anegón (2007) apontam para uma redução crescente do uso da busca por assunto nos SOC. Isso reflete que os usuários não encontram na busca por assunto o que almejam, ou simplesmente não confiam, ou se identificam com a maneira como a informação é representada.

Hjørland (2012) e Biscalchin (2013) reforçam a necessidade de uma aproximação com o usuário e sua realidade para tentar otimizar esses sistemas e restabelecer a confiança do usuário.

A teoria e a prática têm de se aproximar no "fazer" da classificação, o profissional deve conhecer seu usuário e o contexto em que ele se encontra inserido, suas necessidades de informação e suas estratégias de busca. É o profissional que "sente" o perfil e as mudanças dos usuários e conseqüentemente das áreas científicas e das terminologias utilizadas para a busca de informação.

12. adotar a abordagem de desenvolvimento simultâneo do vocabulário controlado nas diferentes línguas a serem representadas;

13. utilizar a estrutura assimétrica entre os termos na construção de vocabulários controlados multilíngue representativos dos valores sócio, histórico, cultural das terminologias de áreas científicas especializadas;

14. relevar o aspecto comunicativo e a flexibilidade na construção de vocabulários controlados multilíngue consistentes para a recuperação de informações pertinentes de áreas do conhecimento.

A dinamicidade das ciências hoje é enorme, conceitos tradicionais por vezes são contestados e derrubados, alteram profundamente a estrutura de algumas áreas do conhecimento. Isso implica em dois fatores extremamente importantes na CI. Primeiro, os sistemas de classificação devem ser passíveis de atualizações constantes. Segundo, tendo em mente a grande velocidade de mudança nas ciências em geral, os sistemas devem seguir parâmetros claros e bem definidos quanto a política de atualização terminológica, estrutural e conceitual das áreas do conhecimento, para que respeite valores éticos e científicos.

Destaco esses dois itens pois no caso do segundo em particular, não podemos incorrer no erro de ficar alterando a todo o tempo definições, sempre que forem questionadas ou afirmadas por um grupo. Desse modo a política de indexação e de construção desses sistemas deve ser bem clara aos critérios definicionais de como e quando representar essas mudanças.

Por mais que os algoritmos dos buscadores da *web* se atualizem conforme as demandas de busca de seus usuários, a classificação por meio de SOC ainda é mais precisa e eficiente para a ciência, e vou explicar porque acredito nisso.

Além dos algoritmos a intervenção humana e o seu crivo na classificação devem ser considerados, mesmo que também se encontrem passivos de erros e preconceitos. Seguir um conjunto de parâmetros, como os apresentados por Biscalchin (2013) pode reduzir imposições e super posições de informação.

Ao nos concentrarmos em buscadores como o Google corremos o risco de localizarmos informações tendenciosas a uma perspectiva de mundo, ou patrocinadas por um grupo, não sendo muitas vezes isentas de patrocínios, visões unidirecionais e por vezes preconceituosas e/ou excludentes. Se realizarmos no Google uma busca de imagens com o termo "mulher bonita", teremos como retorno o esteriótipo da mulher branca e ocidental como padrão de beleza. Esse estereótipo, que é de um grupo social, não necessariamente (e com certeza não é) de todos os grupos. Essa exclusão de minorias empobrece pesquisas e direciona resultados.

Hjørland (2002:310-311) afirma que

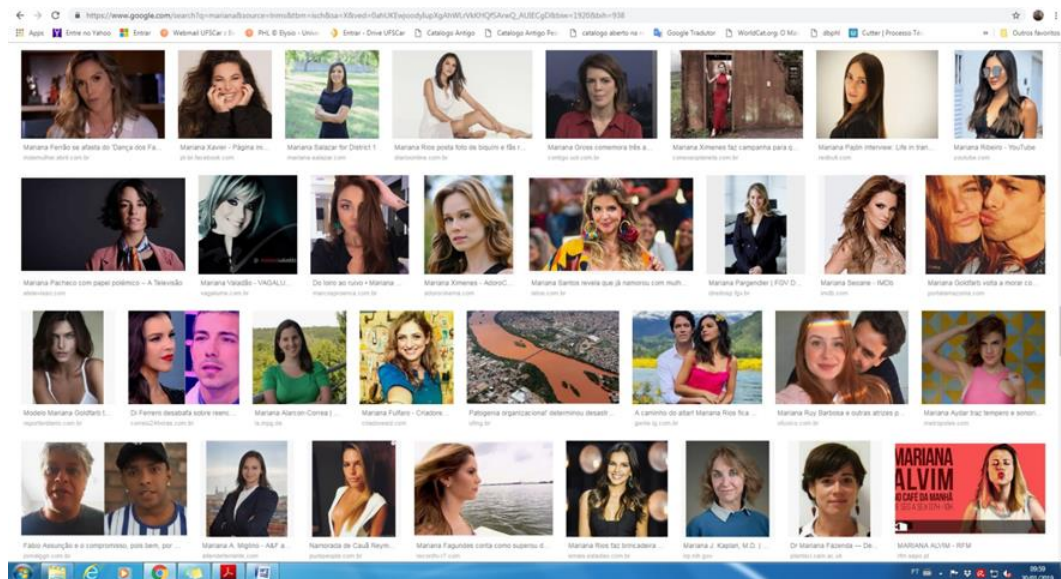
If classifications are to be relevant, they must consider and negotiate between different views and interests. In order to support these given views, classifications must enable IR according to the relevance criteria associated with them. To classify should be to make relevant distinctions in relation to the goals of the system, and therefore implies a consideration and negotiation of different views and interests. Google and other similar IR systems are certainly impressive, but how do they classify and prioritize the relevant information? We tend to think of such systems as neutral and objective tools, but they cannot be.

A ética deve ser discutida e no caso desses sistemas não vejo uma discussão ética e muito menos uma neutralidade no modo como a informação é organizada e recuperada.

O sistema é "desenhado" para atender as necessidades e interesses da organização ou do usuário/área de especialidade a qual ele se propõe a atender. Se você buscar imagens de

Mariana no Google - haverá milhares de fotos de mulheres e artistas com esse nome, mas fotos da cidade homônima mesmo após esse novo desastre com a barragem em Brumadinho, só depois de "rolar" muito a tela do navegador, e mesmo assim elas se apresentam em meio a fotos de mulheres.

Fig. 2 – Print da tela de pesquisa realizada no Google com o termo "Mariana" em 30 de janeiro de 2019



Fonte: Google (2019)

Esse exemplo reforça a importância de questionarmos tais sistemas, pois nesse caso os algoritmos não apresentam o assunto principal desse momento, o desastre da Vale na cidade de Brumadinho. Tal assunto é diretamente relacionado ao fato que ocorreu em Mariana, e que assim como esse trouxe grande comoção a nível mundial.

Mesmo buscando por Brumadinho, o sistema quando se busca por Mariana, continua com as fotos das mulheres, o que demonstra que os algoritmos não acompanham o interesse de pesquisa do usuário.

Um SOC que siga parâmetros para a representação da informação, com políticas bem definidas se torna sem dúvida potencializador na disseminação da informação para a sociedade.

3. Considerações finais

A CI e os SOC se encontram mediante uma oportunidade desafiadora e única. A busca por informação e a geração de novos conhecimentos nunca foram tão intensas e rápidas como na atualidade.

Os buscadores da *web* trouxeram uma grande contribuição para a busca de informações no meio digital. Porém, estão imersos em visões de grandes corporações e limitados a estereótipos por vezes preconceituosos. Tais sistemas ainda carecem de competências que podem ser encontradas em SOC que sigam parâmetros de construção que busquem conhecer seus usuários e sua realidade sócio-histórica-cultural.

Políticas de atualização e recolha terminológica são também pilares na construção desses sistemas, assim como também a ética. Se a CI avançar por esse caminho na construção de SOC, conseguirá colaborar de maneira significativa para o avanço da ciência em nível global.

Referências bibliográficas

BISCALCHIN, R.

2013 *Construção de vocabulário controlado multilíngue: um estudo de possibilidades no contexto da garantia cultural e pela perspectiva da Terminologia*. [Em linha]. 2013. [Consult. 4 fev. 2019].

Dissertação de Mestrado em Ciência, Tecnologia e Sociedade - Universidade Federal de São Carlos. Disponível em:

<https://repositorio.ufscar.br/bitstream/handle/ufscar/1107/4969.pdf?sequence=1>.

HJØRLAND, B.

2012 Is classification necessary after Google? *Journal of Documentation*. [Em linha]. 68:3 (2012) 299-317. [Consult. 12 jan. 2019]. Disponível em:

<https://doi.org/10.1108/00220411211225557>.

HODGE, G.

2000 *Systems of Knowledge Organization for digital libraries: beyond traditional authority files*. [Em linha]. Washington, DC: Council on Library and Information Resources, 2000. [Consult. 4 jan. 2019]. Disponível em:

www.clir.org/pubs/reports/pub91/contents.html.

LARSON, R. R.

1991 The Decline of subject searching: long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science*. 42:3 (1991) 197-215.

MOREIRA, W.

2018 *Sistemas de organização do conhecimento: aspectos teóricos, conceituais e metodológicos*. 2018.

Tese de Livre Docência em Sistemas de Organização do Conhecimento – Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP).

VILLEN-RUEDA, L.; SENSO, J. A.; MOYA-ANEGON, F.

2007 The use of OPAC in a large academic library: a transactional log analysis study of subject searching. *The Journal of Academic Librarianship*. 33:3 (2007) 327-337.

Ricardo Biscalchin | ricardo_biscalchin@hotmail.com

Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP) / Universidade Federal de São Carlos (UFSCar), Brasil